Article

Multiple Stakeholder Interaction to Enhance Preservice Teachers' Language Assessment Literacy

Jiyoon Lee [1,*] **, Yuko Goto Butler** [2] **and Xiaolin Peng** [3]

[1] Department of Education, University of Maryland Baltimore County, Baltimore, MD 21250, USA
[2] Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104, USA; ybutler@upenn.edu
[3] Lauder Institute, University of Pennsylvania, Philadelphia, PA 19104, USA; xiaolin@wharton.upenn.edu
[*] Correspondence: jiyoon@umbc.edu

Abstract: Conducted in a U.S. English-to-Speakers-of-Other-Languages (ESOL) preservice teacher education program, this case study aimed to explore a dynamic process of preservice teachers' development of language assessment literacy (LAL). By inviting multiple stakeholders, namely preservice teachers, an inservice teacher and her ESOL students, and their course instructor, this study closely examined the interaction among the stakeholders during a semester-long language assessment development project as a process to develop LAL. The project, which was composed of planning, development, implementation, and reflection stages, was innovative in that it: (a) involved the multiple stakeholders; (b) focused on their dynamic interactions and multi-directional influences on all the participants' enhancement of LAL; and (c) was conducted in an online format. By employing thematic analyses on interactions among the stakeholders, this study described and analyzed how preservice teachers contextualize their assessment while negotiating the needs of the inservice teachers and their students with assistance from the course instructor. The paper identified practical benefits and challenges of professional training where multiple stakeholders are involved. It also highlighted the non-linear dynamic process of preservice teachers' development of LAL.

Keywords: language assessment literacy; teacher education; preservice teachers; inservice teachers; ESOL; multiple stakeholders; interaction

1. Introduction

This case study explores the benefits of multi-directional interaction in developing English-to-Speakers-of-Other-Languages (ESOL) preservice teachers' language assessment literacy (LAL). The multifaceted concept of LAL refers to the knowledge of language assessment principles, skills of developing and using language assessment, and understanding of the consequences of language assessment (Davies 2008; Fulcher 2012; Inbar-Lourie 2008; Lee and Butler 2020). The increased importance of language assessment in education and society as well as teachers' key roles in assessment has motivated extensive attention to be paid to teachers in LAL studies. A traditional way of enhancing teachers' LAL focused on providing knowledge of language assessment from experts (i.e., testing professionals) to novices (i.e., teachers). The concern regarding this approach is the limited possibility of incorporating teachers' voices in the development of their LAL (Inbar-Lourie 2017). More reciprocal and reflective ways would allow all stakeholders to share their strengths and experience in teacher education to enhance their LAL (Yan and Fan 2021).

Inquiring into more inclusive ways to enhance preservice teachers' LAL, we designed a semester-long language assessment development intervention project that involved multiple stakeholders as key informants. The stakeholders included not only a course instructor and preservice teachers, but also an inservice ESOL teacher and her middle-school ESOL students. Composed of planning, development, implementation, and reflection, this language assessment development project conducted in a U.S. public school required

preservice teachers to contextualize their assessment design and implementation while negotiating the needs of the inservice teacher and her students with the help from the course instructor. This study examined all involved stakeholders' LAL enhancement through multiple data sources of their reflections captured in each stage of the language assessment development project and reflection papers and email correspondence.

With the aim of reflecting the diverse pedagogical and assessment needs of multiple key players in teacher education, this teacher education project was innovative in that it: (a) involved not only preservice teachers and a course instructor, but also an experienced inservice teacher and her ESOL students; (b) focused on their dynamic interactions and multi-directional influences on all the participants' enhancement of LAL; and (c) was conducted in an online format. The paper describes practical benefits and challenges of professional training where multiple stakeholders are involved.

2. Literature Review

LAL is a subject-specific assessment literacy that was first conceptualized by Brindley (2001) in designing language teacher education modules. He emphasized the importance of language teachers' competences in assessment with a special focus on its social roles and context-specific interpretation. His emphasis on teachers' understanding of assessments' roles was further developed in subsequent LAL models. The specific details of LAL vary in different models; however, most of the theoretical LAL models have three major components (Butler et al. 2021; Giraldo 2019): (1) assessment theories, principles, or knowledge; (2) skills to design, use, or interpret the assessment results; and (3) understanding of cultural, ethical, pedagogical, political, and social consequences of language assessment (Davies 2008; Fulcher 2012; Inbar-Lourie 2008; Lee 2019). In another approach, Taylor (2013) suggested eight attributes of LAL, namely "knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision-making" (p. 410). She stated that depending on the stakeholders' roles in language assessment, a relative emphasis on such attributes can be tailored for the LAL training; for instance, teacher LAL training should emphasize language pedagogy more than knowledge of theory, principles and concepts, or scores and decision making.

The multi-faceted LAL model reflects teachers' diverse LAL needs and the diverse assessment practices on which they spend extended time (Stiggins 2007). The assessment practices that teachers engage in range from state-mandated, high-stakes, teacher-made exams to formative assessments such as giving exit quizzes or frequent comprehension questions to their students. Language teachers' extensive involvement in language assessment is well documented in research studies as well. Vogt and Tsagari (2014) and Fulcher (2012) surveyed language teachers' LAL needs. Recruiting 863 and 273 inservice teachers, respectively, both studies revealed teachers' diverse and high LAL needs. A great number of language teachers who participated in the surveys reported that they had not received sufficient training on assessment; instead, they enhanced LAL while they were teaching. The diverse and high LAL needs are also shared by preservice teachers; it was revealed that teacher education programs do not provide sufficient opportunities for preservice teachers to enhance their LAL (Csépes 2014; Lam 2015). Reviewing a Hungarian teacher education program (Csépes 2014) and interviewing newly hired inservice teachers in Hong Kong (Lam 2015), Csépes and Lam, respectively, showed that preservice teachers need much more information to fully contextualize assessment principles into their own assessment practices.

Preservice teachers' needs of contextualized assessment practices are not sufficiently reflected in language teacher education. In a review of teacher education on language assessment, Davies (2008) presented that knowledge of psychometrics and statistics was highly emphasized. His review showed recent textbooks on language assessment included more information about assessment skills and understanding of the cultural, pedagogical, and social consequences. However, traditional approaches to teacher education on lan-

guage assessment provides limited experience for preservice teachers to be fully prepared for their own assessment practices.

Traditional approaches to teacher education that emphasize the transmission of knowledge from experts (i.e., teacher educators or testing professionals) to novices (i.e., preservice teachers) have left mixed results. A survey conducted by Volante and Fazio (2007) showed that even after taking an assessment course, preservice teachers expressed their lack of confidence and needs for a further understanding of assessment purposes and types. However, DeLuca and Klinger (2010) showed that those who took an assessment course felt more confident than those who did not. In addition, among the activities they worked on in the assessment course, the preservice teachers found discussions of assessment issues in the real world, assessment portfolios, and reflection papers particularly effective to become more confident in assessment practices.

Recent teacher education on LAL has emphasized more contextualized, reflective, and hands-on approaches. Scarino (2013) reported that preservice teachers who engaged in extensive reflections during a language assessment course were able to increase their awareness of assessment purposes and functions. She argued that LAL could be obtained through individualized experience that integrated stakeholders' experience and beliefs about assessment practices. Koh et al. (2018) observed that hands-on experience of language assessment critically helped teachers of Chinese language develop assessments that could effectively elicit language learners' performance. Engaging in a language assessment development project, preservice teachers reported increased awareness of language assessment purposes and functions and greater confidence in applying theoretical knowledge of assessment principles to develop a language assessment.

Documenting the interaction between teachers and language-testing professionals (i.e., workshop facilitators), Baker and Riches (2017) presented collaboration between stakeholders as an alternative approach to teacher education regarding language assessment. They reported that teachers developed a deeper understanding of the connections between teaching and assessment and their roles as supporters for their students' successful experience in assessment. Meanwhile, language-testing professionals gained an understanding of the specifics of the context and target learners. In a similar vein, interviewing three different stakeholder groups, Yan and Fan (2021) suggested an "apprenticeship-based, experience-mediated model of LAL development" (p. 238). They emphasized collaboration and reflection in assessment activities for enhancing stakeholders' LAL. They suggested that language teachers' LAL could be enhanced through contextualized collaborative practices among different stakeholders.

Based on the previous studies' suggestions, we designed an exploratory case study in which multiple stakeholders were engaged in a language assessment development project by sharing their experiences, beliefs, and expertise. The primary purposes of this study were: (1) to describe and understand interactions that happened among the multiple stakeholders during the project; and (2) to examine how the interactions assist mutual learning of LAL among all the participating stakeholders, not just the preservice teachers. Thus, our research questions are as follows:

(1) How do multiple stakeholders interact during a language assessment development project in teacher education?
(2) How do such interactions assist preservice teachers' LAL as well as their mutual learning among all the participating stakeholders?

3. Study

3.1. Participants

The participants of this study included four groups of stakeholders: (1) seven ESOL preservice teachers (hereafter preservice teachers); (2) one ESOL inservice teacher (hereafter inservice teacher); (3) nine middle-school ESOL students; and (4) one course instructor. The preservice teachers were in a Teaching English to Speakers of Other Languages (TESOL) master's program at a mid-sized public university in the Mid-Atlantic region of the United

States of America at the time of the data collection. ESOL preservice teachers are defined as those who are currently in a teacher education program to obtain a teaching credential or an advanced degree in language education (Lee 2019). The preservice teachers were varied in their educational experience and career goals. Among the seven preservice teachers, two of them (Becca and Mia) were pursuing ESOL certification in order to teach at a public school, two (Helen and Tim) were interested in teaching adult learners, two (Diana and Tiffany) had open options, and one (Yuri) was an international student who planned to teach in his home country upon completion of the degree. Because this professional development course was primarily designed for teachers who were relatively new to teaching, we call them "preservice teachers" in this paper; however, their actual teaching experience varied from none to approximately two decades. The preservice teachers were grouped to work with either seventh-grade students (Becca, Mia, Helen, and Tim) or eighth-grade students (Yuri, Diana, and Tiffany). Pseudonyms were used to refer to the preservice teachers.

The inservice teacher was educated in Russia and in the U.S. and had taught elementary and middle-school ESOL students in the United States since 2002. She is one of the leading teachers in the school district of the city where this study was conducted. In addition, the inservice teacher's students—four seventh-grade students and five eighth-grade students—participated in this study. Their first languages were either Bengali, Italian, Nepali, Spanish, Swahili, or Urdu. ESOL students in this state take the World-Class Instructional Design and Assessment (WIDA), a standardized proficiency test, and can exit the program once they reach level 5 (out of 6). At the time of the study, the average scores for the participating seventh and eighth-grade students were 2.2 and 3.2, respectively, meaning they were at the low–intermediate to intermediate levels by WIDA standards.

Finally, the course instructor was a teacher educator who holds a doctoral degree in educational linguistics specialized in language assessment. She has worked with diverse preservice teachers at multiple institutions in the United States. She has been teaching a graduate-level language assessment course in both online and face-to-face format since 2012.

3.2. Data

The data for this study are twofold: the stakeholders' interaction and preservice teachers' final papers. We relied on multiple data sources, including (1) preservice teachers' reflections throughout the project, (2) their responses to the course instructor and the inservice teacher's feedback, (3) their feedback to the ESOL students, (4) their final papers, and (5) all correspondences between the course instructor and the inservice teacher (i.e., virtual interviews and email exchanges). The data were collected through the virtual learning management system Blackboard and personal correspondence between the course instructor and the other stakeholders. Data triangulation achieved through the multiple data sources strengthened the validity of the arguments this study made.

3.3. Data Analysis

The researchers adopted deductive thematic analysis, which allowed the researchers to analyze the data based on the LAL theoretical models presented in the earlier section. The themes were derived from the three major components of LAL: (1) knowledge, (2) skills, and (3) principles/consequences. Following Nowell et al. (2017), the three researchers independently immersed themselves in the data first. After becoming familiar with the data, the authors discussed potential themes that emerged from the data and came up with potential codes. Approximately 20% of the data were initially selected, and the researchers coded the data independently. Then, they compared their coding for inter-coder reliability. They discussed the discrepancies in their coding until they reached 100% agreement. After 100% agreement was reached, the researchers independently coded the rest of the data. The researchers met three more times to compare and discuss their coding until the inter-coder reliability reached 100%.

As indicated in Figure 1, under the three components for LAL described in the literature section, eight themes were identified: construct, validity, reliability, format, practicality purposes, general reflection, and group work reflection, which were further specified with codes. The codes were as follows: construct of questions, rubric, inter-rater reliability, reducing subjectivity, assessment for learning, item appropriateness, and technology.
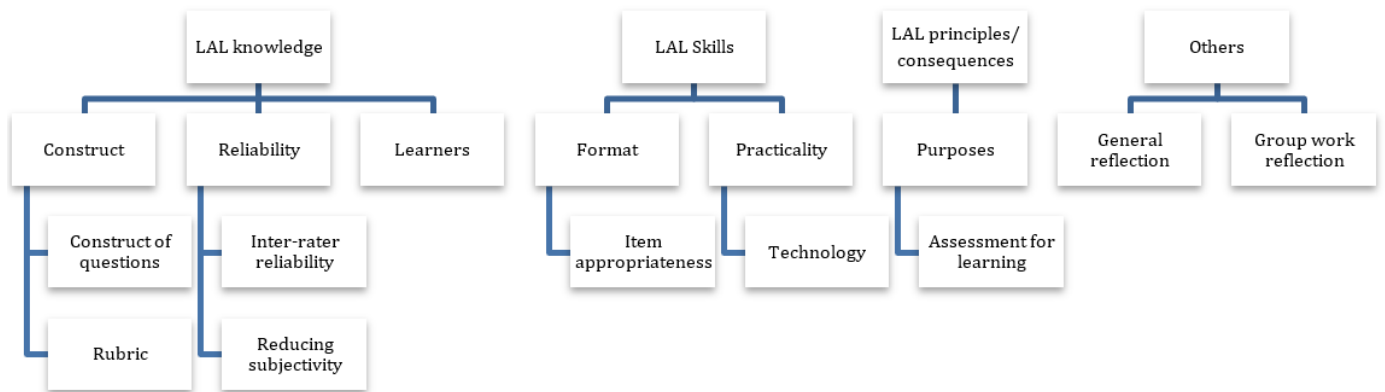

Figure 1. Themes and codes.

4. Results

4.1. Research Question 1

How do multiple stakeholders interact during a language assessment development project in teacher education?

The first research question guided us to examine the nature of the interaction among the stakeholders in the project. As noted in Figure 2, the interaction among the stakeholders in this project was highly interconnected and iterative. The dynamic interaction among the stakeholders was unique compared to other teacher education models that seek to enhance teachers' LAL. Active engagement among all stakeholders in the interaction was required to complete the project successfully.
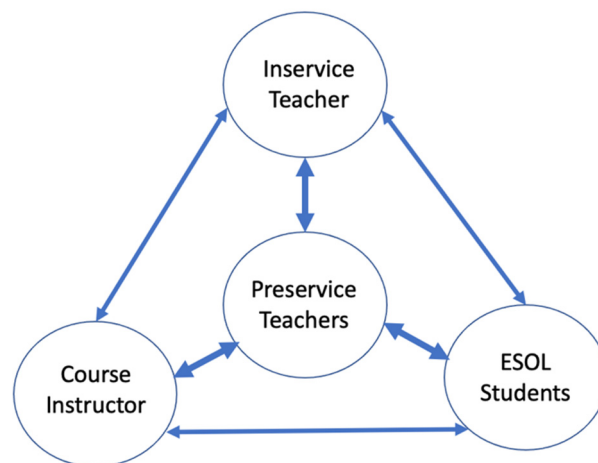

Figure 2. Interaction among the stakeholders.

The project had four stages: (1) planning; (2) development; (3) implementation; and (4) reflection (adopted from Bachman and Palmer 2002). Below, we describe each stage in detail along with the figures that schematize the stakeholders' interaction at each stage. In those figures (Figures 3, 4, 7, and 8), a solid line indicates direct interaction, and a dotted line refers to indirect involvement. The arrows indicate the direction of information flow.

4.1.1. Planning Stage

At the planning stage, as noted in Figure 3, the main interaction happened between (1) the preservice teachers and the course instructor, (2) the inservice teacher and the course instructor, and (3) the preservice teachers and the inservice teacher. The major activity at this stage for the preservice teachers was to obtain information about the ESOL students regarding their backgrounds and needs as well as the school curriculum through the help of the inservice teacher and the course instructor.
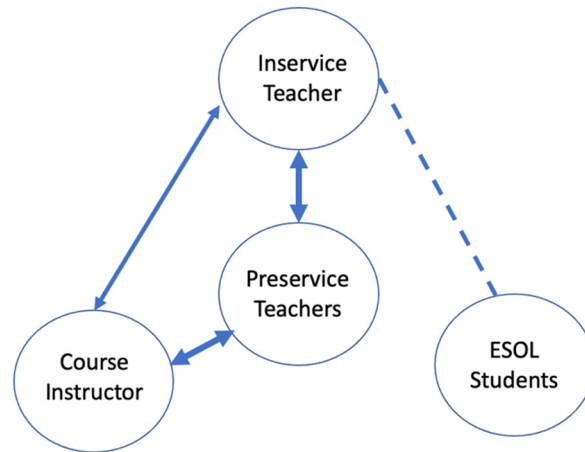

Figure 3. Planning stage.

More precisely, through a series of Skype meetings with the inservice teacher that were arranged initially by the course instructor, the preservice teachers first conducted interviews with the inservice teacher. At the interviews, the inservice teacher shared her instructional strategies and materials with the preservice teachers. Based on the information collected in the interviews, the preservice teachers wrote reports on the test-takers, indicating their WIDA scores, first-language background, educational experience, curriculum, as well as the purpose of their assessment.

Before the interviews were conducted, the course instructor and the inservice teacher spent substantial time setting up collaboration and preparing a few logistics. The logistics included selecting target ESOL student groups, setting up virtual interviews, implementing the assessment, and soliciting the inservice teacher's and ESOL students' feedback.

4.1.2. Development Stage

Figure 4 shows the interaction at the development stage. The interaction was focused between the preservice teachers and the course instructor.
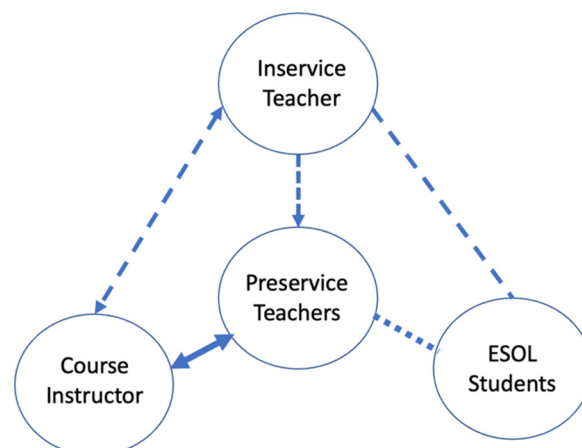

Figure 4. Development stage.

The preservice teachers read assigned papers and watched recorded course lectures in order to enhance their basic background knowledge about assessment development. They spent several weeks reviewing the inservice teacher's lessons, defining target skills (i.e., what aspect of listening and writing skills they would assess in their assessment), identifying constructs, writing assessment items, and developing a rubric for the performance portion of their assessment. The course instructor met with the individual groups multiple times and provided them with feedback on their work-in-progress. The preservice teachers first worked on developing a listening assessment, followed by a writing assessment.

The preservice teachers revised their construct definitions and assessment items based on the course instructor's initial feedback. With respect to the listening assessment, one group recorded their own script, and the other group adopted an audio file that was freely available online. As noted in Figure 5, the preservice teachers built their assessments using Google Forms. Google Forms is widely used in public schools in the state where this study was conducted; therefore, it was necessary for the preservice teachers to become familiar with the platform. Furthermore, the instantaneity of sharing results in the Google Forms made the interaction between the preservice teachers and the ESOL students effective. That is, once the ESOL students completed their writing assessment on Google Forms, the platform tallied the ESOL students' performance for review (Figure 6). The process was almost instant.
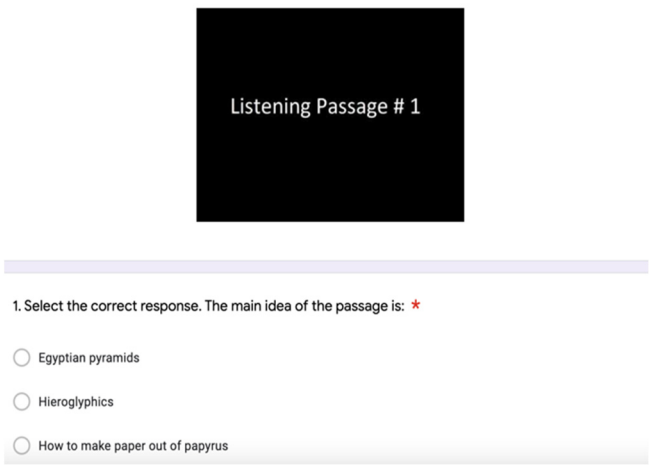

Figure 5. Assessment example. Note: * indicates that a test-taker's response is required.
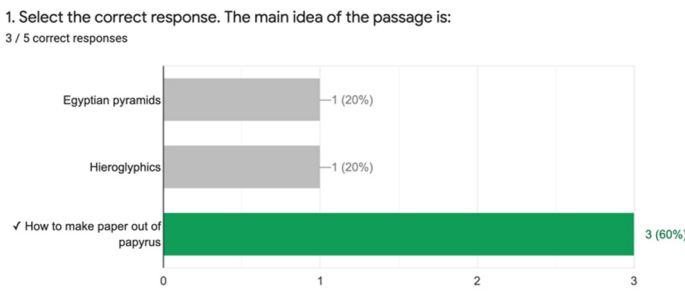

Figure 6. Performance example.

4.1.3. Implementation Stage

At the implementation stage, the interaction was led by the inservice teacher (Figure 7). The inservice teacher administered the listening and writing assessments to the ESOL students.
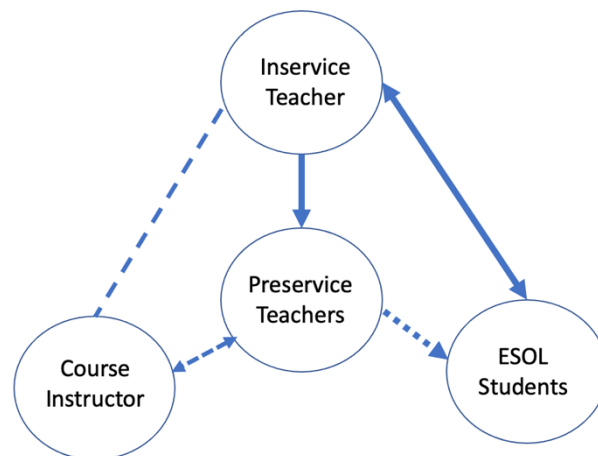
Figure 7. Implementation stage.

The inservice teacher provided detailed feedback on the listening assessment regarding the format and technical challenges, and she also shared her concerns about the assessment with the preservice teachers. A couple of weeks later, writing assessments were administered to the ESOL students via Google Forms. The inservice teacher implemented the writing assessment and provided her feedback on the writing assessment.

4.1.4. Reflection Stage

As noted in Figure 8, extensive interaction happened between (1) the preservice teachers and the inservice teacher, (2) the preservice teachers and the course instructor, and (3) the preservice teachers and the ESOL students at the reflection stage.
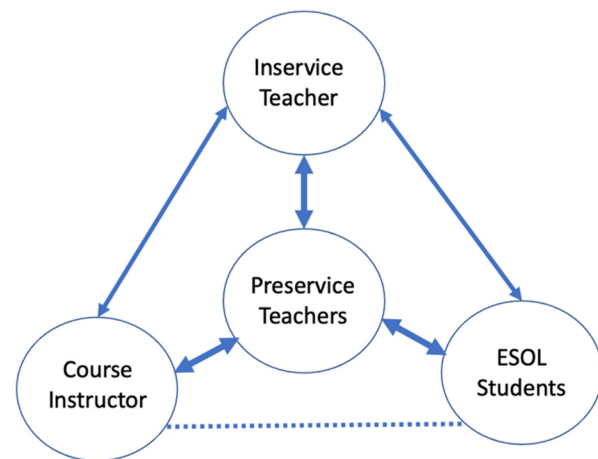

Figure 8. Reflection stage.

At this stage, the preservice teachers were asked to "defend" their assessment through their response to the inservice teacher and to the instructor. The preservice teachers evaluated ESOL students' listening and writing performance and wrote feedback for the ESOL students. The ESOL students in turn shared their experience with the preservice teachers and gave them suggestions from the assessment takers' points of view. Sharing their reflections on the assessment implementation, the course instructor and the inservice teacher also discussed what each stakeholder gained from the language assessment development project.

The language assessment development project required integrated interaction among the stakeholders, in which each stakeholder needed to contribute their expertise and experiences. Compared to unidirectional teacher education models, multiple stakeholder involvement in this model was unique and allowed each stakeholder's voice to be heard (Inbar-Lourie 2017). In this model, not only were the inservice teacher's and preservice teachers' voices integrated, but also the ESOL students'. This closely interconnected interaction provided opportunities for all stakeholders to share their LAL needs. It eventually helped researchers better understand each stakeholder's LAL.

4.2. Research Question 2

How do such interactions assist preservice teachers' LAL as well as their mutual learning among all the participating stakeholders?

As we can see below from their responses to the inservice teacher and the course instructor, as well as from their reflection papers, it was evident that the language assessment development project provided unique opportunities for the preservice teachers to enhance their LAL. While engaging in the language assessment development project through systematic and structured interaction with other stakeholders, the inservice teacher also observed her own LAL development. It was also noted that her ESOL students used the engagement as an opportunity to critically reflect on their assessment experience. The course instructor was able to enhance her LAL on the ways assessment is integrated into learning contexts for ESOL students as well as preservice teachers. In the following, we first discuss the preservice teachers' LAL enhancement in the three domains (i.e., knowledge, skills, and principles/consequences) followed by the other stakeholders' LAL enhancement.

4.2.1. Preservice Teachers' LAL Enhancement 1: Assessment Knowledge

While the preservice teachers improved in all areas of their LAL, including language assessment knowledge, skills, and consequences/principles, the area that the preservice teachers expressed that they had enhanced most was their understanding of assessment knowledge. The preservice teachers' reflections on theoretical aspects of language assessment were one of the most frequently addressed themes in their reflection papers and responses to the course instructor and inservice teacher; the theoretical aspects of language assessment accounted for 62% of the coded data. Notable enhancements of knowledge were found through recognizing awareness of the importance of assessment theories, contextualizing the theories and practicing them in context.

Realizing the Importance of Theories in Assessment Development

The preservice teachers came to acknowledge that theoretical knowledge is critical for them to develop language assessment. For instance, Becca's comment reflects her understanding of what roles that theories would play in language assessment development:

As a result of taking part in this process, I feel that I am better equipped at creating assessments that are more deliberate in the constructs to be tested, as well as more confident in what to include in those assessments to let my students demonstrate what they know.

Even for Helen, who had decades of experience teaching English to adults, developing theoretically sound language assessment was a new experience. In her reflection paper, she said:

The greatest learning for me throughout the process was in determining the constructs to be used in the assessments. I became aware that many of the tests that are created lack one or more aspects of the theories. I realize that it is important for me to consider all attributes in my own future test development. This course has given the tools needed [to do this].

During the language assessment development project, the preservice teachers had ample experience to apply the assessment knowledge into practice. Developing a rubric and experiencing a norming practice with her group members, Diana reported her learning of inter-rater reliability. She and her group members had different scores while they used the same rubric they had developed together:

> *I experienced first-hand how scorer reliability can impact test results. As stated previously, we had varying interpretations of the test results. This disagreement caused our students to have different scores depending on the individual who scored their assessment. Even with a rubric, the scores were different.*

Due to the discrepancy among evaluation, the group members brought much discussion among themselves. Diana believed that this experience further enhanced her understanding of how inter-rater reliability works. The preservice teachers' reflections as well as their interaction with the course instructor showed that the hands-on practices of developing language assessment helped them consider assessment theories less esoteric and technical.

Contextualizing Assessment Theories: Reflecting Test-Takers' Needs

Our data revealed that the preservice teachers enhanced their understanding of the importance of context in assessment practice, which is part of assessment knowledge through the interaction during the language assessment development project.

The following quote exemplified in Yuri's learning path. In his earlier reflection after one of the Skype interviews with the inservice teacher, Yuri stated:

> *During the Skype session today, the teacher talked about a lot of things of interest for the students (eighth graders), but the following are the ones that caught my attention: The use of texts of the students' interest and make them respond to the comprehension questions, and discussion can be a very good activity for them.*

> *The students have been working on the passive voice in present tenses; so, any assessment using the passive voice should be worded in those tenses.*

> *They have been reading about civilization and especially the ancient Egyptians and the Mayans, and they were amazed at discovering that so distant and different civilizations have used the same techniques of communication among themselves and others in writings.*

> *Assessments for the students should be for improving their learning, not for just testing their language ability.*

Figure 9 is an example of Yuri's group's listening assessment. Reflecting the ESOL students' interests in ancient Egyptians, their listening assessment's text was about the interview of an Egyptologist.

After finishing developing assessments, Yuri reiterated his understanding of the importance of incorporating context-specific information to design a valid and reliable language development:

> *The first major learning experience was about the importance of context specifics of our tests (the students' profile—age and L1 background, their language proficiency levels and their familiarity with different testing modes and formats). I learnt that all these factors are to be considered to make more valid, reliable and practical tests with a good likelihood of positive backwash on the students' learning.*

Practicing Assessment Theories: Rethinking Constructs

A series of interactions with the inservice teacher and the course instructor provided the preservice teachers with opportunities to enhance their operationalizing constructs of their own assessment. For example, the inservice teacher realized that her students were confused with one of the open-ended assessment items developed by the preservice teachers (see Figure 10). The inservice teacher asked the preservice teachers to clarify what they intended to measure by the item.

Directions

Listen to this interview with Egyptologist Chris Naunton. After listening, please answer the following questions.

What is the interview about? *

○ Why kids should use their spare time to study and go to college.

○ Why history is interesting and how kids can become Egyptologists.

○ How long it takes to become an Egyptologist.

Figure 9. Listening assessment example (eighth graders). Note: * indicates that a test-taker's response is required.

3. Try to answer the following question in a complete sentence. What else can papyrus be used for?

Short answer text

Figure 10. Listening assessment example (seventh graders).

The preservice teacher responded to the inservice teacher:

In terms of what other things could papyrus be used for? We were hoping to stretch the more advanced students to inference other ideas and make the connection of paper. One student did so with the answer of paper airplanes. For our next assessment, we are titling it with the passages. As suggested, this is more in line with what the students are accustomed to.

The preservice teachers explained what they intended to measure (i.e., the construct). They shared how one of the students understood the assessment item correctly and provided ways to improve the assessment item. In their response to the course instructor, they admitted that the assessment item did not elicit what they expected to measure:

For one of our open-ended questions, one of our test takers provided both a creative answer and a creative false orthography for a poorly constructed test-item ("What else can papyrus be used for?"). For this question we received a range of responses that demonstrated that the students may not be at a strong level for the critical thinking skill inferencing, which was the construct we were aiming to test.

The preservice teachers included "inferencing" as part of their construct. According to WIDA, inferencing is introduced in the ninth-grade WIDA standards (WIDA 2020). While

the preservice teachers wanted to test the ceiling (higher proficiency level) of learners, they could see only one student was able to answer the question. These examples illustrated that the contextualized language assessment development project allowed opportunities for the preservice teachers to internalize assessment theories.

4.2.2. Preservice Teachers' LAL Enhancement 2: Assessment Skills

The data indicated that interaction with multiple stakeholders helped the preservice teachers to develop their assessment skills. One of the inservice teacher's goals in this assessment development project was to help her ESOL students become familiar with online assessments. As the inservice teacher was well aware of her students' assessment experiences including WIDA, she was able to provide substantial feedback on the format of the assessment in particular.

Making Use of Pictorial Aids

The inservice teacher repeatedly suggested that adding some pictorial aids would be helpful for her ESOL students:

> *I feel there was an opportunity to include a picture of the papyrus in the background. It would have given the students the scaffold of comprehensible input.*

> *The questions need pictorial support.*

> *Again, adding a picture of an anime character would help the students in providing a comprehensible input scaffold.*

> *Having some options to choose from pictures would have been beneficial for emerging bilinguals at the beginning stages of language development.*

According to the inservice teacher, her students mentioned the lack of pictorial aids in the listening assessment. They could have noticed the differences between the WIDA listening assessment and the preservice teachers' assessment, as the WIDA listening assessment mostly includes pictorial aids.

While the preservice teachers did not include pictorial aids in their listening assessment, based on the inservice teacher's suggestion, the preservice teachers discussed among themselves and agreed to incorporate pictures in their writing assessment that they developed subsequently. They responded to the inservice teacher:

> *As per your suggestions, pictures would have given the students a scaffold of comprehensible input. This may have better prepared them. This was true in both passages. As a result, we have incorporated illustrations in our writing assessment.*

However, the preservice teachers' response to the course instructor was slightly different:
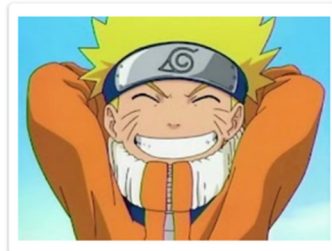
> *After administering the test, the teacher had suggested we use more visuals along with the prompt in order to give more context to draw from when answering the question.*

> *We felt that because this is a listening assessment, we wanted to focus on auditory input rather than provide too much support that would mask their true listening ability.*

In their response to the course instructor, the preservice teachers tried to justify their choice of not including the pictures in the listening assessment; they wanted to avoid a construct-irrelevant factor or distraction from the students. The course instructor agreed with the preservice teachers' justification of not including pictures in the listening assessment. However, she also advised her students (i.e., preservice teachers) to reflect the ESOL students' familiarity with the WIDA exam. In the end, responding to the inservice teacher's comment about the need for pictorial aids, the preservice teachers decided to follow her suggestion in the writing assessment (Figure 11). As the preservice teachers mentioned, the seventh-grade writing assessment included more pictorial aids. The preservice teachers did not express that adding pictorial aids would have impacted the ESOL students' writing performance. Nonetheless, there were no further discussions of the impact of pictorial aids in the writing assessment on the ESOL students' perception of the assessment. It could

have been a fruitful experience if the stakeholders had had a chance to discuss the role of pictures in assessment more explicitly.
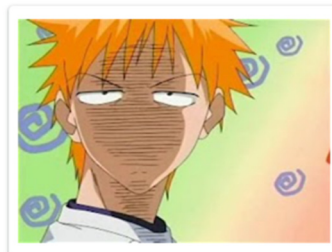

Figure 11. Example of a writing assessment item. Note: * indicates that a test-taker's response is required.

Motivating Learners While Avoiding Construct-Irrelevant Factors

Related to the inclusion of pictorial aids, another important theme that emerged from the data was how to make the assessment more attractive or fun for the ESOL students. The preservice teacher group who developed a listening assessment for the eighth-grade students experimented with using a cartoon-style voice, although they were unsure if it was a good decision or not. The inservice teacher shared her observation with the preservice teachers that it worked well with her students by saying, "the first passage used a cartoon-style voice. The students said it sounded fun and made them interested in listening to it".

The ESOL students also directly reported to the preservice teachers that "listening tasks were fun to listen to". The preservice teachers responded to the inservice teacher:

> *We thought the silly voices in the listening prompt would have been a distraction, but it appears the students enjoyed the voices; this surprised us, but we're glad it kept the students engaged.*

Receiving direct feedback from the inservice teacher and her students gave the preservice teachers a chance to reflect their assessment format and gave them some confidence in their decision.

On another occasion, however, the preservice teachers' attempt to make the assessment more accessible and friendly to the ESOL students was not successful. As shown in Figure 12, the preservice teacher group for the seventh graders included the title of the assessment as "Rockin' Mschool Listening Assessment".

Figure 12. Listening assessment for the seventh-grade. Note: * indicates that a test-taker's response is required.

The group believed that using part of the ESOL students' middle-school name would make the assessment more friendly. However, it caused confusion among the ESOL students when they answered the last question, "what would be the appropriate title of this passage?". The inservice teacher gave feedback to the preservice teachers that "the title had the name of the group in bold letters. The students were a bit confused by it since

they are more used to having story titles at the top of the pages". The preservice teachers responded to the course instructor as follows:

> *A suggestion from the teacher that we would like to take into consideration for replicating this type of assessment would be to omit having the test name in bold at the top of the page while also asking the students to create their own name for the story. The teacher had informed us that the students were confused by our question "What would be an appropriate title for this passage?" because they were accustomed to having story titles at the top of the page, and for our assessment we had the name of our group project in bold letters in that space.*

This interaction helped the preservice teachers realize the unexpected challenges in making an assessment more interesting and accessible for the students while avoiding construct-irrelevant factors. In addition, they realized that it is critical to take test-takers' familiarity of testing formats into account when designing valid assessments.

4.2.3. Preservice Teachers' LAL Enhancement 3: Assessment Principles/Consequences

The collaboration in the language assessment development project also gave opportunities for preservice teachers to enhance their understanding of the consequences of assessment on their educational experience. Through interaction and feedback, the preservice teachers had ample first-hand experience of how assessment can impact the ESOL students' learning. Having direct interaction with the inservice teacher and the ESOL students made the preservice teachers realize how important it is to make the assessment learning-centered, or *assessment for learning* (e.g., Assessment Reform Group 2002; Fulcher 2012).

Assessment for Learning

At the initial Skype interview, the inservice teacher expressed that she wanted her students to experience new types of assessment as part of language learning. Her suggestion motivated the preservice teachers to design a formative assessment, or assessment for learning, rather than a summative assessment. Closely following the inservice teacher's lesson objectives and receiving feedback from the inservice teacher and her students, the preservice teachers experienced how they could use assessment for learning purposes. Although the concept of assessment for learning is often introduced in assessment courses, preservice teachers often do not have the chance to apply this concept through authentic assessment practices:

> *The biggest takeaway that I took from the test development process was understanding how to create ways that enable students to learn through assessment… Seeing real life responses from real students made it easier to envision how my choices in determining what to include in an assessment affect student learning.*

The direct two-way interaction between the preservice teachers and the ESOL students also helped the preservice teachers enhance their awareness of assessment principles. The ESOL student received feedback from the individual preservice teachers; thus, each seventh-grade student received four pieces of feedback, and each eighth-grade student received three pieces of feedback from the preservice teachers, respectively. Below is an example of feedback from Becca to Bowen, a seventh-grade ESOL student. Her feedback started with focusing on his strengths in the listening assessment. She then identified the area that Bowen needed to work on and suggested ways to improve it:

> *In your listening assessment you showed that you are skilled at finding the main idea and specific points after listening to a passage. An area of improvement for this assessment is inferencing, which means that you would benefit from practicing figuring out what would happen in a situation or place that you have not seen before based on experiences you have had in the past. To practice this skill, I would recommend playing a game with your classmates called 20 questions, where your friend picks an object, and you ask questions to figure out what that object may be based on the answers you receive.*

Bowen also received feedback about his listening and writing performances from three other preservice teachers (Helen, Tim, and Mia) independently. The following quote is Tim's feedback to Bowen's writing performance:

> *Your answers for the Writing Assessment were great. The way you answered the first two questions was how I imagined them being answered. First you gave a simple sentence. Then, you added onto this sentence to create a compound sentence. You did a good job describing Pikachu.*

> *Your writing on Dark Ben was creative and gave a lot of information. I could see what he looked like, until he goes invisible of course! I could not have asked for a better writing. You wrote introductory and concluding sentences and provided extra information on his powers and future hopes. Furthermore, I only found a single punctuation issue. You wrote, "the character's nickname would be Dark Ben". Can you tell where the error is? Your writing skills are strong enough that I think you know where the mistake is. Your writing will only get better with more practice Bowen. Keep up the good work.*

Tim's feedback also started with what Bowen did well. In his feedback, instead of identifying the punctuation error, he directed Bowen to find the error by himself. Tim's feedback ended with encouraging and motivating words. At the end of the project, Bowen shared his responses to the experience, "this makes me feel so special!!!!! I am so happy to read this!!! It is not just "Good Job!" Although he must have received feedback from his teacher before, he enjoyed the extensive and individualized feedback from the preservice teachers. In Tim's reflection, he emphasized the importance of avoiding negative backwash in assessment by using a well-prepared rubric and feedback to students.

Our data revealed that the preservice teachers were able to internalize assessment theories and contextualize them through the interaction. They experienced the importance of reflecting learners' perspectives in their language assessment and learned how to design more effective and motivating assessments for learners. Finally, they observed how assessment can be used for learning and the functions of feedback.

4.2.4. Inservice Teachers' and ESOL Students' LAL Enhancement

The language assessment development project also assisted the inservice teacher to enhance her LAL, especially in the domain of assessment knowledge. When the project was initiated, the inservice teacher expressed her interest in assessment knowledge and its application in her own practice:

> *I realized that my own teacher education program (20 years ago) did not offer this information [language assessment theories]. I wanted to learn more about the application of the framework to creation of the assessment experience for my students, based on the content I was teaching.*

Engaging in the language assessment development project, the inservice teacher also had a chance to learn about assessment knowledge and observed how the knowledge was applied in actual language assessment development. In the correspondence with the course instructor, the inservice teacher mentioned that she learned about language assessment theories and application during the language assessment development project.

> *It was beneficial to see the assessment ideas and materials that the pre-service teachers created based on their study of the assessment theory. It was interesting to see how the target content and language skills were reflected in their assessment exercises.*

The inservice teacher also mentioned her new understanding of the impact of language assessment on her students. She stated,

> *The pre-service teachers were providing written feedback to the students, and this practice had a great impact on the student's motivation. In the day-to-day learning and assignments completion with the ubiquitous use of rubrics and computerized responses the feedback tends to be too general and yields limited further learning. When my students got specific feedback praising their effort and success in language structures, as*

well as opportunities for growth, they were motivated to learn further and appreciated the attention.

As her reflection indicated, individualized and focused feedback provided by the preservice teachers to each ESOL student motivated the learners.

She shared that assessment is one area that she needed to improve, which was identified by an advanced certification accreditation party in the U.S. (i.e., National Board for Professional Teaching Standards (NBPTS). At the end of the project, the inservice ESOL teacher stated that she gained more information about assessment than when she first joined the project. She mentioned that through interacting with the other stakeholders, she learned how the preservice teachers incorporated language assessment theories and her curriculum and lesson objectives. The experience helped her to have a better sense of the connections between theories and practices in language assessment.

In addition to enhancing her LAL, the inservice teacher stated that the collaboration helped her enhance her leadership skills. She valued the opportunity to connect the preservice teachers to a real ESOL classroom. She felt a sense of responsibility and pride in providing her insights from the classroom for the preservice teachers and the course instructor. She mentioned, "it was valuable to have the opportunity to influence the creation of the assessments that students could respond to successfully and learn from with greater motivation".

This assessment project was also seen as a learning opportunity for her ESOL students. The inservice teacher shared, "this collaborative work viewed students' contribution as an asset, and they felt important participants". The inservice teacher reflected that while her students were actively engaged in the assessment activities, they developed metacognitive skills by explaining what they liked in the assignments and which aspects were hard for them. She also observed that her students showed interests in a profession of teaching multilingual children. In one of the comments, the inservice teacher stated:

It made school initiatives such as American Education Week and Career Day more relevant because the students gained the experience of contributing to a professional field and saw representation of the ESOL field in practice.

Although the ESOL students did not use jargon, we also found that they were highly capable of articulating their responses to the language assessments that the preservice teachers developed. The inservice teacher had a chance to ask her students about their experience of the new assessment. One student, Bowen, a seventh grader, mentioned, "some questions were confusing, like having multiple answers, or the question asked about stuff that was not in the listening piece". The inservice teacher also asked whether the assessments measured their actual language abilities. Abe, a seventh-grade ESOL student, disagreed, stating, "It does not really show my ability because I was getting distracted by thinking about Fortnite. I could do better". Referring to the writing topic (i.e., anime), Gavin, an eighth grader, said he wanted to write "about soccer" because "I can write a lot, this topic, not, so it is not showing what I can do". Gavin's comments confirmed that assessment content's relevance to a student's interest is an important factor in their motivation when taking part in. Finn, an eighth grader, thought that the assessment successfully measured their language ability. For example, Finn said, "yes, [it measured my ability successfully,] because I tried my best and spent time on it". Another student, Sammy, an eighth grader, also said the same thing. Finn and Sammy's perception that the assessment successfully measured their writing ability was demonstrated in the fact that they were able to fully answer the prompts. When asked about the use of assessment results, Felix said, "I think it should let you know what you should work on", clearly referring to the purpose of formative assessment.

In conclusion, the inservice teacher reported that she was able to refine her knowledge of assessment theories and learn more about formative assessment and online assessment. Her middle-school ESOL students received constructive feedback from assessment developers (i.e., preservice teachers) and shared their feedback to the new form of assessment

with preservice teachers. According to the inservice teacher, her students were motivated to share their voices with the preservice teachers.

4.2.5. The Course Instructor's LAL Enhancement

The collaboration helped the course instructor enhance her LAL as well. While course instructors are influential stakeholders in teacher education, less information about their LAL is known (Bøhn and Tsagari 2021; Jeong 2013). In the current study, the authors confirmed that through interaction with the inservice teacher, the course instructor also developed further understanding of the ways to contextualize language assessment theories and practices. Although the course instructor had been involved in teacher education for an extended period of time, the collaboration with the inservice teacher and her students allowed the course instructor to re-experience the dynamics of classroom teaching and changes of assessment policies.

The language assessment development project also helped the course instructor to observe the gradual but non-linear enhancement of LAL among the preservice teachers. While language assessment theories and practices were provided through lectures, readings, discussions, and hands-on activities to all preservice teachers, the preservice teachers' LAL development varied. In the beginning of the course, the preservice teachers struggled to internalize theories, skills, and understanding of consequences of language assessment. However, once the language assessment development project was launched, they gradually applied what they had learned in the course. Not all components of LAL develop in the same manner, either: some preservice teachers developed their understanding of assessment consequences earlier, but others obtained the knowledge once the project was completed. Some found assessment skills relatively straightforward to get, while other preservice teachers struggled a bit to modify their language assessment formats per their test-takers' preference or interests.

4.2.6. Challenges

While the language assessment development project provided an extensive learning experience to all stakeholders, there were also several challenges. The first challenge was associated with the difficulties in conducting group work among the preservice teachers. Becca reflected on the group work as follows:

> *Working with peers that had a different way of approaching planning pushed me to remember to pay attention to little details while also being confident enough in my own methods to speak up when I feel that we need to work in a different direction.*

> *Having the ability to respectfully disagree and be strategic in working towards a common goal on a team will be a critical skill in the workforce in the field of teaching, so this experience in test development was definitely beneficial in working towards my future career goals.*

Becca's experience was not an isolated example. Diana, who planned to teach adult learners upon completion of her master's degree, also reported, "one of the most important learning experiences I had comes from working with two other individuals". Straightforward and timely communication was certainly a challenge in her group. One of the group members shortened part of an audio file in their listening assessment, but the other member did not remove a few listening questions that were associated with that part of the listening assessment. It caused confusion among the ESOL students.

The second challenge was to help stakeholders understand the iterative and interconnected processes in the language assessment development project. Some preservice teachers felt that the processes were too extensive; however, they also recognized that language assessment development requires extensive time and effort. Becca, who was certified as an elementary school teacher but was pursuing a master's degree in TESOL, mentioned, "as I reflect through the process, I am somewhat overwhelmed by the many steps necessary from the initial planning to the feedback to the students. What became very clear was that test preparation requires a great amount of time". Finally, mediating

the constant interaction among the stakeholders required the course instructor's extensive time commitment. Connecting the preservice teachers and the inservice teacher and transmitting the right information to each group required close attention. The course instructor also made great effort to create a supportive and collegial environment among the group members.

5. Discussion

Despite the logistical challenges, the current language assessment development project is a stellar example of an inclusive approach to enhance all involved stakeholders' LAL. The interconnected interaction during the language assessment development project helped each stakeholder broaden their perspectives on language assessment. It also invited all stakeholders as key informants, which allowed them to share their respective experiences as well as expertise. The online interaction through Skype and the instructional platform also showed the potential of virtual teacher education to still connect preservice teachers with real-world classrooms. The dynamic and interconnected interaction documented in this study is rather unique considering the findings in earlier studies. Language assessment course surveys and textbook analyses confirm that assessment development activities that connect all stakeholders are rare in teacher education (e.g., Brown and Bailey 2008; Davies 2008). We found that the interconnected interaction among the stakeholders could provide opportunities for authentic and contextual teacher education.

5.1. All Stakeholders as Key-Informants to Enhance LAL

This language assessment development project was originally designed to enhance the preservice teachers' LAL. Nonetheless, our results revealed that interconnected interaction among stakeholders during this project positively influences not only the preservice teachers but also all other stakeholders.

First of all, it was evident that the preservice teachers enhanced their LAL in all three of its components (i.e., assessment knowledge, skills, and consequences/principles). Our findings revealed that among the factors of LAL, the preservice teachers particularly made prominent enhancement of their assessment knowledge. Through the interconnected interaction, the preservice teachers operationalized constructs for their assessment that met the ESOL students' needs and contexts. They experienced how concepts such as validity, fairness, and reliability could be integrated in assessment design and interpretation. They also improved their assessment skills, with which they could design assessments that reflected the constructs and the ESOL students' familiarity with the content. The preservice teachers were able to understand how their assessment was received by middle-school ESOL students, how they should present their assessment, and what they should select as constructs to measure.

In addition to the preservice teachers, other participants could benefit from this project. The inservice teacher observed that her students had positive reactions to both the assessment and extensive feedback provided by the preservice teachers. Through this observation, the inservice teacher enhanced her understanding of assessment knowledge and how her students perceived the feedback from the preservice teachers. Even the middle-school ESOL students were able to share their observations of the preservice teachers' assessment (e.g., their familiarities, what was measured (i.e., constructs, not using jargon), preferences, etc.). The ESOL students experienced how their reaction to the assessment could be reflected, which could develop their critical perspectives of their assessment experience. Finally, the course instructor enhanced her understanding of the public school ESOL practices and the way her preservice teachers internalized LAL.

5.2. Inclusive Approaches to LAL Enhancement

By expanding the role of negotiations and discussions in developing stakeholders' LAL as documented in Baker and Riches (2017), our study shows that systematic and structured interaction among stakeholders can enhance their assessment experience. Furthermore, the

interconnected interaction in this study suggests a reciprocal, mutual experience-mediated approach to LAL enhancement (cf. Yan and Fan 2021). Actively seeking each stakeholder's input invites a unique and equal contribution to all stakeholders' LAL enhancement. This approach is possible by valuing each stakeholder's unique assessment experience. In other words, none of the stakeholders were illiterate in language assessment when they joined the current project (Kremmel and Harding 2020; Pill and Harding 2013). This approach also helps us avoid conceptualizing LAL training from a deficit perspective. Instead, this approach provides more inclusive ways of enriching their LAL enhancement experience.

The current language assessment development project, which actively sought out stakeholders' involvement, could be used as a tool to develop their critical perspectives of assessment practices as well. The critical perspectives of assessment practices can empower each stakeholder so that they use assessment to advance their educational experience. The researchers were able to understand what the ESOL students thought about language assessment and their language assessment experience through this project. Although some researchers have expressed their concerns about test-takers' inability to share their assessment experience (e.g., Malone 2017), our study confirms that as long as a guideline is provided and there is an opportunity to invite ESOL students' thoughts on assessment, they can certainly share their assessment experiences (e.g., Butler et al. 2021; Vogt et al. 2020).

5.3. Importance of Assessment Knowledge and Its Application

Our data confirmed that the preservice teachers and the inservice teacher identified assessment knowledge as one of the critical aspects of LAL that they needed and made much enhancement. This was unexpected considering the course instructor purposefully tried not to overemphasize the theoretical knowledge of language assessment, as suggested by Taylor (2013). The preservice teachers' perspectives on assessment knowledge seem contradictory to what Taylor (2013) suggested, who hypothesized that assessment knowledge could be less emphasized than other attributes of LAL for practicing teachers. However, the pre- and inservice teachers in this study valued assessment knowledge as the information was newly presented to them. As the preservice teachers absorbed more and more of the course content, they started using more precise terminology to describe their assessment knowledge. Their understanding of theoretical concepts such as validity, reliability, fairness, contexts, and constructs became more precise in their later reflections than in earlier reflections.

In addition to valuing assessment knowledge, the stakeholders identified applying the knowledge to develop language assessment for actual learners as an important way to enhance their LAL. The hands-on practices provided opportunities to the stakeholders to apply assessment knowledge and assessment skills that they gained to a real-life scenario. The hands-on practice also helped the preservice teachers understand the importance of context in assessment practices. Contextualizing language assessment per test-takers' needs and characteristics is critical to help test-takers use the assessment experience for their own educational growth. After working with the inservice teacher and her students, the preservice teachers were able to contextualize their understanding of language assessment theories.

5.4. Limitations

This study is not free from limitations. While the researchers had the privilege of analyzing rich data, which was filled with unique voices of the stakeholders, this research was still based on one teacher education setting. Although it has limited generalizability, it is hoped that detailed descriptions of the project process and outcome will be informative to educators who are interested in professional development on LAL. In addition, admittedly, all the participants in this project, including the ESOL students, were highly collaborative and motivated to support preservice teachers' professional development; they were willing to share their perspectives on language assessment. Without such positive attitudes, the successful mutual learning experience recorded in this study would not have been possible.

6. Conclusions

The findings revealed that the interaction among the stakeholders was dynamic and interconnected at each stage of the language assessment development project. All stakeholders with their distinct expertise in LAL (including ESOL students as assessment takers) were indispensable informants during the project. The multi-directional interaction among the stakeholders provided learning opportunities to all the stakeholders involved in the language assessment development project. In particular, the interconnected interaction presented in this language assessment development project provided unique opportunities for all stakeholders to share their voices in their language assessment experience. This approach to teacher education confirms that every stakeholder who was involved in this project has some degree of LAL from their education and assessment experience. Although we certainly need more research, the professional development model involving multiple stakeholders discussed in this paper can be a promising approach as an alternative to the traditional knowledge transmission model of professional development.

References

Assessment Reform Group. 2002. *Assessment for Learning: 10 Principles*. Cambridge: Faculty of Education, University of Cambridge.

Bachman, Lyle F., and Adrian S. Palmer. 2002. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Baker, Beverly, and Caroline Riches. 2017. The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing* 35: 557–81. [CrossRef]

Bøhn, Henrik, and Dina Tsagari. 2021. Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research* 12: 222–33. [CrossRef]

Brindley, Geoff. 2001. Language assessment and professional development. In *Experimenting with Uncertainty: Essays in Honour of Alan Davies*. Edited by Cathie. A. Elder, Alison Brown, Kathryn Hill, Noriko Iwashita, Tom Lumley, Tim McNamara and Kieran O'Loughlin. Cambridge: Cambridge University Press, pp. 126–36.

Brown, James Dean, and Kathleen M. Bailey. 2008. Language testing courses: What are they in 2007? *Language Testing* 25: 349–83. [CrossRef]

Butler, Yuko Goto, Xiaolin Peng, and Jiyoon Lee. 2021. Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 0265532221992274. [CrossRef]

Csépes, Ildikó. 2014. Language assessment literacy in English teacher training programmers in Hungary. In *Studies in Honour of Nikolov Marianne*. Edited by József Horváth and Peter Medgyes. Pécs: Lingua Franca Csoport, pp. 399–411.

Davies, Andrea. 2008. Textbook trends in teaching language testing. *Language Testing* 25: 327–47. [CrossRef]

DeLuca, Christopher, and Don A. Klinger. 2010. Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice* 17: 419–38.

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

Giraldo, Frank. 2019. Language assessment practices and beliefs: Implications for language assessment literacy. *HOW* 26: 35–61. [CrossRef]

Inbar-Lourie, Ofra. 2008. Constructing an assessment knowledge base: A focus on language assessment courses. *Language Testing* 25: 385–402. [CrossRef]

Inbar-Lourie, Ofra. 2017. Language assessment literacy. In *Language Testing and Assessment, Encyclopedia of Language and Education*, 3rd ed. Edited by Elana. S. Shohamy, Stephen May and Lair G. Or. Berlin: Springer, pp. 257–68.

Jeong, Heegeong. 2013. Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing* 30: 345–62. [CrossRef]

Koh, Kim, Lydia. E. Carol, Ann Burke, Allan Luke, Wengao Gong, and Charlene Tan. 2018. Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Resaerch* 22: 264–88. [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]

Lee, Jiyoon. 2019. A training project to develop teachers' assessment literacy. In *Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom*. Edited by E. White and T. Delaney. Hershey: IGI Global, pp. 58–80.

Lee, Jiyoon, and Yuko Goto Butler. 2020. Where are language learners? Reconceptualizing language assessment literacy. *TESOL Quarterly* 54: 1098–111. [CrossRef]

Malone, Margaret E. 2017. Unpacking language assessment literacy: Differentiating needs of stakeholder groups. Paper presented at East Coast Organization of Language Testers, Washington, DC, USA, October 20.

Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods* 16: 1–13. [CrossRef]

Pill, John, and Luke Harding. 2013. Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing* 30: 381–402. [CrossRef]

Scarino, Angela. 2013. Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing* 30: 309–27. [CrossRef]

Stiggins, Rick. 2007. Conquering the formative assessment frontier. In *Formative Assessment Classroom: Theory into Practice*. Edited by H. McMillan. New York: Teachers College Press, pp. 8–28.

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

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, Dina Tsagari, Ildiko Csépes, Anthony Green, and Nicos Sifakis. 2020. Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly* 17: 410–33. [CrossRef]

Volante, Louis, and Xavier Fazio. 2007. Exploring teacher candidates' assessment literacy: Implications for teacher-education reform and professional development. *Canadian Journal of Education* 30: 749–70. [CrossRef]

WIDA. 2020. *WIDA English Language Development Standards Framework, 2020 Edition Kindergarten—Grade 12*. Board of Regents of the University of Wisconsin System. Madison: WIDA.

Yan, Xun, and Jason Fan. 2021. "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing* 38: 219–46. [CrossRef]