

DEVELOPMENTS IN ESP TESTING PRACTICE

Radadiana CALCIU*
Liliana KIRIȚESCU**

Introduction

In this article we aim to present some developments in the testing of English for Specific Purposes (ESP) in Romania after 1990, using data which were collected from teachers in higher education institutions in several university cities all over Romania.

Before 1990 there was a lack of a coherent teacher training system in Romania, particularly in the area of teaching ESP at tertiary level. The PROSPER project [1], initiated in 1990, offered training opportunities in almost all areas of language teaching, among which testing played an important role. ESP teachers from different universities all over the country attended shorter or longer testing courses, including a distance learning module in Assessment in Language Learning with Manchester University, carried out over a period of one year, which proved to be the most effective in this sense.

As a result of the skills developed through these training courses, ESP teachers have become more aware of the different aspects involved in test design and administration, and have changed their ways of assessing the students' performance. These changes were recorded and analysed for the Impact Study of the PROSPER Project, whose results were published in 1999 [2, pp. 167-188]. The investigation was based on a comparative analysis of a sample of PROSPER tests (i.e. designed and/or administered by teachers trained within the PROSPER project), PRE-PROSPER tests (i.e. designed and/or administered before 1990), and NON-PROSPER ones (i.e. designed and/or administered in institutions which were not included in the PROSPER project) [2, pp. 167-170].

We must give full credit to the contribution of our colleagues Doina Comanetchi and Sorin Baciuc from the 'Politehnica' University of Bucharest, with whom we collaborated in carrying out the data collection and

interpretation for this chapter of the Impact Study. In what follows we will present the main findings of the study concerning the changes in ESP testing practice, which supported and complemented the general upgrading of ESP teaching in Romanian universities. We shall therefore focus on the data provided by the comparison of PROSPER and PRE-PROSPER tests, which illustrates most clearly the developments that have occurred in this area of ESP teaching due to PROSPER.

The PRE-PROSPER test profile

With a view to offering a clear picture of the starting point, namely the kind of ESP tests which used to be administered to students in the early 90s and before, we have outlined a PRE-PROSPER test profile, based on a number of sample tests as well as on focus group discussions and interviews carried out in several PROSPER institutions. The main characteristics we have identified are the following:

- PRE-PROSPER tests were mainly concerned with the testing of grammar and vocabulary either through specific test items or via writing essays. The range of test methods used was rather limited, the most popular being translation both for grammar and vocabulary.
- The testing of skills was neglected as a direct consequence of the absence of communicative methodology and materials from the English class. Particular interest could be noticed in the testing of specialist knowledge, rather than in the students' ability to use that knowledge when communicating in English in lifelike situations. Consequently, the texts and tasks were, with very few exceptions, non-authentic and non-contextualised.
- In spite of the fact that essay writing and translation were widely used, the test authors showed no concern for developing marking criteria that might increase objectivity in these complex areas.

* Lecturer, Department of Germanic Languages and Business Communication, ASE Bucharest

** Lecturer, Department of Germanic Languages and Business Communication, ASE Bucharest

- Test authors were not aware of the different kinds of tests, the only type being achievement, administered as end-of-year or/and end-of-term test. Usually there was no common testing policy or co-operation in test design between teachers in the institutions considered.

Results of the study

The developments in ESP testing have to be considered within the broader view of innovation and change in language teaching in Romania, where both training and new materials (i.e. textbooks) had a strong impact on the teaching methodology and affected testing substantially. The shift in focus on meaning and the communicative function of language have turned the classroom into a springboard for real-life activities, enabling students to improve their language performance through realistic language tasks.

Authenticity of text and task

ESP tests reflect this shift by using predominantly authentic texts from a wide range of sources such as articles in magazines and newspapers, business documents, interviews from the radio, lectures, etc. It is also important to note that the tasks give the students a real-life purpose by using the language in realistic situations and/or situations of professional relevance. This is achieved through contextualised tasks, which provide a context for using the language and a clear task environment, through more elaborate and explicit rubrics which describe the situation, the participants in communication, the roles, the purpose. This change is well illustrated by the situation revealed by studying the test sample (Table 1):

PROSPER	PRE -PROSPER
Authentic: 52	Authentic: 1
Non-authentic: 21	Non-authentic : 23

Table 1: **The use of authentic and non-authentic texts**

The quantitative difference, in favour of PROSPER tests, in relation to the use of authentic texts is significant the ratio being 52:1 To be noted, however, that non-authentic texts are still being used by PROSPER teachers. As far as the task type is concerned, the difference between the two kinds of tests, PROSPER and PRE-PROSPER, is even more relevant, showing a dramatic change, in the sense that while PRE-PROSPER tests had practically no contextualized tasks and extremely few authentic ones, PROSPER tests have a large number of both authentic and contextualized tasks

(Table 2).

PROSPER	PRE -PROSPER
Authentic: 109	Authentic: 3
Contextualised: 94	Contextualised: 0

Table 2: **The use of authentic and contextualised tasks**

Focus on skills testing

As a consequence of the use of communicative materials and of a change in teaching focus from grammar structures to skills, PROSPER tests show a move towards skill testing, with concern for all four skills. This represents the crucial difference from PRE-PROSPER tests, which mainly tested grammar and vocabulary (speaking, in particular, was very rarely, if ever, tested). This does not mean that PROSPER tests have abandoned grammar and vocabulary altogether, but rather that PROSPER tests focus on all four skills, as well as on grammar and vocabulary. This situation is clearly illustrated by the results in Table 3 below:

PROSPER			PRE-PROSPER	
Test focus	Large set	Manchester Graduates	Test focus	
Writing	74.1%	80.9%	Grammar	100%
Reading	67.2%	76.1%	Vocabulary	93.3%
Grammar	58.6%	38.0%	Writing	26.6%
Vocabulary	53.4%	47.6%	Reading	26.6%
Listening	37.9%	57.1%	Listening	0.0%
Speaking	27.5%	28.5%	Speaking	0.0%

Table 3: **Skills and language testing**

It is interesting to note that the different exposure to training can be easily traced in the test sample. Thus, after analysing the sample tests we have come to the conclusion that grammar and vocabulary assessment still prevail with the teachers who have had no or little exposure to training, while Manchester graduates display a much wider range of aspects tested. *Grammar* is still assessed in PROSPER tests, but the percentages indicate a marked decrease in PROSPER compared with PRE-PROSPER, where each and every test included in the data set tested grammar. The PROSPER tendency of replacing the testing of grammar with that of other skills is reinforced by the restricted data set where the percentage for grammar is 38% compared to 58.6% in the large set. Likewise, *vocabulary* testing occupies the second place in frequency in PRE-PROSPER (93.3%) tests, after grammar. With PROSPER tests, reading and writing come first, while grammar and vocabulary rank third and fourth, with lower percentages. This development emphasises the trend towards the testing of skills.

The figures show that with PROSPER tests all four skills are tested, with higher percentages for *writing* and *reading*, which are the top preferences. It should be noted that the figures simply show whether the tests contain a section in the area indicated. They do not indicate section weighting or importance. A possible explanation for the high percentages of grammar and vocabulary tests in relation to those focusing on the four skills is the fact that a considerable part of the PROSPER sample was provided by PROSPER teachers who had only attended a 3-hour Basic Teacher Training Session on testing and who still feel more comfortable with traditional testing techniques. Compared to the overall set, the data for Manchester graduates indicate a notably smaller percentage for grammar testing and a higher one for testing listening skills.

The situation is totally different for PRE-PROSPER tests, with *grammar* and *vocabulary* top of the list followed by *reading* and *writing* scoring a much lower percentage.

As can be seen from the table, a significant change refers to the testing of *speaking* and *listening* which were absent altogether from the tests belonging to the PRE-PROSPER category.

Range of testing methods

Another criterion we used when analysing the two categories of tests (i.e. PROSPER and PRE-PROSPER) was the testing method. We were thus able to determine the fact that the range of testing methods is much wider in the case of PROSPER tests.

A variety of methods are used in PROSPER tests for testing different skills (*information transfer, multiple choice, multiple matching, cloze, note taking, summarising, letter and essay writing, role play, etc.*). With PRE-PROSPER tests the testing methods are overwhelmingly *translation, fill in the gaps with the correct form, sentence transformation*, which assess disembodied language elements, this being due to the fact that emphasis was laid on language usage rather than on language use, on accuracy rather than on fluency.

In order to illustrate this conclusion, let us take a closer look at the situation in the case of reading tests. The only testing methods used for PRE-PROSPER tests were multiple choice and comprehension questions, while for PROSPER tests there is a variety of testing methods, of

which the most frequent are *true/false, matching elements, cloze, chart/table completion*.

The methods for assessing students' **reading** skills show a focus on testing meaning comprehension, not memory. Methods like *paragraph reconstruction from jumbled sentences, structuring text into paragraphs, completing table/chart with relevant information* show that tests focus on processing and selecting information. *Matching* includes a variety of elements: *sentence to picture, heading to paragraph, multiple matching*. The presence of the *cloze* in reading tests indicates that integrated language is tested in PROSPER tests. The relatively big number of testing methods indicates the fact that PROSPER teachers are willing to innovate and try out a variety of methods.

As far as the testing of **writing** is concerned, we could say that this is the skill most frequently assessed by both categories of tests. Although the testing of this skill is widely present in the sample analysed, the range of methods used for testing writing is much wider in the case of PROSPER tests and there are considerable differences between PRE-PROSPER and PROSPER tests.

13 out of the 16 methods in the PROSPER category involve writing whole texts with focus on discourse and register elements. These methods give a context and a purpose for writing and are relevant for professional purposes (*process description, project work, diagram interpretation*), with less emphasis on specialist content and more on communication. In the set of sample tests belonging to Manchester graduates, the range of methods is fairly similar, with *essay* and *letter/memo writing* in top positions, which means focus on testing writing of whole texts where organising information and message is the main marking criterion, as shown in the discussion of marking criteria below.

The PRE-PROSPER sample includes tests which assess knowledge of specialist content (*definitions of specialist terminology, information transfer, answering questions on content*). There is also extensive testing of grammar through *translation, sentence completion, sentence formation from jumbled words, etc.* No PRE-PROSPER test involves writing a professional document; whenever whole texts are produced, it is through *essay writing* with no real purpose (e.g. "Write about Hamlet"). A full list of the methods used in testing writing, indicating the number of occurrences, is given in Table 4 below.

PROSPER	PRE-PROSPER
<ul style="list-style-type: none"> - Essay/Paragraph writing (presenting opinion): 11 (integrated with reading prompts: 1) - (integrated with listening + vocabulary: 2) - Letter writing: 6 - Rewriting with change of register: 3 - Information transfer: 3 - Translation: 3 - Text completion (clozed letter): 3 - Sentence formation: 2 - Summarising: 2 - Question formation: 2 (testing grammar) - Project work: 2 - Report writing: 1 - Writing a dialogue: 1 - Writing responses to a given situation: 1 - Arranging and punctuating a letter: 1 - News report on a given topic: 1 - Process description: 1 - Sentence writing based on diagram interpretation: 1 - Definition writing: 1 	<ul style="list-style-type: none"> - Answering questions on/ Giving definitions of specialist terminology: 3 - Sentence writing based on diagram interpretation (similar to information transfer, testing specialist content): 1 - Writing responses to a given situation (dialogue): 1 - Literary essay: 1 - Summary writing: 1 - Translation*: 1 - Sentence completion*: 1 - Sentence formation from jumbled words*: 1 <p>* all testing grammar</p>

Table 4: Test methods for testing writing

A great variety of methods is also used by PROSPER teachers in testing **listening**, a novelty in PROSPER tests, as listening skills were not tested in any of the PRE-PROSPER tests analysed. The test methods used require students to concentrate on gist or specific information, rather than to understand every word. The most frequent method is *table completion*, which requires processing and selecting relevant information, but all other methods used focus on the message and in one case integrate the testing of listening with that of grammar (Table 5).

PROSPER
<ul style="list-style-type: none"> - Chart/Table completion with specific information: 10 - Blank filling with specific information: 4 - True/False: 2 - Project work: 2 - Ordering pictures: 2 - Comprehension questions: 2 - Matching elements/multiple matching: 2 - Identifying mistakes in pictures: 1 - Identifying mistakes in text: 1 - Reporting conversation: 1 (integrated with grammar) - Note-taking: 1 - Multiple choice: 1

Table 5: Test methods for testing listening

As shown above, the sample of PRE-PROSPER tests we analysed did not offer examples of **speaking** tests; nevertheless, the teachers who took part in group discussions mentioned that this skill used to be tested through monologue, by asking the students to speak on a given topic. The data presented below (Table 6) are indicative of the fact that the testing of speaking, besides being a new development which is characteristic of PROSPER tests, also uses a great variety of methods. Although in our sample speaking is the least frequently tested skill, the methodology of testing it is interesting and complex. The range of speaking skills tested is quite wide including *oral presentation skills, debating, asking for and giving information*. Notable is student-student interaction (through *role play, conversation, debate*), where the teacher assesses as an observer, without interrupting the students. Only one test contains *open questions testing specialist knowledge*, a remainder of PRE-PROSPER testing practices. The same features are presented by the restricted set of data (i.e. Manchester graduates), with preference for *role play* and *information gap*.

PROSPER
<ul style="list-style-type: none"> - Oral presentation: 4 - Information gap (asking for/giving information): 4 - Role play (problem solving; negotiation; meeting): 3 - Free discussion on a given topic: 2 (1 integrated with reading) - Project work: 2 - Dialogue construction on a given topic: 1 - For and against discussion: 1 - Discussion (problem solving): 1 - Open questions testing specialist knowledge + language elements: 1

Table 6: Test methods for testing speaking

As far as the testing of **grammar** is concerned, PROSPER tests show a tendency to test it in context, through the use of more elaborate techniques (*cloze, project work*). Unlike PRE-PROSPER tests, where *translation* is used mainly for testing grammar structures, PROSPER tests provide a professional context, which is both realistic and relevant for the learner as translating various office documents can be a task which the students will do in their future professional life (Table 7).

PROSPER (58.6%)	PRE-PROSPER (100%)
<ul style="list-style-type: none"> - Gap filling: 24 - Transformations: 8 - Question formation: 6 - Multiple choice: 6 - Sentence formation: 5 (integrated with listening) - Translation: 5 (in fact it tests grammar + vocabulary + discourse) - Word ordering: 3 - Correcting mistakes: 2 - Cloze: 2 - Sentence completion: 2 - Project work: 2 - Writing the date correctly: 1 - Joining sentences: 1 - Identifying categories: 1 	<ul style="list-style-type: none"> - Translation: 12 - Gap filling: 10 - Transformations: 9 - Sentence formation: 4 - Multiple choice: 3 - Question formation: 2 - Word ordering: 2 - Grammar rules explanation: 1 - Identifying sounds: 1 - Sentence completion with specialist information: 1

Table 7: Test methods for testing grammar

PRE-PROSPER tests focused on grammatical accuracy and metalanguage. Students were asked to explain/provide grammar rules, identify categories and/or to use them in non-authentic, non-contextualised tasks, having no real-life purpose, the rubrics being short and mechanical.

The methods used to test **vocabulary** are listed in Table 8 below:

PROSPER (53.4%)	PRE-PROSPER (93.3%)
<ul style="list-style-type: none"> - Gap filling: 10 (one integrated with reading) - Multiple choice: 5 - Word derivation/family: 5 - Nominal compound formation: 2 - Identifying the meaning of words in context: 2 (integrated with reading) - Translation: 2 - Explaining terminology: 2 - Project work: 2 - Matching words with definitions: 1 - Cloze: 1 - Using words in own sentences: 1 - Brainstorming on specific topic + using in texts: 1 - Labeling diagram: 1 	<ul style="list-style-type: none"> - Translation: 5 - Rephrasing: 3 - Matching synonyms: 2 - Word families: 1 - Providing synonyms: 1 - Word and phrases explanation: 1 - Using words in own sentences based on specialist knowledge: 1 - Gap-filling: 1

Table 8: Test methods for testing vocabulary

Some of the PROSPER test methods, such as *word and*

phrase explanations, word families, word derivation are similar to the ones used in PRE-PROSPER tests, assessing vocabulary in isolation. It should be noted, though, that there is a strong element of vocabulary in context, focusing on language use (*cloze, gap filling, labeling diagrams, identifying the meaning of words in context*).

The vocabulary test items produced by Manchester graduates are sometimes input for other tasks (e.g. brainstorming on a specific topic can result in a list of words to be used in a writing test, or a gapped text is first used for a ‘fill in the blanks’ task, and once completed it is further used in a reading test item).

Discourse elements are tested implicitly by PROSPER teachers. This can be seen from the marking criteria for writing (which include coherence and cohesion, organisation and sequencing of ideas), as well as from the methods testing other skills: listening (*note-taking, reporting conversation*), speaking (*oral presentation, for and against discussions, response to a given situation*) and reading (*text cohesion through sentence insertion in text, text reconstruction from jumbled paragraphs, joining sentences with cohesive elements* etc.). The tests focusing on discourse elements confirm the general trend towards testing integrative language. It is interesting to note that the whole sample of tests which deal with discourse and functions comes from Manchester graduates.

There was no such concern on the part of PRE-PROSPER authors, who tested discrete point items, especially through grammar and vocabulary.

Marking criteria and procedure

The analysis we carried out also focussed on the **marking criteria** used by teachers before and after being trained in the PROSPER Project. The conclusion we have drawn is that the marking criteria for PROSPER tests are more diversified and that there is a concern for reducing subjectivity in marking.

Interviews with groups of teachers were carried out, where the participants were invited to comment freely, to recollect past practices and to present current ones.

The PRE-PROSPER situation is characterised by the lack of clear marking criteria, with accuracy as the main or sometimes only criterion in marking writing. Composition writing criteria associated with organising the message clearly and putting forward an argument were totally disregarded. The lack of criteria also meant that no distinction was made among mistakes as regards their seriousness and the final marking was totally impressionistic being guided by ‘the amount of red on a page’ as one of the teachers in Brasov put it.

As the interviews show, the teachers felt the need for a

more objective system of marking and for a more complex set of criteria for assessing the students' work. Teachers are aware of the need to work on two main issues: one is establishing a wider range of criteria for marking, the second is creating/adopting band descriptors which describe language performance in more detail and can contribute to more accurate marking. Although the assessment of speaking and writing cannot be objective, teachers are aiming towards improved subjectivity. The marking criteria PROSPER teachers mentioned in the group interviews are:

- organisation of information/message, overall text coherence
- appropriacy of vocabulary, register
- text cohesion through the use of appropriate connectors
- communicative competence in a wider sense, not only linguistic competence
- fluency, ability to keep communication going (strategic competence)
- layout of letters and written documents
- creativity, originality, rather than memorisation
- ability to cope with authentic discourse in testing listening (e.g. recognition of function-words/fillers).
- accuracy

Apart from the diversification of marking criteria, PROSPER has also brought about a diversification of the **marking procedures**, with two main procedures mentioned in the group interviews:

- The teacher provides feedback on the most frequent mistakes which occur in a test by discussing them with the group after the test results are given. This procedure can contribute to increasing the students' perception that tests help them to improve rather than penalise them for the mistakes they make.
- The students are sometimes involved in assessing the product of their colleagues, which again contributes to increasing their responsibility for language learning.

Awareness of different test types

All the teachers who participated in the group interviews agreed that in PRE-PROSPER testing there was no test-type differentiation. Although in theory teachers made the distinction between formative (progress) and summative (achievement) tests, this difference was made depending on the time when the test was administered and not on specific features of different test types. All interview participants admitted that, before being trained through

PROSPER, they were not aware of the various types of tests, the purposes for which they were administered, as well as of design issues in connection with each type. In practice there was no use of placement, progress, achievement or proficiency testing. Irrespective of the test purpose, there were two types of tests administered: grammar tests focusing on discrete point language and essay writing, impressionistically marked.

All teachers agreed that the training offered by PROSPER, however limited in the case of the teachers who did not take the one-year distance learning module, raised awareness of different purposes for testing and the corresponding test types. The most frequent issue mentioned was the difference between achievement and proficiency tests, of which few teachers were aware before PROSPER.

All groups interviewed confirmed the necessity to correlate testing with teaching methodology as a working principle. In PRE-PROSPER times this correlation was done automatically as the grammar-translation methodology was the only available option. Now, due to the effect of PROSPER, teachers have become aware of the necessity to correlate testing with teaching and of more complex issues, like the washback / washforward effects which need to be considered with respect to the teaching-testing relationship. Although teachers do not always design their own tests (with the exception of Manchester graduates) but prefer to adapt existing ones, they are careful about testing what and how they taught.

A quotation from the group interviews is illustrative with respect to the issues presented above:

“In a way, we can say that we correlated teaching and testing before in the sense that both were based on grammar-translation. So we did it automatically as we were not aware of other options. We did not think about the necessity to test in the same way you teach in order to offer the student a coherent approach.”

In three of the group interviews teachers expressed their on-going concern for test quality/reliability: in the process of administration, tests are continuously redesigned and improved, as regards both form and content. In the respective institutions the same test is administered by several teachers and afterwards discussed and improved in order to increase reliability.

The PROSPER test profile

Instead of a conclusion, we would like to outline a profile of the PROSPER test. The results of the analysis have confirmed the fact that in-depth

developments have taken place in the field of ESP testing in Romania since 1990, which can be summarized as follows:

PROSPER tests reflect a concern for assessing the four skills through a variety of methods. Even though Listening and, particularly, Speaking are less frequently tested, they include a wide range of test methods and a variety of authentic texts from many different sources. These are new developments which were virtually non-existent in PRE-PROSPER tests.

PROSPER tests show a move towards assessing the students through authentic and contextualised tasks. Students are not only required to make full use of their language competence, they are given a real-life purpose as well, as the test items focus on meaning and the communicative function of language.

However, old testing strategies and methods can be identified in PROSPER tests in varying degrees. As stated earlier, they point to the varying degrees of

exposure to PROSPER influence and change. The analysis of the group interview data has revealed that, along with the four skills, vocabulary and grammar are still widely tested, due to requests from students, and to the fact that grammar and vocabulary tests provide a clear track record of progress for both teachers and learners.

PROSPER test designers have diversified their marking criteria and show concern for more accurate and reliable marking, by trying to increase objectivity and reduce subjectivity in test marking.

PROSPER teachers are aware of and administer different types of tests (placement, diagnostic, achievement, proficiency). They are also aware of the necessity to correlate teaching methodology with test design and the complexity of aspects involved in this, and are able to match test content to test purpose. In several of the institutions under consideration, English Departments have a common testing policy, and the teachers co-operate in designing, administering and improving tests.

NOTES AND REFERENCES

1. "Project for Special Purpose English in Romania", initiated by Romanian ESP teachers and co-ordinated by the British Council Romania. For a detailed description, see BARDI, M. et al. (ed.) 'What is PROSPER?' in *Innovation in Teaching English for Specific Purposes in Romania – A Study of Impact*, Bucharest, The British Council & Cavallioti Publishing House, 1999, pp.5-10
2. BACIU, S., COMANETCHI, D., IVANESCU (CALCIU), R., KIRITESCU, L. - 'PROSPER Tests' in BARDI, M. et al. (ed.) *Innovation in Teaching English for Specific Purposes in Romania – A Study of Impact*, Bucharest, The British Council & Cavallioti Publishing House, 1999