PERSPECTIVES OF LANGUAGE ASSESSMENT TRAINING FOR
TEACHERS AND TESTING PROFESSIONALS

BY

HEEJEONG JEONG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Fred Davison, Chair
Professor Jennifer Greene
Professor Katherine Ryan
Professor Melissa Bowles
Abstract

Even before the field of Language Testing emerged as an independent field in applied linguistics, language assessment courses had been taught in various forms and by different instructors. Currently, these courses are being taught by professionals who have majored in the area of Language Testing (LTs) but also by others who come from different majors (non-LTs). This study seeks to investigate the effect instructors bring in shaping the characteristics (i.e., content, structure) of language assessment courses and to what extent the student-teacher are satisfied with the course. To get the full picture, the characteristics and satisfaction of the courses are researched from 4 different lenses; instructors who teach the course (Language Testers vs. Non-Language Testers) and different grade levels of language student teachers (Adult vs. K-12) who have taken the course.

A large scale on-line survey, in-depth follow-up phone interviews and syllabi document review have been done for the study. A total of 384 instructors and student teachers completed the on-line survey (instructors n=140, student-teachers n=244). Based on the survey results an in-depth phone interview has been conducted with 13 instructors from 5 different countries and 5 student-teachers.

Survey findings show there are significant differences in the content of the course depending on the instructors’ background in six areas; test specifications, test theory, basic statistics, classroom assessment, rubric development, and test accommodation. Interview results confirm non-LTs are less confident in teaching technical assessment skills compared to LTs and have a tendency to focus on more classroom assessment issues. Student teachers are overall satisfied with the course but wanted more activities that were directly related to the course. The study ends by predicting the future of language assessment courses and why it is important for
both LTs and non-LTs to communicate and work actively to develop a better course that fulfills the needs of the student teachers.
For the 384 instructors and student teachers who participated in the study
Acknowledgments

So many people took part in completing the dissertation. First and foremost, I would like to show my deepest gratitude to every single participant who took part in the study. Without their help there would never be a dissertation. Each person generously took the time to take part in the study. Thank you. Also I express special thanks to the 18 interview participants who came from different countries. I truly enjoyed talking with each and every one of you.

I could have never written this dissertation without the guidance of my advisor Dr. Davidson. Through the dissertation writing process there were times when I was lost and unsure of my next step. During those difficult times you were always there to guide me and gave endless support. You encourage students to expand their curiosity and creativity. It was a great honor to be your student.

Dr. Greene, thank you so much for your support. You were there every step of the way; my early research both qualifying exams and of course the dissertation. I have learned how to become a researcher from you.

I would also like to thank my parents. Their unconditional love and support brought me here. Mom, who is my biggest fan, was with at all times. Thank you and I love you. Friends, who had been there to give feedback, are those who I cannot forget. Especially, Hongling, who took the time to read my draft, give feedback and who was such a good friend.

Finally, I would like to thank the Hardie Dissertation Award and TOEFL Small Grants for Doctoral Research in Second or Foreign Language Assessment for providing funding for my study.
Table of Contents

Chapter I. Introduction ................................................................................................................1
    Teacher Training in Language Assessment- Students Dislike Assessment, So Do Teachers ...............................................................1
    Importance of the Study.............................................................................................................6
    Research Questions .................................................................................................................7
    Summary of the Problem .........................................................................................................7
    Organization ............................................................................................................................9
    Terms ......................................................................................................................................9

Chapter II. Literature Review ......................................................................................................11
    Assessment Literacy ..............................................................................................................12
    Impact of Assessment Training in General Education .........................................................13
    Concerns with Assessment Courses .....................................................................................14
    Teachers’ Past Assessment Experience and Attitudes to Assessment .................................17
    Field of Language Assessment ............................................................................................18
    Teacher Training in Assessment for Language Teachers .....................................................19
    Language Assessment Textbook, Past and Present ..............................................................20
    Research on Language Assessment Courses .......................................................................22
    Statistics and Language Assessment ...................................................................................27
    Importance of Public Policy in Language Assessment Training .........................................28
    Summary ................................................................................................................................29

Chapter III. Methodology ...........................................................................................................30
    Research Questions ..............................................................................................................30
    My Mixed Methods Stance ..................................................................................................30
    The Mixed Methods Approach for this Study .................................................................32
<table>
<thead>
<tr>
<th>Phase I: Preliminary Syllabi Review</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase II: The Pilot</td>
<td>35</td>
</tr>
<tr>
<td>Pilot Study Results</td>
<td>38</td>
</tr>
<tr>
<td>Implications for the Main Study</td>
<td>41</td>
</tr>
<tr>
<td>Limitations of the Pilot</td>
<td>42</td>
</tr>
<tr>
<td>Phase III: The Survey</td>
<td>43</td>
</tr>
<tr>
<td>Research Questions for the Survey</td>
<td>44</td>
</tr>
<tr>
<td>Participants</td>
<td>45</td>
</tr>
<tr>
<td>Survey Pre-test</td>
<td>46</td>
</tr>
<tr>
<td>Survey Administration Process</td>
<td>47</td>
</tr>
<tr>
<td>Sampling</td>
<td>49</td>
</tr>
<tr>
<td>Survey Data Analysis Plan</td>
<td>50</td>
</tr>
<tr>
<td>Phase IV: Follow-up Interviews</td>
<td>55</td>
</tr>
<tr>
<td>Interview Participants and Process</td>
<td>55</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>56</td>
</tr>
<tr>
<td>Integration of Quantitative and Qualitative Data</td>
<td>57</td>
</tr>
<tr>
<td>Maintaining Data Quality</td>
<td>58</td>
</tr>
</tbody>
</table>

**Chapter IV. Results** .....................................................................................................................59

Research Question 1: What are the characteristics of introductory language assessment courses?..............................................................................................................60

Study Participants .......................................................................................................................60

Who teaches the courses? Instructor Survey Demographics .......................................................60

Interview Instructor Demographics .............................................................................................69

Who takes the course? Student Teacher Survey Demographics......................................................76

Survey Student Teacher Demographic Summary.............................................................................86

Interview Student Teacher Demographics......................................................................................87
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is being taught?</td>
<td>90</td>
</tr>
<tr>
<td>Research Question 2: What constitutes evidence of an “effective” language assessment course for classroom teachers?</td>
<td>96</td>
</tr>
<tr>
<td>Which topic is most helpful for student teachers?</td>
<td>96</td>
</tr>
<tr>
<td>Which topic is most difficult to teach?</td>
<td>99</td>
</tr>
<tr>
<td>Important Topics for Classroom Teachers</td>
<td>101</td>
</tr>
<tr>
<td>Expectations for the Course</td>
<td>111</td>
</tr>
<tr>
<td>What is being used? The Materials</td>
<td>113</td>
</tr>
<tr>
<td>Changes in the Course</td>
<td>116</td>
</tr>
<tr>
<td>Assessment Courses vs. Other Courses: Is there coordination?</td>
<td>117</td>
</tr>
<tr>
<td>Language Assessment Courses Around the World</td>
<td>118</td>
</tr>
<tr>
<td>Summary</td>
<td>119</td>
</tr>
</tbody>
</table>

**Chapter V. Discussion** .............................................................................................................122

<table>
<thead>
<tr>
<th>Research Question 3: How effective are language assessment courses for language teachers’ everyday classroom practices?</th>
<th>122</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Theory</td>
<td>122</td>
</tr>
<tr>
<td>Classroom Assessment</td>
<td>124</td>
</tr>
<tr>
<td>Alternative/Performance Assessment</td>
<td>127</td>
</tr>
<tr>
<td>Test Specifications/Item Writing</td>
<td>126</td>
</tr>
<tr>
<td>Rubric Development</td>
<td>128</td>
</tr>
<tr>
<td>Statistics</td>
<td>128</td>
</tr>
<tr>
<td>Ethics</td>
<td>134</td>
</tr>
<tr>
<td>Test Accommodation</td>
<td>136</td>
</tr>
<tr>
<td>Assessment Literacy: Only for student teachers? What about the instructors?</td>
<td>137</td>
</tr>
<tr>
<td>Can Non Language Testers Teach a Language Assessment Course?</td>
<td>138</td>
</tr>
<tr>
<td>Characteristics of Successful Language Assessment Courses</td>
<td>140</td>
</tr>
<tr>
<td>Importance of Pre/In-Service Courses</td>
<td>143</td>
</tr>
</tbody>
</table>
Limitations of the Study........................................................................................................144
Summary..............................................................................................................................145

Chapter VI. Conclusion........................................................................................................147

Summary of Main Findings .................................................................................................147
Revisiting the Purpose ........................................................................................................148
My Own Bias, Misconceptions of Language Assessment Courses.................................149
Suggestions for Future Research .......................................................................................150
Future of Language Assessment Courses ..........................................................................151
Wishes and Dreams............................................................................................................152

References..........................................................................................................................153

Appendix A Pilot Interview Questions/Follow-up Interview Questions............................158
Appendix B List of Commonly Covered Topics in Language Assessment Courses...........162
Appendix C Description of Pilot Study Participants..........................................................163
Appendix D Description of Pilot Study Courses ..............................................................165
Appendix E Original Survey Instrument .........................................................................169
Appendix F Revised Survey Instrument .........................................................................184
Appendix G List of Professional Organizations Used for the Survey..............................199
Chapter I
Introduction

Teacher Training in Language Assessment- Students Dislike Assessment, So Do Teachers

Being a good teacher is a challenge; training teachers to become one is even more.

November 15, 20x
Faculty Search Committee Meeting

It is an early winter morning at Central States University. Charles Sagon, chair of the search committee, and three other faculty members (David Kent, Jina Lee, Mia Horwitz) met to discuss the search process of hiring a new faculty member for the program. The department has just recently got approval from the dean to hire an additional faculty member.

Sagon: Good morning everyone. Today is our first search committee meeting. As every knows, our department received approval to hire a new faculty member to teach the language assessment course. Due to requests from students and also considering the importance of the subject area, in our last faculty meeting, we all agreed it is about time to offer a separate language assessment course. David has been teaching this course as part of the material development course and stressed the importance of developing an independent course.

Kent: Yes. There is a limit in teaching language assessment part of another course and assessment is such an important topic to these student teachers it should be offered as a separate course.

Sagon: In today’s meeting, I wish to discuss who to hire to teach the course. We need to make decisions quickly otherwise there is a high possibility the budget will go to another department.
Horwitz: Is there really anything to discuss about who to hire? Don’t we all know? Shouldn’t it be a person who majored in Language Testing?

Kent: Mia, I don’t think it is that simple. Our department is always very short of faculty and hiring a person who can only teach one course may not be the best choice. We should think of other possibilities.

Horwitz: I’m sorry David, but I do not understand. Didn’t you just say assessment is a specialized area and doesn’t that mean it should be taught by a specialist?

Lee: David, Mia, I understand both of your concerns. But let’s remember our goal is to first hire someone and begin this process as soon as possible. Let’s do some research on what type of people teach the course. There are hundreds of schools in this country who offers the course. We should do a bit of search on these schools.

Sagon: Jina has a good point there. Before we make any hasty decisions we should take some to think this over.

Horwitz: I honestly think that will be a waste of time. Why do we need to spend time and money on something we already know the answer to? A person who majored in language testing will be the BEST to teach the course!! I don’t want students complaining all over again about the course. That course had bad evaluations since the first semester we offered it and I think it is time to bring a change.

Sagon: Mia, yes I agree that our focus should be offering a better course. However we should keep in mind hiring a new faculty should be done with care. Let’s think of inviting both types of people for the interview. It’ll be nice if we could get some resources that could help us develop the first round interview questions.
Lee: You know, I think there could be a person who could help us with these questions.

Sagon: Who?

Lee: A former student of mine who is currently in a doctoral program is actually writing her dissertation on this topic. Her interest is in introductory language assessment courses and it would be helpful to invite her to one of our meetings.

Sagon: Okay, that could be a start. Her study could be a good reference for us. Why don’t we send her the questions raised on today’s meeting and invite her to present her dissertation studies in our next meeting?

Horwitz: I’m still not too thrilled with this. It does not seem to be a necessary process.

Lee: Mia, don’t worry. It won’t delay the process too much. I’ll contact her and see what her schedule is.

Stiggins (2002) states there are two purposes for assessment: one is to “inform decisions” (p.760) and the other is to “motivate student learning” (p.760). The former he calls assessment of learning and the later as assessment for learning. High-stakes assessment of learning, “informs our decision about accountability” (Stiggins, p. 759). Through large-scale test results teachers and other stakeholders monitor students’ progress and make important decisions. Teachers are expected to have the training and ability to interpret standardized tests to communicate the results with students and other stakeholders. Another purpose for assessment which is assessment for learning, assessment is used to promote student learning. Teachers use assessment results in their instruction to encourage students’ learning through positive washback. Often teacher-developed, small-scale classroom assessments are used for this function.

Logically, the two different purposes of assessment can also shape the structure and content of language assessment courses. Language assessment courses can cover both large-
scale, assessment-related topics in addition to small-scale, classroom assessment topics. The degree of attention paid to the two areas in a language assessment course differs depending on external (e.g., curriculum of the course, status of the course within the program) and internal factors (e.g., instructor educational background, student background).

Standardized assessment has always been prioritized among other topics in the field of assessment and has received extraordinary attention, especially for assessment courses targeted for primary and secondary teachers. Even with pre-service courses, where the target student audience are adult language learners, classroom assessment could be, in many cases, treated as a secondary topic. It could be often covered at the end of the course or less attention was given to classroom-related issues compared to standardized assessment. Assessment courses are filled with knowledge from large-scale measurement perspectives rather than hands-on, teacher-developed, small-scale tests. For this study, I will focus on how and to what degree classroom assessment-related content is covered in language assessment courses and how beneficial the knowledge is to teachers’ everyday classroom assessment practices.

Teachers may spend up to one third of their professional time involved in assessment-related activities, yet few teachers have been taught standards of good assessment practice in undergraduate education programs (Stiggins, 1999). They are often intimidated by the statistical manner in which standardized test results are presented (Stiggins, 1991).

When I taught assessment, this was the case with my student teachers. They were afraid of numbers and could not find a strong connection to their student teaching. When student teachers take classes in assessment, many, especially pre-service teachers, do not feel the urgency to fully comprehend the information at the time of the course since it feels distant from classroom teaching. However it is important that student teachers understand assessment to
participate in the language learning culture (Shepard, 2000). When teachers lack the competence to effectively use assessment information, this raises concerns, since the consequences of test results are considerable. Depending on how tests are made, interpreted, and used, they can have a significant impact on students’ lives. Placement in schools, acceptance to higher education institutions, and getting jobs are all based on students’ test results.

One critical problem arises when student teachers complain about lack of practicality in language assessment courses. Student teachers are not happy that instructors are not familiar with the so-called “real” class situation and that the materials taught inside these professional courses are not always directly applicable in practice. This gap may come from different views instructors and teachers have in looking at the assessment culture and also from the training and educational background of the instructor. The requirements that instructors value as important for effective language assessment (e.g., test theory) could be dissimilar from what student teachers think are helpful in the classroom (e.g., rubric development). Thus, the course objectives instructors believe are important may not be as useful as the instructors think in the classroom context.

In one study, teachers who self-reported that they were confident about their assessment skills indicate that their assessment competence was acquired through their past assessment experiences and trial and error rather than language assessment courses (Jeong, 2007). Student teachers’ reliance on developing classroom assessment skills based on trial and error could be a dangerous practice which might be detrimental to the students. Hence, a detailed diagnosis of current assessment courses is essential to know how assessment courses are taught and how effective they are to classroom language teachers. All four teachers in Jeong’s study took assessment courses, but only one commented on the impact on classroom assessment. This
shows that assessment classes did not have much influence on these student teachers and implies that there are doubts about the effectiveness of the course. The ideal language assessment training course should respond to teachers’ needs while conveying the theoretical rigor of language testing from the instructor’s point of view.

**Importance of the Study**

Teacher education in language testing is crucial because testing is a bridge between teaching and learning, and classroom tests are regarded as mirrors in which teachers and students can see their reflections clearly (Koksal, 2004). In order to become a teacher, student students go through a pre-service teaching program. They take courses in multiple areas, experience student teaching, take a test, and become certified. Among the many courses pre-service teachers are required to take, an assessment course is usually taught at the final stage of the pre-service program (Stiggins, 1993) and it attempts to combine all the knowledge learned prior to the course.

When teaching a pre-service assessment course, I discovered that student teachers were not happy with the course content or method. The student teachers felt the content covered in the course did not really apply to their everyday teaching and were more interested in developing assessment tools that they could actually use in the classroom. This led me to wonder how language assessment courses are currently taught and what could be done to make training in language assessment more meaningful and relevant to future and current language teachers. There is a lack of research about how student teachers digest assessment knowledge and process it in their classrooms. There is a paucity of empirical studies that have actually researched assessment courses, let alone language assessment classes (Kleinsasser, 2005; O’Loughlin, 2006).
Previous studies mostly relied on participants’ self-reports and there were no studies that have researched language assessment courses taught by instructors who were not, themselves, trained as language testers. Many student teachers are trained in language assessment by instructors who did not major in language testing, yet we have little knowledge about how the courses are taught and what teachers have learned from them. It is difficult to know how these courses differ in terms of content and structure compared to classes taught by language testers. Another gap in the literature is the lack of knowledge on how effective these courses are in classroom teaching and what constitutes “effectiveness” in a language assessment course. Through this study, I expect to explore the aforementioned gaps in the literature regarding the characteristics and effectiveness of language assessment courses for language teachers’ everyday classroom assessment practices. I hope this study will contribute to the early body of work in language assessment research.

**Research Questions**

1. What are the characteristics (i.e., content, structure, objectives) of introductory language assessment courses? (What factors determine the differences among courses?)

2. What constitutes evidence of an “effective” language assessment course for classroom teachers? (Is there a difference between instructors\(^1\) and student teachers\(^2\)?)

3. How effective are language assessment courses for language teachers’ everyday classroom practices? (Is there a difference depending on the teachers’ years of teaching experience, final academic degree, and teaching environment?)

**Summary of the Problem**

---

\(^1\) Instructors are people who teach language assessment courses.

\(^2\) Student teachers are people who take language assessment courses.
Assessment is an everyday practice for almost every teacher. However, many feel they have not received sufficient training and acquired enough knowledge in preparing themselves to be proficient with assessment needs. This study seeks to discover the degree to which current assessment training is beneficial to language teachers’ everyday classroom practices. The purpose of the study is to provide information to teacher educators in language assessment to reflect the factors that most influence teachers in developing assessment skills. The end result of the study is to improve the quality of teacher training by helping language teachers develop assessment literacy and positive attitudes toward assessment. This will encourage student teachers to be confident in their assessment tools and help them in effectively applying assessment results to their teaching. The study also has implications for policy, research, and practice, and contributes to the needed literature on language assessment courses.

The basic rationale for the study concerns the end-state of an assessment class. Through language assessment courses, teachers should be provided the education they need to be confident in their classroom assessment practices. After exiting the course, student teachers should have acquired the ability to clearly explain the purpose, process and results of their assessment tools to their students. For student teachers to fulfill their role as good assessors in the classroom, the first step is to provide student teachers with relevant and practical classroom assessment related information in the introductory assessment courses. Most likely a language assessment course they take as an Undergraduate student or Graduate student is the only course in assessment they will ever take. Rarely do student teachers have the luxury of taking an advanced course or workshop in language assessment. Since there is a high possibility that the introductory language assessment course will be the one and only assessment course teachers
will take, it is essential for the instructors and course developers to accurately assess the needs and wants of teachers in classroom assessment practices.

**Organization**

The dissertation will begin by defining assessment literacy in general education and narrowing down its meaning into language assessment. Using previous literature as a backbone, I will then present the core issue: the characteristics of current language assessment courses and their effectiveness in improving classroom teaching. Finally, the findings from the study will produce a suggestion for teacher educators in language assessment in developing a helpful course for language teachers.

**Terms**

**Assessment.** This study will use “assessment” instead of “testing” to include the paradigm shift of the field. The term assessment includes a variety of formal and informal assessments for multiple purposes. The shift looks into assessment as not only being technical and psychometric but also as a social practice (O’Loughlin, 2006).

**Assessment Literacy.** Teachers’ understanding of what assessment methods to use and when to use them) and having the ability to communicate the assessment results to students and other stakeholders (Stiggins, 1999).

**Language Assessment Literacy.** Teachers’ knowledge about assessment combined with language-specific competency (Inbar-Lourie, 2008).

**Assessment Culture.** This means, “educational evaluation practices that are compatible with current ideologies, social expectations, attitudes and values” (Inbar-Lourie, 2008, p. 285).
Language assessment course. This refers to courses that cover assessment for language teachers. All pre-service and in-service assessment-related courses fall under this category.

Instructors. Instructors are people who teach the language assessment courses.

Student Teachers. Student teachers are the students who take the language assessment courses. Both pre-service and in-service teachers are referred to as student teachers in this study.

K -12 Student Teacher. These are student teachers who are currently or will teach in a K – 12 setting.

Adult Student Teacher. This refers to student teachers who are currently or will teach in adults.

Language Testers (LTs). These are instructors whose main interest area is in Language Testing. Language testers have received their advanced degrees in this area and majority of their work is related to testing.

Non-Language Testers (Non LTs). Non-Language Testers are instructors whose primary research interest is not in language testing but who have experience in testing related areas and have taught or are currently teaching language assessment courses.

Stakeholders. All people who are influenced by language assessment; students, teachers, parents, test developers, college instructors, school administrators, policy makers…etc.
Chapter II

Literature Review

This chapter establishes a background of scholarly literature and theory about teacher training in language assessment. It first begins with how teachers are trained and taught about assessment in general education and moves on specifically to the area of language assessment. To help in understanding the overall structure of this chapter, a literature-review graphic organizer is presented (Figure 1).

Figure 1. A Literature Graphic Organizer
Assessment Literacy

Assessment literacy is not a new concept. A typical language teacher understands that he or she needs to be familiar with the basic principles of assessment in order to develop classroom assessments and interpret standardized tests. However, many teachers think that being literate in assessment requires a special knowledge, even though it is more about understanding everyday knowledge of measurement (Popham, 2006). Stiggins (1999) defines assessment literacy as involving teachers’ understanding of what assessment methods to use and when to use them so teachers can gather reliable information/data about student achievement. Assessment literacy also incorporates teachers being able to communicate the results of assessments effectively to students, parents, and other educational professionals (Stiggins, 1999). Inbar-Lourie (2008, p. 389) writes of assessment literacy as “having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results.”

Even though the term is defined differently depending on the researcher, conceptions of assessment literacy typically include the following common features: (a) a basic understanding of alignment of curriculum standards, assessment, and instruction; (b) an understanding of discrete elements of data that teachers utilize to impact instructional decision making; (c) a demonstration of competency in which appraisal methods to use and when to use them; (d) the ability to collect reliable information on student achievement; and (e) the skill to communicate results clearly to stakeholders (Black & William, 1998). For language teachers, being assessment literate in language assessment means possessing assessment literacy skills combined with language-specific competencies (Inbar-Lourie, 2008). In order for teachers to be assessment
literate in their classroom assessment practices it is important they are provided with the appropriate teacher training in assessment.

**Impact of Assessment Training in General Education**

Evidence exists in the literature (Green & Stager, 1986) that assessment training improves the quality of teacher-made assessments. Teachers who have had more training in the field report developing higher-order critical thinking skills items compared to those who do not have the training. However the content and format of assessment training offered to teachers is quite diverse.

Even though teachers are aware they should be assessment literate, many lack sufficient knowledge to appropriately apply assessment literacy to their own teaching. One of the main reasons why teachers are not confident with assessment literacy is because of insufficient educational courses in this area. Schafer (1993) states that in many programs, testing and measurement courses are embedded within other teacher education courses rather than being offered as a separate course. Pre-service teachers enrolled in such programs are most likely learning about educational assessment from instructors with little expertise in the area (Quilter, 1998). Many teachers finish their undergraduate training without completing a course in testing and measurement. Of those teachers who complete a course in educational assessment, few are well prepared to deal with the realities of interpreting, evaluating, and using test results for a variety of purposes (Jett & Schafer, 1992; Ward, 1980).

The initial purpose of assessment courses is to prepare teachers to construct and to use appropriate teacher-made tests for their instructional purposes and to understand how to interpret and to communicate student performance on standardized tests administered in schools (Borg, Worthen, & Valcarce, 1986).
Fifty years ago, Ebel (1961) voiced the following concerns about teacher-made tests. A logical question is whether or not progress has been made on these issues:

1. Teachers rely heavily on their own subjective standards in judging pupils’ achievement.
2. Classroom tests are poorly planned, too short, and lack adequate sampling of the content domain.
3. Test items are ambiguous or contain irrelevant clues to the correct response that are not reviewed by colleagues.
4. Teachers tend to overlook or underestimate the effects of errors on test scores thereby assuming the individual test scores are absolutely accurate.
5. Teachers rarely undertake analyses of their own tests to check their effectiveness in assessing pupils’ progress.

The errors Ebel (1961) stated are because of the lack of training in assessment and having little chance to familiarize themselves with item writing guidelines and rubric training. When teachers choose assessment methods, they usually do so based on convenience and experience rather than on strong theory. Teachers may not always choose the best assessment method that represents the content, but instead pick methods (e.g., short answer, matching, multiple choice, true and false) that are easy to administer, familiar to both themselves and the students, and appear to be objective measurement tools to administrators and parents (Quilter, 1998).

**Concerns with Assessment Courses**

In Stiggins (1993) he noted reasons why teacher training in educational assessment is currently inappropriate. First, Stiggins believes assessment course are focusing on the process rather than the product. Thus, teachers who take the course are more focused on completing the
course rather than developing a meaningful assessment tool. Another reason could be that compared to other courses teachers perceive assessment courses to demanding and difficult. Stiggins states most assessment courses are offered at the end of the program which can give extra burden to the teachers who take them. In addition, he noted teachers have a tendency to make the achievement targets relatively vague to avoid conflicts with parents and other stakeholder. For many teachers, assessment methods are also predetermined by the state or school district which leaves little room of freedom for teachers. Finally Stiggins adds for teachers, assessment courses have been long construed to be irrelevant to classroom teaching and testing.

Although many teachers have had some exposure to test and measurement training either pre-service or in-service, both teachers felt there is still a lack of training in this area, and teacher education programs must take the initiative to provide their students with skills and knowledge in assessing student performance (Plake, Impara, & Fager, 1993).

To resolve the problem with teacher training courses in assessments, experts in this field have called for more pragmatic training in assessment (e.g., Stiggins, 1991; Stiggins & Conklin, 1992). An effective way to educate teachers in assessment literacy is to use a blend of workshops, learning teams or cadres, and individual study (Stiggins, 1999). Adults learn most effectively when sharing lessons learned in collegial learning and support groups (Stiggins, 1999). Assessment is best practiced when teachers work as a team. When a team of teachers meets regularly, this provides an opportunity for interaction and growth among members (Arter, 2001; Stiggins, 1999). Through teamwork, members can find out what worked and why and determine the group’s level of assessment literacy. Opportunities to work collaboratively must take place in a safe learning environment so teachers will take the risk of putting their work in front of their colleagues for the purpose of improving teaching and learning. For teams to work
effectively, they should set clear goals, assess members’ level of assessment literacy, and provide risk-free opportunities to practice over a long-term approach (Arter, 2001). Preference for teamwork was also confirmed in Jeong’s (2007) study. In a survey of 100 secondary English teachers in Korea, teachers wanted to develop test items through teamwork with other English teachers who taught the same grade level and the same textbook. The teachers in the study also reported that open discussion with other teachers was the most helpful activity in developing their assessment skills. “Collegial or collaborative cultures create an environment where teachers can examine and evaluate instructional plans and assessment strategies and then review, discuss, and inquire into each other’s ideas” (Hyer, 2006 p. 25). However, the dilemma with teacher teamwork is that even though teachers value opportunities to work together, exchange ideas, share strategies, and reflect on practices, teachers’ work takes place in an autonomous environment where they rarely share the wealth of their experiences with one another (Guskey, 2003). In some cultures, teaching itself is done in a very private manner—let alone assessment. Even when teachers are working as a team, seniority in the teaching society presents another difficulty. Beginning teachers may feel restrictions imposed by their senior colleagues. This phenomenon may be culturally strengthened, as for example in Korea, where respect for seniors is more valued than other cultures (Jeong, 2007). Professional development for teachers should be intentional, ongoing, and systemic so that it brings about positive change and improvement in teacher practice and student learning (Guskey, 2000).

In addition to the lack of teacher training opportunities in assessment, teachers are also frustrated about the difficulty of applying the content they learned in assessment courses to a classroom situation (Campbell, 2000). Teachers feel that the abstractness of certain topics (e.g., test theory, statistics) makes certain practices taught in the assessment courses seem nonessential
to student learning. Many assessment courses fail to deliver their initial purpose of conveying assessment literacy to teachers because of the disconnect between assessment and the everyday classroom, which as previously noted, is a major motivating force for the current research. In order for teacher education courses to impact the classroom assessment in a meaningful way, ideas must be presented to teachers so they can be easily assimilated or readily applied to their practice in classroom (Hopkins, 2002; Stringer, 2004). “Teacher training in assessment literacy must meet the needs of teachers, be presented in a way that’s immediately applicable to the classroom, and be used for the purpose of improving both teaching and learning in the classroom” (Hyer, 2006, p. 28). Practice and theory need to be integrated, and both students and teacher educators need to be involved in how materials would be used rather than just discussing the issues of materials development. In a study done by Estey (1999) regarding pre-service assessment courses, teacher educators were neutral or uncertain as to whether pre-service courses in measurement provide the kinds of assessment skills teachers actually use in the classroom.

**Teachers’ Past Assessment Experience and Attitudes to Assessment**

Quilter (1998) argues that teachers’ personal experiences with testing play an important role in their current attitudes toward assessment. Teachers who begin with an unfavorable perception about assessment and who have less knowledge about assessment literacy have a tendency to have a negative point of view towards all forms of assessment. Therefore, if teacher educators address teachers’ perceptions about assessment early on in the training course, they may succeed in having a more positive and effective impact on teachers (Quilter, 1998). This finding illustrates the importance of doing needs analysis at the beginning of the course.

In Quilter’s study, teachers who had a positive experience towards testing were more open and encouraging about testing their own students. Teaching experience and assessment
literacy had very small correlations (.26). This indicates that years of teaching do not influence the degree of assessment literacy, on the other hand is more influenced by professional education in assessment. However, assessment literacy correlated with positive attitudes toward classroom assessment. Teachers who have stronger knowledge about assessment literacy were more favorable towards classroom assessments. Instructors should consider teachers’ backgrounds and culture prior to designing a course in assessment. Learners’ previous experiences (either positive or negative) have a direct impact on how they perceive the content in an assessment class (O’Loughlin, 2006).

**Field of Language Assessment**

Language assessment is an area that derived from applied linguistics and measurement theory. Early on, there was not a field specifically labeled as language testing, and as Spolsky (2008) writes, no one was educated as a language tester. Spolsky states, the increase of students from English-speaking countries who want to study internationally has created a need for a language test that could evaluate the language proficiency of the candidate wanting to study in a foreign country. Although the history of how language testing began is beyond the scope of this dissertation, the history influences how language assessment courses are taught these days (for more information on the history of language testing, refer to Spolsky, 2008). Inbar-Lourie (2008, p. 394) states, “The general purpose of language testing and assessment courses is to train language experts in language assessment concepts, skills and strategies, or in language assessment literacy.”

The range of topics covered in language assessment courses is wide since language itself is taught using all four senses. Students read, write, listen, and speak the language, and assessors need to test all four skills either separately or simultaneously. Another difference is the presence
of multiple published language tests. There are many standardized, large-scale tests of language (e.g., ACTFL, TOEFL, IELTS) used for different purposes. Teachers are influenced by commercially made language tests in their teaching and assessment.

**Teacher Training in Assessment for Language Teachers**

Nowadays language teachers are entailed by an increase or demands on in student assessments. While language teachers may or may not like the idea of feeling more pressure regarding assessment, most will agree that very few feel confident about assessment, whether it is developing their own materials or interpreting test results. Under the expanding role of teachers as assessors, there should be a considerable amount of investment in developing resources in terms of professional development and institutional support (Brindley, 1997). As more educational systems place greater assessment demands on teachers, the more support should be provided, and the first step is to provide teachers with appropriate professional development courses.

Emphasis on teacher education in assessment also applies to language classes. Nowadays many pre-service TESOL (Teaching English to Speakers of Other Languages) programs and other foreign language teacher training programs include an assessment class. An assessment course is a requirement for these programs, since it is part of a language teacher’s role. Developing and administrating tests is an important component in a daily classroom. Well-developed tests are expected to accurately measure students’ language ability and, in the long run, promote student learning. Teacher education in language assessment is crucial because assessment is a bridge between teaching and learning, and classroom tests are regarded as mirrors in which teachers and students can see their reflections clearly (Koksal, 2004). Even with the high demand of quality assessment these days, it is difficult to find research done on teacher
education in language testing, which makes it challenging to find the basis for research in this area.

**Language Assessment Textbook, Past and Present**

McNamara (2006) states that in the last decade language assessment has taken the “social turn”; thus, the field is looking into the accountability, standards, ethics, and the role of tests in the society. This epistemological shift has been a major turning point in the field of language testing and also has an impact in the textbook and the language assessment courses. A textbook review of language assessment books done by Malone (2006) revealed the changes of topics from 1967 to 2005. Malone (2006) states that earlier books cover test theory, basic statistics, item development and practicality; more recent versions include current assessment methods, such as portfolios and assessment samples that could be readily used in the classroom. The author concludes that text books evolved with the change of the assessment culture. Although there is not an empirical study in relation to the changes on the characteristics of language assessment courses other than the Bailey and Brown (1996) and Brown and Bailey (2008) survey it is likely that they evolved throughout the years.

In another article, Davies (2008) reviewed past and current trends in the publication of language testing textbooks. By reviewing the textbooks in language testing, we can picture the content taught in the classroom. In this article, Davies talks about two trends in language assessment textbooks. The first is to develop “all-in material,” (p. 328) thus, having all necessary topics covered in one single textbook without referring to outside materials. This trend is consistent with Jeong’s (2008) syllabi review of 30 language assessment courses. The study states that the majority of language assessment courses had one required textbook. The second trend Davies states is the move from the *skills + knowledge* (p. 328) approach to the *skills +
knowledge + principles approach. Davies notes that skills are the methodology (item writing, statistics, tests analysis, test reportage); knowledge is the theory (e.g., validity, reliability, different models of language testing, etc.) in language testing; and finally, principles concerns the proper use of tests regarding ethics and professionalism. Davies’ analysis is consistent with Jeong’s (2008) syllabi review, which shows increasing interests regarding principles. Through the syllabi review, Jeong found that although courses were offered under various departments and course names, there were common features in what a language assessment course should look like. Most courses covered test theory, test critique on productive and receptive language skills, and developing one’s own testing tool as a final project. In this study, Jeong looked into the differences in the course content as a result of the instructors’ educational and professional backgrounds. Instructors who came from language-testing backgrounds included more test-specific content (i.e., test specifications) and statistics, while non-language testers focused more on classroom-based assessment. The result of the syllabi review shows that the crux of language assessment courses is developing a student-made test. This concurred with Davies’ (2008) finding, which considered students carrying out small-scale test projects as the core requirement for an assessment course at the graduate level for a teaching course.

The topics covered in language assessment courses are diverse, but there are areas more preferred by language teachers. According to Borg et al. (1993), topics that teachers favored were practical issues; such as developing classroom assessment tools and evaluating publisher-offered tests for classroom needs. Teachers disfavored topics that were not directly related to classroom needs, such as statistics and standardized tests, a finding that further supports the goals of the current research.
Regardless of the instructors’ backgrounds and the impact of other external or internal factors, common features were found in almost every language testing course (Jeong, 2007). First, most of the courses (70%, n= 21) taught test theory. Second, language tests that assessed productive (speaking and writing) and receptive (reading and listening) language skills, the fours skills, were covered (four skills, 57%, n=17). And the majority of the courses required students to develop their own assessment tools (item writing and test specifications, 40%, n=12). Despite the common features, there was also a lot of diversity among the classes. Courses were taught by instructors from various backgrounds, which resulted in a range of different topics. There is a lack of research on how this diversity impacts the teachers who take the courses in their teaching practices.

Research on Language Assessment Courses

One of the earliest studies done regarding the characteristics of language assessment courses was by Bailey and Brown (1996, and again replicated in 2008). This was a survey with a limited sample of “self-identified and self-selected instructors of language testing courses(p.352).” The purpose of the survey was to “investigate the instructors’ backgrounds, the topics they covered, and their students’ apparent attitudes toward those courses (Brown & Bailey, 2008, p. 351).”

Brown and Bailey’s (2008) study was definitely a starting point, but still gives a quite narrow view of what the courses really look like. First, one limitation of the study is due to the participants. Unlike Brown and Bailey’s survey, which reported that almost all of the respondents had experience in language testing, 40% of the instructors from Jeong’s (2008) syllabi review did not have experience related to the field. This difference may be due to the characteristics of the mailing list Bailey and Brown used for their study. For their survey, they
used the LTRC (Language Testing Research Colloquium) mailing list. LTRC members are likely to be people who come from a language-testing background. In effect, Bailey & Brown (both 1996, and 2008) committed a sampling error: they did not survey people who actually teach language testing courses, but instead, they relied on a survey of language testers who were presumed to provide such instruction. In addition, as Inbar-Lourie (2008) noted, Bailey & Brown’s methodology used for the research has weaknesses. A survey does not seem to be the appropriate method for researching this topic matter, and Inbar-Lourie suggests interviews and observations as a more suitable method. Next, the content of the questionnaire does not fulfill the researchers’ purpose. Brown and Bailey (2008) state that the questions were developed based on discussion with a small number of colleagues in language testing, and feedback was obtained from the same people. This explains why their questionnaire is overly quantitative and highly theoretical. Jeong’s (2008) study reveals that half of the instructors who teach language testing courses did not major in language testing. To these instructors, the questions are likely to be irrelevant; furthermore, such people would not have been captured by the LTRC mailing list solicitation. The general flaw of the Brown and Bailey (2008) study is the belief that language testing preparation courses are taught by language testing specialists. Finally, their questionnaire consists of six subdivisions: hands-on experience, general topics, item analysis, descriptive statistics, test consistency, and test validity. The content covered under item analysis, test consistency, and test validity is advanced and appears to be more appropriate for a measurement course rather than a beginning language assessment course.

In Brown and Bailey’s (2008) study, for test consistency, general theory regarding reliability and general strategies for estimating test reliability was included in most courses, but the degree of coverage varied and detailed methods used for calculating internal consistency
reliability were not covered in 30~40% of the courses. Especially for the open-ended item analysis section, respondents indicated that the content covered in this section is not appropriate for a basic language assessment course. Instructors who teach this course commented that the focus of the course was on practical application rather than statistical analysis.

Despite the limitations, the findings of the students’ perceptions of the language testing courses in Brown and Bailey (2008) gave insight for studies in this area. Instructors felt the students came with a general dislike for tests, which fostered a negative attitude for the course. Contrastingly, instructors stated students developed positive attitudes toward language testing after taking the course. The findings about students’ attitudes would have been more meaningful if they had come directly from the students’ point of view. One of the limitations stated by Brown and Bailey (2008) is in the design of the questionnaire. They admit the biases they bring to the study might have impacted the topics and organization. The questionnaire is highly quantitative, which could be inappropriate for instructors who come from a more qualitative background. As some respondents commented, some topics were better suited for a quantitative research course.

The items of the questionnaire do not appear to be a good match to fulfill the study’s (Brown and Bailey, 2008) initial purpose, which was to investigate characteristics of language assessment courses. The survey consists of five main parts; general topics, item analysis, descriptive statistics, test consistency, and test validity. Among these parts, it was hard to detect a part that was only pertinent to a language assessment class. This implies that the questionnaire can be applied to other subject areas, and a language testing course does not carry many unique characteristics.
A study done by Jeong (2007) looked into the impact of teacher education in classroom assessment. In a survey with more than a 100 teachers in South Korea, the study found that among teacher education opportunities (e.g., discussion with other teachers, textbooks, workshops, etc.) the teachers were the least satisfied with pre-service assessment classes. Teachers found the classes to be impractical and highly theoretical. In terms of the degree of satisfaction, this differed depending on the quality of the course, the teachers’ teaching environment, and teachers’ teaching experience and education. Teachers with an advanced degree (Master’s and higher) who were taught by language assessment professionals expressed higher satisfaction with the courses.

Other studies about language assessment courses have been done by instructors who have taught assessment courses. An important study by Kleinsasser (2005) covers challenging aspects of language assessment courses from the instructor’s perspective. Kleinsasser states that one of the biggest difficulties in teaching a language assessment course is connecting theory with practice. He notes, “The bridge between the (theoretical) class discussions and the final (practical) test/assessment product, however, was not well constructed. Challenges in getting the students to move from theoretical issues to practical ones often surfaced (p. 82).” For the test developing process in his class, students were discouraged about making language tests from the beginning. The students felt the time spent defining constructs, developing and piloting assessment materials, and rewriting and rethinking the various assessment tasks and items was quite burdensome, since many felt this is not the typical process they go through in a real classroom situation. However, the group work process encouraged them to include various stakeholders’ perspectives in test development and widen their views of testing.
Along with Kleinsasser’s study, another empirical study looked directly into language assessment classes. In O’Loughlin’s (2006) study, the author examined two students in a postgraduate language assessment class. The students came from different backgrounds (one was from China, and the other was a local Australian student). The researcher shows how students’ prior experiences in assessment can impact their learning attitudes in an assessment class. This study is meaningful in the sense that it gave an in-depth description of how the students digested the content of language assessment. Wei-Yen, a graduate student from China who has been assessed under a large-scale standardized test, is concerned about the objectivity of the test and is also interested in using test results in making definitive judgments. On the other hand, Roula, who is a local student, is more familiar with classroom-based assessment in the Australian context and is resistant to the notion assessment, should be “objective” as possible (p. 78). She feels some types of assessments, especially teacher-based assessments, are always subjective to some degree. This study offers a clear contrast between the influence of individuals’ cultural background and professional experience in assessment.

Although the study succeeds on giving a student’s perspective of assessment in general, especially in terms of the cultural differences, it does not go into depth in describing how the course itself influenced students’ perspectives of assessment and how it will impact their future teaching. It is rather focused on describing the student’s previous thoughts of assessment. The course does not seem to directly impact the improvement of language assessment skills, but acts as a channel to refresh their memories of previous assessments. The two students raise topics covered in the classroom (e.g., assessing listening), but do not write about any activities done in the classroom. The students in O’Loughlin’s study point out that one of the biggest achievements of the course was helping students develop a critical eye regarding tests. Before the assessment
class, the students did not challenge the validity or reliability of standardized tests; however, through the class they learned that every test has limitations. O’Loughlin writes that through the course, the participants achieved the main aims (i.e., developing a sound understanding of key concepts in language assessment, critically evaluating existing assessment instruments, and developing or adapting an assessment instrument). The evidence for the students’ achievement, except for the assessment instrument, relies heavily on the students’ self-reports, which threaten the quality of the evidence. The study would have been stronger if it measured the students’ literacy skills in addition to self-reports. Also, one of the weaknesses of the study is that it is challenging to know if the content acquired in the course impacted the students’ assessment in a classroom situation.

**Statistics and Language Assessment**

It seems difficult to think of assessment without raising the topic of statistics. Testing is generally associated with numbers, and so are the assessment courses. However, even though there is little doubt about the power of statistics in assessment, there is a concern regarding whether to teach this in an introductory language assessment course. It has been often said by language testers that the role of statistics in testing has been overemphasized. Because of this, most teachers think of assessment courses as having a large statistics component. However, this was not always the case. Most of the instructors who came from non-testing backgrounds did not cover statistics in their classes (Jeong, 2007). Then there still remains the question of whether or not to teach statistics in the introductory course and to what degree to cover. Unfortunately, no studies have been published regarding this issue.
Importance of Public Policy in Language Assessment Training

No matter how hard we try to make changes within academia, changes in language testing will not occur without changes in public policy (Stansfield, 2008), who writes, “To maximize our reach and influence, we [language testers] need to accept social and political responsibilities” (p. 323). Language testers and teacher educators should be more involved in public policy, join in public policy debates, and publish in popular press, not just technical journals that are not read by the public or policy makers. Stansfield argues, “We need to be available to the mass media, not just to each other. If we speak out by proposing and advocating for fair and just rules, we will increase the prestige of our profession. This will allow us to reach our full potential, both as professionals and as contributors to an organized civil society” (p. 323). The language evaluation process is a socially constructed activity embedded in the local context with teachers, students, and other community members (Leung, 2004; Lynch, 2001; Lynch & Shaw, 2005; McNamara & Roever, 2006). According to Spolsky (2008, p. 301), language testers are constrained by the understandable institutional demands of the bureaucracies they serve to produce clear unidimensional scores that can be quickly and easily interpreted for categorical decisions about the fate of test takers.

As Filer (2000) states, assessment is a “social practice and a social product”; so are courses offered in universities. Teacher education courses are all outcomes of social practice and a social product. Assessment course are offered because of a need in society that views teachers as assessors. One of the unique characteristics of the assessment culture is the contradictory nature of the testing culture. Inbar-Lourie (2008, p. 388) writes that these two cultures are non-compatible, especially in the case of the practitioner who promotes classroom assessment and also is concurrently required to accept the testing culture (i.e., high-stakes standardized testing)
laid out by external authorities. Inbar-Lourie calls the current phenomenon the “dual complex reality.”

**Summary**

The literature related to language assessment courses emphasizes the importance of assessment classes for pre-service and in-service teachers. It also documents the premature status of assessment courses. All research found in the literature comments on the content and method distance of assessments used in the classroom. Most of the research is limited to a few case studies, which fail in giving a broader view of language assessment courses.

The literature also stated the discrepancy between teachers’ and instructors’ points of view on what should be taught in assessment classes. Depending on the instructors’ educational and professional backgrounds, some focused heavily on statistics and educational measurement rather than topics that were directly related to everyday classrooms. Likely, most teachers find theoretical topics irrelevant to teaching and preferred more practical content (Schafer, 1993; Stiggins, 1991).

Although it is nearly two decades since Stiggins (1993) stated the problems with assessment training, “It [assessment training] is regarded as irrelevant, technically complex, academically demanding, and a waste of valuable credit hours” (p. 40). Unfortunately, the situation has not changed much now. Many teachers still finish assessment courses with a negative notion similar to those more than 17 years ago.
Chapter III

Methodology

The study was conducted using a mixed methods approach; with the use of different paradigms as guidance develop specific methods of mixing. The study consisted of four phases with a combination of qualitative and quantitative methodologies. Phase 1 (syllabi review) and 2 (pilot interviews) were conducted as pilot study, and Phases 3 (main survey) and 4 (follow-up interviews) were the main studies of the dissertation.

Research Questions

1. What are the characteristics (i.e., content, structure, objectives) of introductory language assessment courses? (What factors determine the differences among courses?)

2. What constitutes evidence of an “effective” language assessment course for classroom teachers? (Is there a difference between instructors and teachers?)

3. How effective are language assessment courses for language teachers’ everyday classroom practices? (Is there a difference depending on the teachers’ years of teaching experience, final academic degree, and teaching environment?)

My Mixed Methods Stance

For this study, a mixed methods approach derives from both a philosophical and methodological sense. My study began with a set of research questions rather than a particular theoretical framework which reflects the alternative paradigm stance (Greene, 2007). I choose the method that best answered my research questions which is based on the form of pragmatism (Johnson & Onwuegbuzie, 2004). The small number of previous empirical studies made it difficult to detect an established framework regarding my research topic. None of the previous studies (Bailey & Brown, 1996; Brown & Bailey, 2008; O’Loughlin, 2006; Kleinsasser, 2005) fully described the characteristics of Language Assessment Courses, and I believe one of the
reasons to this limitation was the restrictive methods used for the studies. Therefore, it was only natural and appropriate to investigate the study using a mixed methods approach. However, the mixed methods approach is not only about the methods. While collecting and especially analyzing the data, I found I was also using the mixed methods design in a philosophical sense as well. My understanding of the data, interacting between both qualitative and quantitative data built a research philosophy that required a mixed methods way of thinking. The mixed methods design first began for practical purposes but also had an effect in the way I view the world and do research.

My rationale for mixing is to better understand the social phenomena (p.20, Greene 2007) of Language Assessment Courses. The nature of the study required a quantitative component. To understand the overall features and characteristics of Language Assessment Courses, a large scale survey was a must. Also, to encompass different groups of participant’s thoughts and ideas of the course, a quantitative research method was needed. Each assessment course exists upon unique conditions; every classroom context is different, the instructor comes with different values and so do the student teachers. The complexity of the language assessment course and the people associated with it makes it difficult to fully understand the course without a deeper investigation for each situation. The combination of both quantitative survey and qualitative interview helped to understand the nature of the Language Assessment Courses. From this research perspective, I believe I approached the mixed methods design with a complementary strength sense (Greene, 2007) in addition to the alternative paradigm stance. Researchers who favor the complementary stance believe different paradigms are perceived to be important in different ways “with a respect for the importance of maintaining the integrity of any given methodological tradition” (Greene, 2007, p. 77). For my study I value the importance of
both the quantitative characteristics of the survey and qualitative aspects of the interview by utilizing each method’s unique features and strengths.

The Mixed Methods Approach for this Study

In this study, a mixed methods approach is used for purposes of development and complementarity. The type of mixed methods design employed in this study is an integrative design (Greene, 2007). The study takes the form of an iterative design where all phases interact with each other throughout the process. Although the implementation of each phase was done separately, the analysis was closely linked to the next phase of the study. Since one purpose of mixed methods is developmental, constructing the form and structure of the next instrumental tool was the first priority in the data analysis stage. For example, while analyzing the pilot interview data, at the same time, I began developing the survey questionnaire. The findings from each phase became the base for the development of the next phase. The developmental nature of the design strengthened the instruments used for this study. It is important to go through this developmental process because little of the published literature discusses the nature of language assessment courses, and even fewer published survey or interview questions currently exist in this subject area.

For the final phase of the study (follow-up interviews), the purpose of using mixed methods is for complementarity. “In a complementarity mixed-method study, qualitative and quantitative methods are used to measure overlapping but also different facets of a phenomenon…” (Greene, Caracelli, & Graham 1989, p. 258). The findings from the interviews complemented the findings of the survey and the two methods were conducted sequentially.

Since there is a paucity of research regarding language assessment courses, the first step (Phase I) of the study was to investigate how these courses were taught. The most accessible and
appropriate way was to conduct in-depth participant interviews similar to those that would be conducted in the main study. For the interviews, I focused on collecting as many different perspectives as possible from both student teachers and instructors. I interviewed instructors who were physically available using face-to-face interviews; for those who were not, I conducted email interviews. The student teachers that were interviewed had different teaching experiences, genders, nationalities, and were taught by different instructors at different time stages. The initial purpose was to ensure variety and then to look for common areas among interviewees.

Through a blend of methods, the study seeks to elaborate, enhance, illustrate, and clarify the results from one method (quantitative) with the results from another method (qualitative) to increase meaningfulness (Greene, 2007). In a complementarity study, two methods look into overlapping areas from different facets of a phenomenon. By collecting both qualitative and quantitative data, a more complete and contextual explanation for the research questions could be derived (Greene, 2007).

Table 1

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Methods</th>
<th>Purpose</th>
<th>Mixed-methods purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>Syllabi Review:</td>
<td>To get an overview of the characteristics of</td>
<td>Development</td>
</tr>
<tr>
<td>(May 2008)</td>
<td>Qualitative</td>
<td>language assessment courses</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Generate the population to interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Generate topics and questions to answer for pilot</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(continued)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interview</td>
</tr>
</tbody>
</table>

Table 1 (continued)
Table 1 (continued)

<table>
<thead>
<tr>
<th>Phase II</th>
<th>Pilot Interview: Qualitative</th>
<th>Generate survey questions from the target population</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>(October 2009)</td>
<td></td>
<td>Increase validity and reliability for the survey</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase III</th>
<th>Survey: Quantitative</th>
<th>Investigate the characteristics of language assessment training</th>
<th>Complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Summer 2010)</td>
<td></td>
<td>Find the effectiveness of language assessment training</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cross-check the results with the survey</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase IV</th>
<th>Follow-up Interviews: Qualitative</th>
<th>To explore and build on survey findings</th>
<th>Complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Fall 2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Phase I: Preliminary Syllabi Review**

The preliminary syllabi review (n = 30) was done in May 2008. The purpose was to find out the topics covered in language assessment courses and the external (department, position in the program, and geographic factors) and internal (instructor, target student population) factors that influence the content and structure of the course. A convenience sampling method was used for the collection of the syllabi and was gathered through the Google Internet search engine. A combination of key words ‘language, testing, assessment, syllabi, course’ were used to search syllabi. Through this search 30 syllabi were collected. They were from 27 different schools (one school had more than 1 language testing course and another was taught by different instructors) and 30 different instructors. Twenty-four were from English speaking countries (U.S.A., Canada,
UK) and six were from non-English speaking countries (Korea 3, Japan 2, Hong Kong 2, Puerto Rico 1). All the syllabi were written in English.

The results of this syllabus-review study showed that only half of the language assessment courses were taught by professionals who specialized in language assessment. The others were taught by instructors who had backgrounds in foreign languages (e.g., Spanish, French, German, etc.), special education, bilingual education, quantitative research methods, and rhetoric. Different educational backgrounds of the instructors translated into different structures of the course. First, courses taught by LTs offered a variety of topics, such as computer adaptive testing, ethics in testing, statistics, and performance assessment. When these were compared to courses taught by non-LTs, these instructors without testing backgrounds focused more on delivering the basic theory of testing and emphasized classroom assessment. Statistics or measurement was taught by 13 instructors, 10 of whom had testing backgrounds. Only three non-Language-Tester instructors covered statistics or measurement in their courses.

Findings from the preliminary syllabi review informed the type of instructors and teachers I needed to interview for the pilot. The distinction of Language Testers and non-Language Testers made it clear that it was important to interview both groups and also their students.

Phase II: The Pilot

Following up on the results from the preliminary survey, in October 2009, I conducted a pilot interview study with four instructors and ten student teachers. Two of the instructors were language testing professionals, and the other two were from different areas of language teaching (Second Language Acquisition and Second Language Literacy). Three of the instructors taught at a large Mid-West university and the other one in Korea.
This research topic is a relatively unexplored area in language testing; therefore, it was necessary to use the results of the pilot interview to design a stronger subsequent phase of the study. There is only one published survey (Bailey & Brown, 1996; Brown & Bailey, 2008) regarding language assessment courses; thus, I approached the topic from a bottom-up approach. The pilot interviews raised issues and concerns about the topic and also increased the reliability and validity of the survey. Another reason for the interviews was to get a perception of the participants’ thoughts and attitudes towards the courses. A researcher cannot directly observe the thoughts of the participants, and the closest way to approach a person’s perception is through an interview (Patton, 2002).

**Pilot Study Participants.** For the pilot, I interviewed subjects who had characteristics similar to those of subjects in the main study. Ten current or past ESL/EFL teachers from a large Midwest state university were interviewed. I targeted subjects for variety in the pilot; therefore, years of teaching experience varied from less than half a year to 12 years. Teachers came from diverse backgrounds in native language (e.g., Arabic, Korean, Russian English), and the time period during which they took the course. Combinations of purposeful and convenient sampling were used for the pilot. The majority of the participants (n = 7) were current graduate students either in the MA TESOL program (n = 3) or a doctorate program (n = 4). Three were full time language teachers. All teachers except two were graduates of or currently enrolled in the MA TESOL program. All teachers except one had formal teaching experience before attending the language assessment course, and all teachers taught after exiting the course. Many teachers (n = 9) took the course while they were concurrently teaching an ESL/EFL class.

**Instrument.** The pilot interview questions were developed based on the results of a previous review of literature and the preliminary syllabi review. Three doctoral students
majoring in Language Testing reviewed the interview items for content, and another doctoral student and current ESL teacher reviewed the items for content and language. Revisions were made based on the reviewers’ feedback.

**Pilot Process.** Participants answered around 12 semi-structured interview questions (Appendix A). Copies of the current syllabus from Fowler and the past syllabus from Clarkson were provided to the teachers who took his or her class, respectively. I could not provide the syllabus for Lee and Morgan at the time of the interview. Although Fowler’s current semester syllabus may not exactly match the syllabus teachers used at the time they took the course, participants commented that it was helpful in recalling the course. In addition to the course syllabus, a list of commonly covered topics (Appendix B) in language assessment courses that was developed from the preliminary syllabi review was also presented to the participants at the time of the interviews. Table 2 below provides a description of the instructors and teachers. For a full description of the participants, refer to Appendix C. A full description of each course and its instructors are in Appendix D.

**Table 2**

<table>
<thead>
<tr>
<th>Instructors (4)</th>
<th>LT/non LT</th>
<th>Main target audience</th>
<th>Level</th>
<th>Teachers (number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fowler³</td>
<td>LT</td>
<td>MA TESOL</td>
<td>Grad/UG</td>
<td>Justin, Brian, Sumin, Eman, Kathy, Derrick, Debra (7)</td>
</tr>
<tr>
<td>Clarkson</td>
<td>Non LT</td>
<td>MA TESOL</td>
<td>Grad/UG</td>
<td>Terry (1)</td>
</tr>
<tr>
<td>Morgan</td>
<td>Non LT</td>
<td>Bilingual Ed. MA</td>
<td>Grad</td>
<td>Sumin, Jungmin (2)</td>
</tr>
</tbody>
</table>

³ For confidentiality purposes, the names of the participants are pseudonyms.
Table 2 (continued)

<table>
<thead>
<tr>
<th>Lee</th>
<th>LT</th>
<th>In-service Secondary</th>
<th>Grad</th>
<th>Phil (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>English Teachers in Korea</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pilot Study Results**

**Most Beneficial vs. Most Difficult.** The primary goal of Fowler’s class was to instill confidence in teachers that they can do language testing. Fowler thought the most beneficial part of his course for teachers would be the test specifications workshop. This workshop is typically offered at the middle of the semester. Students work in groups on a topic of their choice about language assessment. Fowler stated that the activity was important since all language teachers, regardless of language, location, or grade level, will have to write tests. He believes this is the activity that really grouped the teachers together. The benefits of the test specifications workshop were also illustrated in interviews with most of the teachers who took the course from him. This was universal regardless of the time the teachers took the course, their levels of teaching experience, and other individual factors.

**Most Difficult Areas.** Validity and statistics were often the two areas teachers had trouble understanding in the language assessment course. Teachers reported that the theoretical concept of validity was difficult to comprehend and felt distant from classroom teaching. Language teachers, who hardly come with a background in statistics, had trouble understanding the concepts. For Fowler’s class, basic statistics (e.g., bell curve) and Item Response Theory (IRT) is taught over a 2-week period. During this period, he mainly lectures the concepts related to language testing. Fowler does not expect the teachers to fully grasp the knowledge, but strongly feels it is important to cover this in a Language Assessment course. He states that
statistics is not as important as other parts of the course (e.g., the test specifications workshop); however, it is a topic that cannot be avoided or overlooked. He believes the way he teaches statistics for the introductory language course is an ideal way to convey the challenging concepts to the teachers. He commented that not including statistics in a language assessment course is a mistake, because it is already out in the teaching environment, and interpretations of the tests are achieved through a statistical determination. Two of the other instructors, Clarkson and Lee, shared Fowler's beliefs regarding teaching statistics. Both believed it was important to teach statistics in the language assessment course. Lee felt strongly about including a rigorous background in statistics for language teachers. She felt it was essential for teachers to know statistical concepts in order to understand articles and to do their own research. On the other hand, Clarkson came with a much softer position. She contemplated teaching statistics for a language assessment course before teaching the class, but felt it was necessary in a limited way. She did not feel teachers need to know measurement theory such as IRT, but that they at least need the skills to interpret correlation. Compared to the three instructors who believed in the need for teaching statistics, Morgan had a different opinion. Morgan was not against teaching statistics, but simply felt there was not enough time to cover it in the course period. She commented that she would take the time spent on teaching statistics to teach other “numerous issues of language assessment in the classroom and schools.” This implies that teaching statistics is not a priority for Morgan.

Assessment Process. During the pilot, teachers commented on how the course was assessed. Teachers learned from the instructor’s assessment process and applied it to their own teaching. For an assessment course, it is important for instructors to be aware of the methods they use in their course, because teachers learn and critique the assessment used for the course.
For instance, in his class, Fowler allowed teachers to re-write their assignments. This is clearly stated in the syllabus, “Re-grading will never lower your grade.” This statement in the syllabus gave comfort to the teachers and helped them improve their work. Eman, a writing ESL teacher, mentioned after the assessment class that she gave the re-write option to her students. Eman found this to be an excellent learning opportunity for her students. According to Fowler, the reason for allowing re-write in his class is to promote “positive washback.” Justin also liked the opportunity of the re-write option.

**Gap Between Student Teachers and Instructors.** One of the common reactions across all teachers was that the language assessment course was more theoretical than expected. Teachers came with an expectation of developing tests that would be actually used in their classrooms; however, none of the courses fulfilled this need. Teachers expressed that the course had more weight on theory rather than practice and was challenging to connect with their classes. Clarkson mentioned that teachers had difficulty making the connection of each piece in her class, especially those who did not have prior experience in assessment courses (which were the majority) who did not clearly understand the purpose of the content. She remarked that teachers struggled to contextualize it in relation to the classes they teach. This was a universal reaction among teachers. Teachers from Fowler’s class also reported that the content of the course was abstract and murky. One teacher even indicated that the course felt more like a philosophy class rather than a language assessment class. Teachers appreciated knowing the philosophy of testing, but did not necessarily find it helpful in their teaching practices.

All teachers wanted more hands-on, practical activities for the course. They were interested in topics that could be directly applied to their teaching, such as giving constructive feedback to students, developing writing rubrics, piloting the tests they developed, revising
teacher-made classroom tests, etc. In contrast to teachers’ needs, instructors were not highly interested in making their courses more practical. They were aware that teachers find their courses theoretical, but did not feel the need to change directions. When asked about changes they want to bring to the course, none of the instructors raised issues regarding practicality. This shows a gap between the needs of the teachers and the wants of the instructors. There could be several reasons for this discrepancy between the two groups. First, instructors may not be aware of the needs of the teachers. In the interview, Clarkson said she did not know what the teachers wanted from her course. Next, instructors may believe it is important to teach theoretical concepts even though teachers cannot directly find the connection to classroom teaching. The learning process, in many cases, is not very explicit. Teachers may not realize the benefits of learning certain topics (e.g., validity, statistics) and how they can implicitly help in their teaching and assessment practices.

Teachers and instructors from the assessment courses shared different ideas of the courses. Eman said that, as a teacher, she thinks of testing as part of teaching; for her, this was not a separate area. However, she felt the instructor approached it from a much bigger picture. She acknowledged that, being a professional in language testing, the instructor naturally emphasizes the importance of testing, but she still felt it is one area in teaching language, not the crux of it.

**Implications for the Main Study**

The pilot interviews brought up valuable information and considerations for the main study. First, teachers may have difficulty in recalling specific information of past language assessment courses. Even during the pilot, a few teachers stated they could not vividly remember the features of the course. This implies that for the main study, the survey should cover current beliefs and needs of language assessment training along with past experience. Next, the pilot
revealed that the definition of what constitutes “effectiveness” in a language assessment course was highly subjective. Three teachers who took the course with the same instructor at the same time had very different thoughts of the class. One teacher exited the course with an extremely negative perception of language testing; on the other hand, the other two found the class beneficial to their classroom teaching. This finding show that researchers cannot purely rely on teachers’ self-reports in determining whether the course was helpful or not. Although a teacher might have disliked the course, it could still have been implicitly helpful in their classroom teaching without their even acknowledging it. Therefore, it is necessary to clarify the definition of “effectiveness” in a more objective way. In addition to using teachers’ self-reports, for the main study, I propose to include a short test that assesses teachers’ language assessment ability. He or she may report negative views of the course, but still could have learned assessment skills that benefited their classroom teaching. In this case, I believe it is difficult to say the course was not effective. Finally, for the main study, I proposed to look into different points of views on teaching statistics. The preliminary syllabi review informed that non Language Testers rarely included statistics in their course curriculum, and this was confirmed to some degree from the pilot study as well. In the main study, I included questions regarding whether it is important to teach and learn statistics in an introductory language assessment course.

**Limitations of the Pilot**

All the teachers that were interviewed were pursuing or had higher degrees. The lowest final degree was an M.A. in progress, and four teachers were pursing doctoral degrees. This may not be representative of the population for the main study. In addition, there is always the issue of the researcher’s objectivity. I had personal connections with the majority (n = 12) of the participants. Some were previous classmates, colleagues with whom I have worked, and professors with whom I have taken classes. I acknowledge the biases I bring to the study. Being
a doctoral student majoring in Language Testing and a current ESL teacher, my personal beliefs about teaching and language testing may have influenced the interview process.

**Phase III: The Survey**

Why an On-line survey? For the third phase, an on-line survey was administered to the participants. The reason for doing an on-line survey is that this is the most efficient way to gather information about language assessment courses offered throughout the world. In addition, Bailey and Brown (2008) successfully conducted an on-line survey for one group of the target subjects: instructors who are language testers. An on-line survey is a cost-effective way to overcome geographical limitations and financial barriers. For this study, it is especially appropriate to conduct an on-line survey since almost all of the participants will be familiar with the technology (i.e., using the internet) and have access to electronic devices (i.e., computers with internet connection).

Despite the benefits of using an on-line survey, there were several concerns. First is the unpredictable response rate. Since little previous research has been done using on-line surveys for the target population, it was difficult to predict how many will actually participate in the study. I tried to overcome this limitation by sending reminder emails for the participants and reaching out to as many professional organizations as possible. Second, it is challenging to screen the participants for an on-line survey since the researcher is not present when the surveys are conducted. In response to this concern, I included a detailed description of the criterion to take part in the study (e.g., Instructors who are or have taught a Language Assessment Course, Student teachers who have taken at least one Language Assessment Course). In addition, I included a question early in survey that checks the qualification of the participant. The survey questions were developed primarily based on the results of the pilot interview. The pilot
interviews consisted of demographics, characteristics of the course, and the effectiveness of the teaching.

**Research Questions for the Survey**

1. What are the characteristics of language assessment courses?

   1.1. Is there a difference in course content between language testers and non-language testers? (Assumption: Courses taught by language testers will focus on large-scale assessment compared to non-language testers.)

   1.2. Is there a difference in the content and structure of the course depending on the years of instructor’s field experience? (Assumption: Instructors who have substantial [10 years or more of] classroom language teaching experience will put more emphasis on classroom based assessment activities)

   1.3. Is there a difference between language testers and non-language testers in teaching statistics in an introductory language assessment course? (Assumption: Non-language testers will not favor teaching statistics.)

   1.4. Would teachers have taken the assessment course if it was not a requirement? (Assumption: Teachers would have not likely taken the course.)

2. What constitutes evidence of an “effective” language assessment course for classroom teachers?

   2.1. Is there a difference in what attributes constitutes as an effective assessment course between instructors and teachers? (Assumption: Teachers will define practicality as the most important factor.)

3. How effective are language assessment courses for language teachers?

   3.1. Is there a difference in degree of satisfaction depending on the instructor’s educational background (language testers and non-language testers)? (Assumption:
Teachers will perceive courses taught by non-language testers to be more effective in classroom teaching practices.

3.2. Does the degree of satisfaction of the course differ depending on the teacher’s background (years of teaching experience, time of the course, teaching environment, personal assessment experience, and final educational degree)? (Assumption: Teachers who have longer teaching experience, recently took the course, teach in a low-stakes teaching environment, possess a positive personal past assessment experience, and have an advanced degree will find the course to be more useful.)

Participants

The participants for the main study consisted of two main groups: instructors who teach language assessment courses and student teachers who have taken a language assessment course. The instructor group was divided into two subcategories; Language Testers (LTs) and non-language testers (non-LTs). Language testers are people whose primary research interest is in areas of language testing. Non Language Testers are defined as people whose primary interest is in other areas of language teaching (e.g., Second Language Acquisition) but have had experience in language-assessment-related activities (e.g., developed standardized tests, worked with a testing agency, etc.)

The student teachers were also divided into two subsections depending on the student population they teach. The first group was K-12 student teachers. The other group is Adult student teachers. This group of student teachers teaches adults in a degree or non-degree setting in any place in the world.

Having both perspectives (instructors and student teachers) provided a more complete picture of the language assessment and its effectiveness. Previous studies (Bailey & Brown, 1996; Brown & Bailey 2008) only viewed the course from the instructor’s eyes, which limits the
perspectives of the course. Effectiveness of the course was asked from both sides as well. The
definition of effectiveness varied depending on the group and also the individual.

Table 3

*Descriptions of Participants: Sample List of Professional Organizations (Full list in Appendix F)*

<table>
<thead>
<tr>
<th>Instructors</th>
<th>Source</th>
<th>Context</th>
<th>Student</th>
<th>Source</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers</td>
<td></td>
<td></td>
<td>Bilingual</td>
<td>TESOL, TESL-L,</td>
<td>Local</td>
</tr>
<tr>
<td>Language</td>
<td>LTEST-L,</td>
<td>Local &amp;</td>
<td>(K–12)/ESL &amp;</td>
<td>AAAL, CALICO,</td>
<td></td>
</tr>
<tr>
<td>Testers/non</td>
<td>MwALT,</td>
<td>International</td>
<td>EFL</td>
<td>CATESOL, GATESOL,</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>ECOLT,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testers</td>
<td>SCALAR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KELTA,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JLTA,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EALTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Survey Pre-test**

As a component of the main study, the survey draft (Appendix D) was pre-tested with 12 people; 10 of the participants were teachers and instructors who took part in the pilot study and two were advanced graduate students who are pursuing a doctoral degree in language testing.

The purpose of inviting participants from the pilot was to do member checking; to ensure trustworthiness and increase the validity of the survey items. Lincoln and Guba (1985) define member checking as having participants who originally took part in the study examine the data, interpretation, and conclusion of the study. They state this step is the “the most crucial technique
for establishing credibility” (p. 314). For the pre-test, participants tried out the survey and commented on the process, content, and organization of the items. They also gave feedback on technical difficulties and degree of user friendliness of the on-line survey. The two doctoral students who have background knowledge in language assessment and quantitative data analysis gave advice on scaling and clustering responses. The survey was revised in terms of content and organization following the feedback from the pre-test.

Survey Administration Process

The survey participants were contacted via a professional organization (sample list in Table 3, full list in Appendix F) they are affiliated with and through personal emails. First, I contacted the personnel who are in charge of managing email lists from professional organizations (e.g., Teaching English to Speakers of Other Languages, ltest-l). I then provided these personnel a brief description of the purpose of my research along with the IRB approval letter and requested an approval to use their email list. I sent a letter requesting approval to the list-serve managers twice with a week window in between. For organizations that allowed members to directly post messages to the members, I sent out the messages via email. Participants were given a month to finish the survey. I sent another reminder email after a week to increase the participation rate. I sent out a final email asking for participation three weeks after the initial email.

Participants especially instructors were also recruited via personal email sent by myself. Using the Google search engine I looked up Language Assessment Courses offered in various universities and searched for the instructors who taught the course. Fifty four personal emails addressed to the specific instructor were sent using this method.
Instructors and student teachers took the survey at a time and location convenient to them. The website that provided the on-line survey was 'Surveygizmo'. A consent form was presented at the beginning of the survey that described the purpose of the study, the risks and benefits, contact information and the duration time of the survey. The participants were allowed to stop the survey at any time without any penalty or loss of benefits. After the survey, the answers were quantitatively analyzed. The survey lasted approximately 15~20 minutes.

Each participant submitted only one survey. Multiple responses from the same participant were restricted based on the participant's IP address. Respondents who submitted multiple surveys using the same IP address were not counted in the final data collection.

The survey consisted of demographic information and 4 point Likert scale items. The scale asked about degree of time (1=Hardly any time, 2=Sometime, 3=Pretty much time, 4=Extensive time) and importance of the topic (1=Not important, 2= Somewhat important, 3=Important, 4=Extremely Important). Items that asked about how much time was spent on each topic during the course were given the option ‘Not Applicable’. This response was coded as missing data for the statistical analysis; however, it was reported separately from real missing data in the descriptive statistics.

The on-line survey was first launched on July 2, 2010 and was closed August 20, 2010. A total of 384 people completed the survey. In addition, there were 342 partial surveys and 386 abandoned surveys. Partial surveys are questionnaires that were not completed. The majority of the partial surveys only answered the first question on the consent page (Would you like to take the survey?) on the cover page. All partial surveys were checked to see to what extent the survey was completed. Surveys that were close to completion (missing only a few items) were analyzed as complete surveys. Abandoned surveys are those that people clicked on the link but did not
answer any questions. These were not included in the analysis. Instructors and teachers were primarily contacted through professional organizations’ list-serves. In the email invitation a link to the survey was provided. Participants were also directly contacted via email obtained through school websites and syllabi available on the web. After the first email, a follow up email was sent two weeks after the initial invitation. Depending on the response rate a 3rd follow up email was sent three weeks after the first invitation.

At the beginning of the survey instructors were asked to fill in a set of demographic questions regarding their education background and experience. The purpose of this was to identify; who were the Non-Language Testers, what did they study, what kind of background experience they have and how long has it had been since they received their final degree. The dominance of the data depended on the type of the research question. For example the main source of the first research question, *What are the characteristics of current language assessment courses?*, is derived from the survey findings. General overview of instructor and student teacher survey provides answers to what a language assessment course looks like. Interview findings support the survey results. For the next two research questions, *What constitutes evidence of an effective language assessment course for classroom teachers? How effective are language assessment courses for language teachers’ everyday classroom practices?*, neither one source of data dominated the other. Both have equal weight of importance. Each data set, qualitative or quantitative is equally informative and can be presented separately or together depending on the focus of the study.

**Sampling**

The participants from the pre-test or pilot were not included in the main study. There is a possibility that people who have already been exposed to the instrument may respond differently than participants who have not had the experience.
Survey Data Analysis Plan

The survey data was entered in SPSS v. 17 for Windows, and first it was explained using descriptive statistics to organize and summarize the data. Next, to investigate the effectiveness of the course, among differences between groups (instructors and student teachers) t-tests were conducted.
Table 4

*Map of Research Questions and Survey Items*

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Teacher Survey</th>
<th>Instructor Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q 1. What are the characteristics of current language assessment courses?</strong></td>
<td>Q 4 ~ Q7, : LT vs. non-LT demographics</td>
<td>Q 17: content covered in assessment course</td>
</tr>
<tr>
<td>1.1 Difference in course content between language testers and non-language testers.</td>
<td></td>
<td>Q 17: content covered in assessment course</td>
</tr>
<tr>
<td>1.2 Difference in the content and structure of the course depending on the years of instructor’s teaching experience?</td>
<td></td>
<td>Q 17: content covered in assessment course</td>
</tr>
<tr>
<td>1.3 Difference between language testers and non-language testers in teaching statistics in an introductory language assessment course?</td>
<td></td>
<td>Q 21: importance of field experience</td>
</tr>
<tr>
<td>1.4 Would teachers have taken the assessment course if it was not a requirement?</td>
<td>Q 3, Q 4, Q 5: requirement or elective</td>
<td>Q 4 ~ Q7, : LT vs. non-LT demographics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q 24: thoughts about statistics</td>
</tr>
</tbody>
</table>

(continued)
Table 4 (continued)

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Teacher Survey</th>
<th>Instructor Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2. What constitutes evidence of an effective language assessment course for classroom teachers?</td>
<td>Q 14 ~ Q 17: most helpful, least helpful topics</td>
<td>Q 16: needs analysis</td>
</tr>
<tr>
<td></td>
<td>Q 20: course expectations</td>
<td>Q 18: most helpful topic</td>
</tr>
<tr>
<td></td>
<td>Q 21: needs analysis</td>
<td>Q 20: important topics for classroom teachers</td>
</tr>
<tr>
<td></td>
<td>Q 26: most important factor for teachers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q 28, Q 29: future topics of interest</td>
<td></td>
</tr>
<tr>
<td>2.1 Is there a difference in what attributes constitutes as an effective assessment course between instructors and teachers?</td>
<td>Q 23 ~ 25: perception of the course, degree of satisfaction</td>
<td>Q 24: important factor for effectiveness</td>
</tr>
<tr>
<td></td>
<td>Q 27, 28: course evaluation</td>
<td>Q 26, 27: course evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(continued)</td>
</tr>
</tbody>
</table>
### Table 4 (continued)

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Teacher Survey</th>
<th>Instructor Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3.1</strong> How satisfied are teachers in their assessment practices? Does the degree of satisfaction of the course differ depending on the teacher’s background (years of teaching experience, time of the course, teaching environment, personal assessment experience, and final educational degree)?</td>
<td>Q 2 ~ Q11: teacher demographics (years of teaching experience, teaching environment, degree of assessment knowledge, final academic degree)</td>
<td>Q 27, 28: course evaluation</td>
</tr>
<tr>
<td></td>
<td>Q 13: time spent on topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q 18: important topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q 23 ~ 25: perception of the course, degree of satisfaction</td>
<td></td>
</tr>
<tr>
<td><strong>3.2</strong> Is there a difference in degree of satisfaction depending on the instructor’s educational background (language testers and non-language testers)?</td>
<td>Q 3, Q 4, Q 5, Q 6: LT vs. non-LT demographics</td>
<td>Q 27, 28: course evaluation</td>
</tr>
<tr>
<td></td>
<td>Q 13: time spent on topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q 18: important topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q 23 ~ 25: perception of the course, degree of satisfaction</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5

**Data Analysis**

<table>
<thead>
<tr>
<th>Research question</th>
<th>Main data source</th>
<th>Analysis techniques</th>
</tr>
</thead>
</table>
| Q1. What are the characteristics of current language assessment courses? | Instructor survey  
Instructor interview  
Follow-up instructor interview  
Follow-up teacher interview | To test the difference in groups (language testers vs. non-language testers) a chi-squared analysis will be used.  
Specific areas that will be focused on are: course goals, course assignments, course content (especially in teaching statistics), position of the course (elective, requirement) |
| Q2. What constitutes evidence of an effective language assessment course for classroom teachers? | Follow-up instructor interview  
Follow-up teacher interview | During the interview, instructors and teachers will be asked to define “effectiveness” and what represents its evidence (e.g., student achievement scores, better developed test items). Responses will be analyzed by subject groups (instructors vs. teachers) and different demographic background (e.g., teaching experience, teaching environment, etc.) to identify patterns. |
| Q3. How effective are language assessment courses for language teachers’ everyday classroom practices? | Teacher survey  
Follow-up teacher interview | To test the degree of effectiveness, a few questions of the survey will be designed to test teachers’ knowledge of language assessment skills. This will be cross checked with teachers’ self-report of effectiveness. A person-product moment correlation will be conducted. |
Phase IV: Follow-up Interviews

For the final phase of this study, follow-up interviews were conducted to explore and build on survey findings.

**Purpose of the Interviews.** The purpose of using a mixed-methods approach is to combine the advantages of both quantitative and qualitative research methods. The survey gave generalizability to the data, and the interviews presented the details to understand the specific educational context of the participants (Creswell, 2002). This study is targeted for a cross-national population; thus, it is important to have the contextual information in which the participants are situated to fully understand the survey results. The purpose of mixing for the final two phases is to complement the findings from each method. The aim of this study is to investigate the characteristics of language assessment courses and explore their effectiveness to ameliorate teachers’ classroom practices.

**Importance of the Follow-up Interviews.** To obtain a better understanding of the nature of language assessment courses, it is important to approach it from both qualitative and quantitative perspectives. Previous studies that researched this topic only used one research method; Bailey and Brown (1996; Brown & Bailey, 2008) used a survey, and O’Loughlin (2006) and Kleinsasser (2005) took the form of an active research case study. In order to overcome the limitations of previous studies and to extend and enrich the scope of the study, it was necessary to use qualitative interviews as well as a large-scale quantitative survey.

**Interview Participants and Process**

The participants for the interviews were selected through a combination of convenient sampling and purposeful sampling. First, the participants were people who expressed interest in taking part in a follow-up interview after the survey. I contacted those
who volunteered, and priority were given to people who represent a diverse population (geographically, teaching experience, course characteristics). Also people who were responsive (e.g., answered open ended questions, provided detailed descriptions) in the survey were given priority. The final version of the interview protocol was developed after analyzing the preliminary survey data. A sample of the follow-up interview questions is in Appendix A. Specific questions varied depending on the participant’s context. Interviews were conducted in English. Before the interview, participants received a copy of the survey response to refresh their memory. For the instructors, I asked them to send a copy of their syllabus and 10 were collected. For these courses, I asked the instructors specific questions related to their course.

The interviews were done over the phone (land line or Skype). Via the on-line survey, participants provided email address by which they could be contacted. Through email correspondence instructors and student teachers shared their Phone Numbers or Skype IDs. The interviews took a semi-structured form, and the guiding questions for the interviews were sent to the participants via email prior to the interviews. A written consent form which has been approved by the University Institutional Board was obtained before the interview process. The interviews took about 40–50 minutes and had around 10 questions. In the interview, participants were asked to describe the reason and elaborate on their choice on specific survey questions.

To ensure accuracy of the interview data, all phone interviews were recorded and selectively transcribed. During the phone interviews, I also took concurrent notes.

**Data Analysis**

The interview data was analyzed following the qualitative analysis strategies proposed by Creswell (2007).
1. Data management. The qualitative data was first cleaned and filed into a readable format. Audio recorded interviews were selectively transcribed. Verbatim transcription was done for ten interviews (6 instructors and 4 student teachers) that gave important insights to the study. For the other eight interviewees, notes were taken while listening to the recorded interviews and a summary of the interview was written based on concurrent notes and transcription notes.

2. Reading and taking notes. Once the data are organized, I read the compiled interview files as a whole and took side notes and wrote up summaries. These memos became the base for developing themes and categories for the data.

3. Classifying and interpreting. At this stage, the data were coded and developed into themes based on previous literature, preliminary survey results and my own interpretation of the data. I paid close attention to interview findings that confirm, diverge from the survey, or new issues that emerged from the interview results. The interview data generated eight important topics (Test Theory, Classroom Assessment, Alternative/Performance Assessment, Test Specifications, Rubric Development, Statistics, Ethics) from Language Assessment Courses.

4. Representing and visualizing. For the final stage of the data analysis, I decided on the form (i.e., text, table, and figure) used to present the data and how to integrate it with the survey findings. Chapter 5, the discussion section was written based on the eight important topics using interview quotes and survey findings.

Integration of Quantitative and Qualitative Data

The mixed-methods design for this study takes the form of an integrated design (Greene, 2007). In this mixed-methods design, the methods, both qualitative and quantitative, interact with each other through each phase of the study (Greene, 2007). Greene (2007, p. 126) states, “The integrative task in an iterative design is to represent the results of one
method in ways that meaningfully inform the desired development of another.” The instruments (survey, interview) used in the study were developed based on the findings of the prior phase; survey questions were developed on the findings of the pilot interview, and the follow-up interview questions resulted from the finding of the survey. The survey that will be administered to all accessible language assessment course instructors and volunteer language teachers will allow for generalization of results. The instructor and teacher interviews will strengthen the contextualized understanding of the quantitative findings and in-depth exploration of the questions (Greene & Caracelli, 1997).

Maintaining Data Quality

To ensure data quality, I researched quantitative, qualitative and also the Mixed Methods literature. My focus was to maintain data quality that fulfilled the most important goal of Mixed Methods study which is to bring a better understanding of the phenomenon.

Objectivity. For this study, objectivity was “supplanted by political ideals of fairness and equity (Greene, 2007)”. To give an equal voice to all participants related to the course, I included both instructors and student teachers. Previous studies only focused one or the other and failed to include all voices. Also, objectivity was fulfilled in the sense of traditional post-positivist view. Survey tools included standardization of measures and administration and maximizing response rates.

W warranting quality of inference. A multiplistic stance (Greene, 2007) was adopted to warrant the interpretations and conclusions of the study. According to Greene (p.167) this “(a) focuses on the available data support for the inferences, using data of multiple and diverse kinds; (b) could include criteria or stance from different methodological traditions ….”. For the study diverse data (syllabi, survey, and interview) were used and criteria from different traditions (quantitative, qualitative, mixed methods) were incorporated to warrant the quality of inferences.
Chapter IV

Results

Nov. 31, 20xx

Faculty Search Committee Meeting

Participants: Charles Sagon (Chair), David Kent, Jina Lee, Mia Horwitz, Heejeong Jeong

Sagon: Thank you for coming to our meeting Heejeong. I hope you can share some of the findings from your research to help us in hiring a faculty member for the assessment course.

Jeong: Well, thank you for your invitation. It would wonderful if my research findings can make a contribution in the hiring process.

Lee: So, Heejeong what are the general characteristics of language assessment courses?

Jeong: Similar to many teacher education courses the majority of the courses consist of two areas: theory and practice. The theoretical portion typically consists of test theory, test history, administration and the practice is usually done through activities like developing test specifications.

Kent: Have you found any differences between the instructors background? We haven’t decided on whether to hire a Language Tester or a non Language Tester. It would be great if you could share any findings that discuss the differences between the two types of instructors.

Jeong: Yes, there were a few interesting findings from my study. And I brought copy of my research findings to help your understanding. This is Chapter 4: Results of my dissertation. It covers the first two research questions: “What are the characteristics of
introductory language assessment courses?” and “How effective are language assessment courses for language teachers’ everyday classroom practices?” I have conducted a large scale survey and follow-up in-depth phone interviews to answer the proposed research questions.

[Jeong distributes the results section of her dissertation to the faculty members. ]

Sagon: Wonderful, let’s take a few minutes to go over the findings.

**Research Question 1: What are the characteristics of introductory language assessment courses?**

**Study Participants**

Participants included the instructors who taught the language assessment courses, and the student teachers who took the courses. Following is detailed demographics report for each participant group. The report is first divided into two groups of participants; instructors who teach the course and student teachers who take the course. After an overview of each group it is divided into specific categories. For instructors, Language Testers (LT) and non Language Testers (non LT). Phase I & II of the study revealed there appeared to be difference in the content and the structure of the course depending on the background of the instructor. Therefore for the survey and interview I focused to find out how the courses were taught and shaped differently depending on the instructor. For student teachers, the focus was on the type of students they teach. From my previous experience as a student in Language Testing and from the previous studies (Phase I & II) it appeared there were different assessment needs for different types of student teachers. Based on who they teach student teachers could have an interest in specific topic areas than other student teachers.

**Who teaches the courses? Instructor Survey Demographics**

60
One hundred forty instructors who have had experience teaching Language Assessment Courses participated in the survey. Next is instructor demographic information gleaned from the survey, including gender, primary interest area, final degree, length of the courses, delivery mode (i.e., traditional face-to-face or online), academic position, target audience for their course, and years of teaching experience by grade level.

**Gender, primary interest area, final degree.** Of the total survey participants, 66.42% (n = 89) were female, and 33.58% (n = 45) were male. For the non LTs, 72.9% (n = 51) were female instructors, and the remaining 27.1% (n = 19) were male instructors. However, for LTs, 58.7% (n = 37) were females, and 41.3% (n = 26) were male instructors. Non LTs had a larger percentage of female instructors compared to LTs.
Figure 2. Survey Instructor Demographics Overview.
For the question which asked if language testing was their primary research area, 42% (n = 66) said “yes,” and 53% (n = 73) said “no.” For instructors who answered that language testing wasn’t their primary research area, an additional question was asked regarding assessment-related activities in which they had taken part in. The top activity non LTs chose was having an experience working with classroom teachers about testing (n = 77) and having experience in developing standardized Language Tests (n = 47), and working as a rater (n = 46).

Instructors received their final degrees in various areas. Applied Linguistics (40.7%, n = 57) was the top choice, with TESOL (Teaching English to Speakers of Other Languages, 22.1%, n = 31) next, and Curriculum and Instruction (11.4%, n = 16) placed as the third choice. Other areas included Linguistics, Education, English Literature, Foreign Languages, and Speech Pathology. There was not that much difference between the two groups regarding their final degrees. For both groups, instructors most reported earning their final degrees in Applied Linguistics (LT = 32, non LT = 21). The only notable difference was from the Education and Foreign Language majors. Nine out of ten education majors identified themselves as non Language Testers, and all Foreign Language majors (n = 3) were non Language Testers. Education majors included Education Leadership, International Development in Education, and Literacy Education.

For the item that asked instructors how many years had passed since they received their final degree, the results were equally distributed over the years with less than 5 years receiving slightly more (30.22%, n = 42) than the others (6–10 years, 25.90%, n = 36; 11–20 years, 23.74%, n = 33; more than 20 years, 20.14%, n = 28).

**Length of the courses and delivery mode.** Most of the courses lasted around one semester or a quarter (75%, n = 102). Some courses were under four weeks (10%, n = 14), and a few extended over two semesters (2.9%, n = 4). The majority of the courses were
delivered in the traditional, off-line, face–to-face mode (73%, n = 98). Yet, there seems to be an increase on the number of on-line courses (3.7%, n = 5) and courses that had a mix of on- and off-line features (20%, n = 27, Wilson’s course from the instructor interview). The delivery mode and length of the courses were similar for both LTs and non-LTs.

Instructors who were currently teaching a one semester course thought a semester was not long enough to fully cover the material and wished to have more time in teaching the course. Open-ended items show time constraint as one of the difficulties in teaching this course. Goldberg an experienced LT states, “It’s quite difficult to get it [the assessment course] right. You have such limited time.”

Academic positions and experience teaching the course. Regarding the academic positions instructors held when teaching their courses, 25% (n = 35) were lecturers, 24% (n = 34) assistant professors, 18% (n = 25) associate professors, and 17% (n = 24) professors. Other answers included: visiting professor, adjunct professor, foreign language advisor, teaching assistant, and teacher trainer. In terms of the frequency of teaching the course, 40% (n = 54) had taught it 4–7 times, 31% (n = 42) fewer than 3 times, and 29% (n = 40) had taught it more than 8 times. Twenty-five instructors had taught the course more than 12 times.

Target audience. The main target audience for the course was student teachers (64.2%, n=86). Some courses targeted regular undergraduates (20.44%, n = 28) and graduate students who were not student teachers (13%, n = 18). This shows that more than half of the courses taught by the instructors were for student teachers.

When the target student audience was analyzed depending on whether the instructors were LTs or non LTs, the majority (73.5%, n = 50) of the target audience of non LTs was student teachers. This percentage was a great deal higher than the percentage of LTs who

---

4 All names given in this dissertation are pseudonyms to assure confidentiality of the participants.
replied that they were student teachers, which consisted of a little over half (53.8%, n = 35) of the target audience. This shows that roughly two thirds of the student teachers took language assessment courses taught by a non LT instructor.

**Teaching experience and certificates.** Even though not all instructors held teaching certificates (64%, n = 88 held teaching certificates; 36%, n = 49 did not), almost all instructors (96%, n = 135) responded that they had language teaching experience. The percentages for those who held a certificate were very similar for both LTs (63.6%, n = 45) and Non-LTs (63.4%, n = 45). In Brown and Bailey’s 2008 survey, they indicated 26% of the instructors held a teaching certificate in ESL or a related field. The likely reason this study has much higher percentage could be because there was no restriction in the subject area of the certificate.

Of the instructors who had teaching experience, 70% (n = 93) had more than 10 years of language teaching experience. This finding shows that the instructors were veterans in language teaching. Looking more closely at the instructors’ teaching experience, most (80.3%, n = 106) answered that the grade level they taught was college, and 19.7% (n = 26) taught K–12. More specifically, 13% (n = 20) of the instructors had experience teaching at the secondary grade level, and only 6% (n = 6) had taught elementary school. Non LTs had more K–12 teaching experience compared to LTs. While 26.9% (n = 18) of the non LTs had taught K–12, only 12.3% (n = 8) of LTs had experience teaching K–12.

Table 6

<table>
<thead>
<tr>
<th>Grade Level Taught</th>
<th>Instructor Total (n = 132)</th>
<th>LT (n = 65)</th>
<th>non LT (67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K–12</td>
<td>19.7% (n = 26)</td>
<td>12.3% (n = 8)</td>
<td>26.9% (n = 18)</td>
</tr>
<tr>
<td>Adult</td>
<td>80.3% (n = 106)</td>
<td>87.7% (n = 57)</td>
<td>73.1% (n = 49)</td>
</tr>
</tbody>
</table>
Most of the instructors (92%, n = 123) thought it was important for instructors to have teaching experience in order to effectively teach the introductory assessment course for student teachers. Most of the student teachers (94.9%, n = 224) valued the importance of field experience of instructors as well.

**Meeting Student Teachers’ Needs.** At the completion of teaching their courses, almost all instructors (89%, n = 116) felt student teachers had gained more proficiency in interpreting, evaluating, and developing language tests. This implies that most instructors felt their course was beneficial for student teachers. Ninety-one percent of the instructors (n = 122) thought their course met the needs of student teachers. Seventy-nine percent (n=187) of the student teachers indicated the course met their needs. Although not significant, this mean (m=3.12) was slighter higher for the instructors compared to student teachers (m=3.06). Overall, 98% (n = 129) of the instructors replied that they felt, their course was helpful for student teachers.

For student teachers, 75.2% (n=179) answered there were changes in the perception of the assessment through the course. Out of the student teachers who experienced changes responded having a more positive perception of Language Testing (71.6%, n=141) after the course.

Student teachers felt they more proficient in interpreting (82.8%, n=197), evaluating (89.5%, n=213), and developing (83.8%, n=197) language tests. The results were very similar for both instructors and student teachers (Table 7). Both groups thought student teachers gained the highest proficiency (m=2.89) in evaluating language tests.
Table 7

*Student Teachers Proficiency after Completing the Course*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Instructors</th>
<th>Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Evaluating Language Tests</td>
<td>130</td>
<td>2.89</td>
</tr>
<tr>
<td>Developing Language Tests</td>
<td>131</td>
<td>2.86</td>
</tr>
<tr>
<td>Interpreting Language Tests</td>
<td>131</td>
<td>2.86</td>
</tr>
</tbody>
</table>

1=less proficient, 2=no effect, 3=more proficient

**Difficulty Level of the Course.** About half of the student teachers felt the course had the same difficulty level (48.3%, n=114) as other courses in the program and 41% (n=97) found it *a little or a lot more difficult*. Only 10.6% (n=25) of the student teachers found it *a little or a lot easier* compared to other courses in the program. Following the survey results, we can see a language assessment course is not an easy course for student teachers. This implies instructors should pay special attention to the content and the structure of the course. They should make sure the topics are appropriate for the level of the student teachers.

**Needs Analysis.** Whether or not instructors do a needs analysis was asked both of instructors and of student teachers. For the instructors, 70.1% (n=96) answered ‘yes’ and 29.9% (n=41) said ‘no’. The percentage to ‘yes’ was higher for non LTs (74.3%, n=52) than LTs (65.2%, n=43). However, student teachers who were asked the same question gave opposite responses. 72.4% (n=171) answered ‘no’, and 27.5%(n=65) replied ‘yes’. The results were similar for K-12 (Yes: 33.3% n=26, No: 66.7%, n=52) and Adult STs (yes: 24.8%, n=39, no: 75.2%, n=118).
Figure 3. Was a Needs Analysis Undertaken in the LT course?: Instructors vs. Student Teachers

What could be the reason for the opposite results? First, there could be a miscommunication between the instructors and student teachers. Instructors may think they asked about student teachers’ needs at the beginning of the semester but the message was not clearly delivered to them. The message may not have been delivered explicitly to the student teachers or the student teachers may not remember it clearly. Another explanation could be instructors may have reported they have done it just because it is good practice which is the case of social desirability bias (Fisher, 1993) in survey research.

Whatever the reason, I believe it is important to do an in-depth needs analysis especially for instructors who are teaching a student population they are not familiar with. For example, if an instructor who needs to teach K-12 student teachers but themselves has little experience teaching in such context could benefit from doing this. A needs analysis will help the instructor identify the particular needs of the student teachers. Instructors should not assume this is a topic required for student teachers but directly ask them what they want in order to deliver an effective assessment course. To effectively conduct the needs analysis it would be better to do it in a distinct manner (e.g., ask student teachers to fill out forms) instead of casually asking it at the beginning of the semester. Also, it could be helpful if a
needs analysis is conducted not only early in the semester but sometime during the middle and end. Student teachers’ needs and wants could change as the semester progresses.

**Survey instructor demographic summary.** Regarding the demographic information, overall the instructors were similar in the academic positions they held, how many times they had taught their courses, the delivery mode, the target audience, the length of the courses, and how they felt about the importance of field experience in order to teach their courses. Descriptive statistics show differences in the gender ratio of the two groups, the target audience the instructors taught, and their previous language teaching experience. Non LTs had more K-12 teaching experience than LTs.

From the demographic information, there was not that much difference in terms of the number of years that had passed since each instructor had received his or her final degree or their academic positions at the time they were teaching their courses. Both groups believed that the most important factor that makes a language assessment course successful for classroom teachers is “balance between theory and practice.”

**Interview Instructor Demographics**

Out of the 140 instructors who participated in the survey, 68 people showed interest in taking part in the follow up interview. Following is demographic information gleaned from the instructor interview. Instructors who included constructive feedback in the open-ended questions and represented a diverse population were contacted via email after the survey. Seven non LTs and six LTs participated in the follow up interview. Eight of the instructors resided in the U.S. and five were currently teaching in South Korea, United Kingdom, Ecuador, Canada and Iran. The nationality of the instructors was not inquired but the site of teaching did not necessarily match with the nationality of the instructor.

The following instructor demographics are written based on the information given in the interview, basic demographic information from the instructor survey and the course
syllabi. Although, sharing the course syllabi was not required for the interview all instructors except for three written in foreign languages (French, Arabic, and Spanish) were provided by the instructor.

**Anderson.** Anderson is a part-time instructor who teaches a Language Assessment Course. Her main job is being the head of an Intensive English Language Program at a West Coast University. Her final degree was a Master’s in TESOL. Anderson self identified herself as a Language Tester in the survey. She is a relatively new instructor in this field (taught the course less than 3 times) and had more than 10 years of language teaching experience at the college level.

The unique aspect of her course is that she gives assignments that may be used in the Language Program she is directing. Every semester, the activities and focus of the course is based on the needs of the program she works at. For example, if the Language Program needs a new Speaking Test, the goal of the assessment class will be to develop an oral test. The test constructed in her class is piloted and used for the program. The main activities of her course include one critique of a high-stakes standardized English test, two test development projects with “a presentation and report that includes descriptive information and analytical data on in-class pilots (Anderson Course Syllabi)”, and an optional final exam.

**Cooper.** Cooper is an associate professor at a Mid-West University. She has taught the assessment course 8–11 times and self identified herself as a Non-LT in the survey. The topic she felt was helpful to classroom teachers was alternative assessment. She noted student teachers should learn to develop their own assessment to regain their power from the assessment equation. The topic that was most difficult to teach was Statistics. Cooper wrote she had a very basic understanding of the topic and did not feel comfortable teaching it.

The purpose of the course is to ‘better understand the purposes of assessment, grasp key concepts and their underlying theories in the field of language assessment, and create
authentic assessment tasks in the classroom (Cooper’s Syllabus). The main focus of the course was to develop and use an authentic assessment tool for English Language Learners. The crux of the course was to develop an authentic classroom assessment tool which was done in the form of a final project. The student teachers had to develop, pilot the assessment for a real classroom.

Ellen. Ellen is an assistant professor at a University in Canada. Her primary research area is Language Testing and she has taught the assessment course 4–7 times. In the survey she thought the most helpful topic for classroom teachers was rubric development, rater training, classroom assessment, and item writing. The topics she had the most difficulty teaching was test theory. She felt it was too theoretical for elementary and secondary school teachers.

The course Ellen teaches is a general assessment course for undergraduate students who will be future elementary and secondary school teachers. The course does not specifically focus on language since the student teachers are content based, but she indicated language is also an important factor in teaching the course. A unique aspect of her course is that it is taught in French. She expressed concerns on the lack of assessment resources available in French. The core activity of the class is developing test specifications as a group. Ellen asks student teachers to go to the educational curriculum and choose the subject and assessment objective. She stresses on the practical features from the very beginning of the course.

Dawson. Dawson is a recently retired Language Testing professor from a Mid-West University. He has taught the course more than 12 times and has published numerous articles and textbooks in the field of Language Testing. One of the required textbooks for his class is a recently published book written by him. In the survey he stressed the importance of teaching test theory and reported test specifications was the most challenging topic to teach.
The purpose of his course is to teach the principles and different approaches to Language Testing. He also covers how to evaluate, interpret and construct language tests. The main activities for the class is developing a group test project and writing two papers; conducting a review of a test and discussing an issue in language testing. One of the unique features of the Dawson’s class was that student teachers were asked to assess their own performance.

**Wilson.** Wilson is an assistant professor at a large Mid-West University. She is a Language Tester who has taught this course 4 ~ 7 times. She teaches the course primarily to pre-service M.A. students. The course is offered with a mix of on and off line features. The course is broadcast live over the internet and student teachers have an option of attending on or off line. The topics Wilson thought were most helpful to classroom teachers were; item design, item evaluation using item difficulty and item discrimination indices, and test ethics. The area she had the most difficulty teaching was advanced statistics. In the survey she commented “99.5% of the class will not use them outside the course” and spends little time on it and focuses more on the concepts. However she also commented she recently changed the textbook that focuses more on statistics. She noted her students always want more practice designing items.

The objective of Wilson’s course was to “construct statistically valid, reliable, and ethical test items and tests (Wilson’s syllabus).” Wilson’s course had many assignments and activities. There were six homework assignments, of which five of required students to do statistical analysis. The core activity of the course was either to do a test development project, test review, research paper, or group research project.

**Goldberg.** Goldberg is an associate professor who teaches at a Mid-West University. She identified herself as a Language Tester in the survey and has taught this course 4~7 times. She teaches this course every year to regular graduate students (not student teachers).
In the interview she stated, half of her work is administrative work in addition to teaching and doing research.

Goldberg states in her syllabus the purpose is to provide “an introduction to the technique, practice, and history of language testing”. The main activity for her class is writing a final paper based on a language testing project of the students’ choice. She requires students to do a data analysis in addition to a literature review for the final paper.

**Karimi.** Karimi, is a visiting professor at a University in Iran. He is currently a Ph.D. student in the field of Language Testing. He has taught the course more than 12 times and the target audience for the course was regular undergraduate students. The area he believed was the most important for classroom teachers was classroom assessment. He noted undergraduate students in Iran have little knowledge about assessment and “classroom assessment and test characteristics are basic issues they must know”. Topics he struggled teaching were statistics because students had difficulty grasping the concepts.

**Olson.** Olson is a non Language Tester who teaches at a Mid West University. He is a visiting professor at this school and his main interest area is Second Language Reading and Second Language Teaching Methods. He just recently began teaching this course and has taught it less than 3 times. The topic he felt was most helpful to classroom teachers was Test Theory. He notes the goal of the course is to make “novice teachers aware of the issues that go along with writing tests”. Topics he had the most difficulty teaching was advanced statistics. He felt it was the least critical topic and students do not have sufficient background to fully appreciate the topic.

The goal of Olson’s course was to “review, evaluate and discuss the components of an assessment program” (Olson’s Syllabus). The activities for his class included developing an assessment tool for one of the four skills (listening, reading, writing, or speaking), giving and article presentation, and writing a final article review paper.
**Hennes.** Hennes is a professor at an East Coast University. She is a non Language Tester whose main interest area is online distance language teacher education. The delivery mode of her course is purely online and she has taught the course 4~7 times. The topic she felt was most important for classroom teachers was test development. She noted this was the most important thing teachers do. She commented student teachers have intuitions on what is good practice but needs research in testing to support their tests. Hennes noted the most challenging topic to teach was also test development because her program is targeted for a wide variety of audience in various countries with various polices.

The goal of the course is to “examine language assessment including proficiency and placement testing, high-stakes and standardized testing, and traditional and nontraditional assessment” (Hennes’ Syllabus). Activities covered in the course include; posting reading discussions, presenting an article critique, doing a paired test development project, and taking two tests that covered the class readings.

**Ryan.** Ryan is an associate professor at a Mid-West University. She self-identified herself as a non Language Tester and her primary research area is in Methodology and Interaction. She has been teaching the course 4~7 times. The main target audience of her course is pre-service undergraduates. The topics that she thought were important were needs assessment, grading practices, charting instruction from assessment (Backwards Design) and creating alternative assessment. The area she had difficulty teaching was statistics. She said she had a weak background and her student teachers will “NEVER” do stats for their classes. Topics she covers in statistics were Item Facility and Cronbach Alpha.

The goal of Ryan’s class was to “select, create, administer, and interpret results of assessment instruments and other assessment strategies” (Ryan Syllabus). The main activities of her course included; creating an alternative assessment, a presentation of a standardized test, and developing an individual assessment tool for a non-native speaker of English. This
case study paper required the student teachers to tutor on ESL student and develop an assessment appropriate for the specific student.

**Smith.** Smith is an associate professor at a university in the United Kingdom. She is a non Language Tester whose primary research area is English for Academic Purposes. She has taught this course 8–11 times and was the person to suggest a language assessment course for the Master’s program curriculum. The topic Smith believes is most helpful for student teachers was classroom assessment since “very few teachers develop large scale tests(Smith interview)”. The area she most struggled in teaching was also classroom assessment. Smith commented almost nothing is written about classroom assessment in language testing therefore “most of what is in testing textbooks are not relevant to them [student teachers]”.

The language assessment course was co-taught with another instructor. The reason for this was not for academic purposes but rather for practical reasons (to be prepared when unexpected vacancies occur by one instructor such as health failure). The goal of the course was to “produce graduates with a great depth of knowledge and understanding of the process of assessment in language teaching and issues relating to it” (Smith’s Syllabus). The expected outcome was to help students develop an in-depth knowledge of the practices and principles relating to language assessment. The main assignment of the course was writing an essay on a topic of test development and construction. Student teachers were given the liberty to choose a topic of their choice.

**Jackson.** Jackson is an assistant professor at a university in South Korea. He identified himself as a non Language Tester in the survey and his main interest area is collaborative learning. He has taught this course less than 3 times. The topics he believed to be important for classroom teachers were history and theory. He stated those areas help understand current assessment practices better. The topic he struggled most was making students develop multiple choice questions. He said, these types of questions were the most
commonly used format in the student teachers’ teaching context but they were very poorly done and practice did not really help them.

The goal of Jackson’s course is to introduce “important concepts and developments in evaluation, assessment, testing, and to foster an understanding of these concepts through reading, reflection, analysis, and design” (Jackson’s Syllabus). The core activities were to post reading reflections and design a testing instrument. Jackson was using a classroom based evaluation textbook that was published 15 years ago. He commented that the book was outdated but could not find a better one to replace it.

Rivera. Rivera is a teacher supervisor for a teacher education program in Ecuador. She is a non Language Tester whose main interest area is in Communication. She has taught this course less than 3 times to undergraduate students. The topic she felt was most important to classroom teachers was rubric development. She commented rubrics help assessment to be objective and can be used by other testers. It also helps students to be aware of the requirements and expectations for the tasks. The topic she most struggled teaching was advanced statistics. She said she did not know much of that topic area.

Who takes the course? Student Teacher Survey Demographics

A total of 244 student teachers participated in the survey. Following is demographic information of student teachers gleaned from the survey, including gender, instructor background (whether or not he or she was a language testing professional), whether the language courses students took were required or electives, the delivery mode of the courses (traditional or online), number of years of teaching, the number of courses taken, final degrees, and length of the courses taken.
Figure 4. Survey Student Teacher Overview.
**Gender and instructor background.** There were many more female student teachers (78%, n= 190) compared to males (22%, n = 53). Of the K–12 student teachers, 88.9% (n = 72) were female and only 11.1% (n = 9) were male. On the other hand, for the adult student teachers, there were a great deal more male student teachers (26.9%, n = 42) compared to K–12 student teachers.

Table 8

*Student Teachers’ Instructor Background: LT or non LT?*

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>LT</th>
<th>Non LT</th>
<th>Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Student Teachers</td>
<td>53.5% (n = 129)</td>
<td>19.9% (n = 48)</td>
<td>26.6% (n = 64)</td>
</tr>
<tr>
<td>K–12</td>
<td>42.0% (n = 34)</td>
<td>17.3% (n = 14)</td>
<td>40.7% (n = 33)</td>
</tr>
<tr>
<td>Adult</td>
<td>59.4% (n = 95)</td>
<td>21.3% (n = 34)</td>
<td>19.4% (n = 31)</td>
</tr>
</tbody>
</table>

For the question if the instructor who taught their course was a language testing professional, Table 8 indicates, 53.7% (n = 131) responded “yes”; 19.7% (n = 48) answered “no,” and 26.2% (n = 64) replied, “I’m not sure.” Analyzing by the student teachers’ grade-level, findings shows that adult student teachers were taught more by Language Testers (59.4%, n = 95) than K–12 student teachers (42.0%, n = 34). More specifically, the percentage of K-12 student teachers who were not sure of their instructor’s background was higher (40.7%, n = 33) than adult student teachers. It is interesting to see more K-12 student teachers did not know the background of their instructor.

Defining who is a Language Tester and who is not can be confusing for the instructors themselves, and there is room for error in that student teachers may not have accurately reported it. The percentage of student teachers who took courses from non
Language Testers (19.4%, n = 47) was much lower than the percentage of language assessment courses taught by non LTs (65.7%, n = 131). This result is different from the instructor survey. One explanation for this discrepancy could be the possibility that the instructors who fall under the ‘I don’t know’ category could come from a non LT background. From the instructor survey non LTs replied 73.5% (n=50) of their target audience were student teachers. Unless student teachers had a particular interest in the instructor, they might have assumed the instructor to be a Language Tester because he or she was teaching the course.

Another reason for the mismatch of the instructor survey results and student survey results is that student teachers weren’t necessarily taught by the instructors who participated in the survey. Participants for both surveys were volunteers who expressed interest in taking part in the study. To secure privacy, school names and instructor names were not asked. Therefore it is impossible to trace back the student teachers’ instructors or the student teachers the instructors taught. The findings from the comparison between Instructors and Student Teachers should be interpreted in light of this information.

The amount of time that had elapsed since the student teachers took the course did not really have an effect on whether they remembered if the instructor was a Language Tester or not. Of the student teachers who had taken the course more than 5 years ago, 27.7% (n = 28) replied that they weren’t sure whether the instructor was a Language Tester or not. This percentage is only slightly higher than the percentage (25.5%, n=13) for those who had taken it more recently.

**Elective or requirement.** The language assessment course was a required course for 70.1% (n = 171) of the student teachers; it was an elective for 26.6% (n = 65). The percentage of student teachers who took the course because it was a requirement was slightly higher (77.8%, n = 63) for K–12 student teachers compared to adult student teachers (70.6%, n =
A total of 84% (n = 200) of student teachers answered that they would have taken the course even if it had not been a requirement; only 16.1% (n = 38) answered that they would not have done so. The percentage for not taking the course was a little higher for the adult student teachers (17.3%, n = 27) compared to K–12 student teachers (13.8%, n = 11). This is quite promising for the future of language assessment courses. It appears that student teachers know the importance of assessment. Reasons for choosing not to take the course were: lack of interest in the topic, feeling it wasn’t necessary for a classroom teacher, not being aware of needing the course at that time, negative feedback from other students who had taken the course, not being aware of the impact of assessments.

In comparison to previous studies (Brown & Bailey, 2008) the results were different for the topic of whether the course was an elective or a requirement. Brown and Bailey, who surveyed 97 language testers in regards to the basic characteristics of the course, found that for 33% the language assessment course was an elective and for 49.5% it was a requirement. The percentage responding that the language assessment course was required is a great deal lower than the 70.1% from my student teacher survey. Since Brown and Bailey surveyed only LTs, I purposely selected and analyzed the findings from student teachers who took the course from a Language Tester. Even with this analysis, the results were not different, with 72.4% (n = 94) of the student teachers who took the course from LT instructors responding that the course was a requirement and only 27.1% (n = 35) answering that it was an elective. One reason for this difference could be because Brown and Bailey’s target audience was different from mine. They asked this question to the instructors, while I asked it to student teachers. It is not clear why there is such a big difference from the previous study.
**Student Teachers’ Primary Language**

<table>
<thead>
<tr>
<th>Primary Language</th>
<th>Student Teachers (n = 244)</th>
<th>K–12 (n = 82)</th>
<th>Adult (n = 160)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>70.2% (n = 173)</td>
<td>56.1% (n = 46)</td>
<td>43.9% (n = 127)</td>
</tr>
<tr>
<td><strong>Foreign Language</strong></td>
<td>17.6% (n = 43)</td>
<td>25.6% (n = 21)</td>
<td>13.8% (n = 22)</td>
</tr>
<tr>
<td><strong>English &amp; Other</strong></td>
<td>11.5% (n = 28)</td>
<td>18.3% (n = 15)</td>
<td>7.5% (n = 13)</td>
</tr>
</tbody>
</table>

**Primary language and time of the course.** The primary language student teachers taught was English (71.2%, n = 173, Table 9). Student teachers also taught languages other than English (17.7%, n = 43), and 11.1% (n = 27) of the student teachers had experience teaching English and also another foreign language.

In more detail, K–12 student teachers had a lot more experience teaching other foreign languages (43.9%, n = 36) than adult student teachers. For adult student teachers, only 21.3% (n = 35) taught other foreign languages. From this result we can see that adult student teachers primarily teach English and K–12 student teachers have more experience teaching other languages than English.

Table 10

**When Student Teachers Took the Course**

<table>
<thead>
<tr>
<th>When the course was taken</th>
<th>Student teachers (n = 240)</th>
<th>K–12 (n = 81)</th>
<th>Adult (n = 159)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within the last year</td>
<td>20.8% (n = 50)</td>
<td>30.9% (n = 25)</td>
<td>15.7% (n = 25)</td>
</tr>
<tr>
<td>2–3 yrs ago</td>
<td>20.4% (n = 49)</td>
<td>23.5% (n = 19)</td>
<td>18.9% (n = 30)</td>
</tr>
</tbody>
</table>

(continued)
Table 10  
(continued)

<table>
<thead>
<tr>
<th></th>
<th>4–5 yrs ago</th>
<th>More than 5 yrs ago</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17.1% (n = 41)</td>
<td>18.5% (n = 15)</td>
</tr>
<tr>
<td></td>
<td>16.4% (n = 26)</td>
<td>42.15% (n = 102)</td>
</tr>
<tr>
<td></td>
<td>27.2% (n = 22)</td>
<td>49.1% (n = 78)</td>
</tr>
</tbody>
</table>

For the question that asked when student teachers took the course, 42.15% (n = 102) replied that they had taken the course more than 5 years ago; 20.66% (n = 50) had taken it within the last year; 20.25% (n = 49) had completed the course 2–3 years ago; and 16.94% (n = 41) had taken the course 4–5 years ago. Almost half of the adult student teachers (49.1%, n=78) took the course more than 5 years ago, which is a lot more than K-12 student teachers (27.2%, n=22). The distribution of K-12 student teachers were roughly even for the time they took the course. More K-12 student teachers took the course recently (within the last year, 30.9%, n=25) compared to adult student teachers (15.7%, n=25).

**Delivery mode.** Most of the assessment courses student teachers took were taught via the traditional, face-to-face, off-line mode (76.76%, n = 185), a few courses consisted of a mix of both off- and on-line components (12.45%, n=30), and a small percentage (7.88%, n = 19) were purely on-line courses. Ten out of 8 adult student teachers took the courses by traditional, face-to-face mode, while 72.4% (n = 55) of K–12 student teachers took the courses off-line.

Although the majority of the courses are still offered off-line, more and more M.A. TESOL courses are being offered on-line now. Instructors should think of ways to effectively teach the course in an on-line environment. For example, one of the important activities in a Language Assessment Course might be developing Test Specifications. To do this activity, tests and other types of assessment tools need to be developed together as a group to increase
the validity of the test. Although it is a challenge for instructors to devise test specifications activities when the course is taught on-line, whether instructors favor it or not, the number of courses offered on line will increase. In order to meet the needs of on-line courses, it is important for language assessment instructors to develop methods that will work in an on-line teaching context.

**Years of teaching and number of courses taken.** Similar to the instructors, student teachers who participated in the survey had many years of teaching experience: 46.89% (n = 113) had taught more than 10 years; 27.39% (n = 66) had been teaching 5–9 years; 12.03% (n = 29) had taught 3–4 years; and 9.54% (n = 23) had taught 1–2 years. Only 4.15% (n = 10) of the student teachers had taught less than a year. There were more adult student teachers (50%, n = 79) who had been teaching more than 10 years compared to K–12 student teachers (40.2%, n = 33).

When asked how many language assessment courses they had taken, 66.3% (n = 159) answered 1 course; 21.7% (n = 52) replied 2 courses, 4.6% (n = 12) had taken 3 courses, and 7.1% (n = 17) had taken even more than 3 courses. The number of courses taken by student teacher grade level was overall similar for both groups, except for the more than 3 courses response. There were more adult student teachers (8.2%, n = 13) who had taken more than 3 courses in language assessment compared to K–12 student teachers (4.9%, n = 4).

Most likely, student teachers who had more teaching experience took more language assessment courses than those who did not; 17 student teachers took more than 3 assessment courses, and 75% (n=12) of them had more than 10 years of experience.

**Student teachers’ final degree.** The student teachers who took the survey were very well educated. For example, 95% (n = 226) of the student teachers had Master’s degrees, doctoral degrees, or either in progress; 17% (n = 41) were in progress of their doctoral degrees, and 14.46% (n = 35) had doctoral degrees. Grouping this by student teacher grade
levels, there were more Master’s in progress (24.4%, n = 20) for K–12 student teachers than for adult student teachers (10.1%, n = 16). Half of the adult student teachers (49.7%, n = 79) had a Master’s degree, while 40.2% (n = 33) of K–12 student teachers had a Master’s degree.

The percentage of students with Doctoral degrees was much higher for adult student teachers (20.1%, n = 32) compared to K–12 student teachers (2.4%, n = 2). The percentage of student teachers with only a B.A. or B.A. in progress was 1.9% (n = 3) for adult student teachers and 13.4% (n = 11) for K–12 student teachers. For both groups, more than 60% of the student teachers had an M.A. or were pursuing one.

Table 11  
Student Teachers’ Final Degree

<table>
<thead>
<tr>
<th>Time of the Course</th>
<th>Student Teachers (N=241)</th>
<th>K-12 (n=82)</th>
<th>Adult (n=159)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.A., &amp; in progress</td>
<td>5.8% (n=14)</td>
<td>13.4% (n=11)</td>
<td>1.9% (n=3)</td>
</tr>
<tr>
<td>Master’s &amp; in progress</td>
<td>61.4% (n=148)</td>
<td>64.6% (n=53)</td>
<td>59.8% (n=95)</td>
</tr>
<tr>
<td>Doctorate &amp; in progress</td>
<td>31.1% (n=75)</td>
<td>20.7% (n=17)</td>
<td>36.5% (n=58)</td>
</tr>
<tr>
<td>Other</td>
<td>1.7% (n=4)</td>
<td>1.2% (n=1)</td>
<td>1.9% (n=3)</td>
</tr>
</tbody>
</table>

It was surprising to see that many student teachers had advanced degrees. This may show that the participants of the survey may not be representative of the actual student teacher population. Student teachers who are willing and interested in taking part of a research study are likely to be interested in doing research themselves. Also, the recruiting for the study was done via list-servers of professional language teacher organizations (TESOL, AAAL, and other). This could have been another factor in limiting the characteristics of student-teachers.

**Length of the course.** Student teachers responded, the assessment courses was usually offered for one semester (79.1%, n=189), some courses were less than a semester,
usually a quarter long (9.2%, n=22) and a few were two semesters long (5.4%, n=13). The results were similar for both K-12 and adult student teachers; 72.8% (n=59) of the K-12 teachers took a semester long course and 82.2% (n=129) of the adult STs had a one semester course. However for classes that lasted less than a semester there were more adult STs (10.2%, n=16) compared to K-12 student teachers (7.4%, n=6). In contrast, for courses that ran 2 semesters long, there were more K-12 (11.1%, n=9) teachers than adult STs (2.5%, n=4). The results show the majority of the language assessment courses were offered as a single semester course and for courses than covered more than 1 semester, there were more K-12 STs than adult STs.

These findings corroborated with the instructor survey results. 73.6% (n=103) of the instructors replied they teach a semester long course and 15% (n=21) answered the course was less than a semester (usually a quarter).

**Difficulty level of the course.** A little less than half (48.1%, n=114) of the student teachers elicited the course was about the same level as other courses in the program. 41.4% (n=98) replied it was a little or a lot more difficult than other courses. The results were very similar for K-12 and Adult STs.

Table 12

*Course Difficulty from Student Teachers’ Perspective*

<table>
<thead>
<tr>
<th>Compared to other courses, what was the difficulty level of the course? (N=236, missing=8)</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>a lot easier</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>a little easier</td>
<td>7.6</td>
<td>18</td>
</tr>
<tr>
<td>about the same level</td>
<td>48.3</td>
<td>114</td>
</tr>
<tr>
<td>a little more difficult</td>
<td>29.7</td>
<td>70</td>
</tr>
<tr>
<td>a lot more difficult</td>
<td>11.4</td>
<td>27</td>
</tr>
</tbody>
</table>
Following the survey results, we can see a language assessment course is not an easy course for student teachers. In Brown and Bailey’s (2008), 33.4% of the instructors reported their student think the course is difficult. This implies instructors should pay special attention to the content and the structure of the course. They should make sure the topics are appropriate for the level of the student teachers.

**Most important factor for effective language assessment courses.** For student teachers the most important factor for an effective Language Assessment Course was “balance between theory and practice” (62.8%, n=150). Practicality was the next important factor (33.0%, n=81). The response was similar for K-12 and Adults STs. K-12 student teachers emphasized the importance of practicality (39.2%, n=31) a little more than Adult STs (31.4%, n=50).

Similar to student teachers, almost all instructors believed the most important factor was balance between theory and practice (82.7%, n=115) and some 16.5% (n=23) stated practicality. The results for this question were not that different for LTs and non-LTs. Student teachers and instructors agreed upon that keeping a balance between practice and theory was the most important factor but STs gave more value on practicality than instructors.

A question related to this topic was given to instructors in the follow up interview. The question asked how much percent instructors devoted to practice vs. theory. According to the survey findings the answer should be close to 50/50. Yet, not all instructors said they had an equal balance.

**Survey Student Teacher Demographic Summary**

Student teachers who take the assessment course came from two different groups; K-12 and Adult. More than two thirds of the student teachers were females and this percentage was higher for K-12 teachers. About half of the STs indicated they took the assessment course from a Language Testers. For most of the STs the course was a requirement and STs
responded they will take the course even if it was not required. The primary language STs taught was English but there were also people who taught other languages. The majority of the course was offered in the traditional face-to-face mode that lasted one semester.

The participants of the survey had many years of teaching experience and the majority (66.3%, n=159) took one language assessment course. For STs final degree, 61.4% (n=148) had either a Master’s degree or were in progress toward one.

The findings of the student teacher survey corresponded with the instructor survey in the areas of course length and course mode but showed discrepancies in the instructor’s background (LT or non-LT) and target student teacher population (K-12 or Adult).

**Interview Student Teacher Demographics**

From the 244 student teachers who took part in the survey, 123 student teachers expressed interest in doing the follow up interview. From these student teachers, fifteen who gave detailed responses on the open-ended questions and represented a diverse population were contacted via email. Out of the fifteen student teachers, five responded to the interview request and took part in the follow-up phone interview. Two teachers were K-12 English teachers and the remaining three taught adults. One of the adult teachers was a pre-service teacher. All the other four were currently teaching. All teachers resided in the United States.

Student teacher demographics are written based on the information given in the interview, basic demographic information from the survey and the course syllabi (if provided by the student teacher). Sharing the course syllabus was not a requirement and was only collected if the student teachers were willing to offer the document. Therefore, the breadth and depth of the description may vary depending on the available sources for each individual student teacher.

**Carol.** Carol is an elementary school teacher who took the course 2~3 years ago by a Language Tester at an upper east coast university. The topics she found helpful to her
classroom teaching were Rubric Development and Alternative Assessment. She felt the
course was a lot more difficult compared to other courses in the program.

Carol was overall very happy with the course and commented she had a really good
instructor. Her instructor had published several textbooks regarding Language Assessment
particularly for K-12 teachers. One of the required textbooks was also written by the
instructor. The target audiences for Carol’s class were candidates for the ESL and Foreign
Language Teacher Licensure Program. The purpose of the course was to provide “classroom-
based assessment of language learning students in ESL, bilingual education, foreign
language, and grade-level classrooms in Grades PreK-12, Adult Education, and University
programs” (Carol’s Course Syllabus). The main activities for the course were; compare
standardized achievement tests to classroom-based assessments, review language proficiency
measures, identify assessment of language learners with special needs, develop assessment
tools for four language skills and draft a student assessment portfolio.

Bell. Bell is K-12 ESL teacher at a Mid West school. She took the course 4–5 years
ago. The course was a requirement part of her M.A. TESOL program. The topic she felt was
most helpful to her classroom teaching was test critique because it directly helped developing
classroom materials. She found all the topics covered in the course to be useful but wanted
more experience on test modifications and accommodations. As a practitioner she discovered
making tests ELL friendly was a major part of her job.

Bell took the course from a well known Language Tester who has done extensive
research in this area and also published several books in this field. The goal of the course was
to “examine purposes and types of language tests in relation to theories of language use and
language teaching goals” (Bell’s Syllabus). The activities covered in the course included
three mini-papers, an ungraded critique of a testing company or agency website, and a small
group test development project which is the core activity of the course.
**Sarah.** Sarah teaches writing courses at a university language program. She took the course more than 5 years ago and was not sure if the instructor was a Language Tester or not. The topic which she said was most helpful was Test Theory. She noted this was because she has personal interest in the topic. Test proctoring was the topic she felt was the least helpful because she thought the process was evident. Overall she enjoyed the course and found it relevant to her teaching. She commented, “Testing is an integral part of language acquisition”.

**Mark.** Mark currently teaches part time at a community college while completing his Master’s Degree at an upper west coast university. He took the course within the last year from a Non-Language Tester. In the survey he elicited test critiquing and item writing activities were the most helpful topics and alternative assessment was the least helpful since it was difficult to implement them.

The course he took was for both undergraduate and graduate students. The purpose of the course was to “understand fundamental concepts and principles of psychometric measurement as well as strategies for alternative assessment, construct language tests for various skills and measurement purposes, and score an interpret test results (from Mark’s Syllabus)”. Mark enjoyed this class because the instructor focused on discussion, group activities and hands on work rather than lectures. A unique aspect of the course that was especially appealing to Mark was the on-line assessment project done with an English class in Taiwan. The students of the assessment course had to “develop, administer and grade an Integrated Performance Assessment (IPA) test” (Mark’s Syllabus).

The main activities for the course were group presentation for the IPA project, contributions to IPA project, reflections and giving feedback to students in Taiwan. For graduate students like Mark two additional activities; article presentation and quantitative analysis of interpretive task results were required.
**Jenna.** Jenna is currently on the job market after finishing her Master’s Degree. She is interested in teaching Adult ESL. She holds a Ph.D. in German Literature but decided to pursue another M.A. in TESOL for job purposes. The course she took was a required class for the Master’s program. It was taught by a famous Language Tester but she recalls that very few people actually benefited from the course. In the survey she said she would have not taken the course if it was not required since she did not learn much from the course. She felt none of the topics she learned in the class was helpful and could have learned more by reading books on her own. Jenna stated she had a more negative perception after teaching the course.

The instructor spent most of the time reading his manuscripts in class. Jenna felt the instructor was proficient in his area but did not possess any teaching skills and did not care about teaching. His focus was mainly on research. Jenna wrote a critique of a major test but no feedback was ever given to the student teachers. There was little guidance in class activities. The group activity which was helpful was to look at test items together and critique them but was only covered once in the quarter. The comprehension exam for the class was recycled from previous years and could not be done without additional sources. Some of the items on the exam were not covered in the assessment class. Jenna noted the only thing she took away from the class was that she can say she took a Testing class.

**What is being taught?**

To get a general overview of what is being taught, instructors were asked to report how much time they spent on 14 topics, choosing one of the following four responses on a Likert-scale item (1 = hardly any time, 2 = a little time, 3 = some time, 4 = extensive time). Independent t-tests were used to identify the significant difference between the Language Tester and Non-Language Tester group. The findings for the Likert-scale items are first reported for the whole group (instructors and student teachers) and more specifically divided
into subgroups (LTs vs. non-LTs, K-12 STs vs. Adult STs). The Bonferroni adjustment (p<.0035) was used to avoid Type 1 error.

**Instructors.** The top five areas instructors spent the most time teaching were test theory (m = 3.33), classroom assessment (m = 3.31), alternative performance assessment (m = 3.17), test specifications (m = 3.09), and rubric development (m = 2.93).

Topics that received similar time coverage by both instructor groups were; test critique, history of language testing, and rater training. Instructors spent the same amount of time on test ethics, history of language testing, and advanced statistics. There were six areas of significant differences between the two groups. The greatest area of disagreement is regarding test accommodation. Language testers ranked this category towards the bottom (13th out of 14 categories, m = 1.64), while non-language testers ranked it 9th (m = 2.39) out of the categories. This shows that, compared to non LTs, LTs spend very little time covering test accommodations.

Topics that received similar amount of time coverage by Language Testers and non-Language Testers were test critique, history of language testing, and rater training. They spent the same time on test ethics and history of language testing. As for advanced statistics, the t-test shows that both groups spent a similar amount of time. However, this should be interpreted with caution. The response rate for non-language testers is only half (n = 35) of the total (n = 70) size. This implies that many non-language testers either did not answer the question or checked Non Applicable since it was not relevant to their class content.
Table 13

*Topic Time Coverage from Instructors’ Perspective*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Instructors</th>
<th>Language Testers</th>
<th>Non Language Testers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  M  SD  Rank</td>
<td>n  M  SD  Rank</td>
<td>N  M  SD  Rank  t  p</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>1. Test Specifications</td>
<td>133 3.09 0.811 4</td>
<td>61 3.26 0.656 2</td>
<td>68 2.93 0.919 6 2.364 .020*</td>
</tr>
<tr>
<td>2. Test Admin</td>
<td>131 2.38 0.872 9</td>
<td>58 2.26 0.828 10</td>
<td>69 2.43 0.899 8 -1.140 .260</td>
</tr>
<tr>
<td>3. Test Critiquing</td>
<td>133 2.89 0.813 6</td>
<td>59 2.83 0.769 5</td>
<td>70 2.94 0.866 5 -0.772 .450</td>
</tr>
<tr>
<td>4. Test-taking skills or strategies</td>
<td>122 2.09 0.996 12</td>
<td>54 1.94 1.071 12</td>
<td>65 2.22 0.927 12 -1.479 .143</td>
</tr>
<tr>
<td>5. Test Theory</td>
<td>137 3.33 0.749 1</td>
<td>63 3.48 0.618 1</td>
<td>70 3.19 0.839 3 2.251 .022*</td>
</tr>
<tr>
<td>6. Basic Statistics</td>
<td>128 2.42 1.054 8</td>
<td>58 2.62 1.056 7</td>
<td>66 2.23 1.034 11 2.092 .023*</td>
</tr>
<tr>
<td>7. Advanced Statistics</td>
<td>78 1.51 0.802 14</td>
<td>41 1.49 0.779 14</td>
<td>35 1.51 0.818 14 -0.144 .897</td>
</tr>
<tr>
<td>8. Test Ethics</td>
<td>130 2.53 0.837 7</td>
<td>60 2.53 0.812 8</td>
<td>66 2.53 0.863 7 0.020 .940</td>
</tr>
<tr>
<td>9. History of Language Testing</td>
<td>123 2.13 0.905 11</td>
<td>59 2.10 0.865 11</td>
<td>61 2.15 0.946 13 -0.277 .990</td>
</tr>
<tr>
<td>10. Classroom Assessment</td>
<td>135 3.31 0.796 2</td>
<td>62 3.15 0.938 3</td>
<td>69 3.46 0.632 1 -2.300 .024*</td>
</tr>
</tbody>
</table>

(continued)
Table 13 (continued)

<table>
<thead>
<tr>
<th>Topic</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Alternative Performance Assessment</td>
<td>131</td>
<td>3.17</td>
<td>0.824</td>
<td>3</td>
<td>58</td>
<td>3.02</td>
<td>0.908</td>
<td>4</td>
<td>69</td>
<td>3.30</td>
<td>0.754</td>
<td>2</td>
<td>-1.948</td>
<td>0.051</td>
</tr>
<tr>
<td>12. Rubric Development</td>
<td>131</td>
<td>2.93</td>
<td>0.879</td>
<td>5</td>
<td>61</td>
<td>2.74</td>
<td>0.814</td>
<td>6</td>
<td>66</td>
<td>3.09</td>
<td>0.924</td>
<td>4</td>
<td>-2.278</td>
<td>0.020*</td>
</tr>
<tr>
<td>13. Rater Training</td>
<td>128</td>
<td>2.30</td>
<td>0.968</td>
<td>10</td>
<td>59</td>
<td>2.36</td>
<td>0.943</td>
<td>9</td>
<td>65</td>
<td>2.25</td>
<td>1.000</td>
<td>10</td>
<td>0.627</td>
<td>0.488</td>
</tr>
<tr>
<td>14. Test Accommodation</td>
<td>122</td>
<td>2.08</td>
<td>1.033</td>
<td>14</td>
<td>53</td>
<td>1.64</td>
<td>0.901</td>
<td>1</td>
<td>66</td>
<td>2.39</td>
<td>1.021</td>
<td>9</td>
<td>-4.208</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

Note: The response scale was as follows: 1 = hardly any time, 2 = a little time, 3 = some time, 4 = extensive time.
Note: Multiple independent t-tests were used to identify the significant difference exists between the two groups. Using the Bonferroni adjustment required significance at the level (p<.0035).
**Student Teachers.** In the survey, student teachers reported that the topics that received the most coverage in the course were Test Theory (m=3.18), Classroom Assessment (m=2.82), Test Critique (m=2.74), Test Specifications (m=2.69), and Rubric Development (m=2.66). In terms of statistically significant differences, the results were similar for both K-12 STs and Adult STs. The only topic that showed a significant difference was advanced statistics (p=.016). The rankings for the top 2 topics (Test Theory and Classroom Assessment) were the same but the remaining topics showed some variance. K-12 teachers indicated there was more time coverage on classroom assessment related topics such as rubric development (#3) and alternative assessment (#4). On the other hand, Adult student teachers thought more extensive time were spent on test critique (#3) and test specifications (#4).

One interesting finding in regards to the means of the groups is that adult student teachers reported a higher mean score on 8 out of the 14 topics than K-12 student teachers. The topics that K-12 student teachers scored higher than Adult student teachers were topics more closely related to classroom teaching; test taking skills, classroom assessment, alternative performance, rubric development, and test accommodation.

Table 14

**Topic Time Coverage: Instructors vs. Student Teachers**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Instructors</th>
<th>Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n  M   SD</td>
<td>n  M   SD</td>
</tr>
<tr>
<td>Test Specifications</td>
<td>133 3.09 .811</td>
<td>223 2.68 .921</td>
</tr>
<tr>
<td>Test Admin</td>
<td>131 2.38 .872</td>
<td>200 1.61 .856</td>
</tr>
<tr>
<td>Test Critiquing</td>
<td>133 2.89 .813</td>
<td>225 2.73 .912</td>
</tr>
<tr>
<td>Test taking skills or</td>
<td>122 2.09 .996</td>
<td>204 1.81 .954</td>
</tr>
</tbody>
</table>

(continued)
Table 14 (continued)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Instructors</th>
<th>Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>5. Test Theory</td>
<td>137</td>
<td>3.33</td>
</tr>
<tr>
<td>6. Basic Statistics</td>
<td>128</td>
<td>2.42</td>
</tr>
<tr>
<td>7. Advanced Statistics</td>
<td>78</td>
<td>1.51</td>
</tr>
<tr>
<td>8. Test Ethics</td>
<td>130</td>
<td>2.53</td>
</tr>
<tr>
<td>9. History of Language Testing</td>
<td>123</td>
<td>2.13</td>
</tr>
<tr>
<td>10. Classroom Assessment</td>
<td>135</td>
<td>3.31</td>
</tr>
<tr>
<td>11. Alternative Performance</td>
<td>131</td>
<td>3.17</td>
</tr>
<tr>
<td>Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Rubric Development</td>
<td>131</td>
<td>2.93</td>
</tr>
<tr>
<td>13. Rater Training</td>
<td>128</td>
<td>2.30</td>
</tr>
<tr>
<td>14. Test Accommodation</td>
<td>122</td>
<td>2.08</td>
</tr>
</tbody>
</table>

1=Hardly Any Time, 2=Sometimes, 3=Pretty Much Time, 4=Extensive Time

**Instructors vs. Student Teachers.** Responses for Instructors and Student teachers were compiled together (N=384) to analyze mean differences between the groups. Both groups were asked the same set of questions in a different survey questionnaire. The results show a significant difference in 9 categories (Table 15). Teachers gave a consistently lower mean for almost all categories except for Advanced Statistics.
There could be several different interpretations to this result. First, instructors who teach the course may feel they are covering a variety of topics for a much longer time compared to student teachers. This could be connected with social desirability bias in survey studies. Instructors could have given a more favorable response to the question compared to student teachers. Student teachers may think the topic was not covered in depth as the instructor intended. Another way to interpret the findings is that student teachers could have forgotten the length of time spent for each topic. They could have marked a significantly less time than it was actually covered. Whatever the reason, there is a clear difference in time coverage of topics for instructors and student teachers.

**Research Question 2: What constitutes evidence of an “effective” language assessment course for classroom teachers?**

**Which topic is most helpful for student teachers?**

For the open ended question -- “which topic do you think would be most helpful for student teachers” -- instructors’ (valid: n=119, missing: n=21) top answer was Classroom...
Assessment (34.5%, n=41), Test Specifications (22.7%, n=27) and Test Theory (15.1%, n=18). Some reasons for classroom assessment being the most helpful topic for student teachers were given, for instance, it is ‘the most practical application for their immediate use in the classroom, this is what most teachers need to be able to do on their own or with relatively little guidance’. For Test Theory reasons were because, ‘[using Test Theory] they can review tests that are mandated for their students and provide solid reasons for their conclusions’. Other instructors commented that being the most helpful topic depends on the context. This appears to vary based on the student teachers’ needs and teaching environment. Student teachers can come from different testing cultures. For example, an instructor who is from Finland commented ‘In my country, the teacher is solely responsible for assessment in compulsory education, and also has a major role in upper secondary, as my country (Finland) has no high-stakes, school-external examination, except at the end of upper-secondary schooling.’ A student teacher who teaches in such a teacher driven assessment environment may have had different needs compared to student teachers who have to follow external standards and implement standardized tests.

There were not that many differences between LTs and non LTs regarding this question. Both replied that Classroom Assessment was the most important topic. Areas that showed clear differences were Test Specifications and Alternative Assessment. 29.8% (n=17) of the Language Testers picked Test Specifications as the most helpful topic for classroom teachers whereas the percentage (16.4%, n=10) was a lot lower for non-LTs. For Alternative Assessment, the two instructor group gave opposite responses. 16.4% (n=10) of the non-LTs found this to be the most helpful topic while only 3.5% (n=2) of the LTs said it was. The findings for Test Specifications correspond with the results of time spent on each topic. There was significant difference for Test Specifications between the two groups (t=2.364, p=.020).
This same open ended question was inquired to the Student Teacher group. The results were quite different.

Table 15

*Top 5 Most Helpful Topics for Student Teachers (Open-Ended Question)*

<table>
<thead>
<tr>
<th>Topic</th>
<th>N</th>
<th>%</th>
<th>Topic</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Assessment</td>
<td>41</td>
<td>34.5%</td>
<td>Test Specifications</td>
<td>59</td>
<td>26</td>
</tr>
<tr>
<td>Test Specifications</td>
<td>27</td>
<td>22.7%</td>
<td>Test Theory</td>
<td>31</td>
<td>12.7</td>
</tr>
<tr>
<td>Test Theory</td>
<td>18</td>
<td>15.1%</td>
<td>Rubric Development</td>
<td>30</td>
<td>12.3</td>
</tr>
<tr>
<td>Alternative Assessment</td>
<td>13</td>
<td>10.9%</td>
<td>Classroom Assessment</td>
<td>27</td>
<td>11.1</td>
</tr>
<tr>
<td>Rubric Development</td>
<td>7</td>
<td>5.9%</td>
<td>Alternative Assessment</td>
<td>27</td>
<td>11.1</td>
</tr>
</tbody>
</table>

Student Teachers replied the most helpful for the given 14 topics was Test Specifications, Test Theory, Rubric Development, Classroom Assessment, and Alternative Assessment.

There is an interesting discrepancy between the instructors and student teachers. Although student teachers think that Classroom Assessment is the most important topic for a classroom teacher (Table 16), when they pointed out the topic from what they learned the most helpful to their classroom teaching this was ranked #4 out of the 14 topics. This difference can be related with the quality of the topic taught by the instructor. Student teachers could have found Test Specifications to be the most helpful to their classroom teaching because it was the topic they got the most out of. However they could have believed classroom assessment was an important topic but was not as helpful because instructors failed to teach it effectively compared to the other topics. This can be connected with how confident instructors feel about the topics. There could be a gap between, knowing the importance of the topic and having the skill to teach it successfully.
Which topic is most difficult to teach?

For topics, instructor had the most difficulty teaching was Advanced Statistics (28.6%, n=32) and Basic Statistics (21.4%, n=24). Other answers included Test Theory (14.3%, n=16) and Test Specifications (12.5%, n=14). Noticeable areas that showed difference between LTs and non-LTs were regarding statistics. Although both groups indicated this was the most challenging area to teach, LTs expressed more difficulty in teaching Advanced Statistics (32.1%, n=18, for non LT 23.6%, n=13) while non-LTs said it was challenging to teach Basic Statistics (25.5%, n=14, for LT 17.9%, n=10). This result could be due to Advanced Statistics is more taught by LTs than non LTs. Statistics is a complicated subject for both group of instructors but the areas that are having difficulty in were different.

Table 16

*Instructors Top 5 Most Difficult Topic to Teach (N=112, missing=28)*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Topic</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Advanced Statistics</td>
<td>32</td>
<td>28.6</td>
</tr>
<tr>
<td>2</td>
<td>Basic Statistics</td>
<td>24</td>
<td>21.4</td>
</tr>
<tr>
<td>3</td>
<td>Test Theory</td>
<td>16</td>
<td>14.3</td>
</tr>
<tr>
<td>4</td>
<td>Test Specifications</td>
<td>14</td>
<td>12.5</td>
</tr>
<tr>
<td>5</td>
<td>Classroom Assessment</td>
<td>5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Reasons for why it was challenging (n=91 valid, n=49 missing) were because the topic is difficult for students to understand; test theory (53.3%, n=8), especially concepts about validity fell under this category, also it was due to the limited instructor’s knowledge (31.9%, n=29). Instructors wrote that student teachers have difficulty grasping test theory because ‘the concepts and be abstract and unfamiliar, are often confused with validity and
reliability, the validity framework is very complicated and deep, and providing examples for these concepts is difficult.’

For statistics, particularly for Advanced Statistics, instructors noted they themselves were not familiar with the topics (55.6%, n=15) let alone the students. Instructors commented, ‘I have little training in this area, but I also believe that it is not something that will be helpful to classroom teachers.’ For Basic Statistics instructors thought it was difficult to teach because students had trouble understanding the topic (45.5%, n=10) and due to limited instructors’ knowledge of the topic (36.4%, n=8). Teaching basic statistics is a challenge since, ‘the students are very resistant, they lack interest and background knowledge’. Other answers included the topic was irrelevant to Student Teachers’ practices (14.3%, n=13) and due to time constraints (8.8%, n=8).

Table 17

*Student Teachers: Least Helpful in Teaching (N=206, missing n=38)*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Topic</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Advanced Statistics</td>
<td>36</td>
<td>14.8</td>
</tr>
<tr>
<td>2</td>
<td>Basic Statistics</td>
<td>35</td>
<td>14.3</td>
</tr>
<tr>
<td>3</td>
<td>History</td>
<td>30</td>
<td>12.3</td>
</tr>
<tr>
<td>4</td>
<td>All Helpful</td>
<td>23</td>
<td>9.4</td>
</tr>
<tr>
<td>5</td>
<td>Test Admin</td>
<td>19</td>
<td>7.8</td>
</tr>
</tbody>
</table>

For student teachers, Statistics related topics (Basic and Advanced) were thought to be the least helpful to teaching (29.1%, n=71). Student teachers felt statistics was ‘not applicable to classroom assessment, it did not help with immediate [classroom needs], the information was over my head, it was not something used on a week to week basis’. History
was also a topic teachers did not find very helpful in the classroom. However, there were a few teachers (9.4%, n=23) who thought all the topics covered in the class was helpful.

The main reason for topics that were thought not to be helpful was that it was not practical to teaching. Other answers included student teachers thought it was self-explanatory (e.g., ethics) or they were already familiar with the topics. There were topics (e.g., Statistics) that were not helpful in their teaching but student teachers found it useful for their research. In addition, there were cases when the information covered in the course did not apply to the student teachers teaching situation. One student teacher commented, ‘I have no choice in administering these tests, so ethical concerns [were] not valid’.

In terms of which topic is most difficult to teach and which was least helpful in teaching, both instructors and student teachers agreed it was statistics.

**Important Topics for Classroom Teachers**

Along with how much time was spent teaching the course, a second set of Likert-scale items were asked to instructors to identify how important the topics were for classroom teachers.

**Instructors.** For instructors (N=140) the top 5 most important topics (Table 19) for classroom teachers were; Classroom Assessment, Alternative Assessment, Test Specifications, Test Theory and Rubric Development. Five areas showed a significant difference between LTs and non-LTs; Test Specifications, Test Taking Skills and Strategies, Alternative Assessment, Rater Training, and Test Accommodation. Out of the these five, two topics (Test Specifications, Alternative Assessment) were in the top 5 most important topics for classroom teachers and topics that got the most coverage in teaching. Similar to earlier results, Test Specifications received a significantly higher mean (m=3.54) from Language Testers compared to Non-LTs (m=3.23). Language Testers ranked this the second most important topic for classroom teachers while Non-LTs ranked it #5. For alternative
assessment, the results were the exact opposite. Non-LTs pointed this as the second most important topic (m=3.63) but LTs placed it #4 (m=3.37).

Table 18

*Instructors Topic Comparison: Time Coverage vs. Important Topics for Classroom Teachers*

<table>
<thead>
<tr>
<th>#</th>
<th>Topic</th>
<th>M</th>
<th>#</th>
<th>Topic</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test Theory</td>
<td>3.33</td>
<td>1</td>
<td>Classroom Assessment</td>
<td>3.76</td>
</tr>
<tr>
<td>2</td>
<td>Classroom Assessment</td>
<td>3.31</td>
<td>2</td>
<td>Alternative Assessment</td>
<td>3.51</td>
</tr>
<tr>
<td>3</td>
<td>Alternative Assessment</td>
<td>3.17</td>
<td>3</td>
<td>Test Specifications</td>
<td>3.37</td>
</tr>
<tr>
<td>4</td>
<td>Test Specifications</td>
<td>3.09</td>
<td>4</td>
<td>Test Theory</td>
<td>3.36</td>
</tr>
<tr>
<td>5</td>
<td>Rubric Development</td>
<td>2.95</td>
<td>5</td>
<td>Rubric Development</td>
<td>3.35</td>
</tr>
</tbody>
</table>

In comparison to the earlier set of questions which asked the extent of time spent on teaching the course, we can see the top 5 topics are the same but in different order for importance. Test Theory which was ranked #1 had a lower ranking in topics important for classroom teachers.

Table 19

*LT vs. Non-LT (Time Coverage for Teaching Topics, Important Topics for Classroom Teachers)*

<table>
<thead>
<tr>
<th>Language Testers</th>
<th>Non-Language Testers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currently teaching topics</td>
<td>Currently teaching topics</td>
</tr>
<tr>
<td>Important student teacher topics</td>
<td>Important student teacher topics</td>
</tr>
<tr>
<td>Test theory</td>
<td>Classroom assessment</td>
</tr>
<tr>
<td>Test specifications</td>
<td>Test specifications</td>
</tr>
<tr>
<td>Classroom assessment</td>
<td>Classroom assessment</td>
</tr>
<tr>
<td>Alternative assessment</td>
<td>Alternative assessment</td>
</tr>
<tr>
<td>(continued)</td>
<td></td>
</tr>
</tbody>
</table>
The above table compares the results of the two Likert-scale items in regards to the different instructor groups. The findings are very similar for Non-LTs. There is only a slight difference in the ranking for Rubric Development. However for LTs, the ranking is different for 3 out of the 5 topics. Language Testers spent the most time in teaching Test Theory but ranked #3 for important topics for classroom teachers. Rubric Development which ranked #5 for all instructors was not included in the top 5 most covered topics for Language Testers. This category ranked #6 for LTs.

![Figure 6. Language Tester: Time Coverage for Topics, Important Topics for Classroom Teachers.](image_url)
These findings can imply that non-Language Testers who indicated more than 73.5% (n=50) of their target audience are student-teachers teach the topics they think are important for student teachers. On the other hand, LTs who have more regular students (non student teachers) 46.2% (n=30) are teaching in a way that is targeted for general students and also student teachers. Since the results of the topics instructors think are important for classroom teachers are similar, LTs may teach the course differently if they had only student teachers.

However, another way to interpret these results is that for LTs, more than half of (53.8%, n=35) the target audience are student teachers; therefore, I feel there is too much focus on Test Theory. By devoting extensive time to one particular area can sacrifice the time to cover other topics (e.g. Rubric Development) that are more directly helpful to classroom teachers. I believe it is important for LTs to give more attention on topics that are important for student teachers if they are their main target student population.
## Table 20

**Important Topics for Classroom Teachers: Instructors’ Perspective**

<table>
<thead>
<tr>
<th>Topic</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Specifications Item Writing</td>
<td>139</td>
<td>3.37</td>
<td>.735</td>
<td>3</td>
<td>67</td>
<td>3.54</td>
<td>.611</td>
<td>2</td>
<td>71</td>
<td>3.23</td>
<td>.814</td>
<td>5</td>
<td>2.555</td>
<td>.012 *</td>
</tr>
<tr>
<td>2. Test Admin</td>
<td>138</td>
<td>2.93</td>
<td>.727</td>
<td>8</td>
<td>66</td>
<td>2.95</td>
<td>.711</td>
<td>9</td>
<td>71</td>
<td>2.92</td>
<td>.751</td>
<td>8</td>
<td>.312</td>
<td>.756</td>
</tr>
<tr>
<td>4. Test-taking skills or strategies</td>
<td>138</td>
<td>2.72</td>
<td>.781</td>
<td>11</td>
<td>66</td>
<td>2.58</td>
<td>.842</td>
<td>11</td>
<td>71</td>
<td>2.85</td>
<td>.690</td>
<td>10</td>
<td>-2.039</td>
<td>.042 *</td>
</tr>
<tr>
<td>5. Test Theory</td>
<td>137</td>
<td>3.36</td>
<td>.674</td>
<td>4</td>
<td>66</td>
<td>3.47</td>
<td>.661</td>
<td>3</td>
<td>70</td>
<td>3.27</td>
<td>.679</td>
<td>4</td>
<td>1.723</td>
<td>.087</td>
</tr>
<tr>
<td>6. Basic Statistics</td>
<td>137</td>
<td>2.89</td>
<td>.863</td>
<td>9</td>
<td>66</td>
<td>2.92</td>
<td>.865</td>
<td>10</td>
<td>70</td>
<td>2.86</td>
<td>.873</td>
<td>9</td>
<td>.450</td>
<td>.653</td>
</tr>
<tr>
<td>7. Advanced Statistics</td>
<td>137</td>
<td>1.74</td>
<td>.795</td>
<td>14</td>
<td>65</td>
<td>1.74</td>
<td>.853</td>
<td>14</td>
<td>71</td>
<td>1.76</td>
<td>.746</td>
<td>14</td>
<td>-.161</td>
<td>.872</td>
</tr>
<tr>
<td>8. Test Ethics</td>
<td>136</td>
<td>3.12</td>
<td>.751</td>
<td>7</td>
<td>65</td>
<td>3.03</td>
<td>.684</td>
<td>7</td>
<td>70</td>
<td>3.19</td>
<td>.804</td>
<td>6</td>
<td>-1.209</td>
<td>.229</td>
</tr>
<tr>
<td>10. Classroom Assessment</td>
<td>139</td>
<td>3.76</td>
<td>.494</td>
<td>1</td>
<td>67</td>
<td>3.69</td>
<td>.556</td>
<td>1</td>
<td>71</td>
<td>3.82</td>
<td>.425</td>
<td>1</td>
<td>-1.541</td>
<td>.126</td>
</tr>
</tbody>
</table>

(continued)
Table 20 (continued)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Instructors</th>
<th>Language Testers</th>
<th>Non Language Testers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  M  SD</td>
<td>N  M  SD</td>
<td>N  M  SD</td>
</tr>
<tr>
<td>12. Rubric Development</td>
<td>138 3.35 .751 5</td>
<td>67 3.31 .656 5</td>
<td>70 3.40 .824 3</td>
</tr>
<tr>
<td>13. Rater Training</td>
<td>136 2.89 .795 9</td>
<td>66 3.03 .744 7</td>
<td>69 2.74 .816 12</td>
</tr>
<tr>
<td>14. Test Accommodation</td>
<td>137 2.60 .781 12</td>
<td>65 2.42 .659 12</td>
<td>71 2.76 .853 11</td>
</tr>
</tbody>
</table>

Note: The response scale was as follows: 1 = hardly any time, 2 = a little time, 3 = some time, 4 = extensive time.
Note: Multiple independent t-tests were used to identify the significant difference exists between the two groups. Using the Bonferroni adjustment required significance at the level (p<.0035).
Table 21

*Important Topics for Classroom Teachers: Student Teachers’ Perspective*

<table>
<thead>
<tr>
<th></th>
<th>Student Teachers</th>
<th>K – 12</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1.</td>
<td>Test Specifications</td>
<td>Item Writing</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Test Admin</td>
<td>237</td>
<td>2.45</td>
</tr>
<tr>
<td>3.</td>
<td>Test Critiquing</td>
<td>238</td>
<td>3.11</td>
</tr>
<tr>
<td>4.</td>
<td>Test-taking skills or strategies</td>
<td>237</td>
<td>3.04</td>
</tr>
<tr>
<td>5.</td>
<td>Test Theory</td>
<td>240</td>
<td>3.10</td>
</tr>
<tr>
<td>6.</td>
<td>Basic Statistics</td>
<td>240</td>
<td>2.48</td>
</tr>
<tr>
<td>7.</td>
<td>Advanced Statistics</td>
<td>237</td>
<td>1.59</td>
</tr>
<tr>
<td>8.</td>
<td>Test Ethics</td>
<td>236</td>
<td>2.85</td>
</tr>
<tr>
<td>10.</td>
<td>Classroom Assessment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Table 21 (continued)

<table>
<thead>
<tr>
<th>Topic</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Rank</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Alternative Performance Assessment</td>
<td>239</td>
<td>3.48</td>
<td>.703</td>
<td>3</td>
<td>80</td>
<td>3.73</td>
<td>.477</td>
<td>2</td>
<td>158</td>
<td>3.3</td>
<td>.767</td>
<td>4</td>
<td>4.495</td>
<td>.000*</td>
</tr>
<tr>
<td>12. Rubric Development</td>
<td>239</td>
<td>3.51</td>
<td>.709</td>
<td>2</td>
<td>79</td>
<td>3.52</td>
<td>.658</td>
<td>3</td>
<td>159</td>
<td>3.50</td>
<td>.737</td>
<td>2</td>
<td>.162</td>
<td>.872</td>
</tr>
<tr>
<td>13. Rater Training</td>
<td>232</td>
<td>2.72</td>
<td>.964</td>
<td>9</td>
<td>78</td>
<td>2.68</td>
<td>.987</td>
<td>10</td>
<td>153</td>
<td>2.74</td>
<td>.958</td>
<td>9</td>
<td>-.439</td>
<td>.661</td>
</tr>
<tr>
<td>14. Test Accommodation</td>
<td>237</td>
<td>2.64</td>
<td>1.005</td>
<td>10</td>
<td>79</td>
<td>3.03</td>
<td>.905</td>
<td>8</td>
<td>157</td>
<td>2.45</td>
<td>1.003</td>
<td>11</td>
<td>4.424</td>
<td>.000*</td>
</tr>
</tbody>
</table>

Note: The response scale was as follows: 1 = hardly any time, 2 = a little time, 3 = some time, 4 = extensive time.
Note: Multiple independent t-tests were used to identify the significant difference exists between the two groups. Using the Bonferroni adjustment required significance at the level (p<.0035).
Student Teachers. Similar to the instructors, a set of Likert-scale items that asked the importance of topics for a classroom teacher was given in the survey. The top 5 topics Student teachers thought were important were Classroom Assessment (m=3.69), Rubric Development (m=3.51), Alternative Assessment (3.48), Test Specifications (m=3.36) add Test Critique (m=3.11).

Table 22

Student Teachers’ Topic Comparison: Time Coverage vs. Important Topics for Classroom Teachers

<table>
<thead>
<tr>
<th>#</th>
<th>Topic</th>
<th>M</th>
<th>#</th>
<th>Topic</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test Theory</td>
<td>3.18</td>
<td>1</td>
<td>Classroom Assessment</td>
<td>3.69</td>
</tr>
<tr>
<td>2</td>
<td>Classroom Assessment</td>
<td>2.82</td>
<td>2</td>
<td>Rubric Development</td>
<td>3.51</td>
</tr>
<tr>
<td>3</td>
<td>Test Critiquing</td>
<td>2.74</td>
<td>3</td>
<td>Alternative Assessment</td>
<td>3.48</td>
</tr>
<tr>
<td>4</td>
<td>Test Specifications</td>
<td>2.69</td>
<td>4</td>
<td>Test Specifications</td>
<td>3.36</td>
</tr>
<tr>
<td>5</td>
<td>Rubric Development</td>
<td>2.66</td>
<td>5</td>
<td>Test Critiquing</td>
<td>3.11</td>
</tr>
</tbody>
</table>

From the student teachers perspective, regarding the top 5 most topics covered in the assessment class, the topics were in a different order of importance. The most noticeable finding from this comparison is about the different thoughts towards teaching and learning Test Theory and Alternative Assessment. Test Theory which was thought to have spent the most time in teaching was not in the top 5 category for important topics for classroom teachers. In contrast, Alternative Assessment ranked #3 in regards to important topics for classroom teachers but, ranked #6 in the Topic Time Coverage items.

This could mean that Student teachers do not value the importance of this topic as much as the instructors for Test Theory. Student teachers may feel this topic is not directly
helpful in developing classroom teaching and would prefer to spend much more time learning other topics such as alternative assessment.

As the survey results show, Test Theory is definitely an important topic from the instructor’s point of view. The reason why Student teachers do not agree in terms of the importance may relate to how the content is presented to Student teachers. Specific ways in how theoretical topics should be taught will be discussed in detail in specific topics and suggestions for effective language assessment course in Chapter 5.

Topics that showed significant difference for Student teachers groups were, Test Taking Strategies, Alternative Assessment, and Test Accommodation. K-12 student teachers thought all these three topics were much more important than Adult Student teachers.

Table 23

<table>
<thead>
<tr>
<th>Important Topics: Instructors vs. Student Teachers</th>
<th>Instructors</th>
<th>Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>1. Test Specifications</td>
<td>139</td>
<td>3.37</td>
</tr>
<tr>
<td>2. Test Admin</td>
<td>138</td>
<td>2.93</td>
</tr>
<tr>
<td>3. Test Critiquing</td>
<td>137</td>
<td>3.18</td>
</tr>
<tr>
<td>4. Test taking skills</td>
<td>138</td>
<td>2.72</td>
</tr>
<tr>
<td>5. Test Theory</td>
<td>137</td>
<td>3.36</td>
</tr>
<tr>
<td>6. Basic Statistics</td>
<td>137</td>
<td>2.89</td>
</tr>
<tr>
<td>7. Advanced Statistics</td>
<td>137</td>
<td>1.74</td>
</tr>
<tr>
<td>8. Test Ethics</td>
<td>136</td>
<td>3.12</td>
</tr>
<tr>
<td>9. History of Language Testing</td>
<td>137</td>
<td>2.31</td>
</tr>
</tbody>
</table>

(continued)
For Instructors and Student Teachers, there was a significant difference in six areas (Test administrations, Test taking skills and strategies, Test Theory, Basic Statistics, Test Ethics, and History of Language Testing). This implies the two groups have different perceptions on which topics are important for student teachers.

**Expectations for the Course**

For student teachers an additional Likert-Scale item was included that asked their expectations of the course. The topics they most expected to learn were: how to develop Classroom Tests (m=3.31), Making rubrics, Performance Assessment, Test Theory, and Test Critique.

Analyzed by grade level, the results were significant in 4 topics (Table 25). Adult Student teachers had a significantly higher expectations to learn Test Theory (m=3.26, #2) than K-12 student teachers (m=2.90, #5). For the other three topics, Performance Assessment, Test Accommodation, Test Ethics, K-12 student teachers showed more interest in learning the topic.

<table>
<thead>
<tr>
<th>Table 23 (continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classroom Assessment</strong></td>
</tr>
<tr>
<td><strong>Alternative Performance Assessment</strong></td>
</tr>
<tr>
<td><strong>Rubric Development</strong></td>
</tr>
<tr>
<td><strong>Rater Training</strong></td>
</tr>
<tr>
<td><strong>Test Accommodation</strong></td>
</tr>
</tbody>
</table>
Table 24

*Expectations for the Course (K-12 vs. Adults)*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Student Teachers</th>
<th>K - 12</th>
<th>Adults</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>#</td>
</tr>
<tr>
<td>Classroom Assessment</td>
<td>236</td>
<td>3.31</td>
<td>.811</td>
<td>1</td>
</tr>
<tr>
<td>Make Rubrics</td>
<td>237</td>
<td>3.27</td>
<td>.787</td>
<td>2</td>
</tr>
<tr>
<td>Test Theory</td>
<td>237</td>
<td>3.14</td>
<td>.881</td>
<td>4</td>
</tr>
<tr>
<td>Basic Statistics</td>
<td>234</td>
<td>2.72</td>
<td>.886</td>
<td>7</td>
</tr>
<tr>
<td>Test Critique</td>
<td>237</td>
<td>3.06</td>
<td>.765</td>
<td>5</td>
</tr>
<tr>
<td>History of LT</td>
<td>237</td>
<td>2.36</td>
<td>.875</td>
<td>10</td>
</tr>
<tr>
<td>Performance Assessment</td>
<td>237</td>
<td>3.15</td>
<td>.761</td>
<td>3</td>
</tr>
<tr>
<td>Rater Training</td>
<td>236</td>
<td>2.76</td>
<td>.872</td>
<td>6</td>
</tr>
<tr>
<td>Advanced Statistics</td>
<td>237</td>
<td>1.90</td>
<td>.861</td>
<td>11</td>
</tr>
<tr>
<td>Test Accommodation</td>
<td>238</td>
<td>2.52</td>
<td>.926</td>
<td>9</td>
</tr>
<tr>
<td>Test Ethics</td>
<td>232</td>
<td>2.56</td>
<td>.861</td>
<td>8</td>
</tr>
</tbody>
</table>

Note: The response scale was as follows: 1 = hardly any time, 2 = a little time, 3 = some time, 4 = extensive time.

Note: Multiple independent t-tests were used to identify the significant difference exists between the two groups. Using the Bonferroni adjustment required significance at the level (p<.0035).
Survey results infer that K-12 and Adult student teachers have different needs from an assessment course. Thus it will be helpful for instructors to keep this in mind when designing a course for their students. For example, when designing a course for K-12 teachers, the instructor should stress on rubric development activities and include topics on Test Accommodation. On the other hand, Adult Student teachers will not mind learning more about Test Theory compared to K-12 teachers. To teach a successful Language Assessment Course the instructor needs to keep a good balance of topics that would be of an interest for the target student audience.

What is being used? The Materials

Survey results about textbooks: “This textbook sells well, but not because anyone likes it; it’s the least worst one out there.”-Smith-

From the teacher survey, student teachers were very positive towards the use of textbooks. A total of 78.69% (n = 184) agreed (57.59%, n = 135) or strongly agreed (20.94%, n = 49) to the statement that language assessment textbooks are useful for classroom teachers. Only 21% (n = 50) of student teachers had negative impressions towards language assessment textbooks. However, student teachers had different thoughts in the interview.

Interview findings reveal that currently available textbooks regarding language assessment courses were thought to be difficult and not very useful to the student teachers or the instructors. A Non-Language Tester in particular commented that every time she hears a new textbook is published, she gets very excited, but once she goes over the content, she sighs and quickly closes the book. The overall comments regarding the textbooks was that the majority of them were targeted for large scale assessment usage, so the content covered in these books was not directly related to language teachers’ needs. Anderson commented, “I’d like to see [textbooks] more in classroom assessment. Ultimately, that is for the population I work with; that’s what they’re going to need.” She did understand the need for teachers to be
aware of these issues surrounding standardized assessment and felt a lot of the content covered in standardized assessment is applicable in the classroom context. Yet, a veteran Language Tester, who has written several books in this field, commented that there is little difference in theory regarding large-scale or small-scale assessment. He states, “Validity is validity, and reliability is reliability.” Another opinion from a non LT instructor was that textbooks seemed to be appropriate for Ph.D. students majoring in language testing rather than M.A., undergraduate pre-service, or in-service teachers.

Mark (a student teacher who currently teaching at a community college) also thought the textbooks were targeted for large scale assessment context. “[The] textbook was designed for people [dealing with] standardized tests used on [a] large number of people. … Half of the [material was] not feasible for a small-scale test for one teacher to use.” Mark who was very happy with the course itself, felt the textbook that was used was much more geared to a large-scale test. He states, “[Regarding the] analysis of different kinds of [large scale] test items, the analysis [in the book] was insightful and useful. [Regarding the] classroom test, still a lot of useful information in it, were not appropriate, [nor] insightful.”

Textbooks written by Language Testers are rich in depth and content. They cover test specifications, ethics, and a lot more on standardized assessment compared to classroom-assessment-oriented books. However, these books did not seem to be used much by non LTs in teaching their courses. Although Anderson appreciated new books in the field, she commented that they could be over the top for her student teachers. “I refer to it [a textbook] a lot; I read it out loud, and students borrow it and use it, [but it is] too dense and not practical enough [for my student teachers]. I do not think it would be a useful resource [for student teachers] to have them on their shelf. It’s a great [resource] for me, but not the students.” Anderson liked the top-selling Language Assessment Textbook in the market, not necessarily because it was the best in terms of the content, but it was reader friendly and accessible to non-native speakers. As Smith mentioned in regards to the same textbook, “This textbook
sells well, but not because anyone likes it; it’s the least worst one out there.” It appears that because a textbook is a best seller does not necessarily mean it has the best quality.

When writing a textbook, I am sure authors have many factors to consider; the scope and breadth of the content, the target audience, and how it will be used are a few. From the interviews, I found one area that textbook writers rarely consider in writing a book is: whether or not the reader is a native speaker. If one considers the huge population of non-native English teachers in the market, this is definitely something to keep in mind. Books that are written clearly and easily comprehended by student teachers are more appreciated.

It is beyond the scope of this dissertation to evaluate the quality of textbooks, but experienced instructors appear to rely more on their own practice and materials they possess rather than textbooks. Anderson commented, “I’m lucky I already have a lot of resources for my students [teachers] to use.”

From the interviews, it showed that non LTs and teachers were unhappier about the textbooks than the LTs. This could be because the majority of language assessment textbooks are written by Language Testers.

**Why are textbooks important?** Language testers and non-language testers may teach their courses differently, but both groups used similar textbooks; however, the degree of satisfaction varies among instructors. For example, Goldberg used the same book as Carol’s class. Due to such disappointment in textbooks, many language testers expressed interest in writing one or are in the process of writing a textbook.

The majority of the language assessment textbooks in the market are written by professionals in the field. The overall impression from the student teachers and instructors (non-LTs in particular) is that these books are pretty difficult and dense. When writing a textbook, the writers should be clear of the target audience. If it is a book geared towards an introductory-level course, the language and content should match the audience who will use the book. Writers should remember the majority of the audience who will be reading these
books will not be experts in the field, thus, in most cases, the books should not be targeted for graduate students who will be majoring in language testing.

Then how do student teachers feel about language assessment textbooks? Bell indicated that her level of satisfaction for the textbook she used for the course was satisfactory. She thought it provided the baseline to work on. Even though she had never referenced the book since her graduation, she did mention, “But I have thought of it. I have thought of the pages he wrote of eugenics and the army test.” The textbook Bell used didn’t directly help her in classroom teaching, but the concepts she learned from the book and class are always on her back burner.

This comment made me a bit confused. I thought a textbook that was rarely referenced in teaching was not a good textbook, but it appears that there are different types of textbooks: books that are practical and hands-on that can be on the top shelf of a teacher’s desk, and books that student teachers used when taking the course but that have not been used since. Just because a teacher does not go back to the book doesn’t mean it wasn’t helpful. The books could have been useful at the time they took the course. The knowledge acquired from these books can implicitly affect the teacher. This could be one of the reasons why the degree of satisfaction with the textbooks was high in the survey.

**Changes in the Course**

When Ellen began teaching her course, she relied heavily on what she had learned as a student. She spent most of the time discussing validity, test fairness, and other theory that is covered in a traditional language assessment course, then talked about statistics, and at the very end, covered some practical stuff. She recalls that student teachers were overwhelmed when she taught it in this form. Therefore, over the years she has revised the course to be a lot more practical. She states, “At the beginning, it was mostly the classical testing course, but it started to turn out something more practical, because that is what the [teacher] students want. That is what the students need.”
Similar to other teacher education courses, the structure and content of an assessment course changed as time passed. Some changes like Ellen’s case were made by the request of the student teachers, others reflected the development of the assessment environment (e.g., using technology in assessment). Since assessment courses are impacted by both the teaching and testing environment changes in these areas will result into courses that are different from the past.

**Keeping up with the changing classroom situation.** By the time an instructor is in the position to teach a college-level pre-service/in-service course, it is likely that many years have passed since they taught in a language class. From the survey the majority of the instructors have taught a language class, yet their experience would be quite outdated by the time they have received an advanced degree. Then what could be done to overcome the gap between the instructor’s past teaching experience and the current classroom situation?

One way is to build on the student teachers’ prior teaching experience. Experienced student teachers are the experts in the classroom. They are well aware of what’s going on, what the students’ needs are, and what has to be learned regarding assessment. It is important for the instructor to value and build upon what the student teachers bring into the assessment course. This is also directly linked to needs analysis. Instructors should ask what the teachers want in the course and develop it upon their needs. Courses that teachers thought were successful, and courses in which instructors showed passion were those that utilized student teachers’ prior experience, either as teachers or test takers. Student teachers appreciated when they were acknowledged for their prior teaching experience in regards to assessment and were more likely to engage in the course. Instructors who valued teachers’ experience successfully incorporated and pulled together previous knowledge of language assessment.

**Assessment Courses vs. Other Courses: Is there coordination?**

From the interviews, most assessment courses appear to stand alone. There is little interaction or coordination among other courses. The ones that were successfully integrated
with other courses (e.g., Teaching Methodology) were the courses that were taught by the same instructors. Ryan did a good job connecting the Methods course with the Assessment course. The Teaching Methodology course which is taught a semester before the assessment course ends with small assessment component that is connected to the following semester. Ryan states co-ordinations with other courses was possible because she taught both classes.

In rare cases, if the integration was considered from the development stage of the course, connection with other courses was possible. For example, Smith was one of initial members to set the M.A. program she is working for. While building the course she stressed the importance of including an assessment course in the program and carefully considered the relationship between other courses and the assessment course.

The lack of coordination between other courses may not be a problem for only assessment courses. Many courses in the M.A. TESOL program are developed separately and taught by different instructors. There are unseen political and practical reasons why this coordination is not happening. This is a concern from the student teacher’s perspective, because student teachers who learn assessment separately from other courses may end up viewing it the same when they go out to teach. They may teach a lesson and develop a separate test that is not directly related to the content of the course. The backwards design (Wiggins & Mc Tighe, 1998) should be applied not only in a classroom setting, but in the college training curriculum as well.

**Language Assessment Courses Around the World**

The majority of language assessment courses are taught in English and are about English tests. However the depending on the context of the course (where it is taught, who is it for), assessment courses comes in different shapes and forms. For instance, there are courses taught in English, but in a non English speaking country. Jackson, who is a professor in an East Asian country, teaches this course to a group of non-native English speakers. He indicates that, along with the challenges of teaching the content of the course, he also has to
consider the student teachers’ limited language proficiencies. He commented that doing group discussion about assessment is not easy since not all student teachers are comfortable conversing in English.

The courses are also taught in different languages other than English. For example, Ellen teaches this course in French for prospective French teachers in various subject areas. The course is contextualized and reflects the unique educational and cultural environment. And of course, the content of the course depends heavily on the testing environment in which the course is situated. Instructors (Jackson, Karimi, Rivera) who teach the course in non-English speaking countries (South Korea, Iran, Ecuador) that still focus on multiple-choice, short-answer language tests, teach student teachers how to develop good test items for those types of tests.

Since methods and purpose of testing vary so much depending on where, who the student teacher is teaching, assessment courses have difficulty coping with these diversities. This is especially difficult for courses offered on-line, where student teachers come from all parts of the world, and the instructor does not have the background knowledge of the teachers’ situation. Hennes, who is the director of an on-line course, stresses that instructors should come down to the most common denominator when applying testing projects to student teachers. After establishing a common understanding of testing concepts, testing principals should then be applied to student teachers’ specific context. She does this to make the subject applicable to the entire student teacher audience, but at the same time she wants to be individualized to meet their specific needs.

Summary

This chapter presented the findings of the survey and interview comments from instructors and student teachers. The first half of the chapter focused on presenting the overall features of Language Assessment Courses. What topics are being taught, who teaches it, and how the course is different between LTs and non LTs. Also, the chapter presented how
student teachers perceived the course. The remaining parts covered what is thought as an effective language assessment course for student teachers.

**Nov 31, 20xx**

**Faculty Search Committee Meeting Continued**

Sagon: So, what do you think?

Lee: First, in terms of what topics to cover for the course, I think we should pay more attention to the needs of the student teachers.

Kent: I agree. The majority of the students of our program are pre-service adult student teachers so I believe we should focus on what they want in an assessment course.

Sagon: All right, then what about the instructor? Who should we hire?

Horwitz: As I said earlier hiring a language tester is the best choice. They are so much stronger teaching test specific topics such as Test Specifications and Statistics.

Heejeong’s study confirmed this.

Lee: I kind of agree with Mia on this one. From past experience, statistics was a challenging topic to cover for non Language Testers.

Kent: Ladies, did you miss the point where student teachers said statistics isn’t really an area that is much of an importance to the students? If we were to hire a Language Tester because they can teach statistics, then I think we should hire a person who majored in quantitative research methods. There are many who are looking for jobs these days.

Sagon: Folks, I think it is important for us to look this from different points of views. We need to draw out the pros and cons of each person. Also, we have to consider our budget. A person who has a lot of experience could go over our budget. I don’t want to fight with the dean again.
Lee: Heejeong, what do you think? If you were in the search committee who would you hire?

Jeong: This is not an easy decision. I probably will propose to hire a language testing professional. That was how I was educated and became interested in this area. However, my school was a much larger school and we had the luxury to hire different people for different areas. I’m not sure if this is the same for your school.

Kent: You have a good point there. There are many factors to consider here. Out budget, our school, our students…etc.

Lee: What about hiring a Language Tester who has plenty of classroom teaching experience similar to our student teachers. I think we should hire an instructor who knows about the pre-service teachers’ teaching context.

Kent: Jina, don’t you think there is a very slim possibility to find a person who falls under that category? Even from Heejeong’s study there were very few Language Testers who actually had K-12 teaching experience.

Horwitz: I must agree with David. We might be asking for too much in one person. We should be willing to give up a few things.

Jeong: Well, considering the majority of your student teachers will be teaching in a K-12 context, I personally think that is a very important criterion among many other qualifications. Student teachers really appreciated instructors who knew well about their teaching environment and could give practical hands on directions and help.

Sagon: Okay, it appears we have better idea of what to consider in hiring a faculty member. Let’s have another meeting next month to finalize our ideas. In the meantime, Heejeong could you give a brief presentation of the discussion and conclusion chapters from your study in our next meeting?

Jeong: No problem. I’ll also send a copy of the chapters a week before our meeting.
Chapter V
Discussion

Dec 10, 20xx

Faculty Search Committee Meeting

Participants: Charles Sagon, David Kent, Jina Lee, Mia Horwitz, Heejeong Jeong,

Sagon: Thank you for sending the chapters ahead of time Heejeong. It was quite helpful.

Heejeong: I'm glad you liked it.

Lee: Just to refresh our memory, Heejeong could you present a brief summary of the topics that were widely discussed among instructors and student teachers and also the characteristics of courses that student teachers appreciated?

Heejeong: Sure. To being with I will first talk about the eight important topic areas in language assessment courses. These topics were those that showed significant differences between groups and areas that were frequently mentioned in the survey and interview were selected to discuss the effectiveness of the courses. Then I will propose four requirements for a language assessment course to be successful.

Research Question 3: How effective are language assessment courses for language teachers’ everyday classroom practices?

Test Theory

“How often do we refresh ourselves on theory?” – Bell-. Out of the 14 topics, instructors spent the most time teaching test theory. For specific groups, LTs ranked this #1 (m = 3.48), and non-LTs ranked this #3 (m = 3.19). Both group of instructors thought it was important to teach test theory, but results show that Language Testers value its importance more than non LTs. Student teachers also answered that instructors spent the most time in
teaching Test Theory ($m = 3.17$). Although instructors and student teachers will not doubt the importance of teaching or learning theory in an assessment course, the degree of importance of this topic varied depending on the group.

Teaching a theoretical topic is always tricky for instructors. In the open-ended survey question instructors ranked this as the third (#3) most difficult topic to teach. Student teachers felt the theoretical concepts to be abstract and unfamiliar and it was also hard to find practical applications of Test Theory for student teachers. Despite these difficulties, there were instructors who claimed success in teaching this topic. For example, Ellen noted in the interview, theory should be presented in “little doses” during practical work. She felt instructors should not dominate the class with theory. Other ways to teach the topic more effectively included presenting the course relevant to the student teachers’ teaching context.

Instructors are aware of what is needed or requested by the student teachers to some extent, but not all instructors present the material in a way that can be effectively understood by the student teachers. Not only should instructors have a good base of knowledge of the content but they should also possess effective teaching skills to be well perceived by their student teachers.

That said, what do student teachers think about learning Test Theory? Bell believes it is important to learn theory, but, as a teacher she doesn’t often use it in her job. She said, “If I wanted to show it off to someone, like present it at TESOL, I might have to go back to [the books] to show what the parts are but for the practical application, in-service teachers, unless we are presenting at a site where there are other language teachers, [theory] is not a necessary thing. Even when we present to colleagues we work with, we don’t have that common jargon.”

The level of theoretical expertise varies among teachers. Some teachers have very little theory behind them and learn on the job while others begin with a strong theoretical background. Bell believes there is a difference between those teachers (who learn on the job)
and a traditionally trained teacher like herself. For example she comments, “I know how to handle when there is [a] pronunciation topic where my colleagues have to go and look something up or let the opportunity pass. I have that mastery.” She believes one way for pre-service teachers to be competitive with those who are currently on the job would be by having a strong theory based understanding. Other student teacher comments from the open-ended questions were, “I felt that too much time was spent on test theory., Although knowing the basics is helpful for writing tests, [the instructor] goes so far in-depth which seemed unnecessary.”

We might think teachers do not like theory, but the interviews and survey findings revealed that teachers know the significance of having a strong theoretical background. They might not use it day to day in their jobs but acknowledge the difference in not having it. The key to teaching this topic well is probably in presentation. Similar to other topics like statistics, student teachers appreciate theory when it is taught directly related to a practical setting.

**Classroom Assessment:** “I know this is important but I don’t teach it much.” -Anderson-

Another important topic that was covered in Language Assessment Courses was classroom assessment. From the instructor survey, classroom assessment ranked the second most covered topic (m=3.31) in Language Assessment Courses. Non-LTs responded that they spent the most time (m = 3.46) teaching Classroom Assessment above all other topics, and LTs ranked this third (m = 3.15). For the topics that were important for student teachers, both instructor groups pointed out that classroom assessment is the most important topic for a classroom teacher. Even if both groups of instructors agreed that classroom assessment is the most important topic for pre/in-service language teachers, for LTs, Test Theory and Test Specifications had more time coverage than classroom assessment.

Why then is there a difference in the time allocated for classroom assessment for LTs and non LTs? Survey findings show that LTs spent significantly less time (p=.024) teaching
this topic to non LTs. There could be a few different reasons for this phenomenon. First, an 
instructor can only teach what he or she knows. Instructors, who have little classroom 
teaching experience or have been away from the classroom for an extended period of time, 
may not be familiar with the current classroom situation. Even if they knew the needs of the 
student teachers, instructors may not feel comfortable teaching it; therefore they resort to 
more theoretical concepts of language assessment, such as Test Theory.

Anderson commented that, the reason she doesn’t cover much of the topic is, “I think 
it’s because we [instructors] haven’t done it [classroom assessment]. I think that is why I 
don’t do it enough. I’ve been out of the ESL classroom [for a long time].” Even though the 
survey findings show that, 96% of the instructors had language teaching experience and 70% 
have taught for more than 10 years, their language classroom experience can be outdated. For 
instructors to best teach this topic, it would be ideal if they had a chance to teach a language 
course from time to time to refresh their memory about classroom teaching and assessment.

Another reason for not extensively covering Classroom Assessment could be because 
of the content of current language assessment textbooks. Though there have been recent 
changes, the available textbooks in the language assessment field are still largely focused on 
large scale assessment. Even for instructors who may want to teach more classroom 
assessment, since there is such a limited resource on the topic, it may not be easy to get 
sufficient information regarding it. This could be particularly the case for non-LTs who have 
a tendency to rely more heavily on textbooks than LTs. Olson just began teaching this course 
relied a great deal on the course textbook. His syllabus and activities closely followed the 
structure and content of the textbook. He even pulled out the textbook during the phone 
interview when asked to describe his course. Instructors like Olson, who had little testing 
experience or background are likely to refer to available resources. If the given sources are 
targeted for large scale tests, then that will be the topic he or she will teach.
Finally, covering practical classroom assessment activities can be hard for instructors who cannot offer a real classroom setting for their student teachers. Without a classroom setting, it is not easy for student teachers to develop authentic classroom assessment activities. From the interviews, language assessment courses that were successful had one particular aspect in common (for example, Mark and Anderson). Mark’s instructor and Anderson provided an actual classroom setting with which student teachers could work with. The instructor gave a set of students they can pilot their test to and provided authentic examples either from the classroom or textbooks used in the class. Student teachers were able to apply the content they learned to students which played an important part in making the course useful. Mark’s instructor had the student teachers develop a test for a set of students in Taiwan and Anderson required her STs to make tests that will be actually used in an Intensive Language Institute.

The concept of classroom assessment is broad, and the depth and breadth of the topic varies on the instructor or student teacher. For example, Anderson who is the director of a language program and teaches an assessment course, focuses more on how the classroom tests meet the standards of the program. Aligning with program standards is an important issue, regardless of the age level in which teachers teach. This is especially stressed on K-12 teachers who are obligated to meet specific standards in their school district. However, are instructors familiar with the district, school or program standards? Are they aware that they need to train their teachers to meet the standards and develop classroom assessments that align with them? For instructors to successfully teach their courses, especially in terms of assessment, it is important that they have this knowledge.

**Alternative/ Performance Assessment**

Alternative Assessment is a topic that was valued by both K-12 student teachers (Important Topics for Classroom Teachers: Student Teachers’ Perspective, #2, m= 3.73) and adult student teachers (#4, m=3.36). Instructors also thought this topic was important and
ranked #3 (m=3.17), in regards to the topic time coverage, but the student teachers did not feel the topic was given as that much attention (Time Coverage from the Student Teachers’ Perspective, #6, m=2.57). This implies it is important for instructors to be more explicit in teaching this topic. It appears that instructors believe they are covering the topic extensively but student teachers do not agree. It is especially important for instructors who teach K-12 student teachers to devote more time in teaching this topic since K-12 student teachers ranked this #2 (m=3.73) out of the 14 topics important for classroom teachers.

From the open-ended survey questions student teachers commented that alternative/performance assessment techniques were important because “[standard] testing does not allow us [teachers] to take into account non-standard students, who may be good practical achievers”. The student teachers believe for English Language Learners an alternative assessment tool could be a better way to measure students’ abilities. There were student teachers who were against the idea of standardized testing and believe there should be “a better, more ethical, more motivating way to evaluate student performance”. Student teachers thought that a paper and pencil test is not always the best indicator to show a students’ language proficiency and preferred to become “better equipped to assess students’ abilities through portfolios, performance tests and other alternative test forms.”

**Test Specifications/Item Writing**

For test specifications, there was a significant difference (p = .020) between LTs and non LTs. The overall ranking of this topic was #4 (m = 3.09), but for LTs, this was the second topic on which they spent the most time (m = 3.26) teaching. However for non-LTs, it ranked #6 out of the 14 topics, with a mean of 2.93, which was significantly lower than the mean for the LTs. From the interview, Bell who took the class from a Language Tester indicated that the most important topic that she learned from her language assessment course was test specifications.
Rubric Development

Rubric development was a topic student teachers valued much more than the instructors. It ranked #2 (m=3.51) for important topics for classroom teachers from the student teachers’ perspective and also ranked the second (m=2.37) most expected topic to learn from a language assessment course. However, instructors did not think this topic was as important as the student teachers. It ranked #5 (m=2.93) in regards to topic time coverage from the instructors’ perspective and #5 (m=3.35) for the item important topics for classroom teachers.

For classroom language teachers, developing good rubrics appear to be one of the main tasks they do as an evaluator. Activities for language classes are often assessed using rubrics. Student teachers from the open-ended survey items commented they work in an e-portfolio and rubric driven school and preferred to see more focus on those topics. They feel that learning more about creating rubrics would be useful in their current teaching situation. The main assessment tool student teachers were using to assess students’ speaking and writing skills were through rubrics.

Statistics

“If we teach testing from the point of view everyone is going to be a psychometrician this will not be helpful for those teachers.” –Smith -. The topic of teaching statistics is quite controversial in introductory language assessment courses. Different instructors had different ideas regarding whether or not to cover this in their course (Jeong, 2009). Some instructors (e.g, Ryan, Smith, Olson) hardly taught any statistics and others (e.g., Wilson) spent a majority of their time covering this topic in their courses. Goldberg who is a veteran LT thinks “By teaching statistics in relation to language testing, this would be more meaningful for the teachers.” Ellen who has been teaching this course for around 5 years, teaches only basic descriptive statistics for her class. She tries to present them
in a simple way that makes sense to her student teachers. She quotes, “It’s over their heads if I go into the advanced interpretive statistics.”

Table 25

*Instructors Thoughts on Teaching Statistics*

<table>
<thead>
<tr>
<th></th>
<th>Instructors</th>
<th>Language Testers</th>
<th>Non-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>136</td>
<td>3.32</td>
<td>.749</td>
</tr>
<tr>
<td>Interpretive Statistics</td>
<td>134</td>
<td>2.96</td>
<td>.778</td>
</tr>
<tr>
<td>Advanced Measurement Theory</td>
<td>138</td>
<td>2.25</td>
<td>.838</td>
</tr>
</tbody>
</table>

*Note:* Scale is as follows: 4 = Strongly Agree, 3 = Agree, 2 = Disagree, 1 = Strongly Disagree

In terms of teaching statistics in an introductory language assessment course, instructors agreed that the course should cover descriptive (m = 3.32) and interpretive (m = 2.96) statistics, but disagreed (m = 2.25) about teaching advanced measurement theory. In regards to group differences, both instructor groups had similar opinions on whether or not to teach interpretive or advanced theory for the introductory class, but there was a significant difference (p = 0.005) when it came to covering descriptive statistics. Language Testers felt much more strongly (m = 3.52) about teaching this compared to non LTs (m = 3.16).

Smith, who is a non LT, believes that statistics are important, but teaching it adds apprehension and elitism to testing. She believes that there are different levels of assessment literacy for student teachers; some will be able to do advanced statistics and measurement, and others, though quite knowledgeable, will not have that level of knowledge. She feels it is important to recognize the level of knowledge student teachers will need regarding statistics. She comments, “If we teach testing from the point of view that everyone is going to be a psychometrician, this will not be helpful for those teachers.”
Smith is not against teaching statistics, but feels language assessment courses should not over emphasize this subject area, since it could scare teachers away. Smith feels, “Statistics has a role in testing but should not go beyond the basic topics (e.g., understanding the concepts of Standard Deviation, skewed distribution, norm referencing).” Her concern for overdoing statistics is that it will sacrifice other important topics that could be taught in that time. “If you spend a lot of time on them, there will be other issues you won’t be able to spend time on. Given [that] most of my students will be classroom teachers [or become the] head of a language center, it is important that they are studying the type of assessment they need to do [rather] than high level statistics.” She feels that at the intro level, there are so many other issues that are important, that going beyond understanding the basics is more than what they need. Student teachers who are really into testing should take other courses. She ends by commenting, “I do cover the basics in the course; I only have [a] limited number hours, [so] we do not go into depth.”

Hennes, who is the head of an M.A. TESOL program, said the reason they changed instructors for the Language Assessment course was because of statistics. This course used to be taught by Hennes and other non LTs, but as head of the program, she felt statistics would be better taught by a language testing professional. Hennes comments that she wants the student teachers to learn statistics as part of their instruction. From this case, we can see the importance of teaching statistics and why it should be taught by a language assessment professional.

**Different Levels of Statistical Expertise.** Although the degree of knowledge instructors have in statistics or quantitative research methods is beyond the scope of this study, it was obvious that the level of expertise varied depending on how the instructor was educated—trained in the area he or she worked in. Some instructors and most of the LTs possessed a very advanced knowledge of the area and were comfortable teaching all levels of
statistical courses. Other instructors stated frankly that they were not strong in this field and did not possess enough knowledge to teach the course.

Then is the quality of the course threatened if the course instructor does not have sufficient statistical background? My answer to this question would be ‘yes’ and ‘no’. Obviously, assessment is associated with numbers, whether student teachers like it or not. We usually cannot discuss the reliability or validity of a test either large or small scale without mentioning the numbers associated with it. There are certain areas in testing that must be interpreted with numbers (i.e., item difficulty, item analysis). Yet, for an introductory assessment course, I believe the extent that should be covered should be at a minimum. One of the preconceptions people have about Language Testers is that they have strong background in using statistics and measurement tools. People assume testers have sharp technical skills which make them stand out from other applied linguists. However this does not apply to all Language Testers. I do agree that Language Testers compared to other applied linguists may have a stronger knowledge in statistics but this also varies depending on how the Language Tester was trained and practiced. That is, a self-identified Language Tester come with different levels of statistical knowledge and their areas of expertise in Language Testing are not the same.

From the interviews, I noticed a subtle hesitation when I asked questions regarding statistics to instructors who commented that they did not have a strong statistical background. Even though, they did not openly say this, I had an impression that they seemed concerned about this and did not feel confident about the fact they were teaching a course without a confident statistical background. Although statistics is important in assessment, this did not seem to be the main factor that determined the quality of the course, by which I mean: to what extent the student teachers were satisfied with the course? None of the student teachers I interviewed or surveyed commented on or doubted the expertise of the instructor in statistics. They were more concerned about topics directly related to teaching, such as classroom
assessment, rubric development and alternative/performance assessment. Student teachers were more interested in how the course content applied to their own teaching rather than looking for an instructor who came with a good statistical background. For student teachers, learning statistics was not of particular interest compared to other topics. Expectations for learning basic statistics ranked #7 (m = 2.72) and advanced statistics (m = 1.90) ranked at the bottom of the 11 topics. Results were similar for the question that asked how important statistics were for classroom teachers. Out of the 14 topics, student teachers ranked basic statistics #12 (m = 2.48) and advanced statistics #14 (m = 1.59). For student teachers, this topic was not crucial or beneficial in their classroom assessment practices.

However, instructors did not necessarily agree with the student teachers on this topic. Same as the teachers, instructors replied that learning Advanced Statistics was the least helpful topic (m = 1.74, ranking #14) for student teachers, but for Basic Statistics they had a different opinion. Instructors felt that basic statistics was an important (m = 2.89, ranking #9) topic for classroom teachers. Although not significantly different, the mean was slightly higher for LTs (m = 2.92) compared to non LTs (m = 2.86).

What do student teachers think of statistics? It is commonly believed that language teachers are not close to numbers. This may be true in some cases, but not all. From the interviews, some teachers (Mark) were not necessarily apprehensive about statistics. Anderson commented that, surprisingly, a few of her student teachers had taken statistics as undergraduates, found it interesting, and were excited about doing statistics. For student teachers who disliked the topic, it was because they could not make the connection between the content and their teaching. If the instructors clearly articulated the reasoning for teaching the content, student teachers were not as resistant.

When teaching statistics, Anderson covers only the very basic topics, such as simple item analysis, discrimination index, and facility index. She comments that her student teachers, who at first were terrified about statistics, realized later that statistics can be really
quite “easy and interesting.” Student teachers find it to be an effective tool to develop good test items for their tests. The reason for her success in teaching this topic is that she applies statistical tools to the tests student teachers have created, administered, and piloted. She asks teachers to plug the results of the 15–16 pilot tests into excel and run the stats. Anderson said that this makes everything more “real.” Making a statistical connection to a test student teachers themselves have created makes the content more interesting, and student teachers easily find the effectiveness of the tool. Anderson notes that when she teaches this topic, she makes sure teachers are not “bogged down” on statistics. She feels teachers should know the basics, but she draws a line regarding to what depth it should be covered in her class.

Regarding statistics, Mark, who teaches ESL at a community college, said, “At first, I didn’t like it; I don’t enjoy running statistics because I can run into problems, but I enjoy looking at the results of the data.” He felt running statistics on a paper-based, multiple-choice questionnaire was not difficult and found out this was one more useful tool for language teachers. Mark enjoyed statistics because it gave him a chance to be reflective about how tests are being used and how tests are meeting specific needs. By running an item analysis, he was able to find out which items were helpful, useful for different kinds of purposes, and became reflective of the tests that were used for his students.

In Bell’s case, she liked the way her instructor talked about statistics, even though she noted that she may never use the formula or specific features in her job. She felt there was a change in how she perceived large-scale tests after learning about test statistics. “Honestly, in all [cases], I will never use it again and I don’t remember much of the [specifics]. But knowing how a big test is normed out, all the things that went into it, knowing the making of the test, in [a] way makes you trust them a little bit more.” She learned that if the test is properly validated, it can be a really useful tool in teaching. Bell felt many ESL teachers who are not aware of the validation process of a big test are likely to have negative feelings towards it, i.e., feel that test developers have no idea what really goes on in
a classroom and what students really know. After discovering the statistical scrutiny that tests go through, Bell came to have a more positive impression about large-scale tests, “If the test is properly validated, it can be a really useful tool, because I know there is a development [process that goes] into it. It is not someone forcing me to do something.”

**Presentation: the key to teaching statistics successfully.** Anderson’s approach to statistics is successful because she provides materials that are manageable for the teachers. “I don’t force them to do math. Maybe that’s the difference. I don’t make them plug the formula in. I give it to them readymade and I go over it.” She does not care for the calculations. Bell also mentioned that she benefited from the statistics because, “He [the instructor] gave just the taste and you understood why it was related, he taught it with a purpose, and he stopped at an [appropriate] level.” Student teachers appreciated learning statistics when it was taught with a purpose and at a level that was appropriate for them.

Knowledge of statistics may not be the core factor in constructing an effective language assessment course, but I do feel it is important for all instructors, especially for those who are not familiar with statistics to gain at least a minimal knowledge of the topic. They may not necessarily end up teaching this to their student teachers, but still having it in their back burner will make them more confident in teaching the course.

**Ethics**

Test ethics is presently receiving a lot of attention compared to the past. The issue of fairness is also a hot topic in the testing society. Instructors ranked Test Ethics #7 (m = 2.53) out of the 14 topics the taught for Language Assessment courses, and the results were similar for both LTs and non LTs. Language Testers ranked it #8, and non-LTs ranked it #7. The mean (m = 2.53) was the same for both groups. However for student teachers, this topic was not thought to be as important as for the instructors. For the question about how important the following topics are to classroom teachers, instructors thought it was a great
deal more important (m = 2.53) than the teachers (m = 2.14). There was a significant
difference in regards (p = 0.000) to the importance of this topic.

So why do instructors think it is crucial to teach this topic? Anderson comments
since that English tests are such a powerful tool, it is critical for students to be fairly
evaluated. She thought test fairness was especially important for people who are taking high
stakes tests that result into making life changes in situations such as immigration, jobs and
access to higher education. She feels that because the population [test takers] does not have
power and does not have the tools to critically evaluate a test, it is “critical for people in the
profession [student teachers] to have a sense of ethical responsibility because the decisions
are so important.” Many test takers are usually the minority and disempowered group in
society. As Anderson said, these people are likely not to complain and often do not have the
language to stand up for themselves. Therefore, it is crucial for instructors to deliver the
information to student teachers regarding test ethics and test fairness, because it will be the
student teachers who will be meeting and communicating with the test takers.

For student teachers, the reason why they are not as aware of the importance of the
topic could be because they may lack the knowledge of the political and social features
associated with testing. One student teacher commented in the open-ended survey, “[For
ethics] you either have them or you don’t. Taking a class on it won’t help make you change if
you don’t want to.” From this comment, I feel the student teacher is not clear what Test Ethics
is. Test Ethics goes way beyond being honest on taking a test and is a much more complex
issue. Test Ethics is different from moral ethics and is a topic that must be taught to STs.
From the student teachers quote, many STs need to have a better understanding of the topic;
therefore it is the instructor’s job to discuss the concern and problems regarding ethics and
fairness of testing.
Test Accommodation

Out of the 14 topics, the topic that showed the most significant difference between instructors was Test Accommodation. Language Testers do not cover this topic to the degree that non LTs do. One reason for this phenomenon is that, more likely, LTs’ target audience is teachers who will be teaching in the higher education context, compared to non LTs, who have a bigger K–12 audience. From the interviews, K–12 teachers Bell and Carol indicated that accommodating and modifying tests is one of the most important roles they have as ESL teachers and wished they knew more about it. Carol, who is an elementary school teacher, said, “It [test accommodation] is extremely important for K–12 teachers. They [the instructors] did make us aware [that] there was such a thing as test accommodations, but they did not really get into.” She wished she knew more about the specific examples on how to accommodate tests. In the interview, she recalled that, due to lack of knowledge in this area, she made a mistake when she first gave students a math test that was above their reading level. During the math test, she read the test out loud for her students, which gave clues to the answers. While she was taking the assessment course, she was never given any information on how to read the numbers and did not know there was certain way to read it that prevents students from identifying the answers. Therefore, when she first gave the test, she did not do this correctly. She wished she had known this prior to giving the test.

Bell, who teaches K - 12 expressed a similar dissatisfaction similar to Carol’s. Bell commented that she had appreciated learning the theory and doing the test specification activity, but later, when she began working she found out that the main task she does, which is sheltering tests, was never taught in her language assessment course. “A big part of my job is dealing with classroom assessment regular education teachers give to my students. I take those tests and shelter the English or make the test friendlier by eliminating choices, or doing less matching, choosing language tasks that are appropriate to their level, [and] eliminating sections of tests.” Even though sheltering tests is the main assessment- related activity she
does for her job, she says, “I sure didn’t learn that from my language assessment course. I learned that from reading about SIOP [Sheltered Instruction Observation Protocol] and being forced to do it. It [Learning about sheltering test questions] would have been a great exercise to experience in a testing course.” For Bell, the biggest frustration regarding the assessment course was that it did not cover what teachers need to do in their daily lives.

Another reason why Language Testers do not cover this topic is because instructors who teach prospective teachers who will teach adults place less importance on test accommodation. However, for K–12 teachers this is one of their most important tasks. Thus, this topic should be covered extensively by instructors who teach K–12 teachers, compared to those who do not.

Teaching test accommodation will be difficult for instructors who teach various groups of student teachers. For example, student teachers could come from different countries, or may be teaching in different languages, thus the diversity of the student teachers is much more severe for adult language teachers than for K–12 teachers. In this case, it will be extremely challenging for the instructor to cover test accommodation that is relevant to each unique teaching environment.

Assessment Literacy: Only for student teachers? What about the instructors?

Previous studies (Stiggins, 1997; 2002) in assessment literacy are concerned with assessment skills for teachers. As teachers, they are asked to participate in assessment related activities regardless of who and where they teach. Teaching and assessment should sing the same song so, it is important that teachers acquire good knowledge of assessment before going out into the field. However, from this study I came with a question if the instructors who teach a language assessment course are all “assessment literate,” let alone the student teachers? How do we know if they have the appropriate level of knowledge to teach the course?
For Language Testers, the evidence is pretty obvious. Professionals in the field will probably have published and/or did major research in the area of Language Testing. Publications and research experiences in the field would have been one of the criteria in selecting faculty members. Then what about the non-Language Testers? Did they go through a similar screening process as the LTs? There is a slim chance non-LTs were asked to provide evidence of research or publication for an area that is not their primary interest.

**Can Non Language Testers Teach a Language Assessment Course?**

Being concerned with the level of expertise of non-LTs, a controversial question arises, ‘Can Non Language Testers teach an Assessment Course?’ I think this is similar to when one looks for a doctor. When you are ill you have the choice of going to a specialist who is expensive and difficult to meet, or a close by family doctor who is accessible. This decision will probably depend on the severity of the illness. Usually people feel comfortable seeking a family doctor for a common cold but prefer to go to a specialist for a more serious sickness. So for Language Assessment, is this a course that can be taught only by testing professionals or can it be covered by non Language Testing professionals as well?

According to the survey, half of the instructors came from a non-testing background. They majored in a variety of different fields. In the interview, I focused on how these non LTs came to teach the assessment course and how they were trained to teach the course. The interview results show non-LTs (e.g., Olson) taught the course on a needs basis. For instance, health failure of the previous instructor required another faculty to teach the course. Others (e.g, Smith, Ryan) developed the course on their own because the curriculum did not have an assessment course, a few were asked to teach the course because there was nobody to teach it. Past experience or education in Language Testing was not a crucial factor in making these decisions and none of the instructors from the interview received additional training after being selected to teach the course. Olson expressed interest in learning more about language assessment if the funding and time was given.
During the interviews with LTs, I asked if they were concerned about the quality of assessment courses taught by non Language Testing professionals. All six Language Testers who were interviewed said that it would be great if assessment courses were taught by those who majored in the field, but they understood the lack of personnel and thought it was acceptable for a non-LT to teach the course. Goldberg commented, “Something is better than nothing. They [Non-LTs] will represent a different take on the knowledge that I would have. The kind of background they [instructors] had really made a difference [in teaching the course], [but] somebody has to do it. There is a demand. There is a need [for language assessment courses]. It would be better than nothing.” Goldberg thinks it is better for non-LTs to teach an assessment course than not to have a course at all, but she expects that the content of the course will be more related to classroom issues which are different from her course.

**Learning from each other.** Each instructor, whether a Language Tester or not, brings a unique feature to an assessment course. It could be life experiences from a rich teaching background, experience in other fields, or from a hard core testing background. What the instructors bring to the course is diverse and I believe there is something to learn from each other.

One of the biggest differences between LTs and non-LTs is the different level of technical knowledge they possess in teaching the course. Language Testers, who were trained and worked largely in the testing field, probably possess a stronger technical background compared to non Language Testers. So, to what extent is the degree of technical knowledge important? How can we define them? In the survey findings, Language Testers spent significantly more time in teaching Test Specifications and Basic Statistics showed compared to non Language Testers. One of the reasons for this result could be that non LTs may not have acquired the appropriate level of knowledge to comfortably teach the topics. In contrast to other topics, Statistics and Test Specifications is an area that requires more professional
knowledge than others. Thus, it is difficult for non-LTs to obtain this information without any professional training. The majority of instructors who have extensive teaching experience can have some degree of knowledge for topics closely related to the classroom (e.g., classroom assessment, rubric development, and alternative assessment) but for Test Specifications it is difficult to teach it without having any instruction in Language Assessment.

I believe this could be an area LTs can collaborate with non-LTs and share their expertise with them. Knowing how to develop Test Specifications is one of the most important areas to be a successful assessor. Student teachers should have experience and acquire the knowledge required to develop good Test Specs and Language Testers should help non-LTs who are less familiar with this topic to be better prepared to teach it.

The language assessment course is taught by Language Testers and non Language Testers. To develop a better course, it is important for both instructors to help and learn from each other. As stated above, LTs can begin by sharing their expertise in areas non LTs need more guidance in and for Language Testers, they can ask about non LTs classroom teaching experience and teaching skills. The level of content knowledge is a concern for non-LTs, but the effectiveness of teaching methods and lack of K-12 teaching experience can be a concern for Language Testers. For example, an instructor who has expert knowledge in the field does not necessarily mean he or she is a good teacher educator. In the interview, Jenn was very unhappy with the language assessment course she took despite the fact it was taught by a famous Language Tester. The instructor cared little about teaching and was not interested in helping the student teachers. To be a good teacher educator, instructors should have the passion in teaching along with the content knowledge.

**Characteristics of Successful Language Assessment Courses**

From my study, I have compiled the common characteristics of successful assessment courses. Courses that are thought to be successful incorporate the following features:
Rich examples. When a student teacher steps into a language assessment class, the topic that is often covered at the beginning is validity. It is usually located in the first few chapters of language assessment course books, and so most student teachers quickly learn that validity is important in testing. However, anyone who has taught or taken an assessment course quickly realizes how difficult this concept is to teach. Even veteran Language Testers who have been teaching their courses for almost 20 years comment that teaching validity is the most challenging area in language testing (Davidson, Personal Communication, 2009). I myself who have taken several language assessment courses as a student and am studying towards on becoming a professional in this field, yet struggle with this question when I’m asked by a teacher, “What is validity in language testing?”

From the study, the best way to teach grand, but important, concepts such as “validity” is to use direct examples. Ellen, who has changed her course from theory based to practice based, said she uses a real test example in teaching this concept. “Even when we talk about validity, I get them to evaluate it with a test from the government, and we look at them and talk about the issues—what issues are there. If I don’t do this, they will not get it.” Ellen, says, “They [student teachers] are mostly like to learn something if you provide them with concrete hands-on material. You cannot speak about it vaguely.” She has discovered that, in order for the teaching to be effective, the content of the course, whether it is theory, statistics, or other, must be presented through examples student teachers understand. If the instructors do this, student teachers are willing to learn even the most complicated topics.

Bringing rich authentic examples and materials in the class was a must for a successful assessment course. One reason, Anderson was teaching a very successful language assessment course was that she brought authentic materials and activities to the classroom. Every time she teaches the course, she has different activities. The activities are based on the assessment needs of the Language Program she is working in. If they need to develop reading tests, that will be the main activity for the course, if the program needs a new rater training
program for its teachers, that will be the core project for the course. By bringing a real life situation, into the course, student teachers will benefit enormously and will be interested in the activity.

Mark, who also experienced a successful class as a student teacher, said that he developed an assessment tool for a real set of students provided by the instructor. As a group, student teachers developed and piloted the assessment tool and reported back on the success of the tool in their final paper.

**Student Teacher course involvement: utilizing past experience.** Instructors who taught good language assessment courses also knew how to use student teachers’ past experiences. Student teachers, who take the course, especially for graduate-level students, come with a very rich teaching background. Some student teachers may have had many years of teaching experience before beginning the program. For the course to be helpful for these student teachers, it is important for instructors to build upon the experiences student teachers bring to the course. Often, instructors indicate the biggest challenge for the course is the diversity of the student teachers. Instructors usually are not in the position to control the diverse background of the students, but they can ask and learn from their student teachers’ backgrounds. Instructors who teach at the university level, even though they do have teaching experience, not many have primary or secondary teaching experience. Smith, who teaches an M.A. course in the UK, identified this as a limitation and thought it was important to know what kind of testing is going on in the student teachers’ teaching context. Smith emphasizes the importance of utilizing teacher experience: “Teachers have a lot of experience. Some of it, we couldn’t imagine ourselves; we have to use them as a resource—try to make sure the course is relevant to them.”

**Flexibility.** Student teachers appreciated instructors who gave flexibility to their students. More often than not, an assessment course will consist of a variety of students who come from different backgrounds and will be teaching at different settings. Even the most
homogeneous class, like Carol’s class which consisted of almost all K-12 teachers who were currently teaching or will teach upon graduation in the same state had teachers who taught different languages. Student teachers liked courses that gave options on final projects and presentation topics. They enjoyed doing activities relevant to their own teaching context.

**Practical Teaching Method.** Student teachers wanted more practice in teaching and disliked the conventional instructor-centered lecturing style of the course. Bell stated, “I think theory and test critique is useful, but I would like to have seen more of the practical application in my M.A. program in general. I was rather frustrated upon exiting the program.” A student teacher wanting more practical application is a universal desire. This is not only for Bell, but a longing from all teachers. Of course, there are those who are successful but there are student teachers who had a very bad experience like Jenn, “The instructor believes in principle but not in practice. I would have liked to get into alternative assessment, portfolio which was never brought up.”

**Importance of Pre/In-service Courses**

The quality of the courses student teachers take in pre-service, in-service courses directly impacts how competitive they are when they are in the job market. It is the instructor’s job to well prepare student teachers for the real teaching situation. Instructors should provide student teachers with strong theory in addition to practical knowledge that can be immediately applied to teaching. Bell states, “If they [student teachers] are prepared in both [theoretical and practical] ways, not only do they have the superior knowledge but they are ready for the practice.” Unfortunately, not all of the current language assessment courses are successfully doing this. Bell comments, “Most of the important things—the things that make me an expert with dealing with tests on a daily basis—are things that I have learned from experience.” Learning from experience is one thing; giving teachers the chance to learn it prior to teaching would be better. Bell expressed that it was sad that she never learned how to modify a content area test for language learners. “I could have been faced with modifying
a content area test. … I could have dealt with it and been ready for it before it even happened.” Many instructors are teaching great courses but there is always room for improvement for both Language Testers and non Language Testers.

Limitations of the Study

Survey Participants. The recruitment for the on-line survey was done through available list-serves in addition to personal invitations. The survey might have recruited participants who might not have necessarily been appropriate for the study. For example, there were instructors who were from Speech Pathology. They also have a language-related assessment course, but it was not the target course I was looking for. Although I clearly stated in my consent letter the survey only targets for instructors and student teachers who have either taught or taken a language assessment course, I cannot escape the possibility of having participants who did not fall under my criteria.

Difficulty in Reporting the Response Rate. The nature of the study made it difficult to determine the sample size. Participants recruited for the study were contacted using list-serves from multiple organizations which often has overlapping subscribers (i.e., there is a high possibility one person was affiliated with more than one professional organization). Therefore, I could only report the number of responses not the response rate.

Sampling. At the design stage of the study Student Teachers were not divided into two groups (K-12 vs. Adult). Therefore, recruiting a similar number of student teachers in each group was not addressed, even though that distinction surfaced as important in the analyses of the results. The survey participants for the teacher instructors were recruited through professional ESL/EFL teaching organizations. The professional organizations (e.g., TESOL) may have attracted more student teachers who teach adults. A lot more Adult student teachers (n=160) participated in the survey compared to K-12 (n=82) student teachers. Also, the time of the data collection was during the northern hemisphere summer break, so K-12 teachers may have been less accessible. For future studies, I recommend that
researchers collect surveys during the school semester, particularly for K–12 teachers. If the researcher is interested in collecting data from this group, I would encourage him or her to contact organizations that are targeted for K–12 teachers.

**Interview Participants.** Unlike the survey participants who come from a variety of different countries and nationalities, the interview participants are more from the local area. Especially all the student teachers reside in the United States and many of the instructors are from the Mid-West region. There was not a preference in location in terms of the interview recruiting process but it seems participants who resided geographically closer to the researcher showed more interest in the study. Therefore, the interview findings from the student teachers may be more relevant to the U.S. context.

**Trade offs for Using Mixed Methods.** Using a Mixed Methods approach brought strengths to the study. It opened doors to do a large scale survey and also an in-depth interview and analysis. Yet, this approach also presented tradeoffs. For example, due to the time constraint I was not able to perform a more advanced statistical analysis of the survey data (e.g., factor analysis). Also, even though there were many instructors and student teachers who showed interest in participating in the follow up interview I was only able to conduct 18 phone interviews.

**Summary**

This chapter focused on answering how effective the language assessment courses are for student teachers. Student teachers in general found the course effective and appreciated topics that were closely related to their everyday teaching practice (e.g., classroom assessment, rubric development, alternative assessment). They also enjoyed learning theoretical topics (e.g., test theory, statistics) if it was presented with practical features. For the language assessment course to be effective the course should be taught with rich authentic examples with a practical hand on teaching method. Student teachers also appreciated
instructors who valued their past teaching experience and offered flexibility in the assignments and activities.

**Dec. 10, 20xx**

**Faculty Search Committee Meeting**

Sagon: So it appears, student teachers appreciated courses that offered rich examples.

Heejeong: Yes, I believe this is one of the most important factors to make a language assessment course successful.

Horwitz: Then, we should select a faculty that knows the importance of using authentic examples and who could provide students with a variety of examples. I’m sure there is a Language Tester who has these qualities.

Lee: And it seems flexibility is another important aspect.

Heejeong: Having flexibility and using student teachers past experience is also a most.

Kent: Hm… in that case, maybe instead of deciding whether to hire a Language Tester or a non-Language Tester it could be better if we look for a candidate who can fulfill these characteristics. What do you think Mia?

Horwitz: Well, I do agree these meeting these requirements are more important than certain labels. Okay this could be a good start. Let’s try to find people who can offer a course fulfilling the proposed characteristics; authenticity, flexibility, practicality, and use of past experience.

Sagon: Well, I feel we have a much better idea of what questions to ask for the first phone interviews. I will review our minutes and come up with a draft of questions and circulate with other committee members. Please feel free to give feedback.

Lee: Thank you Heejeong for sharing your findings. It was truly helpful.
Chapter VI

Conclusion

In this final chapter, the summary of the main findings from previous chapters are first presented. Next, I present my own biases and misconceptions of language assessment courses I had prior to the study. Finally, I end the chapter with suggestions for future studies.

Summary of Main Findings

Characteristics of Introductory Language Assessment Courses. Introductory Language Assessment Courses have two main goals; to teach important concepts and theory in language assessment and to create and implement language assessment tools. Survey and interview findings revealed instructors first covered the theoretical concepts related to testing and based on this foundation moved on to more practical applications which is often in the form of a final project that asks student teachers to develop an assessment tool.

The overall structure and activities of the language assessment courses were similar regardless of who taught the course, however the amount of difficulty and depth incorporated with each course topic varied depending on the instructor. Same topics (e.g., test specifications, test theory, basic statistics, classroom assessment, rubric development, test accommodations) received significantly different attention depending whether the instructor was a Language Tester or not. Instructor background is a factor that determines the structure and focus of the course.

Evidence of an “effective” Language Assessment Course for Classroom Teachers. Both instructors and student teachers agreed “balance between theory and practice” was the most important factor for language assessment courses to be effective for classroom teachers. Specific topics (classroom assessment, alternative assessment, test specifications, rubric development, and alternative assessment) that were perceived to be important for student teachers were similar as well. Although the instructors were aware of the areas that were important for student teachers, they were not necessarily devoting their
time in teaching the proposed topics. For example, from the instructors, the topics that received the most time coverage were ‘Test Theory’ which was not a topic in the top 5 most important topics. There is a difference in what instructors think are important topics from what they actually teach which could be problematic.

For language assessment courses to be effective student teachers wanted the instructors to be more alert to their needs. Student teachers appreciated courses that valued their previous teaching experience and reflected it into the course. Assessment courses that were taught with rich authentic examples and practical hands on teaching method were noted as an “effective” course by student teachers as well.

**How effective are language assessment courses for language teachers’ everyday classroom practices?** The student teachers were overall satisfied with language assessment courses they took. The degree of satisfaction did not show much difference among student teachers’ teaching experience and teaching environment. The majority (79%, n = 187) of the student teachers replied the course met their needs and became to have more positive perceptions (71.6%, n=141) after taking the course. Student teachers also thought the textbooks used for the courses were useful for classroom teachers.

**Revisiting the Purpose**

My overall goal for doing this study was to give insight and suggestions on developing a better language assessment course for student teachers. Being a classroom teacher is a challenging job and I strongly believe it is the role of the teacher educator to fulfill the needs of the student teachers. Since there are such few studies that look into the student teachers thoughts about the assessment courses, I wanted to know how they felt about the courses and what could be done to improve them.

During the data collection process, my heart dropped when I heard horror stories from student teachers who had a terrible experience with language assessment courses. There were instructors who had little interest in teaching the course, who did not respond to
teacher’s interest, who were more interested in his or her own research than teaching. It broke my heart, when the student teacher said that instructor was one of the leading members of the language testing group. I felt bad they had to go through such a bad experience, especially when this was the only language assessment course they took. I was also concerned when I talked with instructors who clearly were teaching an inappropriate course for the student teachers. It was unfortunate the instructor did not seem to realize the mistakes they were making. There were instructors who were pouring technical knowledge to their teachers, who would have preferred to learn other things related to classroom assessment. However, there were many more moments when I was excited and happy to hear wonderful stories of the great courses the student teachers took and the fantastic courses the instructors were teaching. It was delightful to hear how student teachers loved the activities and their instructor. It was a joy to listen and feel the passion instructors had teaching the course. I was happy to hear there were many great courses out there which student teachers enjoyed taking and instructors loved teaching.

My Own Bias, Misconceptions of Language Assessment Courses

The literature I read and on which I reported a year ago focused heavily on the negative aspects of assessment courses, which somewhat corroborated my own personal experience. However, I am happy to say through this study, a lot of my biases proved not to be true. For example, I always thought student teachers did not like theoretical topics and statistics; yet, this wasn’t necessarily the case. Their acceptance to these topics depended heavily on how it was presented. If it was presented at the right level and in a practical way student teachers valued the importance of learning it. Another misconception was that I thought student teachers disliked the course and were unhappy to take it. But to my surprise, student teachers’ satisfaction level was relatively high. I am glad to find that a lot of my biases towards the courses proved to be wrong. I am happy to see the study could fill in the
gaps of previous literature. Overall I am glad to have had the chance to find answers to the
questions that were lingering in my mind since I began this study.

Suggestions for Future Research

Multiple Methods for Language Assessment Course Study. To overcome the
limitations of previous studies, this study expanded the scope and depth of research. It
covered a much larger audience (Language Testers, non Language Testers, K-12 student
teachers, adult student teachers) and applied a mixed methods (survey and interview)
approach. This method helped in defining the characteristics and effectiveness of the
language assessment courses but there are many areas that give suggestions for future studies.
For example, doing a course syllabi review of the assessment courses can project an objective
lens of what is covered in the assessment courses. Both surveys and interview rely heavily on
self report therefore doing a document review (e.g., course syllabi, course material) will add
more validity to the study. Also, a textbook review of the books used for the courses can
provide insight to what is taught in assessment courses. Language testers and non LTs may
have different preferences for textbooks and student teachers may as well. Finally, classroom
observations will also be an effective way to discover new information of the courses.

Definition of the Language Tester. The definition of who is a ‘Language Tester’
brought confusion to interview participants. Instructors had different understandings of who
is a Language Tester. It would be interesting to find out what characteristics defines a
language tester and how it is different depending on the individual. Also, it would be helpful
to know how non Language Testers describe a Language Tester.

Research on Different Forms of Language Assessment Courses and Comparison
with General Assessment Courses. The focus of this study was only on introductory
language assessment courses. Future studies could examine different types of assessment
course such as advanced language assessment courses or on-line course. There has not been a
study that investigated advanced language assessment courses. It would be meaningful to do
research on how these courses are taught and how it is different from introductory language assessment courses. Also, there will be a steady increase in on-line assessment courses. Additional research could look into how group activities such as test specifications are done in an online format. Finally, a comparison study between general assessment courses and language assessment courses could show new directions on how language assessment courses can be taught.

**Revision of Survey Protocol.** To help future researchers, I have developed a revised survey protocol (Appendix F) that reflects the survey results and interview findings. As a researcher, I did my best to develop a survey questionnaire that answered the proposed research questions. Although the questions were carefully developed on previous studies, pilot studies, and literature review, after implementing the survey and looking at the results there were areas that I felt could have been changed, included, or deleted. Every changed item is noted in the revised survey protocol and a commentary box is included to describe the reasons for the change in Appendix F. I hope this could be a help and guidance for future researchers who wish to further research in this area.

**Future of Language Assessment Courses**

**Importance of Teacher Education Courses.** A lot of the findings and suggestions derived from the study relates to teacher educational courses in general. For instructors, teaching a pre-service/in-service course could be among the many duties they do as a scholar, but for student teachers each of these courses whether it is an assessment course, methods course, or practicum are probably the only course they take in that content area before they go out and teach. It is high stakes for the student teachers and they want to be well prepared before facing the challenge of teaching.

For student teachers to be confident about what they do it is important to train them well. The knowledge the student teachers acquire, the practice student teachers go through directly impacts how they teach and assess students. If you consider the influence a single
teacher can have on a student’s life, the role of a teacher trainer is very heavy. Instructors must do a great job. If student teachers aren’t prepared to teach and make constant mistakes on the job, students are the ones who are to suffer.

**Wishes and Dreams**

“I wish there was another professor who comes with a different perspective who is also an assessment specialist.” Ellen “I wish we could have another course in alternative assessment” Ellen

Every teacher, every instructor has hopes and dreams in their field. Instructors wish for more time to teach, more courses to teach, more colleagues to work with. Student teachers desire more practical, real, authentic material. Behind this lies the grand purpose of doing all of this; to develop a better course for instructors and student teachers. I believe the future of Language Assessment Courses is bright because there are many out there who have the passion and interest in further developing this area. Instructors strive to develop better courses for their student teachers, and student teacher work hard to become better assessors in the field.

When I first began the study, I was concerned I will not have enough participants who will be interested in the topic. However, once I began collecting data, more people than I expected showed interest and passion in this topic. It is unfortunate I couldn’t interview all the instructors and student teachers who volunteered to participate in the survey. From their eagerness, the field of language assessment is bright. There are hundreds and thousands of people in the world, who think and practice assessment daily, who work hard to make better tests, who strive to become better teacher educators. It is my limit, I couldn’t share all their stories, but I believe this is only the beginning. The language assessment courses will develop, more and more research will be done in this area, and more textbooks are currently being published and will be published that will benefit the student teachers and eventually benefit the students these teachers teach.
References


Appendix A

Pilot Interview Questions/Follow-Up Interview Questions

Instructor Questions

Demographics

1. What degree do you hold in which area?
2. What is your primary research interest area?
3. Do you have any ESL or language teaching experience? Which grade level? How many years?
4. Do you hold a Teaching Certificate (any grade level, any country or school district, any subject area)?
   (1) Yes (2) No
5. What is your academic position?
6. [Training] How were you trained to teach the assessment course? Could you describe any assessment related activities you have done in the past?
7. What types of student(s) take the testing course? Who is your target audience?
8. Is the course a requirement or optional?

Language Assessment Course

1. [Course Goal] What is your course goal? What do you see as the primary focus of the course you teach?
2. [Core Topics] Of the topics covered in a language assessment course, which topic do you think will be most helpful for classroom teachers?
3. [Unique Features] How is this course different from other courses (e.g., methods, seminars)? What are the unique features of the language testing course?
4. [Uncovered Areas] What are the topics you would want to cover but cannot (due to
practical constraints)?

5. [Course Position] What difficulties do you have with curriculum design of the course? (e.g., Even though you feel there is a strong need, is it difficult to open an advanced course in assessment?)

6. [Challenges] What are areas you have the most difficulty teaching? What topics do your students have the hardest time understanding? How do you overcome this challenge?

7. [Training] Do you believe it is important for teacher educators to have field experience in order to teach the course?

8. [Assessment] How do you measure/assess (e.g., papers, tests, presentation, test development) students’ achievement of the course objectives?

9. [Changes and Development] How has your course changed or developed over the years? What changes have occurred in the content compared to the first time you taught the course? What would you like to do differently if you were to teach the course again?

10. [Assessment Proficiency] When students complete your course, how proficient are they in interpreting, evaluating, and developing language tests compared to when they began the course? (before and after the course)

11. In general, do you think your students find the your course;  
    (1) theoretical (2) practical (3)easy (4) useful (5) difficult

12. Is there anything else you want to know about regarding language assessment courses?
Student Teacher Questions

Demographic Information

1. When did you take this course?
2. What degree do you hold? (Bachelor’s Degree, Master’s Degree, Teaching Certificate, Doctorate in progress, Doctorate)
3. Did you take any other language assessment courses other than the course offered by xx?
4. Including this year, how many years have you taught ESL?
5. What kind of assessments have conducted in the past or what kind of assessment related activities are you currently conducting?

Language Assessment Course

1. [Most Beneficial] Of the topics covered in a language assessment course, which topic was most helpful in your teaching? In what ways was it helpful in your classroom teaching?
2. [Challenges] What areas were the most difficult to understand?
3. [Course Difficulty] Compared to other courses, how was this course different?
4. [Resources] Have you used/consulted books related to language assessment? Did you find those books useful? What do you do if you have a question regarding assessment or testing? How do you find your answers? Do you ask fellow teachers? Search information on the web?
5. [Uncovered Areas] If you were to take this course again or take an advanced course in language testing, what would you like to learn that wasn’t covered (or wasn’t covered thoroughly)?
6. [Change in Perception] Did your perception of testing change throughout the course? For example, unfortunately many students who come to this course come
with a strong negative image of testing. Did this change after finishing the course?

7. [Assessment Literacy] After completing the course, did you feel you became more proficient in interpreting, evaluating, and developing language tests compared to when you began the course?

8. [Different Vision] What is the different perspective of language assessment courses between students and instructors? Do you feel the course met your needs as a classroom teacher?

9. [Effectiveness] What do you think are the requirements for an effective language assessment course? What kind of language assessment course is helpful for in-service or pre-service teachers?

10. In general, did you find the course helpful/practical as a teacher?

11. Is there anything else you would like to comment or ask regarding language assessment courses?
## Appendix B

List of Commonly Covered Topics in Language Assessment Courses

<table>
<thead>
<tr>
<th>Item writing, Test administration, Test scoring, Standardized test score interpretation</th>
<th>Test critiquing, Test analysis, Test-taking skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test theory (reliability/validity), Test Specifications (writing), Basic Statistics</td>
<td>Test ethics, History of language testing, Classroom assessment</td>
</tr>
<tr>
<td>Alternative/Performance Assessment, Developing Rubrics</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix C

#### Description of Pilot Study Participants

<table>
<thead>
<tr>
<th>ESL/EFL Teachers (10)</th>
<th>Instructor</th>
<th>Date of Course</th>
<th>Years Ago</th>
<th>Grad. Student</th>
<th>Native/Non-native</th>
<th>Sex</th>
<th>Teaching Experience (yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justin</td>
<td>F</td>
<td>Fall 2001</td>
<td>8</td>
<td>Y</td>
<td>NS</td>
<td>M</td>
<td>12.0</td>
</tr>
<tr>
<td>Brian</td>
<td>F</td>
<td>Fall 2005</td>
<td>4</td>
<td>N</td>
<td>NNS</td>
<td>M</td>
<td>7.0</td>
</tr>
<tr>
<td>Sumin</td>
<td>F/M</td>
<td>Fall/Spring 2007 /Spring 2009</td>
<td>2/ Within 1 year</td>
<td>Y</td>
<td>NNS</td>
<td>F</td>
<td>3.0</td>
</tr>
<tr>
<td>Eman</td>
<td>F</td>
<td>Fall 2008</td>
<td>1</td>
<td>Y</td>
<td>NNS</td>
<td>F</td>
<td>3.0</td>
</tr>
<tr>
<td>Kathy</td>
<td>F</td>
<td>Spring 2009</td>
<td>Within 1 year</td>
<td>Y</td>
<td>NNS</td>
<td>F</td>
<td>.5</td>
</tr>
<tr>
<td>Derrick</td>
<td>F</td>
<td>Spring 2009</td>
<td>Within 1 year</td>
<td>Y</td>
<td>NS</td>
<td>M</td>
<td>3.0</td>
</tr>
<tr>
<td>Debra</td>
<td>F</td>
<td>Spring 2009</td>
<td>Within 1 year</td>
<td>Y</td>
<td>NNS</td>
<td>F</td>
<td>2.5</td>
</tr>
<tr>
<td>Jungmin</td>
<td>M</td>
<td>Spring 2009</td>
<td>Within 1 year</td>
<td>Y</td>
<td>NNS</td>
<td>F</td>
<td>1.0</td>
</tr>
<tr>
<td>Terry</td>
<td>C</td>
<td>Spring 2008</td>
<td>1.5</td>
<td>N</td>
<td>NS</td>
<td>F</td>
<td>7.0</td>
</tr>
<tr>
<td>Phil</td>
<td>L</td>
<td>Winter 2008</td>
<td>1</td>
<td>N</td>
<td>NNS</td>
<td>M</td>
<td>7.0</td>
</tr>
</tbody>
</table>
## Instructors

<table>
<thead>
<tr>
<th>Instructors (4)</th>
<th>LT/non LT</th>
<th>Main Target Audience</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fowler (F)</td>
<td>LT</td>
<td>MA TESOL</td>
<td>Grad/UG</td>
</tr>
<tr>
<td>Clarkson (C)</td>
<td>Non LT</td>
<td>MA TESOL</td>
<td>Grad/UG</td>
</tr>
<tr>
<td>Morgan (M)</td>
<td>Non LT</td>
<td>Bilingual Ed. MA</td>
<td>Grad</td>
</tr>
<tr>
<td>Lee (L)</td>
<td>LT</td>
<td>In-service secondary English teachers in Korea</td>
<td>Grad</td>
</tr>
</tbody>
</table>
Appendix D

Description of Pilot Study Courses

Fowler’s course

Fowler was the core person to teach the Language Assessment course at a large Midwest university. This was a requirement for the students in the MA TESOL (Teaching English to Speakers of Other Languages) program. Fowler held a doctorate in applied linguistics, and his primary research focus was language testing. He had around 10 years of teaching ESL at the college level and also taught 2 years of EFL before pursuing a professional degree. Fowler teaches the language assessment course every semester in this university. His course primarily consists of MA TESOL students (65%–70%) and also occasionally doctoral students who are interested in language testing, visiting scholars, and a few undergraduate students who take the course to fulfill a teaching endorsement.

Clarkson’s course

Clarkson’s primary research interest is in Second Language Acquisition. She taught the language assessment course once when Fowler was on sabbatical. Since it was originally Fowler’s course, there are many features that resemble his course (e.g., Test Specifications Activity, Course Assignments). Yet the course was also different since Clarkson’s educational background was not specifically in language assessment. She had experience working for a testing company during her doctoral studies, but was not trained as a language tester per se. Similar to Fowler she felt the most beneficial topic to the teachers of her class was the test specifications workshop and the item writing activity. Her purpose for the course was to give a good introduction regarding item writing and deliver the concept of testing theory in both large- and small-scale assessment. Terry, the teacher who took Clarkson’s course, also confirmed that the test specifications workshop was the most helpful for her classroom teaching.
Morgan’s Course

The main target audience for Morgan’s course was teachers returning for their ESL/bilingual endorsement. She occasionally has Master’s and Ph.D. students from Curriculum and Instruction and other programs. Morgan has extensive experience in ESL and Spanish teaching, and her main research interest area is in identifying socio-cultural barriers in developing second language literacy. She has experience in training teachers to administer standardized tests and was part of an expert panel for the development of a Spanish test. Most of her testing experience comes from teaching and work experience. Morgan’s course was very different from those of the other three instructors. First, since the course focused on ESL/Bilingual teachers at the K–12 level, primarily elementary, the course content covered issues that concerned students at the age level. Teachers critiqued tests that were currently administered at schools rather than widely known language tests such as the TOEFL or IELTS. Reading materials contained topics that were of interest to current school teachers (e.g., equity issues/tests accommodations, educational policy, No Child Left Behind Act, content area assessment), especially that discussed concerns with English Language Learners (ELLs) whose first language was Spanish. These topics were not covered in the other language assessment courses. Second, as a requirement, teachers in groups developed a classroom test for an in-service teacher who was taking the course. Jungmin commented that this group project was the most beneficial part of the course. She enjoyed developing a writing test for a real classroom context. One of the teachers in her group brought a lesson plan for his class, and the group members developed a rubric that was actually used to grade student papers. This was different from Fowler and Clarkson’s test specifications workshop. Teachers in that course developed tests but did not pilot or test them in a classroom context. Another unique feature of Morgan’s course was that it did not cover statistics. All of the other three instructors taught statistics to some extent in their courses.
Teachers who took her course enjoyed the hands-on activity of developing tests and rubrics that were implemented in the classroom context. Although Jungmin enjoyed group activities, she did express difficulties working with current teachers. Jungmin, who is a doctoral student and not a current K–12 teacher, felt disconnected when the course covered issues directly related to the school context. She was not familiar with many of the tests that were talked about and the current school policy. She commented that not having a sense of real classroom issues made her feel distant from the other school teachers. Sumin, who took both Fowler’s course and Morgan’s course, said the two classes were totally different. It was interesting to see how different an assessment course could be depending on the instructor and the curriculum it was based on, something which was overlooked in the Bailey and Brown studies cited in Chapter Two.

**Lee’s course.**

Lee’s course was an intensive, graduate-level, 2-week course which was offered during the winter break. The course was taught three hours each day (Monday through Thursday) and was specifically targeted for in-service secondary English teachers in South Korea. Lee’s primary research interest is Language Testing, and she holds a doctoral degree in this area. The course was a requirement for teachers who were pursuing a Master’s degree. In the interview, Lee explained that the goal of the introductory language assessment course was to give an overview of different aspects of language testing: introduction to statistics and measurement concepts, different types of standardized tests, how to test the 4 skills in language, etc. In comparison to the other three instructors, Lee felt very strongly about teaching statistics in the course. She wanted to teach teachers how to use statistical software (e.g., SPSS, SAS) but felt that, unfortunately, she did not have the time to cover it in her course. Lee believed it was important for teachers to conduct research and having statistical skills was the key to reading articles and implementing their own studies. The
biggest challenge she had in teaching the course was delivering the concepts of Item Response Theory (IRT). She felt teachers need to understand the models to interpret standardized test scores.

Phil, who is a current middle school teacher in Korea, took Lee’s class and commented that the statistics part was the most challenging part of the course. Although he felt that having statistical knowledge is essential in learning language testing, he thought it would be more effective if it was taught as a separate course. The part of the course he enjoyed the most was developing the final project, which was writing test items and critiquing them. Areas he wanted to learn more about were performance-assessment-related issues. Phil commented that developing valid and reliable performance assessment items are always a challenge and concern for school teachers and thought it would be helpful if this topic were covered in the course.
Appendix E
Original Survey Instrument

Dear Teacher/Professor,

You are invited to participate in a research project concerning the nature of language assessment courses. The purpose of the study is to provide information to teacher educators in language assessment to reflect the factors that most influence teachers in developing assessment skills. This study is conducted by Heejeong Jeong and Professor Fred Davidson at the University of Illinois at Urbana Champaign.

This study will take approximately 20 minutes of your time. You will be asked to complete an online survey about characteristics and effectiveness of Language Assessment Courses.

Your decision to participate or decline participation in this study is completely voluntary and you have the right to terminate your participation at any time without penalty. You may skip any questions you do not wish to answer. If you do not wish to complete this survey just close your browser.

Your participation in this research will be completely confidential and data will be averaged and reported in aggregate. Possible outlets of dissemination may be conference papers, journal articles and a doctorate dissertation. Although your participation in this research may not benefit you personally, it will help instructors to develop a more effective course that will meet the needs of classroom teachers.

No known physical risks are expected as a result of the study, although there may be some minimal risks concerning the content of the survey. There are no risks to individuals participating in this survey beyond those that exist in daily life.

If you have questions about this project, you may contact Heejeong Jeong (hjeong5@illinois.edu) or Dr. Fred Davidson (fgd@illinois.edu). If you have any questions about your rights as a research participant in the study, please contact the University of Illinois Institutional Review Board at 217-
333-3670 (collect calls accepted if you identify yourself as a research participant) or via email at irb@illinois.edu.

Please print a copy of this consent form for your records, if you so desire.

I have read and understand the above consent form, I certify that I am 18 years old or older and, by clicking the submit button to enter the survey, I indicate my willingness voluntarily take part in the study.

1. Would you like to take the survey?
   ( ) Yes, I agree to participate in the study.
   ( ) No, I am not interested.

---

Teacher Survey

2. When did you take the introductory language assessment course?
   ( ) within the last year
   ( ) 2~3 years ago
   ( ) 4~5 years ago
   ( ) more than 5 years ago

3. Was the language assessment course an elective or a requirement?
   ( ) elective
   ( ) requirement

4. Would you have taken this course if it was NOT a requirement?
   ( ) Yes
   ( ) No

5. Why would you have not taken it?
   ( ) Lack of interest in language testing.
   ( ) Not necessary for a classroom teacher.
( ) Lack of time.
( ) Other, please specify: ____________________________________________________

6. What is the highest level of educational degree you have earned?
   ( ) B.A. in progress
   ( ) B.A.
   ( ) M.A./M.Ed. in progress
   ( ) M.A./M.Ed.
   ( ) Ph.D/Ed.D in progress
   ( ) Ph.D/Ed.D

7. Do you hold a teaching certificate? (any grade, any country or school district, any subject area, including emergency/temporary certification)
   ( ) Yes
   ( ) No

8. How many language assessment courses have you completed?
   ( ) 1
   ( ) 2
   ( ) 3
   ( ) more than 3

9. How many language assessment related workshops have you attended?
   ( ) 1
   ( ) 2
   ( ) 3
   ( ) more than 3

10. Which grade level do/did you most often teach?
    ( ) Pre-elementary/Kindergarten
    ( ) Elementary
    ( ) Secondary (Middle/Highschool)
    ( ) College (Undergrad./Grad.)
    ( ) Adult (Non-Degree)

11. Including this year, how many years have you taught ESL/EFL or another foreign language?
    ( ) less than a year
    ( ) 1~2 years
( ) 3–4 years
( ) 5–9 years
( ) more than 10 years

12. What kind of classroom assessments have you developed or used? Check all that apply.
   ( ) Multiple Choice
   ( ) Short Answer Quizzes
   ( ) Essay Questions
   ( ) Performance Assessments
   ( ) Portfolio Assessment
   ( ) Other

13. Of the following topics covered in the language assessment course, how much time was spent on each topic when you took the course?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extensive Time</th>
<th>Some Time</th>
<th>A little Time</th>
<th>Hardly Any Time</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Test ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. History of language testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Classroom assessment (developing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Alternative/performance assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
14. Of the topics covered in Question 13, which topic was MOST helpful in your classroom teaching?

____________________________________________

15. Why was that topic particularly helpful? Check all that apply.
   ( ) Applied directly to classroom assessment
   ( ) Concepts were easy to understand
   ( ) Personal interest in the topic
   ( ) Helped developing classroom materials
   ( ) Other

16. Of the topics covered in Question 13, which topic was the LEAST helpful in your teaching?

____________________________________________

17. Why was that topic not very helpful in your teaching?
   ( ) Concept was vague
   ( ) Instructor did not provide sufficient explanation
   ( ) Not practical to teaching
   ( ) Lack of my background knowledge on the topic
   ( ) Other

18. How important are the following topics to you as a classroom teacher?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extremely Important</th>
<th>Very Important</th>
<th>A Little Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
19. Compared to other courses in your program, what was the difficulty level of the course?
   ( ) a lot more difficult
   ( ) a little more difficult
   ( ) about the same level
   ( ) a little easier
   ( ) a lot easier

20. What kind of expectations did you have before taking the language assessment course?

<table>
<thead>
<tr>
<th>I thought,</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would be developing classroom tests.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will learn how to make rubrics for classroom tests and assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

174
| I will learn about test theory (e.g., validity). |      |      |      |
| I will learn basic statistical skills. |      |      |      |
| I will learn how to critique previously developed test items. |      |      |      |
| I will learn the history of language testing. |      |      |      |
| I will learn about performance assessment techniques. |      |      |      |
| I will learn how to become an accurate rater. |      |      |      |
| I will learn advanced statistical skills (e.g., Item Response Theory). |      |      |      |
| I will learn how to accommodate tests for special needs students. |      |      |      |
| I will learn test ethics. |      |      |      |
| Other, Please specify  |      |      |      |

21. Did you instructor ask you what you wanted to learn in the course at the beginning of the semester?
   (   ) Yes.
   (   ) No.

22. Did your perception of assessment change through the course?
   (   ) Yes, it did change.
   (   ) No, there were not any changes.

23. If there were changes, how did it change?
   (   ) I have a more positive perception of language testing after the course.
   (   ) I have a more negative perception of language testing after the course.
   (   ) I became more confused about language testing after the course.
   (   ) Other, please specify.______________________________________

24. After completing the course, how did you feel a bout the following compared to when you began the course?

<table>
<thead>
<tr>
<th></th>
<th>more proficient</th>
<th>no effect</th>
<th>less proficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreting Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

175
25. Do you agree or disagree with the following statement?

*The language assessment course met my needs as a classroom teacher.*

( ) Strongly disagree  
( ) Disagree  
( ) Agree  
( ) Strongly agree

26. What is the most important factor to make a language assessment course effective for classroom teachers?

( ) practicality  
( ) theoretical knowledge  
( ) balance between theory and practice  
( ) Other, please specify. ______________________________________________

27. Overall, was the course helpful for you as a classroom teacher?

( ) not at all  
( ) a little  
( ) somewhat helpful  
( ) very helpful

28. In general, you found the language assessment course (check all that apply):

( ) interesting  
( ) too theoretical  
( ) easy  
( ) useful  
( ) too practical  
( ) difficult  
( ) a nice balance between theory and practice

29. If you were to take this course again, or take an advanced course in language assessment, what would you like to learn more about?
30. Why do you want to study that topic?

31. Are there any other comments or questions you have regarding language assessment courses?
Instructor Survey

2. What is your gender?
( ) Female
( ) Male

3. How many years have passed since you have received your final degree (e.g., Ph.D, Ed.D or other).
( ) Less than 5 years
( ) 6~10 years
( ) 11~20 years
( ) more than 20 years

4. Which area did you receive your final degree in?
( ) Applied Linguistics
( ) Linguistics
( ) Curriculum and Instruction
( ) Educational Psychology
( ) TESOL
( ) Other ___________________

5. Is Language Testing your primary research area?
( ) Yes
( ) No

6. If Language Testing is not your primary research area, what kind of assessment related activities have you participated in? (Check all that apply.)
( ) Developed Standardized Tests
( ) Worked as a Rater
( ) Worked with Classroom Teachers on Testing
( ) Other

7. If Language Testing is not your main research area, what is your primary research area?

________________________________________________________________________

8. What is the title of the language assessment course you teach?
9. How many times have you taught this introductory language assessment course?

( ) less than 3  ( ) 4 ~ 8  ( ) 8 ~ 12  ( ) more than 12 times

10. Who is the main target audience(over 50%) of your course?
( ) Regular undergrad.
( ) Pre-service undergrad.
( ) Pre-service M.A.
( ) In-service M.A.
( ) Regular M.A. students (non Teachers)
( ) Ph.D Students
( ) Other ________________

11. What is(was) your academic position when you are(were) teaching this course?
( ) Lecturer(non-tenure track)
( ) Adjunct Professor (non-tenure track)
( ) Visiting Professor
( ) Assistant Professor (tenure track)
( ) Associate Professor
( ) Full Professor
( ) Other ____________________

12. Do you hold a teaching certificate? (any grade, any country or school district, any subject area, including emergency/temporary certification)
( ) Yes
( ) No

13. Do you have any language teaching experience?
( ) Yes
( ) No

14. If you have language teaching experience, including this year, how many years have you taught ESL/EFL or another foreign language?
( ) less than a year
( ) 1~2 years
15. If you have language teaching experience, which grade level do/did you most often teach?

( ) Pre-elementary/Kindergarten
( ) Elementary
( ) Secondary (Middle/Highschool)
( ) College (Undergrad./Grad.)
( ) Adult (Non-Degree)

16. Do you ask your students what they want to learn in the course at the beginning of the semester?

( ) Yes.
( ) No.

17. Of the following topics covered in the language assessment course, how much time did you spend teaching each topic?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extensive Time</th>
<th>Some Time</th>
<th>A little Time</th>
<th>Hardly Any Time</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Test ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. History of language testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. Classroom assessment  
   (developing )

13. Alternative/performance assessment

14. Rubric development (analytic and holistic)

15. Rater training

16. Test accommodation

17. Other, Please specify
   _______________________

18. Of the topics covered in the above question, which topic do you think would be MOST helpful to classroom teachers?

19. Of the topics covered in the above question, which topic do you have the MOST difficulty in teaching?

20. How important do you think the following topics are to pre-service and in-service teachers?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extremely Important</th>
<th>Very Important</th>
<th>A Little Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Basic statistics (e.g., mean, percentile, bell curve)
9. Advanced statistics (e.g., Item Response Theory)
10. Test ethics
11. History of language testing
12. Classroom assessment (developing)
13. Alternative/performance assessment
14. Rubric development (analytic and holistic)
15. Rater training
16. Test accommodation

21. Do you believe it is important for instructors to have field experience in order to teach the language assessment course for pre-service/in-service teachers?
   (   ) Yes
   (   ) No

22. After completing the course, how do your students feel about the following areas?
   | more proficient | no effect | less proficient |
---|---|---|---|
Interpreting Language Tests
Evaluating Language Tests
Developing Language Tests

23. Do you agree or disagree with the following statement?
   My language assessment course met the needs of my pre-service/in-service teachers.
   (   ) Strongly disagree
   (   ) Disagree
   (   ) Agree
   (   ) Strongly agree

24. What is the MOST important factor to make a language assessment course effective for classroom teachers?
   (   ) practicality
   (   ) theoretical knowledge
   (   ) balance between theory and practice
25. Do you agree or disagree with the following statement regarding statistics?

<table>
<thead>
<tr>
<th>Introductory language assessment course should cover descriptive statistics.</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introductory language assessment course should cover interpretive statistics.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introductory language assessment course should cover measurement theory (e.g., Item Response Theory, Generalizability Theory).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

26. Overall, do you think the course was helpful for your pre-service/in-service teachers?

- ( ) not at all
- ( ) a little
- ( ) somewhat helpful
- ( ) very helpful

27. In general, do you think your pre-service/in-service teachers find your course (check all that apply):

- ( ) interesting
- ( ) too theoretical
- ( ) easy
- ( ) useful
- ( ) too practical
- ( ) difficult
- ( ) a nice balance between theory and practice

28. If you were to teach this course again, or teach an advanced course in language assessment, what topics would you like to cover?

29. Do you have any other comments or questions you have regarding language assessment courses?
Appendix F
Revised Survey Instrument

Note: This is revised survey instrument reflecting on the findings of the survey and interview. Items that are changed are deleted are noted using Track Changes function; items that are added are presented in a box. All revised items are followed by a commentary box that describes the reasons for the changes.

Teacher Survey

2. When did you take the introductory language assessment course?

-> ( ) year(s) ago

2. Change response form: Specific numbers are needed for accurate statistical analysis

3. Was the language assessment course an elective or a requirement?
   ( ) elective
   ( ) requirement

4. Would you have taken this course if it was NOT a requirement?
   ( ) Yes
   ( ) No

5. Why would you have not taken it?
   ( ) Lack of interest in language testing.
   ( ) Not necessary for a classroom teacher.
   ( ) Lack of time.
   ( ) Other, please specify: ________________________________________

6. What is the highest level of educational degree you have earned?
   ( ) B.A. in progress
   ( ) B.A.
   ( ) M.A./M.Ed. in progress
   ( ) M.A./M.Ed.
   ( ) Ph.D.Ed.D in progress
   ( ) Ph.D/Ed.D

7. Do you hold a teaching certificate? (any grade, any country or school district, any subject area,
including emergency/temporary certification)

( ) Yes
( ) No

**Additional Question:**
*If you hold a teaching certificate which subject is it in?*
( ) ESL/EFL
( ) English
( ) Other Foreign Language
( ) Other _________________________

: *There could be differences depending on the subject area.*

8. How many language assessment courses have you completed?
   
   -→ (     ) courses

**8. Change response form: Specific numbers are needed for accurate statistical analysis**

9. **Delete item: The survey was concerning language assessment courses not workshops**

10. Which grade level do/did you most often teach?
    ( ) Pre-elementary/Kindergarten
    ( ) Elementary
    ( ) Secondary (Middle/Highschool)
    ( ) College (Undergrad./Grad.)
    ( ) Adult (Non-Degree)

**Additional Question:**
*In which country did you take the course?*
(     )

: *Since the survey was distributed internationally it would be helpful to know the location of the course. There could be differences in the characteristics of the course depending on the location.*

**Additional Question:**
*What language was the course taught in?*

English
Other, specify ________________

: *From the survey and interview, I found language assessment courses are taught in languages other than English. Even though, the majority of the courses are currently taught in English, it would be helpful to know what other languages offer language assessment courses*

11. Including this year, how many years have you taught ESL/EFL or another foreign language?
11. Change response form: Specific numbers are needed for accurate statistical analysis

12. What kind of classroom assessments have you developed or used? Check all that apply.
   ( ) Multiple Choice
   ( ) Short Answer Quizzes
   ( ) Essay Questions
   ( ) Performance Assessments
   ( ) Portfolio Assessment
   ( ) Other

13. Of the following topics covered in the language assessment course, how much time was spent on each topic when you took the course?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extensive</th>
<th>Some Time</th>
<th>A little Time</th>
<th>Hardly Any Time</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Test ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. History of language testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 12. Classroom assessment  
   (developing) |
| 13. Alternative/Performance assessment |
| 14. Rubric development (analytic and holistic) |
| 15. Rater training |
| 16. Test accommodation |
| 17. Technology in Testing |

**17. Change choice description:** Hardly any time, and n/a was really confusing and difficult to interpret statistically so I suggest to change n/a to no time.

*Add additional topic:* From the open ended survey comments participants suggested a question concerning technology should be included.

14. Of the topics covered in Question 13, which topic was MOST helpful in your classroom teaching?

```
____________________________________________
```

**14. Change response format:** Instead of an open-ended question, change it to a drop-down option.

Many participants gave more than one answer to this question which made it difficult to analyze.

15. Why was that topic particularly helpful? Check all that apply.

- [ ] Applied directly to classroom assessment
- [ ] Concepts were easy to understand
- [ ] Personal interest in the topic
- [ ] Helped developing classroom materials
- [ ] Other

16. Of the topics covered in Question 13, which topic was the LEAST helpful in your teaching?

```
____________________________________________
```

**16. Change response format:** Instead of an open-ended question, change it to a drop-down option.

Many participants gave more than one answer to this question which made it difficult to analyze.

17. Why was that topic not very helpful in your teaching?

- [ ] Concept was vague
- [ ] Instructor did not provide sufficient explanation

187
( ) Not practical to teaching
( ) Lack of my background knowledge on the topic
( ) Other

18. How important are the following topics to you as a classroom teacher?

<table>
<thead>
<tr>
<th></th>
<th>Extremely Important</th>
<th>Very Important</th>
<th>A Little Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Test administration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Test critiquing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Test ethics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>History of language testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Classroom assessment (developing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Alternative/performance assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Rubric development (analytic and holistic)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Rater training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Test accommodation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Technology in Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18. Add additional topic: From the open ended survey comments participants suggested a question concerning technology should be included.

19. Compared to other courses in your program, what was the difficulty level of the course?
20. Delete Item: The results and contents of this item were similar to Question 18.

21. Did your instructor ask you what you wanted to learn in the course at the beginning of the semester?
   ( ) Yes.
   ( ) No.

22. Did your perception of assessment change through the course?
   ( ) Yes, it did change.
   ( ) No, there were not any changes.

23. If there were changes, how did it change?
   ( ) I have a more positive perception of language testing after the course.
   ( ) I have a more negative perception of language testing after the course.
   ( ) I became more confused about language testing after the course.
   ( ) Other, please specify.____________________________________

24. After completing the course, how did you feel about the following compared to when you began the course?

<table>
<thead>
<tr>
<th></th>
<th>more proficient</th>
<th>no effect</th>
<th>less proficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreting Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developing Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

25. Do you agree or disagree with the following statement?

The language assessment course met my needs as a classroom teacher.

( ) Strongly disagree
( ) Disagree
( ) Agree
( ) Strongly agree
26. What is the most important factor to make a language assessment course effective for classroom teachers?
   ( ) practicality
   ( ) theoretical knowledge
   ( ) balance between theory and practice
   ( ) Other, please specify. ______________________________________________

27. Overall, was the course helpful for you as a classroom teacher?
   ( ) not at all
   ( ) a little
   ( ) somewhat helpful
   ( ) very helpful

28. In general, you found the language assessment course (check all that apply):
   ( ) interesting
   ( ) too theoretical
   ( ) easy
   ( ) useful
   ( ) too practical
   ( ) difficult
   ( ) a nice balance between theory and practice

29. If you were to take this course again, or take an advanced course in language assessment, what would you like to learn more about?

________________________________________________________________________

30. Why do you want to study that topic?

________________________________________________________________________

31. Are there any other comments or questions you have regarding language assessment courses?

________________________________________________________________________
Instructor Survey

2. What is your gender?
   ( ) Female
   ( ) Male

**Additional Question:**

**What is your final degree?**

   ( ) M.A./M.Ed. in progress
   ( ) M.A./M.Ed.
   ( ) Ph.D/Ed.D in progress
   ( ) Ph.D/Ed.D
   ( ) Other___________

: I began with the assumption instructors who taught the course would have a doctorate degree. From the interview, I noticed there were instructors who did not have one. It would be helpful to ask the instructors what their final degree was. This question was asked to the student teachers.

3. How many years have passed since you have received your final degree (e.g., Ph.D, Ed.D or other).
   ➔ (        ) years

**3.Change response form:** For this question, it would be better to ask the instructors pick or write the specific number of years rather than choose from a radio button. Accurate numbers helps in running the statistical analysis.

4. Which area did you receive your final degree in?
   ( ) Applied Linguistics
   ( ) Linguistics
   ( ) Curriculum and Instruction
   ( ) Educational Psychology
   ( ) TESOL
   ( ) Other_________________

5. Is Language Testing your primary research area?
   ( ) Yes
   ( ) No

6. What kind of assessment related activities have you participated in? (Check all that apply.)
   ( ) Developed Standardized Tests
   ( ) Worked as a Rater
( ) Worked with Classroom Teachers on Testing

( ) Other

6. Ask questions to all: The purpose of this question was to know testing related experience non LTs had. Eventually, LT’s also answered this question. Maybe it is worthwhile asking the same question to LTs as well.

Additional Question:
In which country did you teach the course?

( )

: Since the survey was distributed internationally it would be helpful to know the location of the course. There could be differences in the characteristics of the course depending on the location.

Additional Question:
What language was the course taught in?

English

Other, specify ______________

: From the survey and interview, I found language assessment courses are taught in languages other than English. Even though, the majority of the courses are currently taught in English, it would be helpful to know what other languages offer language assessment courses.

7. If Language Testing is not your main research area, what is your primary research area?

____________________________________________

8. What is the title of the language assessment course you teach?

____________________________________________

9. How many times have you taught this introductory language assessment course?

-> ( ___________ ) times

9. Change response form: It would be better to ask the instructors pick or write the specific number of times they taught the course. Similar to Q3, specific numbers are more useful in analyzing the data.

10. Who is the main target audience (over 50%) of your course?

( ) Regular undergrad.

( ) Pre-service undergrad.

( ) Pre-service M.A.

( ) In-service M.A.
( ) Regular M.A. students (non Teachers)
( ) Ph.D Students
( ) Other __________________

Additional Question:
Which grade level do the majority (over 50%) of your student teachers teach?
a. Elementary  
b. Secondary
 c. Adult  
d. Other ________________________________

: This question was asked in the student teacher survey but not in the instructor survey. From the student teacher survey, there were differences between K-12 and adult student teachers in terms of preferred topics for language assessment courses. Therefore, it would be helpful to have this information from the instructors as well as the student teachers.

11. What is(was) your academic position when you are(were) teaching this course?
( ) Lecturer (non-tenure track)
( ) Adjunct Professor (non-tenure track)
( ) Visiting Professor
( ) Assistant Professor (tenure track)
( ) Associate Professor
( ) Full Professor
( ) Other _______________________

12. Do you hold a teaching certificate? (any grade, any country or school district, any subject area, including emergency/temporary certification)
( ) Yes
( ) No

Additional Question:
If you hold a teaching certificate which subject is it in?
( ) ESL/EFL
( ) English
( ) Other Foreign Language
( ) Other ___________________________

: There could be differences depending on the subject area.

13. Delete item: Almost all the instructors had language teaching experience so the item had little meaning.

14. If you have language teaching experience, including this year, how many years have you taught
ESL/EFL or another foreign language?
- > (    ) years

**8. Change response form: Specific numbers are needed for accurate statistical analysis**

15. If you have language teaching experience, which grade level do/did you most often teach?
   ( ) Pre-elementary/Kindergarten
   ( ) Elementary
   ( ) Secondary (Middle/Highschool)
   ( ) College (Undergrad./Grad.)
   ( ) Adult (Non-Degree)

**Additional Question:**
When was the last time you taught a language class?
(    ) year(s) ago

: To check how recent the instructors teaching experience are.

16. Do you ask your students what they want to learn in the course at the beginning of the semester?
   (   ) Yes.
   (   ) No.

17. Of the following topics covered in the language assessment course, how much time did you spend teaching each topic?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extensive Time</th>
<th>Some Time</th>
<th>A little Time</th>
<th>Hardly Any Time</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Test ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. History of language testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Classroom assessment (developing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Alternative/performance assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Rubric development (analytic and holistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Rater training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Test accommodation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Technology in Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Of the topics covered in the above question, which topic do you think would be MOST helpful to classroom teachers?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Of the topics covered in the above question, which topic do you have the MOST difficulty in teaching?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**17. Change choice description:** Hardly any time, and n/a was really confusing and difficult to interpret statistically so I suggest to change n/a to no time.

*Add additional topic:* From the open ended survey comments participants suggested a question concerning technology should be included.

18. Change response format: Instead of an open-ended questions, change it to a drop down option. Many participants gave more than one answer to this question which made it difficult to analyze.

19. Change response format: Instead of an open-ended questions, change it to a drop down option. Many participants gave more than one answer to this question which made it difficult to analyze.
20. How important do you think the following topics are to pre-service and in-service teachers?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Extremely Important</th>
<th>Very Important</th>
<th>A Little Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test specifications/Item writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Test administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test scoring (e.g., transform numeric scores into letter grades)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standardized testing (interpreting, analyzing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test critiquing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test-taking skills or strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test theory (e.g., validity, reliability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Basic statistics (e.g., mean, percentile, bell curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Advanced statistics (e.g., Item Response Theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Test ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. History of language testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Classroom assessment (developing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Alternative/performance assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Rubric development (analytic and holistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Rater training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Test accommodation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Technology in Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Add additional topic: From the open ended survey comments participants suggested a question concerning technology should be included.*

21. Do you believe it is important for instructors to have field experience in order to teach the language assessment course for pre-service/in-service teachers?
22. After completing the course, how do your students feel about the following areas?

<table>
<thead>
<tr>
<th></th>
<th>more proficient</th>
<th>no effect</th>
<th>less proficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreting Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developing Language Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

23. Do you agree or disagree with the following statement?

My language assessment course met the needs of my pre-service/in-service teachers.

( ) Strongly disagree
( ) Disagree
( ) Agree
( ) Strongly agree

24. What is the MOST important factor to make a language assessment course effective for classroom teachers?

( ) practicality
( ) theoretical knowledge
( ) balance between theory and practice
( ) Other, please specify. ______________________________________________

25. Do you agree or disagree with the following statement regarding statistics?

<table>
<thead>
<tr>
<th>Introduction language assessment course should cover descriptive statistics.</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction language assessment course should cover interpretive statistics.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction language assessment course should cover measurement theory (e.g., Item Response)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
26. Overall, do you think the course was helpful for your pre-service/in-service teachers?

( ) not at all
( ) a little
( ) somewhat helpful
( ) very helpful

27. In general, do you think your pre-service/in-service teachers find your course (check all that apply):

( ) interesting
( ) too theoretical
( ) easy
( ) useful
( ) too practical
( ) difficult
( ) a nice balance between theory and practice

28. If you were to teach this course again, or teach an advanced course in language assessment, what topics would you like to cover?


29. Do you have any other comments or questions you have regarding language assessment courses?


198
Appendix G

List of Professional Organizations Used for the Survey

Language Testing Organizations
LTEST-L: Language Testing Research and Practice Discussion List
MwALT: Midwest Association of Language Testers
ECOLT: East Coast Language Testers
SCALAR: Southern California Association for Language Assessment Research
JLTA: Japan Language Testing Association
EALTA: European Association for Language Testing and Assessment
KELTA: Korea English Language Testing Association

Language Teacher Associations
TESOL: Teachers of English to Speakers of Other Languages
AAAL: American Association of Applied Linguistics
CAL: Center for Applied Linguistics
IATEFL: International Association of Teachers of English as a Foreign Language
CALICO: Computer Assisted Language Instruction Consortium
CATESOL: California TESOL
GATESOL: Georgia TESOL
Sunshine State TESOL OF Florida
Foreign Language Teaching Forum
Wisconsin ELL: Wisconsin English Language Learners
Bilingual Teacher Forum
NYTESOL: New York TESOL