LINELT 2013

Teaching and Testing Are Interrelated. But to What Extent?

Jana Bérešová a*

a Department of English Language and Literature, Faculty of Education, Trnava University in Trnava, Priemyselná 4, P.O. BOX 9, 918 43 Trnava, Slovakia

Abstract

This paper investigates the extent to which teaching and testing are interrelated, particularly as it pertains to testing English at B1 and B2 levels. The paper outlines research conducted within the process of linking the English school-leaving examination to the Common European Framework of Reference in 2013. Previous research, carried out in 2008 and 2012, revealed a gap between teaching and testing. In 2013, workshop participants were provided with sufficient awareness of the criteria to be able to analyse and assess their students’ performances in relation to CEFR levels. The results from the research referring to the comparison of teachers’ judgements of/on test-takers’ performances and test-takers’ testing scores will be presented and discussed. As previous research revealed that the coefficient of correlation between language in use scores and total scores of students was strong, the correlation between a total score (as a dependent variable) and several independent variables (scores achieved in testing listening, language in use and reading) were examined again (using multiple linear models in a regression analysis) and will be commented on and explained in the paper. As the approach adopted by the CEFR views learners of a language as social agents who have tasks to accomplish in a given set of circumstances, testing grammar and vocabulary needs to be considered from at least two perspectives: form-based and communication-based ones. The first perspective will be supported by exact examples from the selected tests.

© 2014 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Selection and peer-review under responsibility of the Organizing Committee of LINELT 2013.

Keywords: teachers’ judgements of test-takers’ performances, test-takers’ testing scores, multiple linear regression analysis;

1. Introduction

In 2005, a new way of testing English language competence in Slovakia was officially approved by the Ministry of

* Corresponding author: Jana Bérešová
E-mail address: jana.beresova@gmail.com

1877-0428 © 2014 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Selection and peer-review under responsibility of the Organizing Committee of LINELT 2013.
doi:10.1016/j.sbspro.2014.05.310
Education. Since then, each year a team of item writers prepare three versions of English tests, which are piloted and later included into the official item bank. When the idea of relating local tests and examinations to the Common European Framework of Reference (CEFR) emerged in many European countries due to the document elaborated by the Council of Europe - A Manual. Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR), item-writers of English tests in Slovakia decided to try to establish a link between their B1 and B2 language tests and the common reference levels of language proficiency in a reliable and responsible way.

With reference to the pilot version of the Manual (2009), the first attempts were made in 2008. The linking process was based on five stages, necessary for claiming links to the CEFR. Teachers who were invited to be panellists for the whole process were trained to become sufficiently aware of the CEFR levels to be able to analyse and assess test tasks and performances in relation to the CEFR levels. The second stage related to defining and describing clearly the test that was going to be linked to the CEFR confirmed consistency between official test specifications and the test administered in 2008, but revealed that there were still many issues which do not fit the criteria of objective testing. The most challenging part of the whole procedure was the standardisation and benchmarking stage, during which we could recognise that Slovak teachers are strongly dependent on the course books and follow the articles and the exercises one by one without taking the national documents (the curricula and the syllabi) into account. Another issue for discussion is that the CEFR is based on an action-oriented approach (2001), and, in our research in 2012, the coefficient of correlation between language in use scores and their total scores was strong at 0.904, as we used three scores (listening, language in use and reading scores) as predictors and regressed the total scores on these. We compared the regression weights for the three predictors, as well as the amount of variance in the total score that is predicted by each. On the basis of this regression, we achieved a sufficiently high degree of prediction and determined the relative importance of language in use to the total score. The whole process of relating the examinations revealed many areas that need to be changed in the future, such as evidence of good practice, internal validity and adequate procedures for all the steps of the test development and administration cycle.

2. Research

In 2013, we decided to get data related to two main focal areas:

1) The correlation between language in use scores and total scores to see if teaching grammar and vocabulary still plays an important role in teaching English in class

2) The comparison of teachers’ judgements and their learners’ achievements, officially proved by the statistics.

In our research we followed the same procedures as in previous years. After a familiarisation stage, using the same familiarisation activities as in the past, we presented the tests and asked the teachers to judge their students’ performances in listening comprehension, language in use and reading comprehension at both levels B1 and B2. The individual judgements were collected, analysed, processed and compared with the official statistical data based on scores achieved by the secondary-school students who applied for English test B1 and those who applied for English test B2.

Comparing the results achieved in the listening and reading sections of English test B1 did not reveal new findings. Moreover, it confirmed the data from previous years that students were better at listening (61.3%) than at reading (53.0%) and there is no correlation between these two sections. B2 candidates achieved the similar results: listening (70.8%) and reading (64.5%). On the other hand, both language in use tests B1 (51.6%) and B2 (54.3%) proved that the students’ results were significantly lower than those achieved in receptive skills tests. Therefore, in this year’s research, we decided to focus on language in use in detail. As noted by Purpura (2004:4), “Currently, knowledge of grammar might be inferred from the ability to select a grammatically correct answer from several options on a multiple-choice test, to supply a grammatically accurate word or phrase in a paragraph or dialogue, to construct grammatically appropriate sentences, or to provide judgements regarding the grammaticality of an utterance”, the item writers chose testing techniques with reference to official test specifications. In 2013, language in use test B1 consisted of two tasks: multiple-choice (10 items) and cloze test (10 items). However, test B2 consisted of three tasks: 21–40 multiple-choice (both grammar and vocabulary are measured), 41-50-word formation and 51-60 cloze tests, as the total number of items in the B1 test is 60 and B2 is 80. Although this approach emphasizes a syntactic-centric perspective of language, where the central feature is to observe and analyse the way in which words are arranged in a sentence, a communication perspective of language, where the observational and
analytic emphasis is on how language is used to convey meaning is a priority in productive skills tests (writing and speaking sections), which are not our concern in this paper.

As mentioned above, after taking part in an intensive familiarisation training session, the teachers were asked to judge the performances of their students. Changes in national legislation meant that students who study at secondary grammar schools (which are still considered as preparatory schools for university education) were obliged to choose English test B2. Therefore, only teachers teaching at that kind of school were invited for familiarisation sessions and later judged their students’ performances. As far as B1 English tests were concerned, two groups of panellists were addressed: teachers from technical schools whose students applied for the English test and teachers from secondary grammar schools owing to the fact that till the year 2013 their students could apply for a lower level test and these teachers are very experienced with/in preparing students for both tests.

What is more, the trained teachers were asked to fill in questionnaires to express their ideas related to testing language competence in English. These opinions enabled one to get an overall impression of the teaching process at Slovak secondary schools. A discussion forum in June 2013, held at the Faculty of Education of Trnava University, completed the whole research process, as the present teachers explained their reasoning in judging testees’ performances, pointing at their different experiences from teaching English under varied conditions.

3. Findings and Results

As official testing is run by a national institute appointed by the Ministry of Education, the official data are made public and are used in this paper to provide evidence that the issues discussed are based on official statistics and completed by further research and in-depth analysis.

3.1 English test B1

The reliability of the B1 test (0.922) can be considered high (Ringlerová, 2013). Focusing on language in use items (20-40), the results revealed that two items were too easy (items 21 and 28, respectively), as the items measured the usage of proper vocabulary in context. The context was clear and students did not have problems to choose a correct word. The table below shows that while teachers could recognise an easy item (21) at the beginning of the test, they could not do the same with item 28, which was more incorporated into the text.

Figure 1: Technical school teachers’ judgements versus official statistics

Figure 1 illustrates that technical school teachers expected more consistent scores than those statistically measured. Owing to the fact that in the Slovak syllabus English grammar and vocabulary are largely focused on, the teachers of technical schools, where English is taught 3 lessons per week and the number of unqualified English language teachers is still too high, expected to an important extent lower performances. These expectations were drawn on/rose from the teachers’ own experiences as they are very disappointed by a low number of English lessons per week and a relatively low interest of their students in learning languages at these schools by their students’ relatively
low interest in learning languages at these schools.

Figure 2: Technical school teachers’ judgements, secondary grammar school teachers’ judgements and official statistics

Secondary grammar school teachers presented a completely different opinion. As it has been mentioned before, these teachers are very experienced as they used to prepare students for a B1 level test in English in the past/during previous years and achieved good results with their students every year. The Ministry of Education officially supported the role of secondary grammar schools in the educational system of the country, but at the same time, analyzing the conditions under which the students are prepared for their future education at tertiary level, set higher goals for those who study at this kind of secondary education. In view of being aware of the test specifications, the secondary grammar school teachers overestimated the students’ performances due to both their previous experience and the level of difficulty set for their own students. To conclude, achieving 51.6% in a language in use section of the test cannot be considered a poor result as a similar score was achieved in the reading section of the test. On the other hand, it is important to state that cloze test seems to be more student-friendly. Two years ago, this testing technique replaced a word-formation technique.

3.2 English test B2

The reliability of the B2 English test was quite high (0.919), when analysed by Cronbach’s alpha as a measure of consistency (Mrva, 2013). Lower achievements of students/Students’ lower achievements in the language in use section (54.3) resulted from the previously mentioned change, introduced by the Ministry of Education, as the students had planned to sit for a B1 test in English, but were forced to choose a B2 test due to a legislative change. So all the students who started to learn a foreign language or continued in/with learning a selected foreign language at secondary grammar schools were expected to achieve level B2 in that selected language. Students had problems with one item in a word-formation task (able – disabled) and two items in a cloze test. The C-test is considered good as a measure of so called ‘lower-level’ knowledge of grammatical elements, while at the same time it also drew on higher-level textual competence (Read, 2009).
Figure 3 shows a comparison between secondary school teachers’ judgements and official statistics. Looking at the graph, we can come to the conclusion that teachers were more consistent in their estimates and many times their expectations were higher than testing scores on the grounds that teachers did not analyze particular pieces of grammar but focused on testing grammar and vocabulary as a whole. As level B2 is an option for students from other secondary schools, those who applied for this test achieved a little higher score (72.4%) than secondary school students (62.9%).

Table 1: Correlation between language in use scores and total scores

|                          | Red. Score - Grammar | Total Score |
|--------------------------|----------------------|-------------|
| Red. Score - Grammar     | Pearson Correlation  | .821**      |
|                          | Sig. (2-tailed)      | .000        |
|                          | N                    | 6983        |
| Total Score              | Pearson Correlation  | .821**      |
|                          | Sig. (2-tailed)      | .000        |
|                          | N                    | 6983        |

**. Correlation is significant at the 0.01 level (2-tailed).

The correlation between a/the total score and scores achieved in testing listening, reading and language in use were/was examined, using multiple linear models in a regression analysis. Table 1 shows the correlation between language in use scores and a/the total score, which was quite high (.821), but lower than previous years. The lowest correlation was between listening scores and a/the total score. Many learners did not address all the items of the reading section, as this section is the last one. According to the students’ questionnaires we could conclude that weaker students spent more time working on language in use items and left out the reading section items, primarily those they considered very difficult or time-consuming in the first sight perspective. This might be a reason that this year reading scores in both B1 and B2 tests are lower than they used to be in previous years.

4. Conclusion

Our concern related to linking the Slovak tests in English to the CEFR revealed that although present tests reflect many characteristics of communicative language testing (such as the use of authentic material, all the items in
context, the use of tasks in a writing test, requiring students to produce language appropriate to a range of specified purposes, etc.), the action-oriented approach, focused on testing both receptive and productive skills, is not completely followed. The idea that a test can yield valid results without including grammar and vocabulary components is considered right by Slovak item-writers, but educational traditions in Slovakia still prevail in the whole educational system. Although communicative language teaching (CLT) was introduced many years ago, teaching English is still influenced by a strong emphasis placed on teaching grammar and vocabulary (even though it is less prevalent than in other foreign languages taught in the country). Since every national document related to teaching target languages comprises a list of words and grammatical patterns to be taught, teachers strictly use their course books page by page and exercise by exercise so they spend most time of their lessons explaining and practising grammar and vocabulary. Needless to say, these lists are not calibrated as ‘can-do’ statements in the CEFR, but authorities do not want to recognise this fact.

Studying the relationship between one dependent variable (the total score) and several independent variables (listening scores, language in use scores and reading scores) using multiple linear models in a regression analysis, the same results were achieved. The correlation between language in use scores and the total score was high (0.821). The analysis of the teachers’ questionnaires has revealed that teachers (88%) follow the national documents very precisely and strictly, and in case their course books do not meet the requirements, represented by the lists of grammatical structures and words, they provide their students with a lot of supplementary material. Teachers of English focus mainly on the grammatical structures, which are different from the native language ones, emphasizing their usage more than the form. Teachers (72%) appreciate that universities and in-service teacher training centres give them an opportunity to be exposed to current trends in CLT and share their teaching experience. On the other hand, they admit that their effort to change their teaching in local conditions is not accepted positively, which discourages them. In spite of the fact that they try to support discovery learning, many other subjects are taught traditionally at their schools, which distracts the learners and many times they return to previous teaching practice.

When a new way of testing English in Slovakia was piloted in 1987, progressive/modern teachers were sceptical about its future, as the change resulted from their opinions presented at many local conferences. They considered testing inconsistent with teaching. After 8 years, this new form was officially recognised by the Ministry of Education and traditional teachers were forced to change their way of teaching. It is necessary to believe that teaching and testing are highly interrelated and affect each other in a positive way. To conclude, if communicative language teaching has needed more than four decades to be/to have been adapted for use in a wide variety of curricular and cultural context, communicative language testing needs to evolve gradually as it can have a very negative impact on students’ future studies or professional careers.

References

Common European Framework of Reference for Languages: Learning, teaching, assessment. (2001). Cambridge: Cambridge University Press.

Manual. Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). (2009). Strasbourg: Language Policy Division.

Mrva, M. (2013). A Final Report from the Statistical Analysis of English Test B2. Bratislava: NÚCEM.

Purpura, J. E. (2004). Assessing Grammar. Cambridge: Cambridge University Press.

Read, J. (2009). Assessing Vocabulary (8th printing). Cambridge: Cambridge University Press.

Ringlerová, V. (2013). A Final Report from the Statistical Analysis of English Test B1. Bratislava: NÚCEM.