



## INNOVATIVE APPROACH TO LANGUAGE TESTING

Hana ROMOVÁ – Tomáš ŘÍMSKÝ

### Abstract:

*This article with the title Innovative Approach to Language Testing deals with a new approach to language testing at the Department of Languages. It focuses on establishing an item bank with individual subtests with the same specification and construction for full-time and combined studies. The language level is based on The Common European Framework of Reference for Languages – specialist vocabulary level B2. The comparison of percentage success rate of full-time and part-time students was followed. The main objective of data collection was supported by information technology. The tests were administered in all types of study and the output of percentage success rate of respondents was found out as well as the possible dependency between the percentage success rate of the groups.*

**Keywords:** *language testing, innovation, empirical study, Security Legal Studies, Security Management, Public Administration, full-time study, combined study, team work*

### INTRODUCTION

This article introduces the innovative approach to language testing of future professionals within the Department of Languages of the Police Academy of the Czech Republic in Prague, which is a state university type institution as well as a supreme educational and scientific institution within the Ministry of the Interior.

There are two faculties, the Faculty of Security and Law and the Faculty of Security Management, whose students are police officers, members of the Fire Rescue Service, public administration personnel, employees of the regions and municipalities, security management experts as well as civilian students. The accredited study programmes offered are Bachelor programmes, Master programmes, PhD programmes and practical training providing and specializations in full-time or combined forms. It is possible to achieve Bachelor's, Master's, Doctoral (Ph.D.) and MBA degrees at this University.

The Police Academy employs a team of lecturers and specialists with rich work experience. Teaching, science and research activities are organized through specialized departments focused on law, security and public administration. The language department offers four term courses within the bachelor studies and a three term course in the Master courses. The languages taught at present at the Academy are English, German, Russian and French.



## 1 LANGUAGE TESTING

The construction of testing items is a didactic, time consuming and demanding activity and that is why a whole team of teachers of the language department has to participate. On the contrary to professional testing workplaces all functions in the process of test construction are implemented as test constructors, moderators, administrators and testers. On the basis of a gradual construction of the item bank there can be created a great number of different tests by the help of combinations of individual subtest with the same specification and construction. The construction of quality and reliable tests depends on whether the items really measure what the authors intent and if they are reliable. Reliability or to be reliable means that they are rendering the same results under the same conditions when repeatedly assigned. Methodologically elaborated bundles containing items reflecting the basic and frequented phenomena of the curriculum are optimal when testing language skills.

The author of the test, the moderator, is a professional in the given language. The accuracy of the tested terminology, difficulty – the usage of language means minimally level B1 according to The Common European Framework of Reference for Languages – specialist vocabulary level B2. The date of input follows. Status of the item (e.g. active, in which test, when used and by whom; bundles (items not active for a longer time period should be removed or reconstructed).

Reliability expressed by the coefficient  $r$  with values from 0 to 1, depending on the number of items  $p$ , median, and disparity  $s^2$  and counted according to the formula  $r = m(p - m) / ps^2$ , expected reliability is  $r = 0,6$  and higher. Items with a low reliability have to be restructured. It is necessary to restructure very easy items (85 % and more tested) and very difficult items (25 % and less tested). The correct answer is indicated in the answer sheet.

The full standardisation would be time consuming and financially demanding and that is why in our work environment these quasi-standardised tests are suitable. The tests assess the level of students' knowledge in several parallel groups at one school. They were preparing for the test in comparable conditions from the same recommended teaching materials. The tests had been implemented and evaluated under comparable conditions. Continuous tests were designed as *achievement tests in the sphere of professional language and grammar*. Achievement tests as test of language competence was designed or constructed as multiple-choice test and true/false to assess the scope of certain topical or current language abilities of the students.

Subtest and phases of test preparation – the achievement test has three subtests – listening comprehension with 10 items, reading with comprehension with 10 items and lexical-grammatical subtest including 60 items.



The subtest listening and reading with comprehension consists of two bundles, each of which contains 5 items of the true/false type and 5 multiple-choice items. The lexical-grammatical subtest comprises 6 bundles of 10 item. Bundles of lexical subtest are oriented on filling in suitable words into mutually unrelated sentences, filling in suitable words into the coherent context, matching the correct term to the given definition and the choice of the correct translation of an expression from Czech into English or vice versa. The grammar subtests consist of one bundle of filling in grammar forms into sentences and there is one bundle of times testing the choice of the correct preposition into phrases or into the most frequented verbs.

The listening part forms 10 % of the test since there is the fact taken into account that part-time students have only a limited possibility to train listening skills and a higher point scale of listening could devalue the whole evaluation. Reading skill with comprehension has been evaluated by 10 % since in grammar and lexical test tasks there are context-based tasks demanding mastering these skills. In view of the great volume of vocabulary which should be tested it is necessary that the choice of terms and professional vocabulary for the tests be very thoroughly and should correspond with the real demands laid on the students in practice and real-life. When constructing the tests we proceeded in the following way:

Definition of the objects – analysis of teaching material, choice of basic phenomena, difficulty. Blueprinting - defining the test content, structure of the testing methods and types of tasks. Assignment to the authors, scanning the answer sheets. Keys, preparation of listening tasks, methodological moderation of the bundles. Construction of true/false items, multiple-choice with 4 distractors and filling into the item bank. Three subtests according to the specification - setting up the pilot test and proof-reading. Analysis and evaluation of the pilot test. Cooperation with a statistical expert, test analysis and scores and median. Elaboration and evaluation of the results by the computer and the analytic elaboration of the results and their interpretation. When constructing tests and the individual tasks and elaborating assignments following basic methodological standards and rules are applied: Methodological, didactic and content aspect of the test has to be relevant according to the assignment and purpose of the test. Students are informed at the beginning of the course about the content of the test, its form, scope covering the subject matter, type of tasks, time limit, way of assessment, announcement of results and feed-back. The construction of the test is valid and reliable follows from the specification of the items.

Parameters of the assignment fulfil the demands on the success rate of the test. Out of practical reasons, especially financial demands on technical equipment a combination of PBT paper-based testing and CBT computer-based-testing with the following characteristic were being used: PBT test with tasks chosen by the tester from the item bank, answer sheet, scanning, measuring and assessment of the answers by the computer.



Team work based on sharing experience at creating items and tasks for the item bank influences its quality and brings a certain surplus in the form of trust of students in a just and objective evaluation than the competitive access of individual testers. By a gradual creation of the item bank there can subsequently be formed a large number of different tests with the same specification and construction. According to the scale of thoroughness and verification of the test and its equipment is this test on the borderline of the standardised and non-standardised test. Such tests are called quasi-standardised by some authors e.g. Harris (2006), White (2014) Handley (2018), since they contain standards for assessment of results.

## 2 TEST CONSTRUCTION

Based on gradual construction of the item bank there can be created a great number of different tests by the help of combinations of individual subtest with the same specification and construction. The creation of quality and reliable tests depends on whether the items really measure what the authors intent and if they are reliable. Reliability means that the tests are rendering the same results under the same conditions when assigned repeatedly. Methodologically elaborated bundles containing items reflecting the basic and frequented phenomena of the curriculum are optimal when testing language skills. The structure of the item bank is illustrated as follows:

Topic or theme not to be doubled in the framework of one test, the author of the test, moderator who is a professional in the given language, in some cases a professional in the given sphere consulting the accuracy of the tested terminology, difficulty – the usage of language means minimally level B1 according to The Common European Framework of Reference for Languages – specialist vocabulary level B2. The date of input follows (chosen bundles and items should not be used for revision tests in the running neither in the coming academic year). Status of the item (e.g. active, in which test, when used and by whom); bundles (items not active for a longer time period should be removed or reconstructed).

### 2.1 DIVISION OF TESTS

Since the terminology of test classification is not unified we limit the division test in the following way:

“Achievement Test“ which is used during the lessons, test items with multiple-choice answers, only one of them is correct. The tested items are based on the curriculum of the given semester. The knowledge of what has really been learned by the students is being tested in comparison to what they should have learned according to the syllabus and study materials.

“Progress Test“ is introduced in the exam period of the second semester. This test is designed to find out which extent the students have really learned or mastered the specialist topics.



A successful student has to achieve at least 70 % of correct answers and this is a prerequisite for getting the credit at the end of the semester.

“Proficiency Test“ is a non-standardized test as a part of the examination. It is a cross-section test and it is designed to find out what the students really know at the present time and how they receptive skills have developed. Forms of tasks are: multiple choice questions (3 distractors – 1 correct answer, there is no point subtracted for an incorrect answer). Each item tests only one sphere for the items not to have two meanings.

Target groups are 1st and 2nd year of full-time and part time students studying Security Legal Studies and Security Management.

The full standardization would be time consuming and financially demanding and that is why in our work environment these quasi-standardized tests are suitable. The tests test the level of students' knowledge in several parallel groups at one school. They were preparing for the test in comparable conditions from the same recommended teaching materials. The test had been implemented and evaluated under comparable conditions. Continuous tests were designed as achievement tests in the sphere of professional language and grammar. Achievement tests as test of language competence was designed or constructed as multiple-choice test and true/false to assess the scope of certain topical or current language abilities of the student.

Subtest and phases of test preparation – the achievement test has three subtests – listening comprehension with 10 items, reading with comprehension with 10 items and lexical-grammatical subtest including 60 items.

When constructing the tests we proceeded in the following way: First is the definition of the objects – analysis of teaching material, choice of basic phenomena and difficulty. Blueprinting - defining the test content, structure of the testing methods and types of tasks. Further follows assignment to the author and scanning the answer sheets. Keys and preparation of listening tasks, methodological moderation of the bundles follow. True/false items, multiple-choice with 4 distractors are constructed and filled into the item bank. Three subtests according to the specification or setting up and the pilot test and proofreading are the next step. Analysis and evaluation of the pilot test. There was ongoing cooperation with a statistical expert, test analysis and scores and median. The last step was elaboration and evaluation of the results by the computer and the analytic elaboration of the results and their interpretation.

The construction of the test is valid and reliable, it follows from the specification of the items. Parameters of the assignment fulfil the demands on the success rate of the test. Out of practical reasons, especially financial demands on technical equipment a combination of PBT paper-based testing and CBT computer-based-testing with the following characteristic were



being used: PBT test with tasks chosen by the tester from the item bank, answer sheet, scanning, measuring and assessment of the answers by the computer.

Team work based on sharing experience at creating items and tasks for the item bank influences its quality and brings a certain surplus in the form of trust of students in a just and objective evaluation than the competitive access of individual testers. By a gradual creation of the item bank there can subsequently be formed a large number of different tests with the same specification and construction. According to the scale of thoroughness and verification of the test and its equipment is this test on the borderline of the standardized and non-standardized test. Such tests are called quasi-standardized by some authors, e.g. Hambleton, Rovinelli, since they contain standards for assessment of results, but the standardization is not complete.

### 3 TESTING IN FULL-TIME AND PART-TIME STUDIES

The comparison of percentage success rate of full-time and part-time students followed. The main objective of data collection supported by information technology was to ensure that each test was administered in both types of study with the aim to compare the output of percentage success rate of respondents and to find out the correlation tendency of these two groups and to formulate following the possible dependency between the percentage success rate of the groups.

We can state after analyzing the data gained that full-time students reach better results than part-time students. This fact can be explained by several factors. In the first case it is the difference in the form of language teaching in courses for full-time and part-time students. The absence of regularity and direct contact with the teacher influence negatively the overall testing score of the respondents. Other factors are not given in the data of the empirical inquiry. On the basis of experience of the team members it can be stated that the profile of the average part-time student, which determine his or her entrance language preparedness and succeeding language acquisition capacity during his study, the time factor of his previous language study, time options for study, environment and regularity and systematic character of individual language preparation. Comparison of percentage success rate of study programs of security and public administration was carried out.

During the empirical study and its interpretation, the success rate of students from both study programs was being compared to find out whether the results are comparable.

### A. Comparison of the subtest Listening

TABLE I.

Subtest Listening	Average success rate (%)	Average difference rate between groups of study programmes (%)	Average success rate Full-time study and Part-time study	Difference of success rates between Full-time study and Part-time study (%)	Number of respondents Full-time and Part-time study	Overall number of respondents
B60 1.sem.	90		96	11	92	200
		25	85		108	
B60 3.sem.	65		92	45	52	134
			47		82	
B71 1.sem.	63		73	20	44	87
		7	53		43	
B71 3.sem.	70		72	4	44	76
		3	68		32	
B71 4.sem.	73		77	7	25	72
			70		47	

### B. Comparison the subtest Reading

TABLE II.

Subtest Reading	Average success rate (%)	Average difference rate between groups of study programmes (%)	Average success rate Full-time study and Part-time study	Difference of success rates between Full-time study and Part-time study (%)	Number of respondents Full-time and Part-time study	Overall number of respondents
B60 1.sem.	87		91	8	92	200
		20	83		108	
B60 3.sem.	67		97	48	52	134
			49		82	
B71 1.sem.	86		80	9	44	87
		9	81		43	
B71 3.sem.	77		80	7	44	76
		26	73		32	
B71 4.sem.	51		38	-20	25	72
			58		47	



### C. Comparison of the subtest Grammar

TABLE III.

Subtest Grammar	Average success rate (%)	Average difference rate between groups of study programmes (%)	Average success rate Full-time study and Part-time study	Difference of success rates between Full-time study and Part-time study (%)	Number of respondents Full-time study and Part-time study	Overall number of respondents
B60 1.sem.	82		89	13	92	200
		30	76		108	
B60 3.sem.	52		80	46	52	134
			34		82	
B71 1.sem.	71		77	13	44	87
		2	64		43	
B71 3.sem.	69		72	8	44	76
		4	64		32	
B71 4.sem.	65		63	-3	25	72
			66		47	

We can state after analyzing the data gained that full-time students reach better results than part-time students. This fact can be explained by several factors. In the first case it is the difference in the form of language teaching in courses for full-time and part-time students. The absence of regularity and direct contact with the teacher influence negatively the overall testing score of the respondents.

Other factors are not given in the data of the empirical inquiry. On the basis of experience of the team members it can be stated that the profile of the average part-time student, which determine his or her entrance language preparedness and succeeding language acquisition capacity during his study, the time factor of his previous language study, time options for study, environment and regularity and systematic character of individual language preparation.

Comparison of percentage success rate of study programs of security and public administration with IT support was carried out.





#### D. Comparison of the subtest Vocabulary

TABLE IV.

Subtest Vocabulary	Average success rate (%)	Average difference rate between groups of study	Average success rate Full-time study and Part-time study	Difference of success rates between Full-time study and Part-time study (%)	Number of respondents Full-time study and Part-time study	Overall number of respondents
B60 1.sem.	83		90	13	92	200
		15	77		108	
B60 3.sem.	68		92	39	52	134
			53		82	
B71 1.sem.	83		87	9	44	87
		2	78		43	
B71 3.sem.	81		83	5	44	76

During the empirical following and its interpretation the success rate of students from both study programs was being compared to find out whether the results are comparable and because of good practice between the study programs.

#### E. Comparison of Average success rate

TABLE V.

Test	Average success rate (%)	The difference of Average success rate Full-time study and Part-time study (%)	Number of respondents
B60 1.sem.	85,5	11,25	200
B60 3.sem.	63	44,5	134
B71 1.sem.	75,75	12,75	87
B71 3.sem.	74,25	6	76
B71 4.sem	65,25	-	72

From the table above follows that the average success rate of tests of the study program has a tendency to greater compactness than the other program which shows a higher tendency towards polarization of percentage success rate.

## CONCLUSION

It can be stated, after comparing the average success rate of both study programs with the aid of information technology, that public administration reaches more compact average values of the success rate and the overall success rate in all tests of both study programs lies in the corridor with the average value 75 % with a difference of ca.  $\pm 10$  %. Generated tests reach in most cases higher values of percentage success rate than intended.



According to this fact it is necessary to apply the above mentioned examples of good practice at the creation of new tests items and evaluated them statistically and prove the expected shift in values of percentage success rate in the set band. The average success percent rate of the tests is in the range between 60 a 65 %, which is considered on the basis of practice and specialist literature as optimal. At the same time this result level corresponds with the average point rate of the scale applied for the assessment of exams.

## SOURCES

1. HAMBLETON, Ronald and ROVINELLI, Richard, 1986. *Assessing the dimensionality of a set of test items*, pp. 287-302, *Applied Psychological Measurement*, Vol. 10, No. 3. <https://doi.org/10.1177/014662168601000307>
2. ALTE Manual for Language Test „Development and Examining for use with the CEFR“ [http://www.coe.int/T/DG4/Linguistic/ManualLanguageTest-Alte2011\\_EN.pdf](http://www.coe.int/T/DG4/Linguistic/ManualLanguageTest-Alte2011_EN.pdf)
3. FULCHER, Glen. *Practical Language Testing*, Routledge. 2013. 343 p. ISBN 13: 978-0-340-98448-2.
4. GREEN, Anthony. *Exploring Language Assessment and Testing*, Routledge Taylor Francis Group, 2014. 260 p. ISBN 978-1-415-88962-7.
5. HANDLEY, Zoe 2014: *Engaging Language Learners through Technology Integration*: Li, S. & Swanson, P. (Eds.). IGI Global, 27 p. Theory, Applications, and Outcomes.
6. Mc NAMARA, Tim. *Language Testing*, Oxford University Press 2000.
7. *English Core Inventory for General English*. British Council/EQUALS. (European Association for Quality Language Services) 2010.

### **PhDr. Mgr. Hana ROMOVÁ, Ph.D.**

Policejní akademie v Praze  
Fakulta bezpečnostního managementu  
Katedra jazyků  
Lhotecká 559/7  
143 00 Praha 12  
romova@polac.cz

### **Mgr. Tomáš ŘÍMSKÝ**

Policejní akademie v Praze  
Fakulta bezpečnostního managementu  
Katedra jazyků  
Lhotecká 559/7  
143 00 Praha 12  
rimsky@polac.cz