

On Some Aspects in Final ESP Testing at University Level

Irena Darginavičienė, Nijolė Burkšaitienė

Abstract. The article considers final ESP testing at tertiary level, which plays a particular role in language teaching. The final testing concludes the teaching process and enables language teachers to make conclusions on the success of the teaching process, effectiveness of the teaching methods and usefulness of the course in general. It measures the students' progress in particular.

The article also analyses theoretical and practical issues of testing and opens to view the state of the art of the final examination in ESP at Law University of Lithuania with description of the final examination type, structure, task types and principles of assessment.

The article is one of the first contributions published which raises the question of the necessity of unification of assessment criteria for the final examination in ESP testing at all universities in Lithuania based on the principals of test reliability and validity and ensuring an equal level of difficulty, comparability and yet subject-specificity.

Introduction

Testing is an integral part of the teaching program at a university level. University programs always involve a certain number of different teaching components, e.g. teaching materials, teaching methods and techniques, testing and assessment, etc. The principal distinction between testing and other components lies in the objectives that testing sets to achieve. Thus, the chief goal of testing is to measure and assess the knowledge gained, while other components of university teaching programs enhance the learning process.

There is a need of **scientific analysis** to develop a critical understanding of the principles and practice of final language testing at tertiary level. First of all, we have to assume that the statements "*the best test*" or "*the best testing techniques*" are totally confusing. In fact a test, which proves ideal for one purpose may be quite useless for another purpose or a technique that may work very well in one situation can be entirely inappropriate in another situation (Hughes, 2000). Besides, in every situation the primary issue is to state the testing problem, which marks the differences in test types and testing techniques. Thus, for instance, what agrees with the goals of testing in large testing corporations can be quite irrelevant in the tests of teaching institutions. In the same way, two universities may require very different final tests, depending on the objectives of their courses and importance of the tests. This leads to a great variety of testing goals, different levels of difficulty and test comparability, as well as different assessment criteria not only among different universities in Lithuania, but also among different ESP courses within one university. This problem calls for the necessity of setting and applying unified approaches to final testing in ESP at tertiary level in Lithuania.

In this respect, the article which is one of the first contributions in this area, deals with the issue of the necessity of **unification of assessment criteria** for the

final examination in ESP testing at all universities in Lithuania, based on the principles of **test reliability and validity** and ensuring **an equal level of difficulty, comparability and subject-specificity**.

The **aim** of this contribution is to reveal the main factors ensuring successful design and administration of final testing in a foreign language at tertiary level, to describe the most effective methodology in designing examination papers and the criteria that are the driving force behind the choice of examination test tasks. For these aims we will **firstly** employ an overview of theoretical aspects of testing and will describe:

1. The relationship between the form and the purpose of the test.
2. The relationship between the content of the test and the course covered by students.
3. Modular test as a suggested test-type for final assessment at university level.

Secondly, we will present the **state of the art in final ESP testing at Law University of Lithuania** based on unified assessment criteria and ensuring test reliability and validity as well as equal level of difficulty retaining subject-specificity among different ESP courses at Law University of Lithuania.

Finally, we will present the conclusions and marking scales.

Overview of Theoretical Aspects of Testing

Our **first task** is to identify the relationship between the purpose of testing and the type of the test. Most linguists (Hughes, 2000; Bachman, Palmer, 2000; Alderson, 1999) and others emphasize that the first thing testers have to be clear about is the purpose of testing in any particular situation. Different purposes will usually require different kinds of tests.

Thus, for instance, the purpose of *proficiency tests* is to measure language proficiency regardless of any language courses that candidates may have followed. This type of tests aims to describe what students are capable of doing in a foreign language and these tests are usually set by external bodies such as examination boards. The content of a proficiency test, therefore, is not based on the content or objectives of languages courses previously followed by test takers.

Diagnostic tests tend to diagnose students' strengths and weaknesses, to identify what they know and what they do not know. Where other types of tests are based on success, diagnostic tests are based on failure as this type of tests is used to find out problem areas and to ascertain what further teaching is necessary.

The purpose of *placement tests* is to assist placement of students by identifying the stage or part of a teaching program most appropriate to their ability, i.e. this type of tests will indicate at which level a learner will learn most effectively in a situation where there are different levels or streams. The aim is to produce groups that are homogeneous in level that will use institutional and teacher time most effectively.

The last type of tests differentiated by most linguists is *achievement tests*. The purpose of achievement tests is to discover how far students have achieved the objectives of a course of study. Two types of achievement tests are distinguished in linguistic literature: progress achievement tests and final achievement tests (Hughes, 2000; Harris, McCann, 1998) and others. Progress achievement tests are administered during courses. They are expected to measure the progress that students are making. These tests are performed after certain blocks of study and they exhibit how well the course content is functioning. Final achievement tests are administered at the end of courses and their aim is to see if students have achieved the objectives set out in the syllabus.

Language testers argue on the essential characteristic of final achievement tests. Thus, our **second task** is to present an overview of two approaches relating to this question and found in the linguistic literature (A. Hughes, L. F. Bachman and A. S. Palmer, etc.). Agreeing on the fact that the content of these tests must be related to the courses with which they are concerned language testers strongly disagree on the nature of this relationship.

The first approach is called *syllabus-content approach*. The testers who advocate this approach keep to the opinion that the content of a final achievement test should be based directly on a detailed course syllabus or on the books and other materials used. A. Hughes indicates both advantages and disadvantages of this approach (Hughes, 2000: 11). The advantage of this approach lies in the fact that such tests are always fair since they only contain what the students have actually encountered in the process of teaching. However, the disadvantages of this approach outweigh the advantages. Firstly, if the syllabus is badly designed, or the books and other materials are badly chosen, then the results of a test can be very misleading. Besides, successful performance on the test may not truly

indicate successful achievement of course objectives. For example, the aim of a course may be the development of conversational ability, but the course itself and the test may require students only to utter carefully prepared statements on the topics covered. In this respect the results will fail to show what students have achieved in terms of course objectives.

The second approach expresses the opinion that the test content must be directly **based on the objectives** of the course. This approach is much to be preferred by different testers as it has a number of advantages. First, it makes course designers to be explicit about objectives. Then, it allows to show just how far students have achieved pursuing those objectives. Finally, it puts pressure on those responsible for the syllabus and for the selection of books and materials to ensure that they are consistent with the objectives of the course.

Recently linguists have started to discuss the question of a modular test as a final test-type. Therefore, our **third task** is to refer to *modular test* in ESP testing. M. West distinguishes three types of modular test: national modular test, e.g. taken by 12-year school students, integrated institutional test, e.g. taken by university students who cover similar programs, and individual modular test, e.g. taken by students of one university or higher school. Modular test seeks to test and measure foreign language skills in three different levels: elementary, intermediate and advanced. The most important characteristic of modular test is its structure. The test should consist of three parts: *1 anchor component and 2 special components*. The anchor component is based on the material of low or medium specificity (popular science), which law people could understand. Special components are based on the original material of high or very high specificity in a certain field.

Research Objectives

In this part of the contribution we **target**:

- to describe the state of the art in final ESP examination at Law University of Lithuania;
- to outline the principles of choosing the test type, its structure, task types and assessment criteria applied at Law University of Lithuania.

Final Examination in ESP: The Present

Appropriate design and administration of final foreign language examination becomes of paramount importance because only such exam can properly check and measure students' proficiency in all macro-skills on the basis of the students target needs and the input the students received throughout the course. The exam is designed on the principles of **final achievement test based on objectives of the course covered and following the standards of modular test**.

At the Faculties of Law and Public Management students have an extensive compulsory course in ESP, mainly English for Law, English for Public Management and English for Public Administration, that lasts for three terms

with the frequency of 4 academic hours per week. The extensiveness of the course, on the one hand, enables the learners to focus on their purpose for learning English for special purposes step by step in parallel with their subject course. This has been achieved through matching the curriculum of ESP (and other foreign languages taught at the University) with the curriculum of legal subjects. On the other hand, availability of time allows a number of various out-of-class activities such as presentations and projects to be introduced and feedback provided to the learners.

At the beginning of the course students are informed about the structure of the final mark so that they are aware of the fact that during the course their classroom performance is going to be continuously assessed and a mark given for it will constitute 25 percent of the final grade. At the end of the first and the second terms students have to pass a credit test. The course finishes with the formal examination at the end of the third term.

We have introduced a separate mark for performance in class as part of the final examination grade because, firstly, continuous classroom assessment done by the teacher is important for the feedback that both the teacher and the learner receive. Secondly, it highly motivates the learners from the very beginning of the course to improve their performance, and thirdly, it has most value as an aid to learning.

Classroom performance is assessed during each term taking into account the marks students are given for written tests and oral activities, where marking for written tests is in most cases objective, that is based on scores for correct answers, whereas oral activities are assessed using holistic approach, i.e. matching band descriptors to grades. At the end of three terms the marks for performance are summed and the averaged mark (that becomes one fifth of the final examination grade) is given. The objective and holistic principals of assessment are also used assessing final Examination Tests.

Final Examination: Test Type, Structure, Task Types, and Assessment

The examination test type, structure, task types and principles of assessment have been made uniform for all foreign languages – English, German and French – taught at Law University. Here we will describe the English variant of final testing.

Examination Test Type

In the process of test development some major decisions had to be made. Firstly, the type of test for the final formal examination in ESP had to be chosen; secondly, the decision concerning skills to be tested had to be made; and thirdly, estimation of skills in percentage in the final mark had to be decided.

Having in mind the fact that the objectives tested should be course objectives and should be covered in the teaching process the decision was made to chose the modular achievement test type as it aims to measure how well the

students will perform in their target language tasks and will assess the whole rather than discrete items.

Considering the skills to be tested the decision was made on the basis of the syllabus, i.e. all macro-skills had to be checked as the syllabus was designed both to develop receptive and productive language skills and was topic based.

The issue of weighting the skills in the final mark was the task the solution for which caused some debate. We had to decide which skill should be given most weight in percentage in the total. Finally, the conclusion was made that all skills as well as proficiency in class throughout the course should be given equal weight, each of them having one-fifth in the total.

Examination Structure

The examination comprises written and oral parts that include four Tests checking both receptive and productive language skills, i.e. an integrated Reading and Writing Test, and Legal Terms In Use Test in the written part as well as Listening Test and Speaking Test in the oral part of the examination (see the table below).

Table 1. Written and Oral Tests

<i>Written Part, percentage in the total</i>	<i>Oral Part, percentage in the total</i>
Reading Test, 20% Writing Test, 20%	Listening, 20%
Legal Terms in Use, 20%	Speaking, 20%

As each test makes an equal one fifth of the final examination grade, every student gets four marks for final formal Tests and a mark for performance per three terms. The final grade is averaged.

Reading and Writing Test

The test checks reading comprehension and writing skills in an integrated form, i.e. students are asked to read a highly specific text of about 1,500 print signs and to write a summary. The test format of writing a summary has been chosen as it is considered to be a very important academic skill and a target need.

As all study-and work-related writing is written for a purpose and about a specific matter, so one characteristic of ESP writing tests, which was mentioned by Dudley-Evans, is the provision of input and specification of purpose and audience (Dudley-Evans, 1998). We have adopted this approach, which in fact means that a very important academic skill of summary writing is tested on the basis of the input the learners received throughout the course for specific purposes reflected in the syllabus.

The text types chosen for this paper meet the learners' target needs, thus students reading law are given authentic legal texts while students from the Faculty of Public Management have to read either authentic texts highly specific in public management or public administration. The timing for reading the text and writing a summary is 45 minutes.

Scoring is based on holistically described levels to which clearly recognizable and appropriate descriptions of performance criteria are assigned. We have worked out a scheme in which each of the components/criterion is given equal weight (see the Table 2 below). Each written paper is checked and rated according to the scale for each of the following: organization, linguistic accuracy, use of legal terms, comprehension, and length. Such scoring proved to be valid and reliable, however it should be stressed that every scale is developed for a particular purpose and should not be expected to work well in a significantly different situation without modification.

Legal Terms in Used Test

It aims at measuring students' proficiency in using legal terms from different highly specific topic areas that had been covered by the syllabus and studied during the course. While writing examination items it was essential that the techniques used should not interfere with the use of ESP terms, i.e. items should not check background knowledge in law or that the context should not add an additionally difficult task on top of the task itself. The text types selected for this examination paper include different authentic sources on law, public management and public administration.

Test formats chosen for this type of test are Multiple-choice and Gap filling (with the terms provided in a word bank where the number of terms is bigger than the number of gaps). The total number of items is 30 (20 Multiple Choice and 10 Gap Filling items). The timing for this test is 50 minutes.

The assessment is objective as the number of correct answers is taken into consideration and the mark is calculated according to the equation:

$$\text{Mark} = \frac{30 \text{ correct answers} - 10 \text{ No of correct answers} \times x}{26 - x} \quad \text{eg } 30 - 10 \times \frac{26}{30} = 8,6 = 9$$

During the examination the testers are presented with the Legal Terms in Use Assessment sheet including the final calculations (see Table 4 in Appendix).

Listening Test

Because listening is a receptive skill, the testing of listening parallels in most ways the testing of reading. The aim of item writing, therefore, must be to measure listening comprehension and to permit highly reliable scoring. However, we have to admit that the act of listening does not always demonstrate its successful performance. In developing the items for this test we follow A. Hughes who outlines that successful listening is possible only if the items involve minimum writing (Hughes, 2000). We decided to test one of the most important macro-skills that is directly related to the learners' needs and to the course objectives, i.e. listening for specific information.

The text types for this paper might be specified as a monologue, a dialogue or multi-participant. The topics are different highly specific in law and are samples of authentic speech. Possible sources are radio broadcasts and

different teaching materials. At the outset students are given approximately 2-3 minutes to familiarize themselves with the items. The length of the recording is 3-5 minutes and it is usually played twice so that the students could check their answers.

The test formats chosen for this paper may be either Multiple Choice, Sentence Completion (involving no writing as it asks students to choose the missing word from the provided word bank) or Ticking the facts discussed in the taped text. The total number of items is 10.

It should also be stressed that scoring a test of a receptive skill no deduction of points for errors of grammar or spelling is made provided that it is clear that the correct response was intended (Hughes, 2000). Thus, the assessment is objective, i.e. based on the number of correct answers and is highly reliable because the test involves minimal writing.

However, we have to admit that listening test is the test that reveals a discrepancy in proficiency assessment and learners' ability of performance as it sometimes happens that students usually performing very well in class do not perform correspondingly well during the listening examination test. This suggests that either more listening texts and a greater variety of tasks or testing extended listening including a passage of 10 minutes and longer that will measure the learners' ability to follow an academic lecture should be introduced.

Speaking Test

Speaking tests are actually assessing spoken interaction, that is, both speaking and listening, as they are interactive. In public examination it is usual to have two or three phases: an introductory one to set students at ease and gain a first impression of their ability to handle social niceties, an interaction based on a given stimulus, and lastly some more open discussion (Dudley-Evans, 1998).

At Law University the testers usually start from a very short introductory phase. We consider it is important to put students at ease at the starting point by being pleasant and by being reassuring throughout the examination because individual oral tests are particularly stressful for students. The second stage of interaction is particularly important for the final output of the test. The most obvious test format for the testing of oral interaction is the interview. We have adopted the format of interview in its traditional form, that is, between the two testers and the student. However, it has at least one potentially serious drawback. The relationship between the tester and the student is usually such that the student speaks as to a superior and is not always willing to take initiative. As a result, only one style of speech may be elicited and many other functions (such as asking for information) may be not represented in the student's performance. In order to avoid this problem the testers should introduce a variety of techniques into the interview situation (Hughes, 2000). Thus, our testers use different test formats – students are asked to compare different legal systems, to explain similarities and differences in terminology, to give their opinion on most presently debated issues in law, to give definitions, to role

play a situation, etc. Students have to interact with two testers – his teacher and the teacher whom the student does not know – in a wide range of specified content. The timing for the oral test is about 20-30 minutes for testing a student as A. Hughes suggests it is unlikely that much reliable information can be obtained in less than about 15 minutes and points that 30 minutes can probably provide all the information necessary for most purposes. However, if it becomes apparent that a student is clearly very weak and has no chance of reaching the criterion level, then an interview is brought to an end quicker.

The assessment is based on descriptions of criteria levels of performance that are attributed to marks (see Table 3 in Appendix). However, there will be a danger of a high level of subjectivity in scoring if speaking test is carried out by a single-scorer. That is why we always involve two testers who can independently assess a student and agree upon the final mark.

Conclusions

The final examination in foreign languages at Law University has been designed having introduced innovative approaches of testing foreign languages at tertiary level. Firstly, the examination has been made uniform for all languages – English, German and French – taught in this higher educational institution. It has a clear advantage in that students reading Law, Public Management and Public Administration face the same language requirements and take a uniform final examination at the end of their

language course. The examination is designed on the principles of **final achievement test based on the objectives of the course covered by the students and following the standards of modular test**, i.e. it comprises four examination tests measuring students' performance in all macro-skills on the basis of their target needs and the input the students received throughout the course. Students' performance during the course is continuously assessed and is given an equal weight in the final mark. The assessment is aimed at reliability, validity and accuracy.

We recommend to employ an innovative uniform modular ESP final examination for higher educational institutions training specialists in law and related areas aiming at standardisation of language requirements for target situations in their future profession.

References

1. Alderson, J. Ch., Clapham, C., Wall, D. (1999). *Language Test Construction and Evaluation*. Cambridge University Press.
2. Bachman, L. F., Palmer, A. S. (2000). *Language Testing in Practice*. Oxford University Press.
3. Dudley-Evans, T. (1998). *Developments in English for Specific Purposes*. Cambridge University Press.
4. Harris, M., McCann, P. (1998). *Assessment*. Macmillan Heinemann.
5. Hughes, A. (2000). *Testing for Language Teachers*. Cambridge University Press.
6. McNamara, T. (2000). *Language Testing*. Oxford University Press.

Irena Darginavičienė, Nijolė Burkšaitienė

Specialiosios anglų kalbos baigiamųjų universitetinių testų aspektai

Santrauka

Aukštosios mokyklos baigiamojo užsienio kalbos egzamino testo sudarymas yra nepaprastai svarbus ir atsakingas uždavinys, kadangi tik tinkamai parengtas baigiamojo egzamino testas, kuriuo testuojami visi kalbiniai įgūdžiai (skaitymas, klausymas, kalbėjimas ir rašymas) atskleidžia studentų kalbos mokėjimo lygį ir sudaro sąlygas objektyviam jo žinių įvertinimui pažymiu.

Baigiamasis anglų kalbos egzaminas Lietuvos teisės universitete parengtas vadovaujantis tokiais principais:

- 1) tarpusavio ryšys tarp testo formos ir testo tikslų,
- 2) tarpusavio ryšys tarp testo turinio ir išeito kurso medžiagos,
- 3) modulinio testo kriterijų panaudojimas.

Straipsnyje pateikiamas baigiamojo anglų kalbos egzamino modelis, taikomas Lietuvos teisės universitete, analizuojami egzamino sudarymo kriterijai. Pažymima, kad baigiamojo užsienio kalbos egzamino testo tipas, struktūra, užduočių tipai ir vertinimo principai yra bendri visoms universitete dėstomoms užsienio kalboms (anglų k., vokiečių k., prancūzų k.). Straipsnio pabaigoje pateikiamos santraukos ir kalbėjimo įgūdžių vertinimo bei teisės terminų vartojimo įvertinimo lentelės.

Straipsnis įteiktas 2002 11
Parengtas spaudai 2003 06

The Authors

Irena Darginavičienė, dr.assoc.prof. at Lithuanian University of Law, Lithuania

Research interests: ESP, Law English.

Address: Lithuanian University of Law, Ateities str. 20, LT-2000 Vilnius, Lithuania.

E-mail: idargin@hotmail.com

Nijolė Burkšaitienė, lector at at Lithuanian University of Law, Lithuania.

Research interests: ESP, Law English.

Address: Lithuanian University of Law, Ateities str. 20, LT-2000 Vilnius, Lithuania.

E-mail: ukk@ltu.lt

APPENDIX

Table 2. Summary Marking Scale

Mark	Criteria Description
10	Complete fulfillment of the task. The main idea/s revealed in a well-organised form. Ideas clearly expressed in one's own words, not copied word by word from the text. Perfect use of key legal vocabulary and lexical structures. No grammar mistakes. Appropriate length.
9	Very good fulfilment of the task. Key idea/s clearly expressed and logically sequenced. Nearly all ideas expressed in one's own words. The text coherent. Simple lexical structures prevail with attempt at more complicated structures. Some key legal words omitted. Occasional grammar mistakes. Appropriate length.
8-7	Key idea/s expressed, but not in the appropriate form. Some sentences are copied word-by-word from the text; the text lacks coherence. Occasionally General English used instead of legal specialism. Some grammar mistakes. Too long/too short in length.
6-5	Main idea/s partly revealed. The form not well organised. Too many sentences/all sentences taken directly from the text. Legal vocabulary and structures poor. Frequent grammar mistakes. Too long/too short in length.
4	Main idea/s not revealed/misunderstood. The text presented not in the appropriate form. Sentences taken directly from the text. Legal specialism not appropriately used/was not used at all/misused. Numerous grammar mistakes. Too long/too short in length.

Table 3. Speaking Assessment: Band Descriptors

Mark	Vocabulary	Grammatical accuracy	Pronunciation	Communication	Fluency
10-9	Excellent to very good: A wide range of appropriate terms, idioms and synonyms	Excellent to very good: The standard of grammatical accuracy is rather high with basic grammar structures, word order in both simple and complex sentences	Excellent to good: Pronunciation is generally accurate, occasional stress mistakes can occur. Intonation corresponds to the meaning of the sentence	Excellent to very good: Communication is effective, free. Adequate response to questions, ability to present one's opinion and to contradict	Excellent to very good: Appropriate speed, rhythm normal, no pauses
8	Adequate vocabulary. Good command of terminology. Lack of complex terms and synonyms	Good: Occasional grammar or word order mistakes can occur in complex structures	Good: Occasional mispronunciation can occur; accent of native language is felt	Good: it may be occasionally necessary to repeat the question or to slow down while speaking	Good: good speed though may experience some difficulties with usage of complex structures and terms
7-6	Fair to satisfactory: amount of appropriate terms though possible difficulties mixing complex concepts and paraphrasing terms using general English instead of terminology	Fair to satisfactory: Grammar mistakes influence the meaning, lack of complex structures	Fair to satisfactory: Mispronunciation may influence understanding of the content of a phrase or a sentence	Fair to satisfactory: Can understand a speaker if s/he is speaking slowly, repeating phrases and sentences	Fair to satisfactory: Lack of fluency, long pauses, slow speed
5	Poor: Poor vocabulary even in general English. Lack of terminology, idioms and synonyms	Serious errors in basic structures, word order mistakes are frequent. Sentences are short.	Frequent errors in pronunciation and stress; intonation patterns typical of the native language	Poor: communication is limited due to poor vocabulary and terminology	Very slow speed, lack of fluency due to numerous pronunciation, stress and grammar errors
4	Very poor vocabulary: Lack of appropriate terms. Response of very limited relevance	Very poor: Gross distortion of basic structures	Very poor: Frequent mispronunciation and native accent make basic communication difficult	Very limited: Can understand the speaker with great difficulty. Needs translation into the native language	Very poor: Speech very disconnected.

Table 4. Legal Terms in Use Assessment Sheet

Mark	Oral formulation of the mark	Number of required correct answers
10	excellent	30-28.5
9	very good	28-26
8	good	25-23
7	good enough	22-20
6	satisfactory	19-17
5	poor	16-15
4	inadequate	14-less