

Alenka Tratnik

# Key Issues in Testing English for Specific Purposes

## Abstract

The article discusses some key issues surrounding testing of English for Specific Purposes (ESP), a wide-ranging field which has now been established as an integral part of English Language Testing, and perceived as an independent activity within the ESP process. The first chapter provides a short general overview of the nature of ESP testing field, bringing into focus that ESP testing has been concerned with measuring specific uses of English language among identified groups of test takers. Next, some defining characteristics and principles common to all ESP tests are explored. The factors of authenticity and washback effect in ESP testing are examined in greater detail. Next, the importance of quality assessment is stressed, and ways of achieving it are examined. The chapter on the English Vocational Matura outlines the format of this ESP public examination in Slovenia, examining the tasks of reading and writing in at greater length.

**Keywords:** English for Specific Purposes (ESP), language testing, authenticity, washback effect, Vocational Matura.

## 1. Introduction

From the early 1960s English for Specific Purposes (ESP) has grown to become one of the most prominent areas of English foreign language teaching. This development is reflected in an increasing number of publications, conferences and journals dedicated to ESP discussions. Similarly, more traditional general English courses gave place to courses aimed at specific areas, for example English for Business Purposes. In addition to the emergence of ESP a strong need for testing of specific

groups of learners was created. As a result, ESP testing movement has shown a slow but definite growth over the past few years. In this article it is my intention to discuss some key issues surrounding the field of ESP testing. I will begin the discussion by addressing the nature and characteristics of ESP testing, underline the principles of good testing practice, and offer examples of reading and writing tasks used in the English Vocational Matura in Slovenia.

## 2. Testing of English for Specific Purposes

This section presents three views of ESP testing. Firstly, the field of ESP testing has been seen as a separate and distinctive part of a more general movement of English language testing, focusing on measuring specific uses of English language among identified groups of people, such as doctors, nurses, lawyers, civil engineers, tour guides, air traffic controllers, and others.

Secondly, ESP testing has been viewed in the broader context of the teaching and learning process. From the perspective of Dudley-Evans and St. John (1998) assessment does not stand alone, but occupies a prominent place in the ESP process, giving an ESP teacher a wealth of information on the effectiveness and quality of learning and teaching. As shown in Figure 1 assessment interacts with needs analysis, and is dependent on course (and syllabus) design.

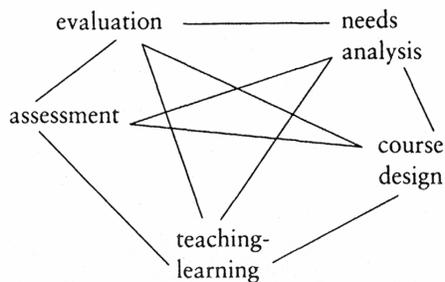


Figure 1: Stages in the ESP process (Dudley-Evans and St. John 1998: 121)

Thirdly, tests enhance the learning process and act as a learning device. Put in the words of Dudley-Evans and St. John (1998: 212) an ESP test is »*an aid to learning*«. Moreover, assessment evaluates the benefits of learning, tests can give learners a sense of accomplishment and a feeling that the teacher's evaluation matches what skills and knowledge have been covered. Along the same lines, Richards and Renandya (2002) suggest that assessment yields an observed judgement of the effectiveness of teaching. Dudley-Evans and St. John (1998: 210) add that assessment »*encompasses benefits such as reinforcement, confidence building, involvement and building on strengths*«.

### 3. Characteristics of ESP tests

Specific purpose testing is primarily concerned with facilitating learners to perform particular communicative tasks, providing feedback on learning, confirming what students have mastered and highlighting those skills needing further attention, encouraging learning, and monitoring progress.

In principle, any ESP test can be classified as a performance test assessing the skills needed to »perform« in the language successfully. Given that language performance in individuals varies according to the context in which the language is used, then the test taker's performance on a test depends largely on the interaction between language knowledge and specific purpose content knowledge. Unlike general English tests, which can be used with young learners (e.g. Cambridge Young Learners Test) or adults (e.g. First Certificate in English), ESP tests are more likely to be used with adults or secondary school learners at intermediate and advanced level, or learners who have already acquired basic knowledge of the language system.

Douglas (2000: 10) cites that ESP tests are »*contrived language use events*« in which, ideally, the test taker's specific purpose language ability and knowledge of the specialist field are measured. If we keep in mind that language ability refers to what a learner can do in or with a language, then what really matters in ESP testing is whether learners can communicate in a specific target language and use knowledge of the field in order to achieve their aims, in order to understand and be understood, in order to get their message across in English.

ESP tests are related in content, themes and topics to particular disciplines, and involve a higher degree of language specificity. Special lexical, semantic and syntactic characteristics of technical language, in addition to its communicative function enable people in a particular academic, professional or vocational field to convey the meaning more specifically. Precision in language is therefore a unique concept in specific purpose language testing.

Not only that, ESP testing enterprise is designed upon the demands of the linguistic characteristics of the specialist area of work or study, as language varies from one situation to another. There are important differences between, say, an English test of engineers and that of tourist guides, as they address the nature of particular varieties of English from a specific vocational and professional field.

Further, one of the prevailing principles of ESP testing is that tests should contain tasks that mirror faithfully those of the candidates' target language use situation. This leads us to a view of ESP testing in which tests are developed on the basis of a detailed analysis of characteristics of context and tasks in target language use situation. The target language profile for assessing a specific group of learners would thus entail samples of communication activities, communication purposes, linguistic features, functions, descriptions of content areas, language skills, etc. For instance, if we were to devise an English test for tourist guides, we would first describe a range

of typical situations in which tourist guides work, typical uses and characteristics of English language that they most often find themselves in.

In that case, if an awareness of the learners' needs shapes the design of an ESP course and teaching materials, then the analysis of the learner's target language use situations is the first and perhaps the most important stage in designing an ESP test. A thorough coverage of typical real-life communicative events would, firstly, result in the selection and creation of suitable test tasks, secondly, facilitate a more integrated and thematically linked assessment, in which test tasks authentically characterize the reality of the learners, and, not lastly, such tests would have higher validity (Douglas 2000; Hutchinson and Waters 1987; Munby 1978).

Accordingly, ESP tests are more concerned to present learners with tasks that involve them in reading, listening to, speaking or writing the target language, and evaluating how well they can do this. Of course, an important component in assessing how well somebody can use English is how accurately they can produce or understand texts written or spoken in the language, but the key to this assessment is to present learners with tasks that resemble in some way the sort of things they may have to do with the language in real life. Therefore, the ESP approach in testing is based on the analysis of learners' target language use situations and specialist knowledge of using English for real communication.

In view of what has been said it is clear that an ESP test is one in which the purpose, the test content and the methods are more narrowly defined. Consider for instance tests in English for Civil Engineers or Military Service devised for the purpose of evaluating occupational language abilities. If we accept the existence of specific purpose tests as described here, the next question might be: How many ESP tests are there? Namely, specific purposes can be wide or narrow, vocational or academic, including anything from English for Airport Controllers, Chemists or Hoteliers to Legal English and beyond. This diversity in itself poses a question: Is ESP testing truly possible between such widely differing fields or does each field of human activity have an ESP test of its own?

To my mind, no human activity and no specific purpose field can simply be classified without overlaps. Thus, the line where one ESP test begins and where another ESP test ends is very vague indeed. A test that is constructed for chemistry engineering and a test devised for assessing lawyers, although belonging to various fields, tend to share a number of features. What is truly fascinating about the practice and testing of ESP is that the vast repertoire of specific tests offered is also underpinned by a common structure of testing principles and techniques. In this sense, then, ESP testing provides a framework relevant across different disciplines.

## 4. Assuring the quality of ESP tests

When devising an ESP test one should aim at creating good and dependable measures of language ability which need to:

- be as authentic as possible,
- provide accurate and reliable measures of language ability,
- have beneficial effects,
- be practical and economical in terms of administration, time, money and personnel.

Most authors (Alderson et al. 1995; Bachman and Palmer 1996; Douglas 2000; Dudley-Evans and St. John 1998) identify authenticity of task, reliability, validity, practicality and economy as the most important factors affecting the quality of a test in a positive or negative way. A detailed account of the above-mentioned factors would go beyond this paper, therefore, I shall briefly define the element of authenticity as one of the salient features of ESP tests.

In case of authenticity, test takers should be engaged in a variety of genuine tasks in which language ability and field specific content knowledge interact in a way which is similar to the student's real-life language use. This interaction between language knowledge and specific purpose content knowledge makes an integral part of the whole ESP testing concept. When authentic language and material are discussed, we usually think of language used in non-test or non-pedagogic texts, language referring to real-life situations and natural communication, and texts close to the readers' needs and language use.

Accordingly, in the ESP testing situation authenticity of task refers the similarity of the test task to the target language use situation, the more authentic a text or task is, the more probable it is for the test taker to carry out the test task in the same way, as he/she would perform it in the actual target situation. Thus, authentic texts and tasks are representative of the specific language use situation or similar to the tasks a test taker actually performs in his/her work.

Another important factor in ESP testing is *backwash* or *washback effect*. The notion of backwash is that of impact, which seeks to investigate the relationship between test use and the ESP situation in which it is used (Hughes 1989). Put simply, it is the effect of testing on classroom instruction, on »*what is taught and how it is taught*« (Dudley-Evans and St. John 1998: 214). Tests, especially those that are important for the test takers, may generate positive or negative backwash.

The research that has been conducted in this area (Karim 2002) indicates that the relationship between testing and teaching is very complex and it is misleading to claim that good tests will automatically have beneficial effects on classroom instruction. Other variables such as teacher competence, motivation and innovation, the climate of the school, socioeconomic status of pupils and teachers combine to exert an equally important influence on what goes on in an ESP classroom.

However, if a test is regarded as important, then preparation for it can dictate teaching and learning activities. More importantly, a public examination might be a signal to students (and teachers) about what is important and what is not. Consider for instance the English *Vocational Matura* in Slovenia, which has been administered for six years now, and does not include a listening section – a skill that encompasses a quarter of the English language syllabus. If test developers responsible for constructing the English *Vocational Matura* keep ignoring listening year after year, they are somehow telling teachers (and students) that this skill is less important and that they can spend less time on it.

We have looked at some cases of harmful backwash. However, backwash need not always be harmful, it can be positive too. An ESP test, which would be based directly on an analysis of English language needs of a specific group of learners as similar as possible to those which they would have to perform in real life, would probably be a more plausible case of beneficial backwash than a case in which no such analysis would be carried out.

In short, what matters in ESP testing is whether learners can communicate in the field specific target language and use knowledge of the field in order to achieve their aims, in order to understand and be understood, in order to get their message across in English. The Common European Framework of References for Languages, devised by the Council of Europe, and increasingly becoming the standard for language curricula, textbooks, and other teaching materials, language examinations and assessment procedures, has similar aims.

## 5. The English Vocational Matura in Slovenia

The *Vocational Matura* in Slovenia is a public examination, in which candidates demonstrate the achievement of standards of knowledge determined by objectives of educational programmes of secondary technical schools, vocational-technical schools and other vocational education. It is a final task-based achievement exam. If candidates pass it, they acquire a secondary vocational education, by which they prove the ability to work in a certain technical area and obtain a qualification for employment directly after completing their education.

An essential feature of ESP formal testing is related to the specific test content. Green (2004) reminds us that background knowledge has an effect on test results. The fact that the candidates sitting the English *Vocational Matura* come from wide-ranging backgrounds poses a question: How do we specify the domain of content if we want to test language ability of students coming from different secondary vocational schools? Alderson (cited in Dudley-Evans and St. John 1998: 216) suggests that one-test solution seems to be the most convenient. Having looked at live English *Vocational Matura* exams, I can assert that the national subject

commission constructing the English *Vocational Matura* normally selects carrier content that is as neutral and comprehensible as possible.

In its six-year long history, the format of the English *Vocational Matura* has not changed considerably. The examination comprises four components: reading, use of language, writing and speaking. I shall examine the tasks of reading and writing more closely to see how they illustrate various features of ESP testing practice.

## 5.1. Assessing reading

In the reading section of the examination there are two independent reading tasks, each based on a different reading passage of up to 650 words. 30 minutes are provided to complete both reading tasks, however, no suggestion is given on how long to spend on each task. We will look at a reading passage and a True/False test method from the *Vocational Matura* live exam administered in June 2006. The instructions for the second reading task run as follows:

Read the text <i>Experts say fairy tales not so happy ever after</i> and decide whether the statements below are true (T) or false (F). Put a tick (✓) in the appropriate box after each statement. If you correct your answers, the corrections should be as clear as possible, or else the answer will be invalid.			
		T	F
1	Ms Baker-Sperry used to go to Purdue University.		
2	In our society women should depend on their intelligence.		
3	The researchers analyzed forty-three fairy tales.		
4	'Little Red Riding Hood' is the only tale which has been reproduced more than 100 times.		
5	Fairy tales suggest women should be and stay beautiful.		
6	One of the stories mentions male looks 114 times.		
7	Fairy tales have a negative influence on young women.		
(7 points)			

Figure 2: Vocational Matura Reading task (2006)

The rubric gives the procedures for responding, along with a general reminder of the need for clear corrigenda. Test takers are told how to correct their responses and are given a reason why it is important to write their answers clearly. On the other hand, the rubric does not contain information about the time allotment, there is no statement about what abilities the task is meant to assess, or any evaluation criteria, apart from the number of points each answer is worth.

The input data for questions 1–7 is a fairly authentic text about fairy tales, as reproduced below.

**Experts say fairy tales not so happy ever after**  
 By Amy Patterson-Neubert  
 11 November 2003

WEST LAFAYETTE, Ind. – Fairy tales about pretty princesses winning Prince Charming are more harmful than parents might think, says a Purdue University sociologist. Liz Grauerholz, an associate professor of sociology, teamed with Lori Baker-Sperry, an assistant professor of women's studies at Western Illinois University and a former Purdue graduate student, to study how beauty is written about in fairy tales and whether stories with beautiful princesses are more likely to be popular. Grauerholz says parents need to be aware that some stories tell children that unattractive people are more likely to be evil and reinforce traditional gender roles that may be confusing for today's young women.

"Fairy tales, which are still read by millions of American children, say it pays to be pretty," Grauerholz says. "It's important to understand the messages our children receive about traditional gender roles, especially during a time when women are encouraged to be independent and rely on their brains rather than beauty. Grauerholz and Baker-Sperry examined 168 Brothers Grimm fairy tales. These stories were written by Jacob and Wilhelm Grimm in the 1800s and were used in central Europe to teach children the roles boys and girls should play, as well as what it means to be good or bad. Of the tales analyzed, 43 percent have been reproduced in children's books or movies.

The five tales that have been reproduced more than 101 times are "Cinderella," "Snow White," "Briar Rose" (also known as "Sleeping Beauty"), "Little Red Cap" (also known as "Little Red Riding Hood") and "Hansel and Gretel." The researchers found that the majority of fairy tales that survived into the 20th century feature characters with young, beautiful princesses. This trend reinforces the message to children that physical attractiveness is an important asset women should aim to achieve and maintain, Baker-Sperry says.

Their analysis showed that 94 percent of the Grimms' fairy tales acknowledged physical appearance, and the average references per story were 13.6. In one story, there were 114 beauty references for women. In comparison, the number of beauty references for men did not exceed 35 per story. "Hearing these messages that were created by an old, patriarchal society may cause women, especially young girls, to withdraw from activities or careers, such as competitive sports or hard labor, because it is not part of being feminine," Grauerholz says. "This continued emphasis on beauty is a way society controls girls and women. Women adopt behaviors that reflect and reinforce their relative powerlessness, which can lead to limiting a woman's personal freedom, power and control."

Article from <http://news.uns.purdue.edu/UNS/html4ever/031111.Grauerholz.taales.html>

Figure 3: Vocational Matura Reading Task (2006)

The sole focus of the reading section is to demonstrate comprehension. As mentioned before there is no field specific purpose or even measurement purpose, apart from the testing purposes, provided for reading this text. A statement of purpose would enhance interactional authenticity of the input data and match an already high level of situational authenticity. The input requires a moderate engagement of background knowledge about Brothers Grimm. The relationship between the input and the response is very direct, in that the information needed to complete the task is included in the text, although as suggested above, a certain amount of background knowledge is needed.

The task appears to be quite close to the general end of the specificity continuum. The text itself though is not intrinsically interesting or challenging, the obvious

weakness of this type of test method is that candidates have a 50% chance of guessing the right answer, which clearly affects the validity of this test method. One way of avoiding guessing is to add another distracter, in this case, »not included in the text« category would reduce the likelihood of guessing to 33%. The second option of reducing guessing is to modify the task by requiring test takers to give a reason for their choice. However, this extra requirement can be problematic, as the candidates have to write the solutions and writing as a skill is not intended to be tested here.

## 5.2. Assessing writing

The writing section contains two tasks, one requiring the test takers to write a short composition for a school newspaper and the other to write a formal or semi-formal letter. Since the second of these tasks consists of a rather straightforward prompt for a 40-minute letter, we will look at the first, somewhat more interesting task. A typical writing task includes the following:

**A short composition (Allotted time: 20 minutes) (10)**

Write a short composition (70 words) for a school newspaper in which you discuss problems in urban communities. Choose to discuss only one of the problems indicated below, examine its harmful effects, and offer at least one piece of advice on how to solve the problem.



The image contains three illustrations. On the left, two people are shown from behind, one holding a glass and the other a bottle, with the word 'alcohol' written below. On the right, a syringe is shown with the word 'drugs' written below. In the center, a person is shown from behind, holding a red circle with a diagonal slash over the word 'GRAFFITI', with the word 'vandalism' written below.

Figure 4: Vocational Matura Writing Task (2006)

We see that time allotment and information about length are specified in the rubric. There is specific information about what to include in the article. Besides, the rubric (in Slovene) includes the content of the task (urban problems), the audience of the writing task (readers of a school newspaper), and the genre (article). Test takers are provided with about 20 lines to write the composition, there is no space for the draft

of the article. There is no information about what the objective of the writing task might be. Further, no information, apart from the total number of points, in this case 10, is given about the evaluation criteria. In the *Vocational Matura* catalogue we are told that candidates are assessed in content, vocabulary and cohesion. This information is addressed to teachers so that they might prepare candidates for the examination, and to students taking English as their third elective subject. The same document informs task takers that dictionaries may be used.

The visual input data does not add much to the authenticity of the task but simply supports the rubric. The task is non-reciprocal, although candidates are given some choice as to what problem to describe. The language required in the response does not require a wide range of vocabulary and syntactic forms; cohesion and organization are more important parts of the construct. In terms of background knowledge necessary for completing this task, the test taker might need some prior knowledge of health or medical problems (harmful effect of drugs, alcohol on people). One must wonder, however, whether the topic might be too sensitive for a teenage candidate.

Overall, writing an article for a school newspaper is a moderately authentic response task, although, as suggested above, information about purpose and some information on what aspects of writing will be evaluated would make the task potentially more interactionally authentic.

## 6. Conclusion

Developing foreign language skills enhances personal and professional fulfilment, and reflects the process of lifelong learning. Therefore, special attention in ESP teaching should be focused on ways of promoting skills, which encourage and enable learners to respond to different day-to-day situations and react properly in special circumstances. Similarly, ESP tests should be based on direct evaluation of language ability in acts of communication (reading, speaking, writing, and listening) and assessment of content knowledge in situations and activities, which are as engaging and realistic as possible for learners. Seen from this point, teachers should devise such test tasks, which enable the learner to see the connection between language uses required in tests (which are mostly simulations) and real-life language uses.

The primary goal of ESP tests is to obtain information about the learner's specific purpose language ability. This information is often very useful and at times even necessary. Modern Europe encourages mobility of labour and of students across the European Union and beyond. In order to be able to take up study places or work opportunities, knowledge of foreign language is essential. In present day, it is increasingly important not only to be able to use a foreign language, but also to be able to demonstrate that one can use it at the level required by employers, schools,

or universities. Considering this, ESP examinations are gaining more and more attention, and are becoming an indispensable tool in the modern educational society.

## Bibliography

Alderson, J. C., Clapham, C., and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Bachman, L. F., and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Čebren, N. (ed.) (1998). *New Challenger in Teaching ESP in Slovenia – conference proceedings*. Ljubljana: Društvo za tuje jezike in književnosti.

Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

Dudley-Evans, T., and St. John, M. J. (1998). *Developments in English for Specific Purposes*. Cambridge: Cambridge University Press.

Green, R. (2004). *Language testing*. Seminar material. Bled, 2–4 March.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Hutchinson, T., and Waters, A. (1987). *English for Specific Purposes: A learning-centred Approach*. Cambridge: Cambridge University Press.

Karim, M. A. (2002). *High Stakes Examination: Washback in the Language Classroom*. Paper presented at the 28<sup>th</sup> conference of the International Association for Educational Assessment Reforming Educational Assessment to Meet Changing Needs. Hong Kong, China.

Munby, J. (1987). *Communicative Syllabus Design*. Cambridge: Cambridge University Press.

Predmetni izpitni katalog za poklicno maturo – angleščina (2005). Ljubljana: Državni izpitni center.