

# Impact of a Proficiency-Oriented, Assessment-Driven, Backward Design Model on Language Teachers' Practices and Professional Identity

Phuong Nguyen\*

*The University of Chicago, USA*

Ahmet Dursun

*The University of Chicago, USA*

Catherine C. Baumann

*The University of Chicago, USA*

## Correspondence

Email: [phuongn@uchicago.edu](mailto:phuongn@uchicago.edu)

### Abstract

Despite efforts to address calls for more language assessment literacy (LAL) training among teachers, few training initiatives emphasize the crucial relationship between assessment and instruction. To address this gap, a comprehensive program was implemented to support and strengthen assessment and teaching across languages at a research-intensive university. This study investigated the program's impact on foreign language (FL) instructors' language assessment knowledge and practices, their teaching methods and strategies, the design and choice of teaching materials, and their professional identities. Following a convergent mixed-methods design, our analyses of survey responses (N = 41) and interview responses (N = 14) from FL instructors revealed that a proficiency-oriented, assessment-driven, backward design model was effective in developing their language assessment knowledge and practices, including a deeper understanding of the testing-teaching relationship and task design. The model was also found to positively impact on the participants' design and selection of materials and activities, and their teaching methods and strategies. Additionally, the instructors reported multiple positive impacts of the model on their professional identity, including enhanced confidence and sense of agency. This study provides important implications for LAL training and professional development for FL instructors.

### ARTICLE HISTORY

Received: 24 April 2025

Revised: 12 October 2025

Accepted: 14 November 2025

### KEYWORDS

language assessment literacy, language teacher training, mixed methods, program evaluation, reverse design

---

### How to cite this article (APA 7<sup>th</sup> Edition):

Nguyen, P., Dursun, A., & Baumann, C. C. (2025). Impact of a proficiency-oriented, assessment-driven, backward design model on language teachers' practices and professional identity. *Language Teacher Education Research*, 3, 1-28. <https://doi.org/10.32038/lter.2025.03.01>

---

### **Introduction**

Professional development (PD) is an indispensable process to support teacher learning with a significant impact on teachers' beliefs and teaching practices (Borko, 2004; Guskey & Huberman, 1995; Kamali et al., 2024). For language instructors, PD can focus on diverse topics. Researchers in assessment literacy (Brindley, 2001; Popham, 2009), both in general and language education domains, have concluded that teachers are not well-prepared to design good tests to make decisions about adjusting instruction to better support student learning. Thus, assessment literacy is an essential component of teacher PD (Black & Wiliam, 1998; Popham, 2009) and has profound impacts on educational settings, especially in terms of high-stakes decisions about language learners' performance and abilities (Davies, 2008; Inbar-Lourie, 2013). Similarly, language testers and researchers have argued for the urgency of preparing instructors with language assessment knowledge so that they can better assess their students and align their instructional practices based on the results (Davies, 2008; Inbar-Lourie, 2013). As many less commonly taught language (LCTL) teachers lack a formal background in pedagogy (Wang, 2009), the need to prepare them to undertake assessment is even more pressing.

A comprehensive program to support and strengthen language assessment and teaching practices, was implemented at a large Midwestern university in the US in 2017. Following a proficiency-oriented, assessment-driven, backward design model, this PD model was informed by findings from a needs analysis on language assessment literacy (LAL) (Dursun et al., 2022). In this systematic model, language instructors identify real-world outcomes, design proficiency assessments and, with outcomes and assessments in place, design and realign curricula. However, to date no formal research has evaluated the impacts of the program on instructors' actual assessment and teaching practices. An evaluation of the program could provide insight regarding approaches to LAL and curriculum design training and its washback on instructors' curricular choices and whether its model could be useful in similar contexts.

### **Professional Development and Assessment Literacy Training for Language Instructors**

Educators have long emphasized PD as a critical career component that benefits both teachers and their learners. These include enhanced teacher agency (Priestley et al., 2015), collaborative communities of practice (Wenger, 1998), and improved subject-matter knowledge, pedagogical practices, and understanding of learners and curriculum (Richards & Farrell, 2005). In a more recent study of a 10-week PD course, teachers reported improvements in their instructional techniques and strategies,

increased confidence and collaborative learning, and a desire for continuous PD (Shin & Kim, 2021).

One critical topic for teacher PD is assessment literacy. As assessment literacy enables teachers to accurately evaluate student progress to improve instruction and learning, it is a fundamental topic (Boyles, 2005; Popham, 2009) and hence pivotal for PD endeavors. Many language assessment researchers have recognized the importance of assessment knowledge and the need to develop resources for pre-service instructors. Boyles (2005) described competencies language teachers need for assessment literacy, including understanding appropriate testing practices, utilizing various means of assessment, interpreting and analyzing results, responding appropriately to results and their meanings, and using results in their teaching. Davies (2008) explained that LAL consists of three fundamental components: skills (practical testing expertise), knowledge (understanding measurement and language), and principles (underlying concepts like validity, reliability, and ethics). Taylor (2009) asserted that a comprehensive grasp of these elements is crucial for effective assessment literacy and advocates enabling professionals to design and utilize tests more effectively. In a study assessing LAL training needs, Fulcher (2012) found language teachers, particularly those who also took on the role of testers, desired a clear, practical guide that incorporates relevant activities. He proposed broadening the LAL definition by integrating theoretical concepts and principles into test development, ensuring that practical experience provides a foundation and embedding theory in a connected, hands-on manner. Discussing the complexity of LAL, Levi and Inbar-Lourie (2020) argued that LAL comprised an amalgamation of different skills: major generic components of assessment knowledge and skills (the generic), newly gained assessment knowhow and existing language-teaching expertise (the language-specific), and teachers' contexts, e.g., familiarity with local environments, national language curricula and testing requirements (the contextual). More recently, Abrar-ul-Hassan (2024) proposed an expanded conceptualization of LAL that integrates ethical, socio-political, and contextual considerations alongside technical skills.

All of the above underscores the bidirectional relationship between assessment and instruction. Despite the consensus among language testers and experts that assessing student progress is essential to improve teaching and learning, many teachers have a limited understanding of assessment fundamentals (Brindley, 2001; Malone, 2013). Well-designed and implemented assessments provide teachers with important information about student performance, enabling them to make decisions regarding their teaching, and, more broadly, to undertake curricular realignment. In turn, changes made to instruction and curricula have a positive impact on the quality of assessments. This interdependence cannot flourish when teachers lack the assessment

knowledge necessary to be able to develop or select and implement tests effectively and subsequently interpret and utilize the results to make informed instructional decisions. In their review of LAL articles, Coombe et al. (2020) outlined key requirements for effective LAL training. Firstly, the assessment knowledge base must be current, reflecting the latest research and policy innovations. Secondly, assessment training should be extensive and sustainable to facilitate deep learning and should integrate the knowledge base with practical contexts. Fourthly, stakeholders' support is crucial. The authors also emphasized the need for more research into teachers' assessment conceptions and practices.

The recognition of the significance of teachers' LAL has motivated studies investigating language teachers' self-reported language assessment competencies and LAL training needs (Giraldo & Yan, 2025; Lam, 2015; Vogt et al., 2020). The findings from these studies have provided recommendations for the literacy required for practitioners to undertake assessment responsibilities. However, fewer efforts on improving LAL among in-service teachers have been documented. One example is the projects done by Malone (2013) who described a tutorial focusing on such topics as reliability, validity, practicality, and test impact without hands-on practice. Montee et al. (2013) reported on a hybrid, 7-week assessment literacy training course which included online sessions on topics such as formative and summative assessment, validity, reliability, practicality, impact, skills assessment, scoring performance assessments and reporting results, and a live session in which participants had an opportunity to develop some classroom tasks. The researchers' analysis of participants' assessment tasks showed that to some extent, they were able to connect instruction and assessment and demonstrated increased awareness of assessment task types appropriate for their programs' age group, language, and proficiency level. Baker and Riches (2018) described a series of workshops for Haitian teachers presenting language assessment concepts and principles and hands-on technical tasks. Brunfaut and Harding (2018) described an LAL training program conducted alongside an actual test cycle and test development project because of the benefits of hands-on test development work in fostering understanding of language assessment principles and theories.

So far, few LAL training projects for instructors in the context of LCTL instruction in higher education have emphasized the relationship between assessment and teaching and incorporated curriculum innovation as an outcome of LAL development. Although the training reported by Montee et al. (2013) provided an opportunity for instructors to reflect on their instructional and program goals, the researchers suggested further focus on the connection between assessment and instruction in future LAL training programs. Additionally, Taylor (2013) speculated that for

classroom teachers, the most important component of LAL is knowledge of language pedagogy. Clearly, integrating language pedagogy into LAL training has the potential of being more impactful. The model described in this study was designed to develop LAL among instructors while emphasizing the essential relationship between assessment and instruction.

### **Proficiency-Oriented, Assessment-Driven, Backward Designed Language Pedagogy**

#### *Theoretical Framing*

Our PD model was framed by three complementary theoretical perspectives. First, backward design (Wiggins & McTighe, 2013) served as the design theory guiding the overall structure, emphasizing the sequence of identifying desired outcomes, determining assessment evidence, and aligning instructional design. Second, the model was informed by the concept of teacher agency (Priestley et al., 2015), which provided a lens for understanding how instructors exercised identity and ownership in shaping their own curricular transformations. Third, collaboration was framed through Communities of Practice (Wenger, 1998), highlighting how sustained interaction, shared practices, and peer learning supported the development of a professional community. Taken together, these three perspectives provide theoretical grounding for how the PD model was organized, enacted, and interpreted.

#### *A LAL Training Model for Language Teachers*

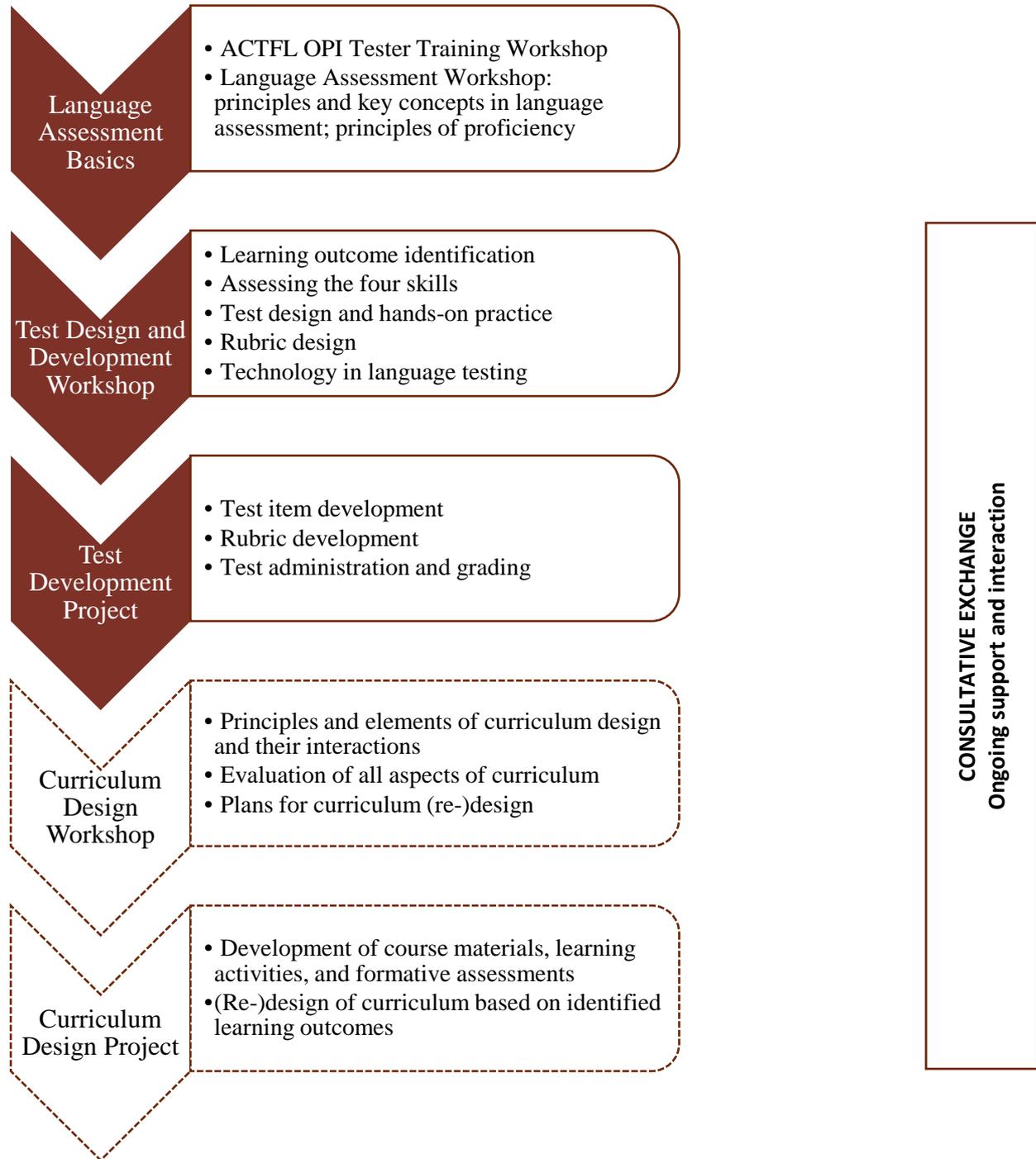
Our PD model was organized around a backward design model that used language assessment to spur language pedagogy innovation and that built a solid foundation for language teaching and testing practices. It was a sequence of language assessment, test design and development, and curriculum design workshops, with financial and PD support at each step. This model was based on five principles. First, it adhered to a backward design framework (Wiggins & McTighe, 2013) utilizing this sequence: identification of desired outcomes, determination of acceptable assessment evidence, and planning of learning experiences. Second, it assumed proficiency-oriented language instruction emphasizing the teaching and learning of language for functional use in real-world contexts. Third, it was assessment-driven in three key aspects: decisions about how and what to assess were informed by and simultaneously embedded in the identified (proficiency) outcomes which in turn informed and drove instructional choices; instructors identified the outcomes and designed the assessments themselves; instructors reverse-engineered their curricula from the proficiency-oriented performance-based summative assessment. Fourth, its bottom-up nature invited and provided opportunities for teacher agency in all phases, ultimately making it possible for teachers to be the agents of their own curricular transformation (Priestley et al., 2015). Finally, it fostered collaborative learning in that

teacher participants were encouraged to enhance their individual learning through multiple instances of collaboration, including interaction between teachers, peer-based learning through mentoring, and the sharing of skills, experience, materials, and solutions to common problems (Richards & Farrell, 2005), all of which resulted in the creation of a community of practice (Baumann et al., 2023).

The model resulted in the development of a PD workshop sequence consisting of five phases (Figure 1). The LAL training sequence, deployed in the first three phases, consisted of Language Assessment Basics, a Test Design and Development Workshop, and a Test Development Project. The first phase, Language Assessment Basics, included an in-person ACTFL OPI® Rater Training Workshop familiarizing participants with the ACTFL proficiency levels and training them to elicit and rate speaking performances, and an in-person Language Assessment Workshop in which participants learned language assessment principles and key concepts as well as additional principles of a proficiency-oriented performance-based test design. In the second phase, participants attended an in-person Test Design and Development Workshop. Working in pre-assigned pairs, they identified learning outcomes for a given course sequence and designed a four-skills performance-based proficiency assessment. They also learned about topics including authentic inputs, rubric design, and technology in language testing, and engaged in hands-on test design activities. The third phase was offered remotely and designed for participants to continue to work in pairs to finish their test development project. Following recommendations from previous researchers (Brindley, 2001; Brunfaut & Harding, 2018; Fulcher, 2012), the LAL training encompassed both theoretical principles and practical test development components, along with ample support to foster greater teacher autonomy and ensure long-term sustainability of the projects.

The LAL training sequence provided participants with the LAL critical to meaningfully engage in curricular (re-)alignment in the next two phases, namely, the Curriculum Design Workshop and Curriculum Design Project (presented in dotted boxes in Figure 1). During the Curriculum Design Workshop, participants who had previously designed an assessment based on their course outcomes using the language assessment knowledge gained from the LAL training sequence examined all aspects of their curriculum and reverse engineered their curricula to ensure that the learners were able to reach the identified outcomes. They attended an intensive 4-day workshop where they learned core principles and elements of curriculum design (e.g., teaching strategies, teaching activities and materials, formative assessments), and then worked on their design, receiving feedback from peers and trainers. In the final phase, participants, upon receiving trainers' approval of the curriculum design, developed course materials, learning activities, and formative assessments.

**Figure 1**  
*Five Phases in the Program*



The program, funded by the institution as an initiative to transform instruction across all languages, has been well attended over the years by 60 instructors, mostly of LCTLs (e.g., Bangla, Catalan, Japanese, Tibetan, and Yiddish). However, no studies have been conducted to evaluate its effectiveness. Investigating the impacts of this program could provide insights into how LAL training programs in similar contexts could be

designed and implemented to bring the most benefit to instructors and their academic programs.

### Research Questions

**RQ<sub>1</sub>:** How has a proficiency-oriented, assessment-driven, backward design model impacted instructors' language assessment knowledge and practices?

**RQ<sub>2</sub>:** How has the model impacted other instructional practices, such as teaching methods, strategies, and materials?

**RQ<sub>3</sub>:** What additional aspects of instructors' professional identity were impacted?

### Methods

#### *Design*

This study followed a convergent mixed-methods design (Creswell & Plano Clark, 2017) where we simultaneously collected qualitative and quantitative data, which were then analyzed separately to reach an overall interpretation to address the research questions. It was selected because of its efficiency (Creswell & Plano Clark, 2017) and ability to provide triangulated findings (Riazi & Candlin, 2014). Participants' responses to Likert-scale items as quantitative data and responses to interview questions as qualitative data were collected and analyzed separately before the results were merged during the interpretation phase to provide comprehensive answers to the research questions.

#### *Study Context*

At the institution where the PD program was implemented, over 50 languages, both commonly taught languages and LCTLs, were taught across various departments. Across the wide array of languages was an equally diverse population of instructors, who varied greatly in expertise and teaching experience. Many LCTL instructors lacked formal training in language teaching, were the sole teachers of their language on campus, and had to rely on outdated or often non-existent textbooks and materials. They continually struggled to maintain enrollment numbers and often had to address the diverse needs of both graduate and undergraduate students, as well as those of heritage and traditional learners. To spur language pedagogy innovation and build a solid foundation for language teaching and testing practices, the institution launched the program which was organized around a backward design model that uses language assessment.

In the first three phases, participants were trained by a certified ACTFL OPI trainer and three language assessment specialists to build their LAL and test development skills. All participants opted into the program and received a \$1,500 incentive upon finishing their test project. They were also able to use their participation in all

assessment-related aspects of the program to fulfill a professional development requirement established by the institution.

### *Participants*

Participants were 48 instructors who were teaching non-English languages at a large Midwest research university. These respondents had completed either the first two (N = 6) or all three phases (N = 42) of the LAL sequence. 34 of them had also participated in the Curriculum Design Workshop and finished a curriculum project.

Fourteen instructors from a range of languages volunteered to participate in interviews describing the program's impact in more detail. The interview participants had either an MA (N = 8) or a PhD (N = 6) in education, foreign/second language teaching, or literature, with teaching experience ranging from 3 – 35 years. Depending on their experience, the participants were categorized into novice ( $\leq 5$  years), mid-career (6 – 15 years), and veteran (15+ years). Regarding LAL training prior to the program, nine stated they had had no previous training, four had some training (e.g., one language assessment course), and only one had more extensive training in their graduate program.

All procedures were reviewed and approved by the university's Institutional Review Board. An informed consent was obtained from all participants prior to data collection, and confidentiality was maintained by removing identifying information from transcripts and survey responses. Participation was entirely voluntary, and instructors could withdraw at any time without penalty. No incentives were given to the instructors for their participation in the survey and interview.

### *Instruments*

#### *Questionnaire*

The questionnaire (Appendix A) collected information about the program's component(s) instructors had participated in, their overall evaluation of the effectiveness of the program and their PD, and the program's impact. The Program's Impact section consisted of twenty-five 5-point Likert-scale items (1 = *strongly disagree*, 2 = *disagree*, 3 = *neutral*, 4 = *agree*, and 5 = *strongly agree*), divided into three categories: Language Assessment Knowledge and Practices (seven items), Instructional Practices (seven items), and Professional Identity (11 items), each with sub-categories. All items were newly developed for this study and drafted to align directly with the study constructs (language assessment knowledge and practices, instructional practices, and professional identity) and reviewed for clarity by the research team.

### *Interview protocol*

The interview protocol included three parts. The first elicited information about the program's impact on instructors' language assessment knowledge and practices before and after participation. The second focused on the program's impact on participants' instructional practices while the last section targeted the impact on other aspects such as instructors' own professional growth or on their language program.

### *Data Collection*

First, the survey was sent to all language instructors using Qualtrics. The qualitative data collection was done separately with 14 instructors who volunteered to participate in interviews, seven of whom had also responded to the questionnaire. They met individually for a 45-minute interview, recorded by MP3 recorder.

### *Analyses*

Because some participants skipped some questionnaire items, the sample size (N) varies slightly across items. Analyses were conducted using pairwise deletion so that all available responses were included. The responses to the Likert-scale items were then checked for abnormalities. Cronbach's  $\alpha$  values were computed to examine the reliabilities for individual scales, which ranged from .82 to .96, indicating a high level of internal consistency (Dörnyei, 2010, p. 94). Average scores and percentages of instructors who selected a particular response category were calculated for each item to reflect the degree to which the participants agreed with the questionnaire statements.

The 14 interviews were transcribed verbatim before thematic coding. In the initial coding, we segmented and coded one randomly selected transcript using the themes in the interview protocol - Language Assessment Knowledge and Practices, Instructional Practices, and Professional Identity. Each theme included several codes as sub-categories. A separate segment was used for each separate comment from instructors related to the nodes in the codebook. Single ideas repeated without adding new information were treated as single segments. After the initial coding, another three transcripts were randomly coded to create new codes that fit within existing themes or refine existing codes. The final codebook consisted of three themes, each with multiple subcategories (Table 1). At this point, any coding discrepancies among the coders were resolved through discussions to arrive at higher inter-rater reliability. The remaining transcripts were divided and coded independently. Krippendorff's  $\alpha$  was reported at .82, indicating high intercoder-reliability (Krippendorff, 2022). The interview coding resulted in 355 segments relevant to the research questions. Twenty-nine valid responses to the open-ended questionnaire items were coded using the codebook created for the interviews through consensus coding. Any comments not

mentioned in the codebook were coded as “Others.” This coding produced 55 relevant segments. Overall, the coding of the qualitative data generated 405 segments. We then counted the instances of occurrence for each code and the total number of instructors who made the comment.

Reliability estimates from both quantitative (Cronbach’s  $\alpha$ ) and qualitative (Krippendorff’s  $\alpha$ ) strands were integrated through a joint display that aligned survey reliability estimates with coding consistency measures, enabling meta-inferences about convergence. Overall, the strands converged, supporting the reliability of both the survey-based ratings and qualitative judgments, and reinforcing the credibility of subsequent interpretations.

### Results

#### *Impact on Instructors’ Language Assessment Knowledge and Practices*

Results from the analysis of the questionnaire responses on Language Assessment Knowledge and Practices are summarized in Table 1 and Figure 2. Overall, the instructors believed that the program had positively affected their language assessment knowledge and practices, as indicated by the very high mean scores above 4.32. Specifically, they agreed that they had gained profound knowledge about language assessment fundamentals, with the mean scores for the individual items between 4.46 - 4.61, and 19.5-32.5% and 58.5-73.2% of the instructors indicated agreement or strong agreement with the positive effect of the program on their language assessment knowledge and practices. They also strongly concurred that the program had helped them define their course outcomes ( $M = 4.45$ , 95% CI [4.19, 4.71]) and align their course target outcomes with their assessment ( $M = 4.49$ , 95% CI [4.23, 4.74]) with approximately 30% and 60% of the instructors selecting *agree* or *strongly agree*, respectively. Additionally, they showed strong agreement on both task design and the testing-teaching relationship, with mean scores of 4.45 (95% CI [4.21, 4.69]) and 4.32 (95% CI [4.02, 4.61]), respectively. For these two items, about one-third of respondents selected “agree” (32.5% and 26.8%), while the majority chose “strongly agree” (57.5% and 56.1%).

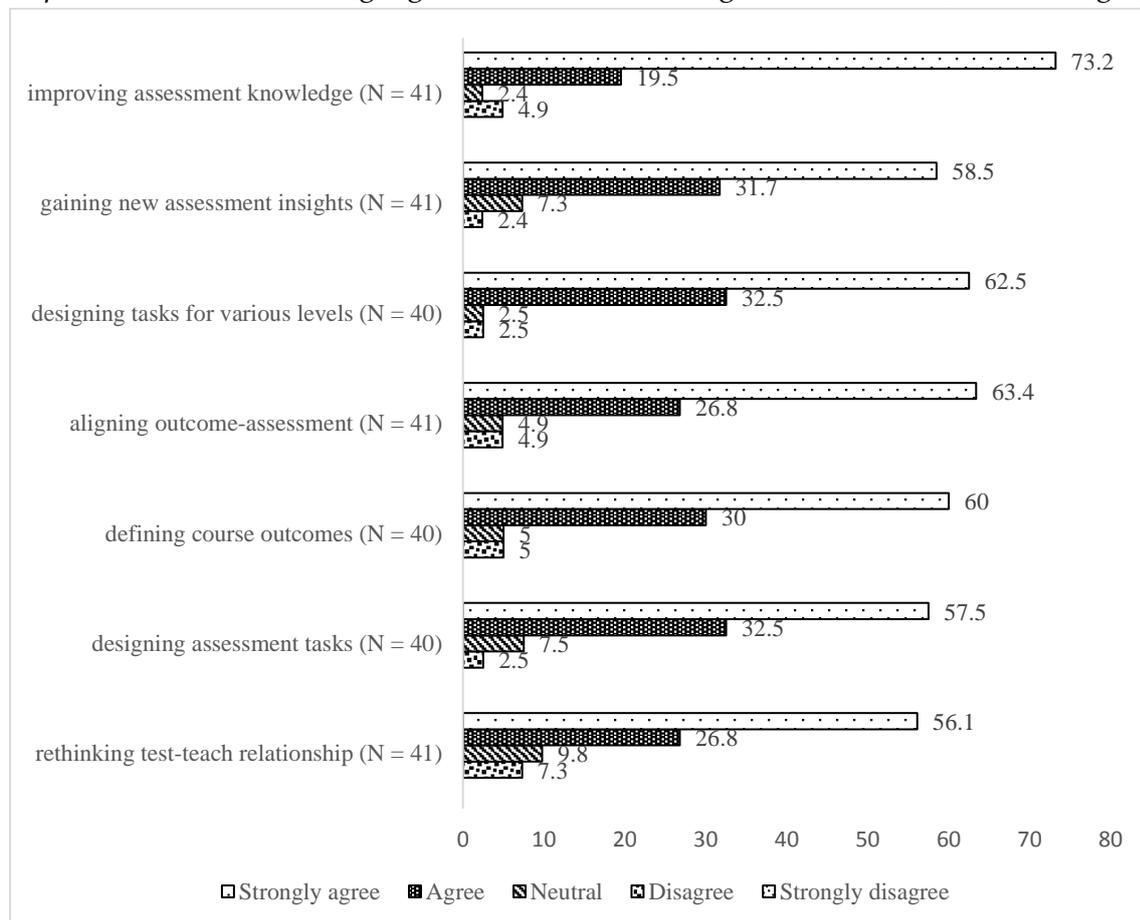
**Table 1**

*Descriptive Statistics for Questionnaire Items on Language Assessment Knowledge and Practices*

Sub-Category	Item	N	M [95% CI]	SD	SE
Language Assessment Fundamentals	improving assessment knowledge	41	4.61 [4.37, 4.85]	.77	.12
	gaining new assessment insights	41	4.46 [4.23, 4.70]	.75	.12
	designing tasks for various levels	40	4.55 [4.33, 4.77]	.68	.12
Outcome-Assessment Alignment	aligning outcome-assessment	41	4.49 [4.23, 4.74]	.81	.13
	defining course outcomes	40	4.45 [4.19, 4.71]	.82	.13
Task Design	designing assessment tasks	40	4.45 [4.21, 4.69]	.75	.12
Test-Teach Relationship	rethinking test-teach relationship	41	4.32 [4.02, 4.61]	.93	.15

**Figure 2**

*Impact on Instructors' Language Assessment Knowledge and Practices (in Percentages)*



These findings were supported by those from the qualitative analysis. Language Assessment Fundamentals received 38 comments from 13 instructors, who confirmed that the program had helped them learn about language assessment essentials, such as the ACTFL Guidelines, differences between test types (e.g., achievement vs.

diagnostic), item design, authenticity, reliability, and technology in assessment. For example, discussing authenticity, one instructor said:

*My assessment was very academic. Now, I think I'm taking my understanding of the assessment to the real world. I feel now that I have to know what the student should know for real life and how can I assess the student to place him or her on the right track. (Instructor 4, LCTL, interview, veteran, no prior LAL)*

It is evident that for this instructor, the program broadened their understanding about using authentic, real-world tasks.

Seventeen instructors also reported the program's impact on aligning their course target outcomes and assessment, with 36 comments on Outcome-Assessment Alignment. For instance, one instructor reported that participating in the program made them aware of the importance of defining course target outcomes, as seen below:

*What I've learned from you makes me really focus on the goals. I'm thinking about the levels where students should be, whether it's the ACTFL guidelines or I'm comparing them a lot with the European guidelines, something that I never considered before your workshops, ever. (Instructor 5, non-LCTL, interview, veteran, some prior LAL)*

Another instructor described how participation in the assessment workshops had helped them redefine their target outcome to attract more students:

*We initially had been teaching it with the goal of reaching Intermediate High, but then after taking the assessment workshop, we decided that to get everybody, we should do Intermediate Mid for the end of second year. And from there, we began to look at the descriptors for the four skills very closely and to align the "Can Do" statements accordingly. (Instructor 12, LCTL, interview, veteran, no prior LAL)*

Additionally, 14 instructors discussed the transformation of their perception of the testing-teaching relationship, with 47 comments. For instance, they recognized the importance of language assessment in teaching, as illustrated below:

*I think the workshop has completely revolutionized my understanding of assessment and how important it is in the teaching of the language. It is a must for every language teacher to understand how your assessment should be, because you are*

*developing a course in such a way that there should be a proper assessment in place.*

(Instructor 6, LCTL, interview, veteran, no prior LAL)

Task Design received the most comments, 49, from 16 instructors, who reported that the training program had transformed test task design. Instructors also described how test tasks designed prior to the training had been more grammar-oriented, while post-training tasks became more proficiency-oriented and performance-based, as explained below by Instructor 9, describing how their current final exams were influenced by the Oral Proficiency Interview (OPI) training:

*I have an oral proficiency interview that I do three times per quarter in the second year, and we do it four times per quarter in the first year, which is heavily based on my training with your OPI and stuff like that.* (Instructor 9, LCTL, interview, mid-career, some prior LAL)

Finally, Grading and Feedback, a new sub-category emerging from the interview data, received 29 comments from 10 instructors. They reported that the program had changed the way they graded or provided feedback on students' performance, as expressed by Instructor 10:

*After the workshop, we got the idea. Sometimes they miss a few words, they miss some punctuation, but you can still get it. So, I've learned how to tolerate mistakes.*  
(Instructor 10, LCTL, interview, veteran, no prior LAL)

The above comment demonstrates that for this instructor, accuracy was no longer the only essential aspect of students' performance and their adoption of other criteria when grading or responding to students' work.

Overall, participants reported that the program had improved their language assessment knowledge and practices. Positive impacts included broader knowledge about language assessment fundamentals, better alignment between teaching outcomes and assessment, deeper understanding of the testing-teaching relationship, and better task design and grading/feedback practices.

### *Impact on Instructors' Instructional Practices*

Table 2 and Figure 3 present results from the analysis of the questionnaire responses on Instructional Practices. Overall, the training program positively influenced the instructors' teaching practices, with mean scores above 4.2 across items.

The strongest agreement was found in the Teaching Methods and Strategies category, where mean scores ranged from 4.46 to 4.5. Between 23–31% of the instructors selected *agree* and over 60% selected *strongly agree* for items such as reflecting on teaching methods (23.1 % agreed, 64.1% strongly agreed), gaining new instructional insights (23.1 % agreed, 65.4% strongly agreed), and implementing new teaching approaches (30.8% agreed, 61.5% strongly agreed).

Materials and Activities also received high ratings, with mean scores of 4.21 (95% CI [3.89, 4.52]) and 4.54 (95% CI [4.23, 4.85]). In particular, “integrating authentic skills” drew strong support, with 26.9% agreeing and 65.4% strongly agreeing. By contrast, Program-Level Change received comparatively lower support. Item means fell between 4.20 and 4.35, with 27.5–30.8% agreement and 50–57.7% strong agreement.

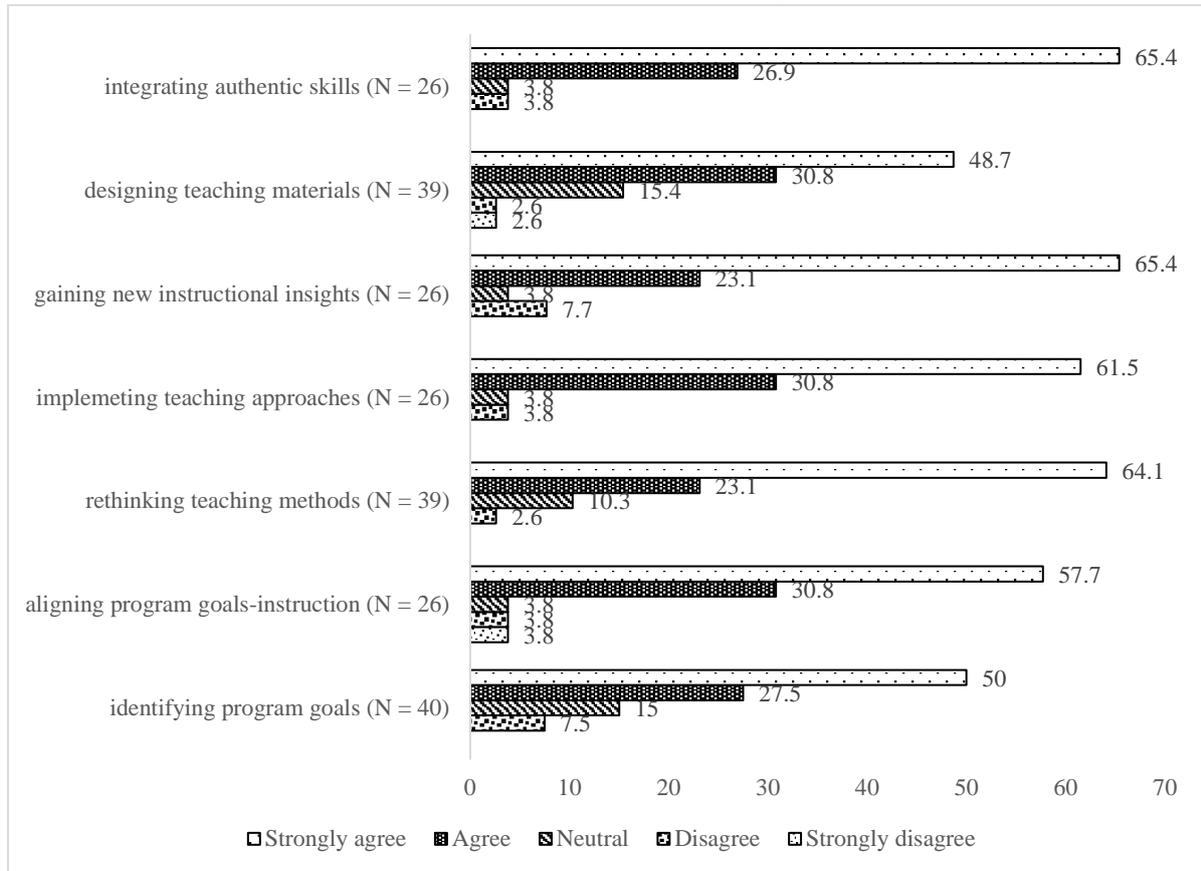
One interesting note is that the mean scores for individual items in this sub-category were lower than those for defining course outcomes and aligning outcome-assessment in Outcome-Assessment Alignment (RQ1), which might indicate that implementing changes on a macro (i.e., program) level might be more challenging than to one’s own course.

**Table 2**

*Descriptive Statistics for Questionnaire Items on Instructional Practices*

Sub-Category	Item	N	M [95% CI]	SD	SE
Materials and Activities	integrating authentic skills	26	4.54 [4.23, 4.85]	.76	.15
	designing teaching materials	39	4.21 [3.89, 4.52]	.98	.16
Teaching Methods and Strategies	gaining new instructional insights	26	4.46 [4.10, 4.83]	.90	.18
	implementing teaching approaches	26	4.50 [4.19, 4.81]	.76	.15
	rethinking teaching methods	39	4.49 [4.23, 4.74]	.79	.13
Program-Level Change	aligning program goals-instruction	26	4.35 [3.94, 4.76]	1.02	.20
	identifying program goals	40	4.20 [3.89, 4.51]	.97	.15

**Figure 3**  
*Impact on Instructors' Instructional Practices (in Percentages)*



The findings from the qualitative analysis corroborated the above results. Materials and Activities was discussed by 17 instructors, resulting in a total of 44 comments, the highest number of comments in this category. They said the training encouraged them to “rethink, reorganize new activities” and “reorganize and find new materials,” as reported by Instructor 8. They also explained that the program had “radically changed” how they selected teaching materials that were authentic and practical for students:

*In the past, it was simple like, “How difficult is this literary text”? as opposed to some of the practical, real-world scenarios that we now involve our students with or try to recreate in our courses in the intermediate and advanced levels. (Instructor 11, non-LCTL, interview, novice, no prior LAL)*

The instructors reported that they had learned to integrate various types of activities, adopting more group work or projects and more tasks developing students’ speaking skills, as shown below:

*And now, I also add time for them to have group work. I encourage all of them to speak. And because they have the group work opportunity, everyone has to talk a lot more than before. (Instructor 10, LCTL, interview, veteran, no prior LAL)*

Fifteen instructors reported on program's effects on Teaching Methods and Strategies, with 35 comments. Instructor 6 remarked:

*I think it has helped me transform myself, not as a traditional language teacher, but more about a modern language teacher going with time, using all kinds of technology, and trying to make a point about how the language should be taught effectively. (Instructor 6, LCTL, interview, veteran, no prior LAL)*

For this instructor, participation in the program transformed teaching methods, facilitating use of technology for instruction. The training had also reportedly transformed the structure of the instructors' grammar lessons, as Instructor 7 reflected:

*We also learned that spending in-class time on grammar explanation is a complete waste of time. There must be a purpose, and grammar cannot be a goal. (Instructor 7, non-LCTL, interview, veteran, no prior LAL)*

The least discussed sub-category was Program-Level Change, with 25 comments from 13 instructors. They commented that the training had helped their language program become "more focused, more organized, and more aware of its objectives" (Instructor 4, LCTL, interview, veteran, no prior LAL). The instructors also discussed the realignment of their language program goals because of the training, as seen below:

*It [The program] has changed dramatically. In the second year, out of the four classes, two are completely redesigned already. And in the first year, I'm getting both instructors to take every possible opportunity because I want to redesign the first years. (Instructor 2, LCTL, interview, veteran, no prior LAL)*

Overall, our analyses indicated that our model positively impacted instructors' teaching practices. They reported changes in teaching methods and strategies and in teaching material/activity design. More importantly, they agreed that the program had changed their language programs, allowing for goal identification and goal-instruction (re)alignment. However, there was also evidence that for the participants, implementing changes at the program-level might be more challenging than making changes at a micro, course-level.

*Impact on Instructors' Professional Identity*

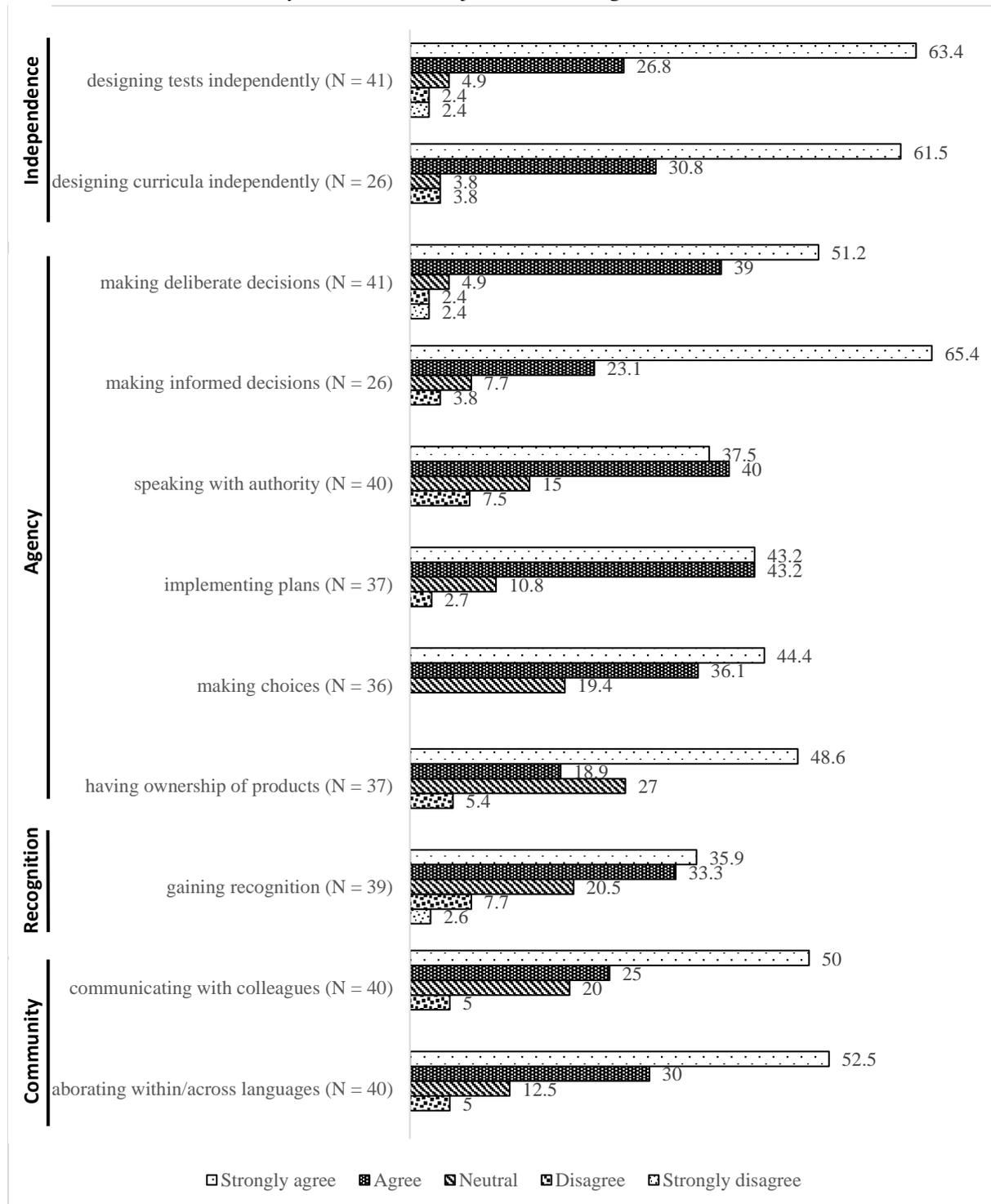
Results from the analysis of the questionnaire responses on Professional Identity are summarized in Table 3 and Figure 4. Our analyses revealed that the program had positive effects on the instructors' professional identity, as evidenced by the high mean scores of around 4.0. Independence received the strongest endorsement from instructors, who agreed that the program enhanced their independence in test design ( $M = 4.46$ , 95% CI [4.18, 4.75]) and curriculum design ( $M = 4.50$ , 95% CI [4.19, 4.81]). Agreement levels were consistently high, with 26.8–30.8% of instructors selecting *agree* and 61.5–63.4% selecting *strongly agree*. Agency also received very positive feedback, with mean scores ranging from 4.08 to 4.50; here, 18.9–43.2% of instructors selected *agree* and 37.5–65.4% selected *strongly agree*. Community showed a similar pattern, with mean scores of 4.20 – 4.30, and more than 25% *agree* responses alongside over 50% *strongly agree*. By contrast, Recognition experienced the lowest endorsement ( $M = 3.92$ , 95% CI [3.58, 4.27]), though disagreement remained minimal, with only 10.3% of instructors selecting *disagree* or *strongly disagree*.

**Table 3**

*Descriptive Statistics for Questionnaire Items on Professional Identity*

Sub-Category	Item	N	M [95% CI]	SD	SE
Independence	designing tests independently	41	4.46 [4.18, 4.75]	.90	.14
	designing curricula independently	26	4.50 [4.19, 4.81]	.76	.15
Agency	making deliberate decisions	41	4.34 [4.06, 4.62]	.88	.14
	making informed decisions	26	4.50 [4.17, 4.83]	.81	.16
	speaking with authority	40	4.08 [3.78, 4.37]	.92	.14
	implementing plans	37	4.27 [4.01, 4.53]	.77	.13
	making choices	36	4.25 [3.99, 4.51]	.77	.13
	having ownership of products	37	4.11 [3.78, 4.44]	.99	.16
Recognition	gaining recognition	39	3.92 [3.58, 4.27]	1.06	.17
Community	communicating with colleagues	40	4.20 [3.90, 4.50]	.94	.15
	collaborating within/across languages	40	4.30 [4.02, 4.58]	.88	.14

**Figure 4**  
*Impact on Instructors' Professional Identity (in Percentage)*



These findings were substantiated by those from the qualitative analysis. In terms of Independence, four instructors, in five comments, explained that they had more freedom to design assessments and courses. Instructor 11 stated:

*It [Participating in the program] also made me a much more independent instructor. Now, I feel much more comfortable designing my own level-appropriate materials, which is obviously necessary for my job. (Instructor 11, non-LCTL, interview, novice, no prior LAL)*

Nine comments from five instructors, when discussing the program's positive impact on their agency, explained that the program had allowed them to make decisions, implement instructional plans, make choices, speak with authority, and feel a sense of ownership of the products created during their participation. For example, Instructor 5 reported being empowered to counter authority imposed by textbook writers or program directors:

*I guess it empowers me, and I feel like I have something valid to contribute. Again, it's this power against the authority, because the authority can be the textbook, or it can be the program director. (Instructor 5, non-LCTL, interview, veteran, no prior LAL)*

However, we also found three negative comments from two instructors about the feasibility of implementing their innovations due to their program administrators as gatekeepers. For example, Instructor 3 described her frustration with the mismatch between the training and her program:

*There is a big disconnect between the top-notch professional development the CLC offers and the everyday practice of some language programs. [...] We are often hampered in any attempts to apply what we've learned when we return to coordinated programs that have dated pedagogical and assessment practices or standards. [...] It makes teaching multi-section coordinated classes within my language program much more frustrating because we are actively hindered from adapting curricular innovations. (Instructor 3, non-LCTL, questionnaire, veteran, some prior LAL)*

Thus, it is obvious that this instructor's plans for implementing changes had been obstructed by a program with "dated" teaching and assessment practices. Although the comments were not directed at the program, one must recognize that when implementing change, instructors might encounter resistance from other stakeholders. It should be noted that this instructor is from a commonly taught language program with multiple instructors. Thus, it is likely that implementing changes is more challenging in a bigger program than in smaller ones.

Instructors also reported that they felt their contributions had been recognized, as stated by five instructors, with eight comments. The instructors felt they had become a more valid member of their professional community, as reported by Instructor 7:

*Before that we were taken less seriously maybe and no one knew what we were doing as language instructors, and there was little bit looked down upon even. It was not considered important. Now, I think we are more appreciated. (Instructor 7, non-LCTL, interview, veteran, no prior LAL)*

An important topic discussed by 12 instructors was Community, with 23 comments. In fact, they remarked that participation in the program had provided them an opportunity to communicate with colleagues and fostered collaboration within and across language programs. For example, Instructor 9 revealed they became more involved in a professional community when participating in the program:

*One of the most positive things for me is just to be involved in a professional community with other people who are working on the same thing, to sort of get an expanded perspective on things, especially people from different languages, not just the language that I teach. So, it has shown me that there's a cohort of people who are doing the same thing, that I can work with, that I can present to, that I can listen to. (Instructor 9, LCTL, interview, mid-career, some prior LAL)*

Two new categories that emerged from the interview data were Confidence, receiving the most comments, 36 from 15 instructors, and Motivation, with 10 comments from six instructors. Eight instructors who participated in the interview expressed that they felt more confident about their language assessment knowledge. Three also reported that they had gained more confidence in their teaching. Additionally, they confirmed that they felt more motivated to teach as a result of their participation, as illustrated below:

*It's really gotten me excited about stuff and new stuff for the first time in a long time. And so, I would love to keep doing more of it because it has really lit a fire under me and been very exciting. (Instructor 3, non-LCTL, interview, veteran, some prior LAL)*

Despite some instructors' negative comments about programmatic support, we found evidence that the program contributed to the development of instructors' professional identities. They became more independent, developed a sense of agency, participated in professional communities, gained recognition for their work, and became more

confident in their language assessment knowledge and more motivated about teaching.

### **Discussion**

The findings from the quantitative and qualitative analyses indicate that our training program had multiple positive impacts on participants. It significantly improved instructors' language assessment knowledge and practices in terms of language assessment fundamentals, task design, grading and rubric design, and the importance of assessment to instructional practices. For the first time, many instructors recognized that assessment was not just a routine curricular component. Instead, assessment was seen as essential for knowing whether their curriculum was working the way they thought it was. Additionally, because the participants had both defined the target outcomes for their courses and designed the tests to measure student performance of those outcomes, they readily began to reconsider their teaching effectiveness based on evidence from their assessments. These considerations became the impetus to implement new teaching practices and strategies. Participants also gained the ability to consider their pedagogy at both the micro/course level and macro/program level and to perceive the limitations of their ability to instigate change. Perhaps most significantly, the training allowed instructors to become independent, confident, and motivated in both testing and teaching practices, echoing previous findings (Giraldo & Yan, 2025). This engendered a sense of agency, empowering them to make their own decisions, implement changes based on evidence, and begin to anticipate the effects of those changes. Finally, it helped instructors build a community of practice with others within and beyond their language program.

The study has important implications for instructors' PD, especially situating LAL as an integral component. Our PD model could be replicated in other institutions with similar contextual factors. The LAL model has been adopted for another larger, externally funded LCTL project across multiple institutions to create cohesive, multi-year curricula, aligned with nationally recognized proficiency levels and assessments and be used for courses available to students across several campuses. Also, the resulting community of practice has spanned many universities in the US as participants shared their work with colleagues from other institutions during conferences. In smaller institutions with limited or no funding, language instructors still constantly revise their curricula and create new materials. Providing instructors with an outcome-focused PD initiative helps to attract them to and keep them engaged in PD programming. In the absence of an in-house language assessment specialist, it may be prudent to initially hire an external expert to conduct the training. Individuals, once adequately trained, could subsequently assume the responsibility of training additional personnel within the program.

This study provides evidence indicating that utilizing an assessment-driven backward design model encouraged language instructors to develop an agency and reverse-engineer effective instruction. Agency is essential because it empowers the instructors to feel both responsible and confident about their ability to make deliberate changes in assessment practices, curriculum development, instructional approaches, and language program design, and to assume leadership roles in language teaching communities. Aside from an overwhelmingly positive impact on the instructors themselves, all of the above contribute to a more effective and engaging language learning environment for students and promote positive outcomes in language education.

We also argue that to effectively embed LAL training in PD around curriculum design and pedagogy, one must go beyond one-off workshops and establish systematic, scaffolded PD programs that are effective and sustainable and result in tangible, deployable deliverables. Our training model focused on both theoretical and practical aspects of language assessment and curricular design, allowing participants to put assessment into practice by first developing and then administering tests, and then considering curricula. This model provided participants with opportunities to apply theoretical concepts and principles to hands-on practice, consequently allowing for the introduction of theory in an embedded and connected manner. We echo recommendations from previous researchers (Brindley, 2001; Brunfaut & Harding, 2018; Fulcher, 2012) that LAL training should include phased scaffolding over a sufficient period of time, building toward greater independence of instructors and longer-term project sustainability. Following Vogt et al. (2020), we also believe that language instructors benefit from collaboration with fellow participants, especially when finalizing test development projects.

The findings also revealed the tension experienced by participants wishing to make curricular changes based on newly gained LAL knowledge and encountering barriers due to a lack of resources or support from language program directors or higher administration. Implementing change is likely more difficult in larger programs than in smaller ones. In fact, language programs do not sufficiently recognize instructors' contributions. This could be acknowledged formally by offering internal awards or invitations to give talks, or informally by simply commenting positively during meetings, creating a positive institutional environment where everyone is a valued member of a community. On the training side, programs with an LAL component resulting in participants' increased agency and independence should consider the educational culture where changes are to be implemented (Coombe et al., 2020; Vogt et al., 2020; Yan et al., 2018). Facilitators can survey participants' current assessment knowledge and practices, help them understand possible mediating effects of local,

institutional policies and mandates on their language assessment practice and interests, and explore available resources and constraints as they guide participants through training process. LAL training courses should prepare trainees for managing change in their own contexts. Due to diverse departmental structures and institutional cultures, a dedicated session in which participants describe their roles and the structure or culture of their department or institution may anticipate obstacles to implementing changes. While some instructors may not be allowed to alter the curriculum or administer their developed proficiency assessments, others may have considerable autonomy. Instructors across this spectrum can exchange ideas on micro-changes, discuss conversations with chairs or deans, and frame changes or assessments as classroom-based research. An open discussion may foster an exchange of ideas, strategies, and support mechanisms.

It should be noted that a limitation of the study was the potential bias caused by our and the interview participants' roles in the training program. Since we were the program facilitators as well as researchers, we are cognizant that those who volunteered for the interview were likely to have positive attitudes towards the program. Although the questionnaire responses were anonymous, the interview data may still reflect this self-selection bias. Beyond this issue, several other limitations should be acknowledged. First, the study was conducted within a single institutional context, which may limit the generalizability of the findings to other universities or program settings. Second, because the research team served as facilitators as well as investigators, there is the potential for bias in data collection, interpretation, and reporting. Third, the study relied heavily on self-reported data from instructors, which, while valuable for capturing perceptions, may not fully reflect actual changes in practice. Finally, the study did not incorporate classroom observations, teaching artifacts, or direct measures of student outcomes, which could have provided triangulation and stronger evidence of the program's impact.

### **Conclusion**

This study aimed to examine the impacts of a proficiency-oriented, assessment-driven, backward design model on language instructors' assessment and teaching practices and on professional identity. Our findings show that the program improved participants' language assessment knowledge and practices, facilitated changes in their teaching methods and strategies to better align their teaching outcomes with assessment, and encouraged them to exercise agency and develop professional identity. A similar model might be applicable to other institutions with similar contexts and parameters, and that a model that combines assessment with instruction and marries language assessment theory and practice is effective and sustainable. Future research on effects of such PD programs could also explore other evidence or

factors. These include test and syllabi artifacts from participants before and after training, classroom observations, effects on student performance, lack of leadership literacy among language program directors, and institutional constraints.

### ORCID

 <https://orcid.org/0000-0002-6799-6397>

 <https://orcid.org/0000-0003-1576-4743>

 <https://orcid.org/0009-0003-0865-8325>

### Acknowledgements

Not applicable.

### Funding

Not applicable.

### Ethics Declarations

### Competing Interests

No, there are no conflicting interests.

### Rights and Permissions

### Open Access

This article is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which grants permission to use, share, adapt, distribute and reproduce in any medium or format provided that proper credit is given to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if any changes were made.

[www.EUROKD.com](http://www.EUROKD.com)

### References

- Abrar-ul-Hassan, S. (2024). Rescoping language assessment literacy: An expanded conceptualization. *Language Testing*, 41(2), 205–226. <https://doi.org/10.1016/j.lantest.2023.101746>
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/0265532217716732>
- Baumann, C. C., Dursun, A., & Nguyen, P. (2023). Building a sustainable LCTL community of practice through assessment-driven reverse design. In E. Heidrich-Uebel, A. Kraemer, & L. Giupponi (Eds.), *Sharing LCTLs in higher education: Collaboration and innovation* (pp. 204–217). Routledge.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. <https://doi.org/10.3102/0013189X033008003>
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 11–15). Iowa State University.

- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.
- 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). Springer. [https://doi.org/10.1007/978-3-319-77177-9\\_9](https://doi.org/10.1007/978-3-319-77177-9_9)
- Coombe, C., Vafadar, H. & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia* 10(3). <https://doi.org/10.1186/s40468-020-00101-6>
- Creswell, J. W., & V. L. Plano Clark. (2017). *Designing and conducting mixed methods research* (3<sup>rd</sup> edition). Sage Publications
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2<sup>nd</sup> ed.). Routledge. <https://doi.org/10.4324/9780203864739>
- Dursun, A., Nguyen, P., Baumann, C. (2022, March 7–11). *Understanding language assessment literacy and training needs of language instructors: A first step in establishing a professional development program* [Paper]. Language Testing Research Colloquium (LTRC), Online.
- Fulcher, G. (2012) Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Giraldo, F., & Yan, X. (2025). Evaluating an online assessment course: Teachers' voices on their language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 27(1), 97–114. <https://doi.org/10.15446/profile.v27n1.114808>
- Guskey, T.R., & Huberman, M. (1995). *Professional development in education*. Teachers College Press.
- Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307. <https://doi.org/10.1177/0265532213480126>
- Kamali, J., Freeman, D., Larsen-Freeman, D., Norton, B., & Farrell, T. S. C. (2024). Editorial: Language teacher education research – key trends, challenges, and questions. *Language Teacher Education Research*, 1, 1–16. <https://doi.org/10.32038/iter.2024.01.01>
- Krippendorff, K. (2022). *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications. <https://doi.org/10.4135/9781071878781>
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/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. <https://doi.org/10.1080/15434303.2019.1692347>
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30 (3), 329–344. <https://doi.org/10.1177/0265532213480129>

- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *National Council of Less Commonly Taught Languages*, 13. <https://www.ncolctl.org/files/LCTL-Teachers-Assessment.pdf>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Priestley, M., Biesta, G.J.J. & Robinson, S. (2015). Teacher agency: What is it and why does it matter? In R. Kneyber & J. Evers (eds.), *Flip the system: Changing education from the bottom up*. Routledge. <https://doi.org/10.4324/9781315678573-15>
- Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511667237>
- Riazi, A. M., & Candlin, C. N. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, 47(2), 135–173. <https://doi.org/10.1017/S0261444813000505>
- Shin, J. K., & Kim, W. (2021). Perceived impact of short-term professional development for foreign language teachers of adults: A case study. *Foreign Language Annals*, 54, 365–388. <https://doi.org/10.1111/flan.12521>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Vogt, K., Tzagari, D., & Spanoudis, G. (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. <https://doi.org/10.1080/15434303.2020.1781128>
- Wang, S. C. (2009). Preparing and supporting teachers of less commonly taught languages. *The Modern Language Journal*, 93(2), 282–287. <https://www.jstor.org/stable/40264062>
- Wenger, E. (1998). *Communities of practice: Learning, meaning and identities*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803932>
- Wiggins, G., & McTighe, J. (2013). *The understanding by design guide to creating high-quality units* (2<sup>nd</sup> ed.). Pearson Teacher Education.
- 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. <https://doi.org/10.1016/j.system.2018.03.003>

**Appendix**

*Professional Development Program Impact Questionnaire*

1. In which aspect(s) of the program have you participated? Check all that apply.
 

<input type="checkbox"/> OPI Training	<input type="checkbox"/> Language Assessment Workshop
<input type="checkbox"/> Test Design and Development	<input type="checkbox"/> Test Development Project
<input type="checkbox"/> Curriculum Design Workshop	<input type="checkbox"/> Curriculum Design Project
  2. How effective do you think the process has been?
 

1 = Very ineffective	2 = Somewhat ineffective
3 = Neither effective nor ineffective	4 = Somewhat effective
5 = Very effective	
  3. Indicate the extent to which you agree with the following statement:  
Participating in the program has contributed to my professional development.
 

1 = *strong disagree*; 2 = *disagree*; 3 = *neutral*; 4 = *agree*; 5 = *strongly agree*
  4. Indicate the extent to which you agree with the following statements.  
1 = *strong disagree*; 2 = *disagree*; 3 = *neutral*; 4 = *agree*; 5 = *strongly agree*
- Participating in the program has...
- |    |  |
|----|--|
| 1  | improved my knowledge of language assessment.  |
| 2  | helped me gain new insights that I can readily implement in my assessment practices.                     |
| 3  | helped me learn how to design proficiency-oriented test tasks for various levels.                        |
| 4  | helped me learn how to design tests that align with the intended outcomes.                               |
| 5  | helped me identify/define outcomes for my course(s).   |
| 6  | changed how I design and/or select tasks for my assessment activities.                                   |
| 7  | shifted my perception of the relationship between testing and teaching.                                  |
| 8  | helped me integrate the learning of specific skills with their real-world use.                           |
| 9  | changed how I design and/or select teaching materials.   |
| 10 | helped me gain new insights that I can readily implement in my instructional practices.                  |
| 11 | allowed me to implement a range of approaches to use in my teaching.                                     |
| 12 | helped me rethink my teaching methods.   |
| 13 | helped my language program align its instruction with its goals.   |
| 14 | helped my language program identify meaningful and realistic goals.                                      |
| 15 | prepared me to design proficiency-oriented assessments independently.                                    |
| 16 | prepared me to design proficiency-oriented curricula independently.                                      |
| 17 | allowed me to make deliberate decisions about my instructional practices.                                |
| 18 | allowed me to make informed decisions about my instructional practices.                                  |
| 19 | helped me speak professionally and authoritatively about my work.  |
| 20 | allowed me to implement what I had envisioned for the project while I was working on my project.         |
| 21 | allowed me to make choices about how to move forward with the project while I was working on my project. |
| 22 | made feel that I have ownership of the test/curriculum that I developed.                                 |
| 23 | helped me gain recognition for my work.  |
| 24 | helped me more easily communicate with other language instructors about instructional practices.         |
| 25 | fostered collaboration within/across languages around the work we are doing.                             |
5. How has your participation in the program impacted your instructional practices?
  6. Additional comments.