

ETHICAL AWARENESS AND FAIRNESS IN LANGUAGE TESTING

Jana Bérešová, assoc. prof. PhDr., PhD
Trnava University/Faculty of Education, Slovakia

Abstract

The submitted paper outlines the impact of political decisions on language testing, predominantly on test scores of the tested population, test design and item-writing. Results from research referring to the comparison of test-takers' performances in the years 2011-2013 reveal a necessity to be sensitive and aware of ethics and fairness in language testing. The research is focused on the professional decision of test developers, item writers and administrators to avoid negative impact and maximize positive washback. The paper focuses on the construct which underlies the B2 English test to provide information about abilities that the test is designed to measure. Testing grammar and vocabulary was supported by using multiple linear regression analysis. The new conditions for test takers in 2012 are analyzed and evaluated in three areas the number of the students applying for tests in English B2, the total scores of students and their grouped distributions and the scores achieved in productive skills and language in use. We conclude by examining the ethical implications of the new conditions for test takers and by demonstrating the effect the ethical dimension had on item writers in their design of subsequent tests. Fairness plays an important yet hard to implement role in test design, and it is often up to the administrators and test designers to cope with external obstacles in promoting this.

Keywords: Communicative language testing, CEFR level B2, ethics, fairness

Introduction

A new theory of language and language use significantly influenced language teaching in the early 1970s, and a decade later, testing as well. The basic idea underlying communicative teaching is that language is used for the purpose of communication, in a particular situation and for a particular purpose, and the important thing is not what a person knows about the language, nor how grammatically correct they are, but if they can use it to communicate in the target language use situation, the real-world situation in which the language will be used (Hymes, 1972). This teaching is to be reflected in testing and therefore communicative tests need to test use of the language rather than usage.

The document, which has increased a communicative view of language in language learning/teaching as well as language testing, *the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*, offers language test designers and item writers the possibility of moving collectively towards a shared language testing system, which conforms to European and international standards of test production.

Since 2001, when *the CEFR* was officially published in English, and 2009, when *the Manual: Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)* and a set of accompanying tools emerged, supported by the Council of Europe, many European countries have re-evaluated their systems of language assessment and designed new national tests, mostly focusing on testing language competence achieved at school-leaving examinations in target languages.

In Slovakia, the reform of the school-leaving examination in English started in 1997, and the new way of assessing language competence was officially recognized by the Slovak

Ministry of Education in 2005. The reform was based on the CEFR perspective which extends the previously narrow linguistic range by focusing on competences either general or communicative. The CEFR became a basis for the elaboration of new English syllabi, describing what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively at B1 or B2 levels. In the beginning, learners who were to complete their secondary school studies could choose the level of the test, however, in 2012, the Ministry of Education required that all the students studying at secondary grammar schools, which are considered preparatory schools for university studies, were obliged to sit for a B2 test.

The English Language Examination at CEFR Level B2

The B2 examination in English comprises five sections, out of which three parts (80 items) are guaranteed by the Ministry of Education, whereas students' performances in two sections (based on testing productive skills) are assessed by appointed English teachers, directly at schools. The test, provided completely externally, consists of these sections:

Listening Comprehension (three tasks/20 items) – 20% of total

Language in Use (three tasks/40 items) – 20% of total

Reading Comprehension (three tasks/20 items) – 20% of total

The test-takers receive their scores in percentage and percentile in an official report. In spite of the action-oriented approach, adopted by the CEFR, which takes into account the cognitive, emotional and volitional resources and the full range of abilities specific to and applied by the individual as a social agent, English grammar and vocabulary is tested directly in the Language in Use section and indirectly in testing writing or speaking.

Productive skills are tested as follows:

Writing (one task) – 20% of total

Speaking (three tasks) – 20% of total

Although the topics of writing are announced officially, the papers are marked at schools, by two assessors using the official marking schemes, provided by the Ministry of Education. The speaking part is still in the hands of the teachers; however, the examination papers are agreed upon and recognised by officially appointed professionals. The examination board consists of three professionals who assess the performances using marking criteria for oral assessment.

An action-oriented approach, adopted by the CEFR, focuses on language ability where language is seen as a tool to perform communicative actions in a social context. The descriptors, which are presented in the CEFR, concerns language skills and linguistic competence defined as knowledge of, and ability to use, the formal resources from which a well-formed, meaningful message may be assembled and formulated (CEFR, 2001:109). Practitioners are recommended to use their frame of reference, identifying the theory, tradition or practice they are following. In the Slovak context, a new look at the assessment of grammar, which emphasises that grammar cannot be treated as an isolated component of knowledge and must be assessed in the larger context of language in communication. In the language in use section, grammatical forms and functions, and vocabulary are tested in context-based texts. In writing and reading sections, a communication-based perspective of language, which views grammar as a set of linguistic norms, preferences and expectations that a test-taker uses to convey pragmatic meanings, which are appropriate, acceptable and natural depending on the situation, prevails. From a communication perspective, Halliday (1994) emphasises that although language can be used to express meaning for a number of social purposes, the language system itself can be reduced to a small set of language functions such as experiential, interpersonal and textual functions in oral and written texts. Cohesion theory focuses on how certain words link grammatical forms to meaning and contextual use. These cohesive ties may be used through grammatical cohesive devices on a sentential level. As far

as vocabulary testing is concerned, the emphasis is put on context-dependent testing which assesses the test-takers' ability to take account of contextual information in order to produce the expected response. The notion of context is included into whole texts as contextualisation means to what extent the test-takers are being assessed on the basis of their ability to engage with the context provided in the test.

The idea to test grammar and vocabulary in the language in use section was based on our research, in which the coefficient of correlation between language in use scores and their total scores was above 0.9 in 2008 and 2012, which can be described as strong. This was discovered while the relationship between one dependent variable (total score) and several independent variables (listening scores, language in use scores and reading scores) was being studied, using multiple linear models in a regression analysis.

Radical Changes and Ethical Awareness

Before 2012, secondary-school leavers were expected to sit a test in one foreign language (English, German, French, Russian, Spanish and Italian) according to their preferences and CEFR levels (B1 or B2) due to their language proficiency. The students were responsible for their options and future professional orientations.

Since 2012, according to a decision made by the Ministry of Education secondary-grammar-school students have been obliged to select a language they want to be tested on, but exclusively at level B2. Since English has become a more and more popular language, replacing German, the expected consequences has been primarily seen in three areas:

the number of the students applying for tests in English B2

the total scores of students

the scores in skills and language in use

The first issue to discuss is a change in the number of applicants for English B2. While in 2011, the number of the students who chose to be tested in English B2 was only 3,268, the decision to sit for a language test at level B2 increased the number of test-takers being tested in English fourfold (15,651). When students were made to choose a higher proficiency level, the number of applicants for each language became logically higher. The preference for English has grown from 83.84% out of all tested population in 2011 to 85.98% in 2013.

Table 1 *Language selection of students applying for level B2 in 2011, 2012, 2013*

Language	2011	2012	2013
English	3,268	15,651	14,819
German	530	2,304	2,033
French	34	59	80
Russian	47	293	265
Spanish	17	27	28
Italian	2	5	5

The second issue which has to be raised is the statistical analysis of the total scores, achieved in English B2. Comparing the achieved total scores in the three-year term revealed that in 2012 a radical change in the mandated language proficiency caused a significant decrease of successful students. While in 2011 the percentage of unsuccessful students was 1.44% (47 students), in the year of the radical change it was tenfold – 10.9% (1,700 students). Nevertheless, it is important to mention that item writers felt all the ministerial changes sensitively, and in order to enhance fairness, they made the test suitable for various competency-equipped students. This approach resulted in positive washback in the year 2013, when the number of unsuccessful students tested in English B2 fell by 4.1% (602 students).

Table 2 Comparison of achieved scores in the years 2011-2013 in English B2

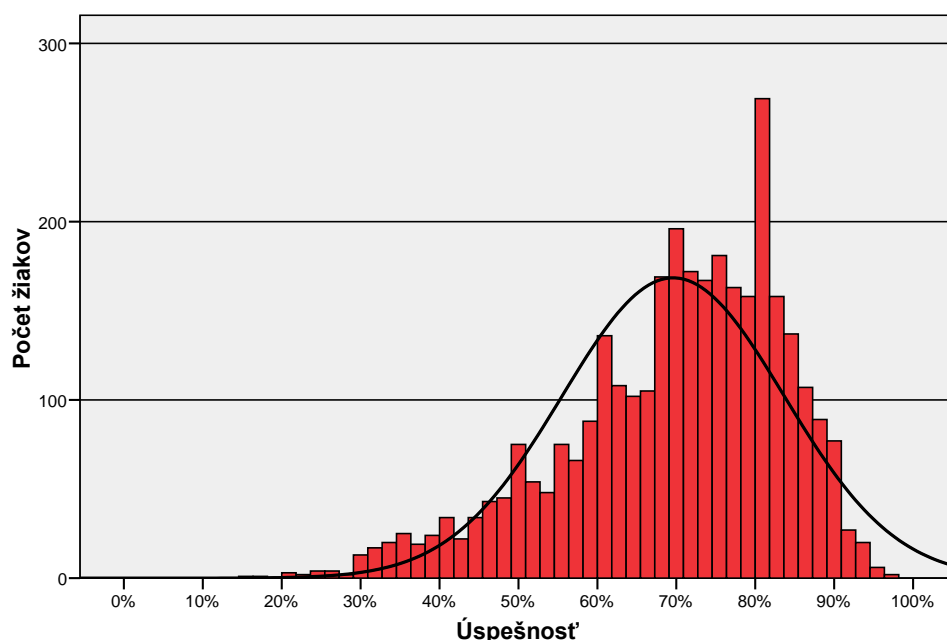
The year	2011	2012	2013
Total scores	69.6	55.4	63.2
Listening	76.7	55.8	70.8
Language in Use	62.1	52.7	54.3
Reading	70.1	57.6	64.5

The mentioned process is clearly seen in the following histograms, which display the grouped distribution of total scores achieved by students in the previous three years.

Analyzing the particular sections of the test, it can be summarized that the difference in achieved score in 2012 and 2011 is more visible in testing productive skills rather than grammar. On the other hand, it should be mentioned that the level of language competence was higher in all three sections in 2013.

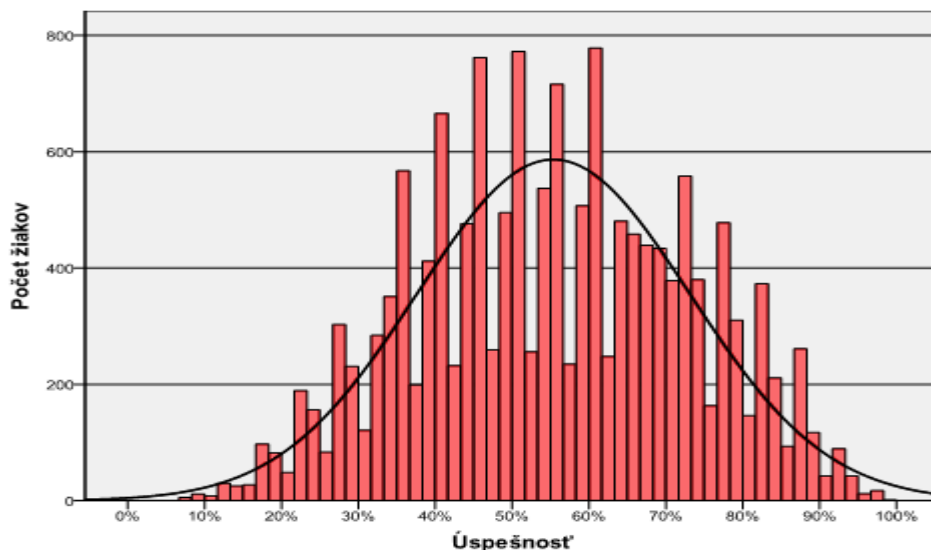
Figure 1 Histogram of frequency distribution of the scores achieved in English test B2 in 2011

Anglický jazyk úroveň B2 2011 - histogram úspěšnosti



The above score distribution (Figure 1) reveals that the achievements of students were quite consistent. The highest score 97.5% was achieved only by one student and 2.264% (74 students) of total test-takers achieved scores in the range from 90.1 to 100 (Urbanová, Ringlerová, 2011). The score of 75.8 was achieved by the highest number of students - 101 (3.09%).

Figure 2 Histogram of frequency distribution of the scores achieved in English test B2 in 2012



Obr. 1 Výsledný histogram úspešnosti

As we mentioned before, item writers sensitively responded to the decision of the Ministry of Education and the impact of their approach to ethics and fairness is seen in the histogram displayed in Figure 2 (Mišová, Mrva, 2012). One of the tested students achieved 100% of total score and the range from 90.1% to 100% was achieved by 203 students (1.297%). The distribution of scores is less consistent as poorer students who considered their English to be more suitably tested at level B1 had to sit for level B2.

Figure 3 Histogram of frequency distribution of the scores achieved in English test B2 in 2013

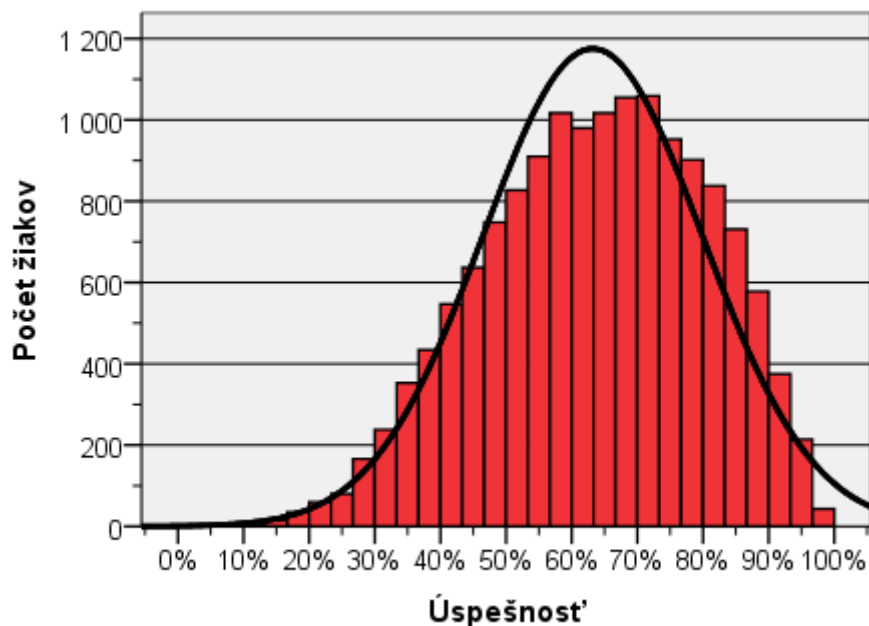


Figure 3 displays the histogram of the grouped score distribution from English test B2, conducted in 2013 (Mišová, 2013). A positive washback of the sensitive approach of item writers from the previous year is reflected in the achievements of the students. One student achieved the score of 99.2, seven students that of 98.3 and nine students 97.5, followed by twenty-six students achieving 96.7. The range from 90.1 to 100 was achieved by 530 students (3.58%). The score distribution is consistent and total grouped scores are proportionally distributed.

Analyzing achieved scores in single sections of the test, it was discovered that in the years when the students could choose their proficiency level to be tested on, the scores

achieved in Listening and Reading Sections were rather comparable. The following years (2012, 2013) reading scores were lower due to a number of left-out items. As the Reading Section is the last section of the national test, we might conclude this phenomenon that poorer students have not developed their reading skills in English so well and need more time for tasks achievement. Surprisingly, the students deal with the language in use items constantly. A strong emphasis put on grammar and vocabulary in teaching English in Slovakia seems to be fruitful. As we mentioned before, the correlation of the total scores and language in use scores is so strong that despite the action-oriented approach of the CEFR, the country adopted the parameters and categories useful for the description of linguistic content in the national curriculum.

Conclusion

The ethical concerns that are discussed in language testing are essential. Decisions made about a person on the basis of a test score can have serious and far-reaching consequences. Ethics in language testing is very important. Spolsky (1977) supported the approach to language testing that requires full justification of all statements based on tests. He pointed out that language testers must be as concerned with the prevention of bad testing as with developing new tests, and that they must be sensitive to the possible educational, social and political consequences of testing. As tests have impact on the lives of test takers, any decision should be done professionally. In our case, the students, who were expected to complete their secondary-school studies, should have been officially informed that they would have to sit for B2 language tests in the first year of their secondary studies. This information would have influenced their approach to language learning and reduced the negative impact on their lives.

On the other hand, item-writers tried to soften the possible consequences and worked consciously in test design and item writing to maximize the possibility of positive washback. This reflected in the following year when the achievements of the students were better and the score grouped distribution was more consistent. The sensitive approach of the administrators, test designers and item writers confirmed the well-known statement that practices must be just and tests must be primarily just and fair for all.

References:

- Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press, 2001.
- Hymes, D. H. On communicative competence. In J. B. Pride and J. Holmes, eds. *Sociolinguistics*. Harmondsworth: Penguin, 1972.
- Manual. Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Strasbourg: Language Policy Division, January 2009.
- Mišová, M. *Správa o výsledkoch externej časti a písomnej formy internej časti maturitnej skúšky z anglického jazyka úroveň B2*. Bratislava: NÚCEM, 2013
- Mišová, M., Mrva, M. *Správa o výsledkoch externej časti a písomnej formy internej časti maturitnej skúšky z anglického jazyka úroveň B2*. Bratislava: NÚCEM, 2012.
- Spolsky, B. 'Language Testing: Art or Science?' In *Proceedings of the Fourth International Congress of Applied Linguistics*, Stuttgart: Hochschulverlag, 1977, pp.7-28.
- Urbanová, M., Ringlerová, V. *Správa o výsledkoch externej časti a písomnej formy internej časti maturitnej skúšky z anglického jazyka úroveň B2*. Bratislava: NÚCEM, 2011