

# SUMMATIVE EVALUATION OF AN ENGLISH LANGUAGE TESTING AND EVALUATION COURSE FOR FUTURE ENGLISH LANGUAGE TEACHERS IN TURKEY

*Çiler Hatipoğlu*

## Introduction

Assessment is one of the cornerstones of the learning process. since it reveals whether the learning process results in success or failure (Dochy 2009; Kozhageldiyeva 2005). What is more, evidence from studies conducted in a range of educational contexts suggests that “the typical teacher can spend as much as a third to a half of his or her professional time involved in assessment-related activities” (Crooks 1988; Dorr-Bremme 1983; Newsfields 2006; Stiggins 1999:23) and that special competence is required to do this job well (Stiggins 1999). Therefore, the issue of future language teachers’ preparation in the field of foreign language testing and evaluation has been a hotly debated topic in the field of education in the recent years (Brindley 2001; Gullickson 1984). Studies focusing on the relationship between teacher training and language testing and evaluation revealed four important results: first, second language assessment is a “notoriously difficult domain of knowledge for students in second language teacher education programs because of the high level of abstraction around its key theoretical concepts, validity, reliability, and practicality, and how they need to be balanced against each other in designing and using assessment instruments” (O’Loughlin 2006: 71). Second, language testing “involves many technologies and developments which are different from language teaching, and yet it interacts closely with most aspects of language teaching” (Johnson & Johnson 2001:187). According to Gronlund (1985: 146), “the construction of good test items is an art” that requires not only field knowledge and clear view of the desired outcomes but also “a psychological understanding of pupils, sound judgment, persistence, and a touch of creativity”. Third, since language teachers are not born testers (Jin 2010:556) they need to be thoroughly trained in language assessment concepts, skills and strategies. What specific concepts, skills and strategies are going to be taught, however, depends on the target audience (i.e., the students attending the course). Inbar-Lourie (2008: 394) states, for instance, that the “shape or size” (i.e.,

the teaching content, methodology and materials) of the language testing course intended for teachers who are responsible for both teaching and assessment should be different from the one aimed at researchers and testing experts. Finally, assessment is “an increasingly important domain of language teachers’ expertise as the professional demands on them to accurately assess their students increases as the theory and practice of assessment continues to mature” (Bailey & Brown 1996; Brindley 2001; Newsfields 2006; O’Loughlin 2006: 71).

Despite the importance of assessment literacy in second/foreign language teacher education and the intricate and delicate nature of the process of preparing language testing and assessment (LTA) courses, very little research so far has been specifically devoted to the discussion of the content and teaching methodology and students’ evaluation of LTA courses (Brindley 2001; Inbar-Lourie 2008; Johnson et al. 1999; Kleinsasser 2005; O’Loughlin 2006).

Among the few studies on language testing courses are the ones conducted by Bailey and Brown (1996), Brown and Bailey (2008) and Jin (2010). All three studies utilised questionnaires as data collection procedures and examined the characteristics of basic language testing courses offered at tertiary level in various countries in terms of instructors, teaching content, materials and methodology as well as the students’ perceptions of the courses. The results of Bailey and Brown’s studies showed that topics such as test critique and test analysis, measuring the different skills, classroom testing practice, item writing (for different skills), item content and item quality analysis, item discrimination, the basic descriptive statistics for central tendency and dispersion, the theoretical issues involved in reliability and the general strategies for estimating test reliability, and the overall strategies for demonstrating validity were extensively covered in the majority of the examined language testing courses. What is more, the majority of the participating lectures

believed that their students thought that LTA courses were interesting, useful and with a nice balance between theory and practice (Brown & Bailey 2008:364). After comparing the results of their 1996 and 2008 studies, Brown and Bailey (2008) concluded that the language testing area is steadily developing. They accepted that new topics (e.g., consequential validity, testing in relationship to curriculum, computer based TOEFL) that had not been included in the language testing curriculum in the earlier years were now deemed essential, but they also pointed out that there were many similarities in responses to items included in both the old and the new version of the questionnaire. Therefore, according to the authors, there is a stable knowledge base related to foreign language testing that is evolving and expanding, rather than shifting radically (Brown & Bailey 2008: 371).

The third study in this group was conducted by Jin (2010) who collected data from 86 instructors teaching at various Chinese universities via a questionnaire, which largely replicated the one utilised by Bailey and Brown (1996). Jin's (2010) study showed that a large number of the topics that were covered in the LTA courses in tertiary level in China are parallel to the ones listed by Brown and Bailey in their studies. This finding led Jin (2010) to conclude that the LTA courses in China cover the essential theoretical and practical aspects of the language testing area. Jin (2010) also reported that LTA lecturers in China think that their students find these courses useful and interesting and that the students were highly motivated to take these courses since they "valued the practical usefulness of the courses" (Jin 2010: 567).

Bailey and Brown (1996), Brown and Bailey (2008) and Jin (2010) collected data from relatively large groups of informants teaching language testing courses, researchers in two more recent studies offered, however, direct and more comprehensive reports related to their language assessment courses. The first of those studies was conducted by O'Loughlin (2006) who focused on a postgraduate elective course titled "Assessment in the Language Classroom" taught in TESOL and Modern Languages programs at the University of Melbourne. The course included both discussions of conceptual themes (e.g., social issues in language testing) and practical components (e.g., designing assessment tools for assessing various skills). Following the narrative contributions of two of

his students for 12 weeks, the author tried to uncover how much students developed their understanding of key concepts in second language assessment and their ability to critically evaluate existing assessment documents and instruments. The researcher reported that both students attained the course objectives; however, the students differed in their willingness and capacity to grasp new ideas in the language assessment area. The author claimed that these differences could be attributed to a combination of factors such as students' cultural backgrounds, prior experience with assessment (both as language learners and teachers), and the characteristics of the input they receive in the language assessment classes. Therefore, he argued, the instructors should adopt a learner-centred approach (which takes into account the abovementioned factors) when planning and conducting the LTA courses.

A second study, which showed how important the collaboration between teachers and students is when the aim is to improve the quality and usefulness of an LTA course, was the one conducted by Kleinsasser (2005). In his study Kleinsasser (2005) discussed how he and his students worked together and transformed an MA LTA course from a content-focused to a learner-centred and teaching content-focused course. At the end of this journey one of the students wrote "This is the best class that I have since my first and second semester" and the reason why the students felt this way was the fact that they had a chance to apply into practice what they had learned and discussed in class.

Kleinsasser (2005) and O'Loughlin (2006) presented readers with direct and detailed reports related to their own language assessment courses, though the number of informants in these studies was very small. In their studies Jin (2010) and Bailey and Brown collected data from a substantial number of lecturers teaching LTA courses but instead of directly asking students about their perceptions related to the courses they asked lecturers what they thought students' general attitudes toward language testing courses were. That is, in Brown and Bailey's own words the studies "describe the topics being taught in language testing courses. But they do not address what the students think their needs are in this particular area of their graduate level training" (2008: 373). Therefore, the authors suggested that future studies should "take the form of a needs analysis that would attempt to determine what a defensible curriculum would be from the point of view of instructors of

language testing courses and their graduate students". Brown and Bailey (2008: 373) also argued that further research should "focus on the international distribution of language testing courses and how such courses differ in different types of institutions, from country to country".

Heeding the advice of Brown and Bailey (2008), the current study aims to be a step towards expanding our knowledge about how language testing is being taught in an undergraduate language teacher-training program and how the students think it should be taught. More specifically, the study focuses on an undergraduate English Language Testing and Evaluation (ELTE) course offered at the Department of Foreign Language Education (FLE) in the Faculty of Education at Middle East Technical University (METU) in Ankara, Turkey, and tries to uncover the students' evaluations related to this course in terms of its teaching methodology and content. The students are chosen as the focal group in the study, since according to Newfields (2006:51), "the biggest challenge in promoting assessment literacy seems to be convincing end-users that the topic is actually worth learning". What is more, work done by Brown and Campione (1996), Dubin and Olshain (1986), Kleinsasser (2005), O'Loughlin (2006), and Posner and Rudnitsky (2006) showed how far a course can be improved if lecturers and students worked together. The current study focuses on a specific ELTE course in a particular institution but it is hoped that the findings and general conclusions will be transferable both to other FLE departments in Turkey and to other institutions in the world where prospective English/second/foreign language teachers are trained.

### Background

Since 1982, the main bodies responsible for training teachers in Turkey are the Faculties of Education (FE). The undergraduate teacher training programs in the FE are four years long and successful graduates can apply for a teaching position at any level in the educational system. Sixty-four state universities in Turkey have FE and 29 of them offer Bachelor's degrees in English Language Education in the Departments of FLE (YÖK 2009) and METU is one of them. All departments of FLE in Turkey follow a standardized curriculum prescribed by The Council of Higher Education (YÖK). That is, there are a number of core courses that have to

be taught in all FLE Departments in Turkey so that their students gain a valid FLT diploma. Short descriptions for each of the core courses is provided by YÖK, however, lecturers teaching these courses are allowed to structure their lessons, choose the teaching materials and the teaching methodology which they believe are most suitable for their particular teaching contexts. Therefore, courses with the same title might sometimes follow different syllabi and/or cover different contents in the different institutions.

Together with these core courses, each FLE Department offers a number of extra (elective) courses whose characteristics depend on the teaching staff availability and/or the traditions of the department (see Appendix A for the curriculum used at METU). One of the reasons why YÖK tries to regulate the curricula of the FLE departments is to ensure parallel education (to a particular degree) in all universities in Turkey since students who graduate from the FLT departments have a wide variety of choices available to them. After completing the four year program, the graduates can apply for a teaching position in state or private primary schools (years 1-8), high schools (years 9-11) or universities (e.g., university prep schools). Since English is the "most studied foreign language and the most popular medium of education after Turkish" (Doğançay-Aktuna 1998:37) in Turkey, FLE department graduates have a great chance to find a job on one of hundreds of private English language teaching courses as well. Usually when appointed to a particular position the language teachers are not only responsible for teaching but also for preparing and administering various tests to their students.

The standardized curriculum prescribed by YÖK comprises three different groups of courses: field courses (i.e., courses directly related to teaching English as a foreign language; e.g., Approaches to ELT, Teaching English to Young Learners, N=29), education courses (e.g., Educational Psychology, N=6) and general knowledge courses (e.g., Introduction to Information Technologies and Applications, Written Communication in Turkish, Non-departmental electives, N=15). As can be seen in Appendix A, among the 29 field courses included in the undergraduate program at the FLE Department at METU, there is only one ELTE course. A small survey scrutinising the curricula posted on the websites of the other 28 FLE departments in the state universities in Turkey revealed that none of the FLE programs

in the country offers more than one ELTE course. That is, university lecturers teaching the ELTE courses in the FLE departments in Turkey are expected to present to the trainee teachers all fundamental principles of language testing and assessment in a single course (i.e., not only the theoretical but also the practical points related to this topic).

How difficult the job of ELTE lecturers in Turkey is becomes even more evident when we take into account the fact that the Turkish education system is very examination oriented. At every level (primary, secondary, high school, university), there are compulsory class evaluations (e.g., quizzes, midterms, take home exams) and end of term exams (e.g., finals) that are usually prepared by class teachers. On the other hand, there are end of the year, centralised tests (e.g., The Level Identifying Exams-SBS, The Student Selection Examination-ÖSS) which determine whether or not students at different levels will gain entry to any of the more prestigious schools or tertiary institutions, and teachers are expected to prepare their students for these exams. In Turkey, the performance of the students, teachers and even schools at each ring of the system is evaluated by looking at how well students perform on various exams. Therefore, graduates of the FLE departments in the country are expected to have a solid knowledge of ELTE, which will help them fit quickly into the system and deliver successful programmes.

### **The English Language Testing and Evaluation (ELTE) Course**

The ELTE course discussed in this study is a compulsory (i.e., every student has to take and pass this course in order to be able to graduate) introductory language testing course (see Inbar-Lourie 2008 for a detailed discussion about the difference between language testing and language assessment courses) placed in the seventh semester of the FLE program at METU (see Appendix A). The course consists of 42 hours of face-to-face instruction. The students can take the course either during the Fall Semester (when they have three hours per week over a 14 week semester) or during the Summer term (seven hours per week over a six week Summer semester). The data for this study were collected at end of the Summer Term of the 2009-2010 academic year.

The main texts used in the course are Genese and Upshur (1996) and Hughes (1989), but other articles and books are also utilised to support the

presentations and discussions of some of the topics (e.g., Bachmen & Palmer 1996; Fulcher & Davidson 2007; McNamara 1996; Woods et al. 1986) (see Table 1 for the topics covered in this course).

When the course is taught during the Summer Term, each week, the first session takes three hours on one day, with the second session taking four hours on a different day. Bearing in mind that the majority of the students (90% of the informants in this particular group) have never studied testing and evaluation before, in the first two weeks, the course lecturer presents and discusses the first five topics (i.e., Teaching and testing relationship, Kinds of tests and testing, Validity, Reliability and Writing Multiple Choice Item Tests). The objectives of these two weeks are to introduce the students to the field and to make sure that the “core knowledge base related to foreign language testing” (Brown & Bailey 2008:373) is covered.

In the following three and a half weeks, the students are divided into groups of approximately 5-6, and they are required to: (i) prepare a 50-minute presentation in which various testing and scoring techniques that can be used to assess a particular skill/type of knowledge (e.g., reading/grammar) are introduced and examined (e.g., techniques for testing grammar: matching, fill in the blanks, multiple choice questions; techniques for scoring writing: holistic, analytical, primary trait scoring). Keeping in mind that students who graduate from the FLE Department at METU can teach at any level in the educational system, the students are also asked to discuss whether or not the techniques they present are appropriate for use with young or adult students, and with beginner, intermediate or advanced level learners;

(ii) prepare midterm or final tests, using the techniques described in their presentations for a specific group of students (e.g., 6-7 year old students who have intermediate level of proficiency in English and who will be attending school in an English speaking country next year);

(iii) choose 15-20 questions from a number of widely used foreign language exams (e.g., TOEFL, IELTS, ÖSS, SBS), and to critically evaluate them (i.e., discuss their good and bad points in class);

(iv) gather all the materials they have prepared (i.e., the presentations, exams and example questions) in one electronic file, ready to be uploaded to online.metu (an e-platform where all materials related to the course are uploaded and from where every student who takes the course can reach them) and to send it to the lecturer.

**Table 1:** English Language Testing and Evaluation: Course Schedule

WEEK	TOPIC	
Week 1	Session 1	<b>Teaching and testing relationship</b> <i>1. Backwash, 2. Inaccurate tests, 3. The need for tests, 4. Testing as problem solving</i>  <b>Kinds of tests</b> <i>1. Proficiency tests, 2. Achievement tests, 3. Diagnostic tests, 4. Placement tests</i>
	Session 2	<b>Kinds of testing</b> <i>1. Direct vs. indirect testing, 2. Discrete point vs. integrative testing, 3. Norm-referenced vs. criterion-referenced testing, 4. Objective testing vs. subjective testing, 5. Communicative language testing</i>  <b>Validity</b> <i>1. Content, 2. Criterion-related, 3. Construct, 4. Face</i>
Week 2	Session 1	<b>Reliability</b> <i>1. Definition, 2. Methods for testing reliability, 3. Sources of unreliability 4. How to make tests more reliable, 5. Calculations (mean, mode, median, range, standard deviation, the standard error of measurement and the true score, the reliability coefficient, item difficulty, item discrimination index), 6. Reliability and validity</i>
	Session 2	<b>Writing Multiple Choice Item Tests</b>
Week 3	Session 1	<b>Testing Skills/Knowledge</b> <i>Testing Overall Ability</i>
	Session 2	<i>Testing Grammar</i>
Week 4	Session 1	<i>Testing Vocabulary</i>
	Session 2	<i>Testing Reading</i>
Week 5	Session 1	<i>Testing Writing</i>
	Session 2	<i>Testing Listening</i>
Week 6	Session 1	<i>Testing Oral Ability</i>
	Session 2	<b>REVISION</b>

**Table 2:** Assessment Procedures

Assessment type	Weight	Topics included
Midterm Exam	30%	Usually a two-hour written examination which covers the topics discussed in the first two weeks and includes two types of questions: (1) questions testing students' knowledge of the theoretical material covered in lectures (2) practical data analysis questions
Final Exam	30%	Usually a two-hour written examination which covers the topics presented by the students and again comprises theoretical and practical questions. The aim of this exam is to ensure that the students do not only know the topics that they presented in class but also study the topics presented by the other groups
Practice/Hand-on activities	35%	Presentation of the testing techniques: 10% Writing midterm/final exams: 10% Critical evaluation of exam questions: 10% Preparation of the final file: 5%
Attendance and Classroom Participation	5%	Attendance Student participation during the class discussions of the topics covered in the course

The assessment procedures employed in this course are presented in Table 2.

**The Study**

This study, which is an example of a summative course evaluation (Cronbach 1963; Dick et al. 2005), was carried out with a total of 81 (M=16, F=65) FLE department students whose age range was 18-25. The majority of the students were in their third (N=39, 49%) or fourth (N=36, 44%) year at the university, but there

was also a small group of informants who were second year students (N=6, 7%). Eighty out of the 81 participants stated that they were planning to work as English language teachers after their graduation, but many of them had not yet decided at what level (i.e., primary school, high school or university) they would like to work.

The data for this study were collected using two data collection procedures: questionnaires and interviews. The questionnaire used in the study was developed by two experts regularly

teaching the ELTE course in the department and was piloted with 45 students who took the ELTE course in the Fall Semester of the 2009-2010 academic year. During the piloting process the students were asked to do three things: first, to answer the questions included in the questionnaire, then, to read the questions again and, if they thought there were any problems with any of the questions (e.g., wording, place in the questionnaire, not a relevant question) to write their comments in the space provided next to each of the questions, and finally, to write down any questions and/or comments they thought should be added to the questionnaire. The analysis of the pilot data revealed that the students did not have any major problems with the questionnaire and the Cronbach alpha coefficient of the scale was .86. Hence, this questionnaire was used in the main study.

Subjects who participated in the current study completed the questionnaire after their final exam. The questionnaire comprised two parts (i.e., A and B). The goal of the first part of the questionnaire was to collect as detailed information related to the participants as possible since previous studies focusing on evaluation of LTA courses showed how important the background of the participants was (e.g., O’Laughlin 2006). The students were asked to provide information related to their age, gender, year at university, permanent place of residence, parents’ level of education and whether they were planning to work as English language teachers after their graduation. In Part B there were the following two questions.

<b>Part B</b>	
<b>B1:</b> First, list five topics, which you have covered in this course and which you think will help you with your future career as an English language teacher. Then, explain why and how each of these topics will help you in your future job.	
TOPICS	REASONS
1.	
2.	
<b>B2:</b> You have completed the ‘English Language Testing and Evaluation’ Course. List 3 things that you think should be changed in relation to this course to make the course better if it is to be taught again.	
THINGS TO BE CHANGED	REASONS
1.	
2.	

No name or any other directly identifying information related to the participants was required in the questionnaire. What is more, the students were informed that the aim of the study

was to critically evaluate and improve the quality and usefulness of the ELTE course for the students who would be taking it in the future, therefore, their sincere opinions were needed. The students were also ensured that in no way would the data provided by them affect their grades in this or any other course they were taking.

The interviews were conducted with 16 volunteer students and took place after the analysis of the questionnaire data was completed. The objectives of the interviews were to triangulate the questionnaire data and to elicit more in-depth information related to the research questions asked in this study.

The qualitative data were analysed in two stages: first, the listed reasons/given explanations were coded into categories and the information in each category was further divided into sub-categories. Then, the data in each sub-category was quantified and analysed using Predictive Analytics SoftWare (PASW).

**Results**

**Question 1:** *First, list five topics, which you have covered in this course and which you think will help you with your future career as an English language teacher. Then, explain why and how each of these topics will help you in your future job.*

The analysis of the data in this section was conducted by matching the topics listed by the students with the topics included in the course schedule outlined in Table 1. The students were asked to list five topics but some of them listed less topics therefore the total number presented in Table 3 (i.e., 339) is smaller than the expected one (i.e., 81X5=405).

**Table 3:** Topics that the students think will help them with their future career as English language teachers

SECTIONS	N	%
Teaching and testing	25	31
Kinds of tests	37	46
Kinds of testing	21	26
Validity	64	79
Reliability	76	94
Writing Multiple Choice Item Tests	32	40
Testing skills/knowledge	84	104
<b>TOTAL</b>	<b>357</b>	

The course evaluated in this study consisted of seven main sections (see Table 1) and as can be seen in Table 3, three topics - Testing skills/knowledge, Reliability and Validity - were mentioned much more frequently than the others by the students.

The most popular topics with the students were the ones discussed in Section 7 (i.e. testing skills/knowledge). All students thought that learning the techniques for testing different skills/kinds of knowledge, the practice of writing exams for testing different skills/knowledge and the critical evaluation of questions from standardised English language tests were practices that they would use when they become English teachers. Some of the students even listed more than one sub-topic (e.g., testing grammar, scoring procedures in testing writing) covered in this Section while answering question B1; therefore, the number of answers in this group is bigger than the total number of students who participated in this study.

When the reasons for listing the topics covered in Section 7 were examined, it was seen that the students' answers clustered mainly around the following four reasons: (i) parallelism between the topics taught in class and the things teachers are doing in FLE classes (F42: Teachers should know how to prepare tests assessing different skills because in Turkish schools skills are taught separately. If we teach skills separately, then we will need to test them separately); (ii) acquaintance with a wide variety of testing techniques (F57: I have learned about test and question types that I had not heard of before); (iii) appropriate tests for different groups of students. The students stated that before taking this course they had not realised that young and adult learners were so different from each other and what a large number of factors should be considered when preparing tests for specific groups of students; (iv) scoring procedures. Informants frequently mentioned the fact that in this course they had learned that different exams require different types of scoring and that students should not be penalised for irrelevant mistakes (F13: ... we should remember that if we test reading we should not penalise students for irrelevant mistakes such as misspelling).

The second most frequently mentioned section by students was **Reliability** (i.e., Section 5). Seventy-six out of the 81 students (94%) thought that the topics that were discussed in this section were the ones that would help them prepare better tests. The students argued that while discussing the topics in this section, they had learned that "F15: Preparing a test is more than putting blanks and choices together. Tests should have some features, especially reliability - to reach their goals". The participants also mentioned the fact that "reliability" was a "really

a new concept" for them but a concept that "completely made sense and seemed a must topic for testing" (F54). Some of them also pointed that now they knew that "without reliability a test is not a test" (F57) and they felt confident about creating reliable tests, since in the ELTE course they had "learned some techniques to increase reliability, which is very important" (F29). Another point frequently mentioned by the students was the realisation that tests require as much work after administration than before. They stated that they were not aware of the importance of the statistical analysis (i.e., the calculation of mean, mode, standard deviation, standard error of measurement, item difficulty, item discrimination index) that were done after the exams. Moreover, they argued that they had a lot of fun doing the calculations and trying to figure out which items they should change and which to keep if the exam was to be administered once again.

The last section which students found very beneficial for their development as future English teachers was Section 4 (i.e., **Validity**). Among the 81 students who participated in this study 64 thought that the topics covered in this section would help them during their future careers. One participant wrote "F31: I think validity is really important because the exams that we give to our students may affect their lives (scholarship, higher education)". One group of students mentioned the fact that even knowing that there were different types of validity (e.g., content vs. predictive) would help them prepare better and fairer exams for their students (F52: I believe it will help me prepare exams more consciously, paying attention to what to include and how to include it). Finally, a group of informants referred to their own experiences as exam takers and how frustrated they were with the spelling, grammar and formatting problems on the exams they had taken, and they argued that the knowledge they gained from this course would help them prepare exams with good face validity.

The examination of the data also showed that the other four sections (i.e., Sections 1, 2, 3 and 6) were not so popular with the students. Section 2 was mentioned by 46% (N=37) and Section 6 by 40% (N=32) of the participants in this study. The least frequently mentioned topics by students were "Teaching and testing" (31%) and "Kinds of testing" (26%). Why students did not find those very relevant to their future careers

became clear when their suggestions for improvement of the course were analysed.

**Question 2: You have completed the ELTE Course. List 3 things that you think should be changed in relation to this course to make the course better if it is to be taught again.**

As seen in Table 4, eight strands of suggestions for change related to the different stages of the course emerged from the students' answers.

**Table 4:** Topics that need change

TOPICS	N	%
1. Course books	4	2.4
2. Topics covered	15	8.9
3. Student presentations	37	21.9
4. Not enough practice	64	37.9
5. Slides before lectures	4	2.4
6. Assessment system used in ELTE course	11	6.4
7. Crowded classes	14	8.3
8. More testing courses	20	11.8
<b>TOTAL</b>	<b>169</b>	<b>100</b>

The first group of suggestions were related to the content and main textbooks of the course. Students (8.9%) argued that the number of topics covered in the ELTE course is too high and that some of them are very abstract (e.g. kinds of testing, teaching and testing) and they would not be able to use them/need them in their classrooms. They suggested that the "abstract topics" were removed and the more practical topics such as testing skills/knowledge that, they believed, would help them "assess their students more successfully" were examined in greater detail in class. A small group of students also recommended that the main texts utilised in the course were changed. They argued that "the language of the books was too heavy, very complicated and very detailed" (F12). Some of them mentioned that they always needed the instructor's or their classmates' presentations to understand the topics better.

The second group of comments (62.2%) focused on the "procedures" (Brown 2001) used in class. They (37.9%) felt that in this course theory and practice were not well balanced and that for a better integration of those two components in the ELTE course: (i) practical exercises showing the relationship between the theoretical material covered in the lectures and the real challenges of day-to-day classroom assessment should accompany or follow the

presentation of each of the topics (16/64); (ii) more time should be devoted to evaluation and discussion of questions from various exams since this practices, according to the students, not only flagged the possible problems that they might encounter while writing exams but also improved their material selection skills (17/64); (iii) students should be asked to write more than one test (10/64); (iv) there should be sessions in which the whole class would be working together on the preparation of various tests (6/64); (v) more time should be allocated for the discussion and evaluation of the exam questions prepared by the students (15/64). The students (21.9%) also argued that changes related to the system regulating the student presentations should be made. Informants stated that "student presentations" were inefficient and boring since students lacked vital knowledge in field of testing (16/37)". They stated that instead of group presentations there should be individual presentations which they believed would help students learn more and develop their skills and knowledge in the testing and evaluation area better than the group presentations (8/37). A group of students believed that "student presentations" should be removed from the course syllabus and replaced with presentations made by the teacher since they "could be time consuming" and since they needed to "hear the topic from the teacher afterwards anyway" (F63). A small group of students (2.4%) maintained that it would be much more beneficial for the students if the lecture notes/slides were available not after but before the discussion of the topics.

The informants also had suggestions related to some of the assessment procedures used in the course (6.5%) (see Table 2 for the details related to assessment). The students argued that instead of two formal exams (i.e. Midterm and Final) in this course "alternative exam procedures" should be utilised. The students proposed that the final exam was replaced with a take home exam or project. They argued that a project asking students to: (i) scrutinise in detail one of the topics covered in class; or (ii) prepare a set of exams (e.g., quizzes, midterms and a final) for a specific groups of students; or (iii) critically evaluate already published exams would be more beneficial for their future careers.

The final groups of comments put forward by students dealt with "administrative issues". They complained that the classes were too crowded and due to this they had difficulties following lecturer's explanations/ presentations/



comments. They also stated that the crowded lectures did not allow for productive class discussions to take place. A group of students (11.8%) thought that just one ELTE course was not enough to cover all the topics in the testing area. They stated that the ELTE course was really important for their professional development and therefore there should be at least one more but preferably a series of testing courses (F27: I think these kinds of lessons should be in the syllabus from the first year onwards. This is a lesson which we will always need; F46: I think one course is not enough because the things that we covered are really important).

### Discussion

The current study aimed to obtain a summative evaluation of an undergraduate ELTE course by trainee English language teachers. It was hoped that this evaluation would help the course instructor to uncover how students viewed the course and whether they believed it prepared them to meet the challenges awaiting them in their professional lives. The results of the study showed why obtaining continuous feedback should be “an integral component of course evaluation and revision” (Daloğlu & Marsh 2007:17). The feedback that the students provided allowed me to see the course and my teaching from a different, more objective and more independent perspective that otherwise might have remained unnoticed. Thanks to the collaboration with my students, I now have empirical evidence based on which I can make some additions, deletions and revisions to the ELTE course in relation to the research questions scrutinised in this study.

First of all, I now understand better why Rohrer (1981: 32) insisted that lecturers teaching ELTE courses should ask and get answers for pivotal questions such as “What will the teachers-in-training be doing when they leave?” and “What do they need to learn about testing to do what they need to do?” before and while preparing the course syllabi. The current study revealed that the “ideal” syllabus which is based on lecturers’ previous experience and expertise might not be “the most” meaningful or the most engaging for the specific group of students. When the teaching content of the ELTE course at METU is compared with the contents of the testing courses around the world (see Brown & Bailey 2008; Jin 2010) it can be seen that it comprises the “core knowledge base related to

language testing”. However, the student evaluations in this course revealed that they found three (i.e., Testing skills/knowledge, Reliability, Validity) of the seven sections of this course very useful while they had more reserved opinions about the other four. In their feedback through the questionnaire and the interviews, the majority of the students stated that they did not think the sections on Teaching and Testing, Kinds of Tests, Kinds of testing and Writing MCIT were relevant to their future careers. One plausible explanation for this evaluation might be the methodology which was followed while presenting these topics. Due to time restrictions, the lecturer defined and presented to the student the main characteristics of the different types of tests and testing and the MCIT but no practical exercises were carried out either in class or as homework in relation to those topics. Therefore, a different teaching methodology will be tested in the following semester. Keeping in mind the importance of these four topics, the diverse interests of the FLE students and the wide range of opportunities they have after completing their degrees, removing those topics from the syllabus might not be the best solution. Instead, in the subsequent semester, the following procedure will be followed. In the first session of the course, those four topics will be briefly introduced to the students and depending on their interests they will be asked to read more on two of those topics at home. Then, in class various exercises allowing students to see the practical side of those topics will be employed (e.g., the students will be asked to find tests exemplifying each kind of test/testing and will be asked to describe when, how and with which groups of students they could be used). It is hoped that these activities will help students to relate theory and practice better.

Based on the students’ feedback some additions to the course will also be made. One of the main additions will be the creation of more time and opportunities for practicing writing and evaluating various exams. In the evaluated course, the students had to prepare one exam assessing a particular skill/knowledge and to critically evaluate 15-20 questions from widely used language tests in groups but the results of the study showed that the students did not feel that these exercises had given them enough opportunity to improve their assessment literacy. Hence, in the subsequent semester each student will be given more chances to work on improving their test evaluation and test writing skills. Following the procedures introduced by

O'Loughlin (2006), the students will be required to write tests assessing each of the examined skills/types of knowledge and to post those on an e-platform (i.e., online.metu) used at the university; and every student will be asked to comment on a particular number of the posted exams. By doing this, the students will not only benefit from the feedback provided by the instructor but also from the comments of their peers. What is more, this practice will give the course instructor a chance to illustrate to the students types of evaluation/assessment that are usually not discussed in the ELTE course (e.g., self-assessment, peer assessment, assessment via electronic means, process evaluation). It is hoped that this continuous practice of test writing and test evaluation will produce more active and more focused foreign language teachers who would be better placed to assess their students' progress. If this new practice is implemented, the suggestions of the students related to changes in the presentation and assessment systems will be realisable as well. The writing and evaluating of exams will then replace the student presentations and one of the currently used assessment procedures (either the midterm or the final exam).

The ELTE course evaluated in this study is offered both in the Fall and in the Summer terms. The summer term is much shorter and much more intense than the Fall semester and the classes are usually more crowded due to the small number of lecturers teaching in the Summer term. At the end of the Summer term the students argued that "important topics were covered in a hurry" and due to this in the FLE curriculum there should be at least two but preferably a series of testing and evaluation courses. Keeping in mind the possibility that the length of the term and the size of the class might have been two important factors affecting students' perceptions, it was decided to replicate this study in the Fall term and after the analysis and comparison of the data from both studies to decide whether or not to propose a new ELTE course.

### Conclusion

Language testing and evaluation "provides goals for language teaching" and "it monitors success in reaching those goals" (Davies 1990:1). "If assessment is not working effectively day to day in the classroom, then the power of assessment at all other levels is diminished" since "standardized tests cannot overcome the dire

consequences for students of the ongoing mismeasurement of their achievement in the classroom" (Stiggins 1999:23). Because of these issues, teacher training departments constantly have to evaluate their programs to make sure that they prepare their graduates well for the challenges of day-to-day classroom assessment (Stiggins 1999). The current study contributes to this strand of research and (as far as the author is aware for the first time in Turkey) tries to obtain a summative evaluation of an undergraduate ELTE course by trainee English language teachers. The study exemplifies how studies which go beyond "yes-no responses" (Stiggins 1999: 24) and include students' opinions in the evolution process can help instructors and administrators obtain a fuller, more detailed, more objective picture of the courses they are offering. The study also shows how by focusing on the needs of a "clear-cut population" (Brown & Bailey 2008) and by using multiple data collection procedures (e.g., combination of questionnaires and interviews) a more "learner-centred and teaching content-focused" even "tailor made" courses that better meet the expectations of the students can be created.

In addition, the findings of the study can be used as a springboard for further research aiming to evaluate testing courses. What was done in this study was to take the first step towards a "spiral evaluation" of a "vital" course in a teacher training program. The next steps would be to include the suggestions of the students in the new course syllabus (as described above) and to follow the current evaluation with a series of needs assessments, and formative and summative course evaluations which will lead to a more reliable and more generalisable picture of what, how and why language testing should be taught at undergraduate level with the 'identified group of students'. In order to see how the course fits into the broader educational contexts, as a subsequent or parallel step, the evaluations of other stakeholders should be sought as well. Cooperation between course lecturers, administrators, material developers, test writers and employers may lead to the creation of ELTE courses that will create teachers that contribute to "excellence in classroom assessment" (Stiggins 1999: 27).

### References

- Bailey, K. M. & Brown, J. D. 1996. Language testing courses: What are they? In A. Cunnings & R. Berwick (eds.), *Validation in Language Testing* (pp. 236-256). Clevedon, UK: Multilingual Matters.

- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Brindley, G. 2001. Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (eds.), *Experimenting with Uncertainty: Essays in Honour of Alan Davies* (pp. 126-136). Cambridge: Cambridge University Press.
- Brown, H. D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy (Second Edition)*. New York: Longman.
- Brown, J. D. & Bailey, K. M. 2007. Language testing courses: What are they in 2007? *Language Testing*, 25/3: 349-383.
- Brown, A. L. & Campione, J. C. 1996. Psychological theory and the design of innovative learning environments: On procedures, principles and systems. In L. Shauble & R. Glaser (eds.), *Innovations in Learning: New Environments for Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cronbach, L. J. 1963. Evaluation for course improvement. *Teachers College Record*, 64: 672-683.
- Crooks, T. J. 1988. The impact of classroom evaluation practices on students. *Review of Educational Research*, 58/4: 438-481.
- Davies, A. 1990. *Principles of Language Testing*. Cambridge, MA: Basil Blackwell.
- Dick, W., Carey, L. & Carey, J. O. 2005. *The Systematic Design of Instruction (Sixth Edition)*. New York: Pearson.
- Daloğlu, A. & Marsh, F. 2007. Teacher-student partnership in evaluating and revising a multidisciplinary sustained-content English language course. *Applied Language Learning* 17/1&2: 15-31.
- Dochy, F. 2009. The edumetric quality of new modes of assessment: Some issues and prospects. In G. Joughin (ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 85-114). New York: Springer.
- Doğançay-Aktuna, S. 1998. The spread of English in Turkey and its current sociolinguistic profile. *Journal of Multilingual and Multicultural Development*, 19/1: 24-39.
- Dorr-Bremme, D. 1983. Assessing students: Teachers' routine practices and reasoning. *Evaluation Comment*, 6/4: 1-12.
- Dubin, F. & Olshtain, E. 1986. *Course Design: Developing Programs and Materials for Language Learning*. Cambridge: Cambridge University Press.
- Fulcher, G. & Davidson, F. 2007. *Language Testing and Assessment: An Advanced Resource Book*. London & New York: Routledge.
- Genese, F. & Upshur, J. A. (1996). *Classroom-based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Gronlund, N. E. 1985. *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Gullickson, A. R. 1984. Matching teacher training with teacher needs in testing. Paper presented at the Annual Meeting of the American Education Research Association, New Orleans, LA (Retrieved July 12, 2010 from <http://eric.ed.gov/PDFS/ED249254.pdf>).
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Inbar-Lourie, O. 2008. Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25/3: 385-402.
- Jin, Y. 2010. The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27/4: 555-584.
- Johnson, K. & Johnson, H. 2001. *Encyclopaedic Dictionary Of Applied Linguistics*. Oxford: Blackwell.
- Johnson, R., Becker, P. & Olive, F. 1999. Teaching the second-language testing course through test development by teachers-in-training. *Teacher Education Quarterly*, Summer 1999: 71-82.
- Kleinsasser, R. C. 2005. Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20/3: 77-102.
- Kozhageldyeva, L. 2005. Importance of Testing and Evaluation in Foreign Language Teaching: A Suggested Syllable for a Testing Course in EFL Departments in Kazakhstan Universities. MA Thesis, Gazi University, Ankara, Turkey.
- McNamara, T. F. 1996. *Measuring Second Language Performance*. New York, NY: Longman.
- Newfields, T. 2006. Teacher development and assessment literacy. *Authentic Communication: Proceedings of the 5<sup>th</sup> Annual JALT Pan-SIG Conference*. May 13-14, 2006. Shizuoka, Japan: Tokai University College of Marine Science, pp. 48-73.
- O'Loughlin, K. 2006. Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1: 71-85.
- Posner, G. J. & Rudnitsky, A. N. 2006. *Course Design: A Guide to Curriculum Development for Teachers*. New York: Pearson.
- Rohrer, J. 1981. Problems and practice in language testing: a view from the Bundessprachenamt. In C. Klein-Braley & D.K. Stevenson (eds.), *Practice and Problems in Language Testing I: Proceedings Of The First International Language Testing Symposium Of The Interuniversitare Sprachtestgruppe* (pp. 31-34). Frankfurt: Verlag Peter D. Lang.
- Stiggins, R. J. 1999. Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, Spring 1999: 23-27.
- Woods, A., Fletcher, P. & Hughes, A. (1986). *Statistics in Language Studies*. Cambridge: Cambridge University Press.
- YÖK (Yükseköğretim Kurulu [The Council of Higher Education]). 2009. *Yükseköğretim Kataloğu*

2009 [Higher Education Catalogue 2009]. (Retrieved September 23, 2010 from

<http://www.yok.gov.tr/content/view/527/222/>).

APPENDIX A: FLE Program at METU					
FIRST YEAR					
First Semester			Second Semester		
<u>Code</u>	<u>Name</u>	<u>Credits</u>	<u>Code</u>	<u>Name</u>	<u>Credits</u>
FLE 133	Contextual Grammar I	(3-0)3	FLE 134	Contextual Grammar II	(3-0)3
FLE 135	Advanced Reading and Writing I	(3-0)3	FLE 136	Advanced Reading & Writing II	(3-0)3
FLE 137	Listening and Pronunciation	(3-0)3	FLE 138	Oral Communication Skills	(3-0)3
FLE 129	Introduction to Literature	(3-0)3	FLE 140	English Literature I	(3-0)3
EDS 200	Introduction to Education	(3-0)3	FLE 146	Linguistics I	(3-0)3
TURK 103	Written Communication	(2-0)2	FLE 178	Second Foreign Language II	(3-0)3
FLE 177	Second Foreign Language I	(3-0)3	TURK 104	Oral Communication	(2-0)2
IS 100	Introduction to Information Technologies and Applications	NC			
SECOND YEAR					
Third Semester			Fourth Semester		
<u>Code</u>	<u>Name</u>	<u>Credits</u>	<u>Code</u>	<u>Name</u>	<u>Credits</u>
FLE 241	English Literature II	(3-0)3	FLE 218	Novel Analysis	(3-0)3
FLE 261	Linguistics II	(3-0)3	FLE 280	Oral Expression & Public Speaking	(3-0)3
FLE 238	Approaches to ELT	(3-0)3	FLE 262	ELT Methodology I	(3-0)3
FLE 277	Second Foreign Language	(3-0)3		Departmental Elective I	(3-0)3
EDS 220	Educational Psychology	(3-0)3	FLE 270	Contrastive Turkish-English	(3-0)3
CEIT 319	Instructional Technology & Materials Development	(3-0)3	FLE 200	Instructional Principles & Methods	(3-0)3
THIRD YEAR					
Fifth Semester			Sixth Semester		
<u>Code</u>	<u>Name</u>	<u>Credits</u>	<u>Code</u>	<u>Name</u>	<u>Credits</u>
FLE 307	Language Acquisition	(3-0)3	FLE 308	Teaching English to Young Learners	(3-0)3
FLE 304	ELT Methodology II	(3-0)3	FLE 324	Teaching Language Skills	(3-0)3
FLE 311	Adv. Writing & Research Skills	(3-0)3	HIST 2202	Principles of Kemal Atatürk II	(3-0)3
	Departmental Elective II	(3-0)3	EDS 304	Classroom Management	(3-0)3
HIST 2201	Principles of Kemal Atatürk I	NC	FLE 352	Community Service	(3-0)3
FLE 321	Drama Analysis	(3-0)3	EDS 413	Turkish Educational System & School Management	(3-0)3
	Non-Departmental Elective I	(3-0)3		Non-Departmental Elective II	(3-0)3
FOURTH YEAR					
Seventh Semester			Eighth Semester		
<u>Code</u>	<u>Name</u>	<u>Credits</u>	<u>Code</u>	<u>Name</u>	<u>Credits</u>
FLE 405	Materials Adaptation and Development	(3-0)3	FLE 404	Practice Teaching	(2-6)5
FLE 413	English Language Testing & Evaluation	(3-0)3	FLE 426	The English Lexicon	(3-0)3
FLE 425	School Experience	(1-4)3	EDS 424	Guidance	(3-0)3
FLE 423	Translation	(3-0)3		Departmental Elective IV	(3-0)3
	Departmental Elective III	(3-0)3			