

Šikolová & Koláčková, 2017

Volume 3 Issue 2, pp. 451 - 464

Date of Publication: 08th September, 2017

DOI-<https://dx.doi.org/10.20319/pijss.2017.32.451464>

This paper can be cited as: Šikolová, M., & Koláčková, L. (2017). Some Aspects of Test Validity in the Czech Military. *PEOPLE: International Journal of Social Sciences*, 3(2), 451-464.

This work is licensed under the Creative Commons Attribution-Non-commercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

SOME ASPECTS OF TEST VALIDITY IN THE CZECH MILITARY

Mária Šikolová

University of Defence, Language Centre, Brno, Czech Republic
maria.sikolova@unob.cz

Ludmila Koláčková

University of Defence, Language Centre, Brno, Czech Republic
ludmila.kolackova@unob.cz

Abstract

The language exams in the Czech military are undoubtedly high-stakes tests, and their results affect the lives of thousands of military professionals. As a consequence of their importance, the quality of the exams is often questioned, leading to a need for continual monitoring and improvement. Test validity should not be solely the domain of testers, but teachers, as well as candidates, should have a voice in the process as well. This concern is why one of the research objectives is to uncover how the exam is perceived by candidates. The instrument for gathering the data is a questionnaire developed by the authors and distributed to uncover the candidates' opinions on the exam, both as a whole and in particular parts. Teachers' views on the tests constitute another piece of evidence of test validity, and these are currently being gathered less formally, but will nonetheless be addressed accordingly. This contribution will illustrate the preliminary results of the data collected from the questionnaire for the exam candidates, as well as the results of a study dealing with a change in assessing testing writing based on the feedback from the teachers.

Keywords

Validity, Language of Instructions, Language Testing, Questionnaire

1. Introduction

Our study is motivated by recognition of the growing importance of standardized language examinations in the military. Almost all military personnel are expected to attain a particular language level by the end of 2019, putting significant pressure on both professional soldiers and teachers. Stakeholders have begun questioning the entire language training and testing system, including exam validity and reliability, the current syllabus, teaching materials, and methods of instruction. We agree with Alimemaj that “Test taking is often very stressful and frustrating for students. Experts tend to focus more on the kinds of tests that work best for students and overlook the emotional aspect of the matter“ (Alimemaj, 2015). These concerns have led the management of the Language Centre to begin reconsidering all areas of language education and to launch projects with the aim of reaching some conclusions about validity. As a part of this process, the authors have decided to address the face validity of the standardized tests currently in use. Firstly, we have conducted a review of the literature which was followed by designing a questionnaire to address the candidates’ perceptions and feelings immediately after completing the exam. At this point, we have gathered at least a sample of these responses which has enabled us to present the results and their interpretation. Nevertheless, during this phase of our research, we realized that there is another gap to be filled, namely that of gathering and analysing teachers’ opinions and beliefs concerning the standardized military exam.

2. Literature Overview

The concept of validity is rather complicated and it is often approached by different authors differently. It is sometimes divided into four different kinds, such as construct validity, criterion-related validity, content validity and face validity. However, Messick considers validity as a unitary phenomenon and in his opinion, validity is not a quality of a test, but it rather refers to the inferences based on the test (Messick in Davidson & Lynch, 2002). Chapelle adds that in her opinion, reliability, otherwise perceived as a separate phenomenon or quality of the test, should in fact be considered a part of validity (Chapelle in Davidson & Lynch, 2002, 134).

Face validity is sometimes treated as not very scientific or relevant. It is commonly addressed by non-experts and, hence, is often looked down on by experts. McNamara defines

face validity of a test as “its surface acceptability to those involved in its development or use” (McNamara, 2008). Alderson et al. see certain importance in face validity – they argue that if the users do not consider the test valid, they may not take them seriously and, as a result, they might not “perform to the best of their ability on ... the test” (Alderson, Clapham & Wall, 2006). They also recommend gathering data based on interviews with students or administering questionnaires. Following these recommendations, the authors have opted for questionnaires from both test takers and teachers.

Hughes presents a very similar stand when it comes to face validity, stating that even if the concept of face validity “is hardly a scientific concept, yet it is very important” (Hughes, 2013). The reasons he gives are almost identical with those mentioned by Alderson (Alderson, Clapham & Wall, 2006).

Face validity is often considered to be less important and quite difficult to measure; according to some specialists it is related to authenticity; in other words, it is “the extent to which the test appeals to test takers and test users” (Bachman & Palmer, 2013).

3. Research Tool – Questionnaire

Although face validity is considered to be on the edge of the interest of testing specialists, on practical grounds, we have decided to make it the focus of our study.

Our research was launched in its preparation phase in October 2016. A structured questionnaire of three parts was constructed. It was based predominantly on positive statements and questions. However, the last part also enabled the respondents to freely express their feelings and impressions.

A small number of respondents was selected for piloting and later the second and the third parts were redesigned to provide clarification to the statements and questions since they were not understood by all respondents; some of the questions were eliminated; some were put into a more logical order. Currently, the third version of the tool is being used.

The candidates were informed about the possibility to take part in our research during the test administration, and the questionnaire (pen-and-paper form) was presented to the respondents immediately after the examination. They were assured that the anonymity of all respondents was guaranteed and we appreciated all the responses as very important, not only for the research purposes but also for practical use. The responses were obtained solely on the basis of students’ good will to cooperate with our team.

The first part consisted of identifying information only – whether the respondent was a soldier or a student of the University of Defence, a student at military secondary school, or a civilian employee.

Part 2, the main part of the questionnaire, was focused on candidates' subjective opinion on the beneficial aspects of the four subtests – listening, reading, speaking and writing. A scale was employed allowing students to mark each of the statements using the numbers from 1 to 5 (*1 – strongly agree, 2 – somewhat agree, 3 – neutral, 4 – somewhat disagree, 5 – strongly disagree*).

In Part 3 of the questionnaire, students were asked six questions regarding their opinion or experience with the language test, use of English at work, preparation for the exam, and on their views about focusing the test toward more of a military context. We also sought to find out candidates' feelings after the completing the exam, as well as their overall perception of the test.

4. Preliminary Results

As is apparent from our stated research goals, the study seeks to address the broader issue of validity, and this part of our investigation focuses on candidates' feedback, i.e. their attitudes to and perceptions of the high-stakes military examination known as NATO STANAG 6001.

As mentioned above, at the moment, the third version of the questionnaire is currently in use, since during the piloting stage we realized that some items were ambiguous and, hence, needed minor changes.

Nevertheless, we have gathered 40 appropriately filled responses, giving us at least a minimum sample to consider and interpret.

Part 2 of the questionnaire deals with the evaluation of the individual subtests. As for the listening comprehension subtest, more than one-third of the candidates consider the speed of speech in the recordings appropriate, while nearly one-third of them believe that the speed was rather high. Similar results were gathered concerning understanding of the items; around one-third of candidates understood them properly, and one-third did not understand them. In the third item connected with the listening comprehension subtest, we wanted to find out if the items are properly constructed and clearly formulated. We used the following statement: *Although I understood the recordings, the correct answer occasionally seemed unclear*. About half of the candidates agreed with the statement, while one-quarter did not agree.

Chart 1: Listening

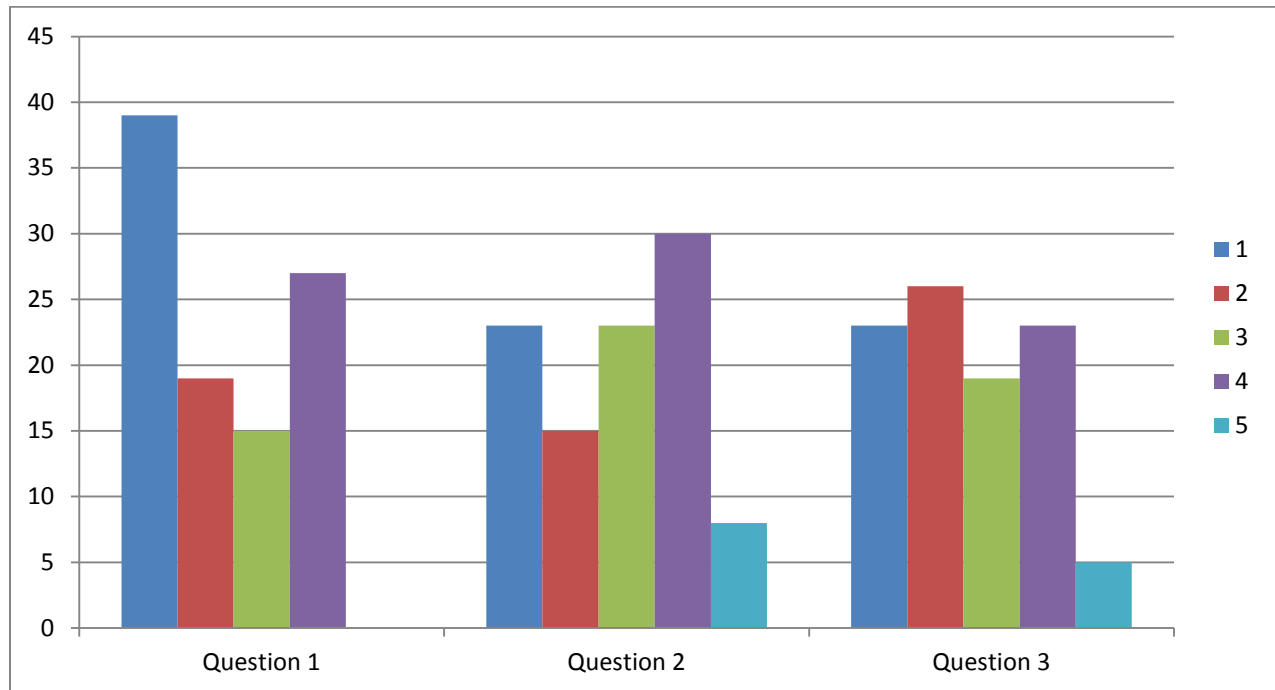


Table 1: Listening

Question 1	The tempo (speed) of the language speakers was appropriate.
Question 2	I understood the content of individual test items without any problems.
Question 3	Although I understood the recordings, the correct answer occasionally seemed unclear.

Referring to the reading subtest, half of the respondents did not understand the content of individual items without problems. Only 7 % of them claim they understood them properly. Another statement: *Although I understood the content of the individual texts, the correct answer occasionally seemed unclear* was agreed with by more than 50 % of respondents, which leads us to consider the need for thorough and deeper reflection on both test construction and teaching methodology.

Chart 2: Reading

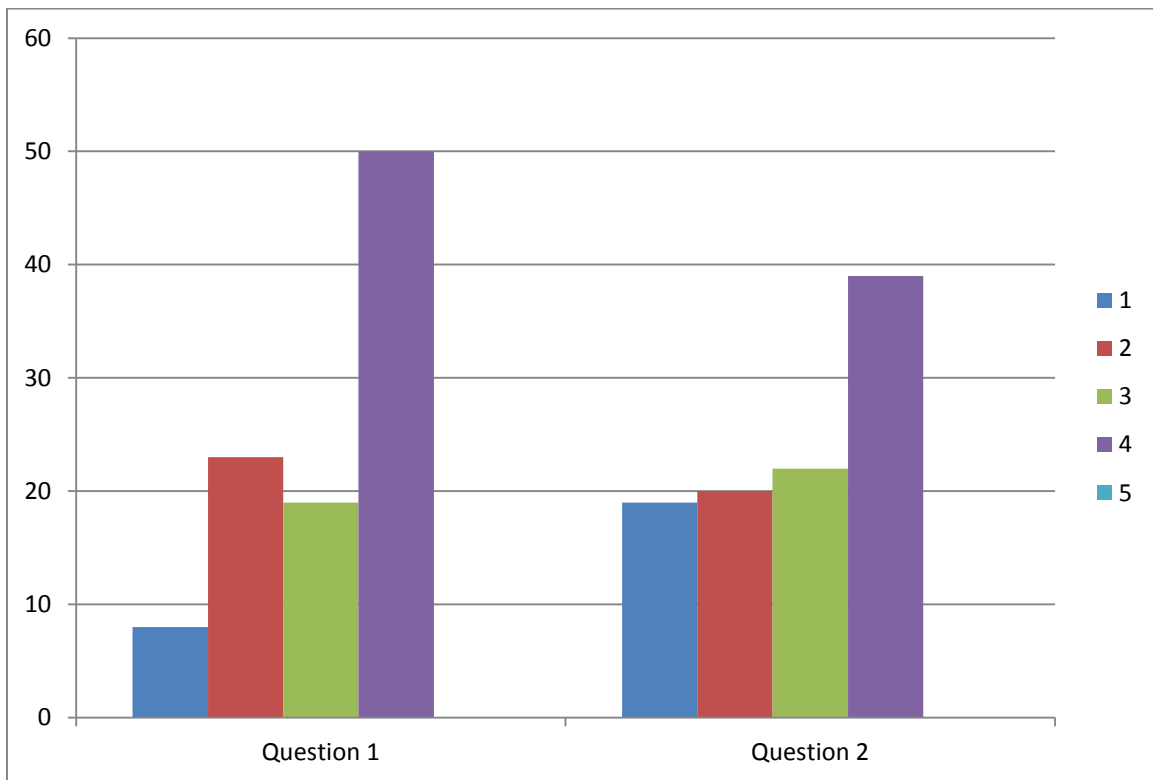


Table 2: Reading

Question 1	I understood the content of individual test items without any problems.
Question 2	Although I understood the content of the individual texts, the correct answer occasionally seemed unclear.

As far as the oral part of the examination is concerned, speaking in the given context was generally perceived as neither difficult nor easy. A relaxed and motivating atmosphere during the examination was reported by more than 60 % of the candidates.

Chart 3: Speaking

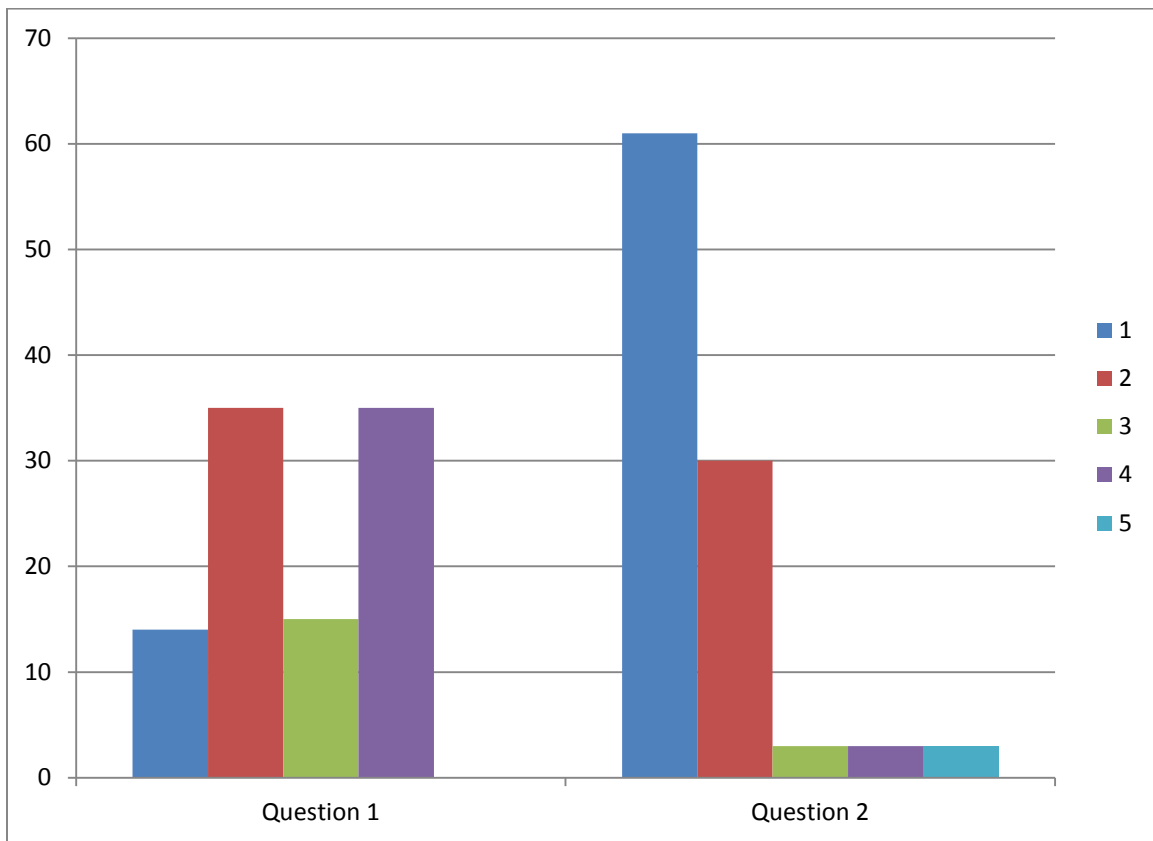


Table 3: Speaking

Question 1	It was difficult for me to speak in this context (unreal situation, lack of knowledge of the subject / information).
Question 2	The atmosphere during the oral part of the exam was relaxed and motivating.

In the writing subtest, around half of the respondents consider the tasks realistic. Similarly, it was neither easy nor difficult for them to write about the given prompts. Typing on the computer does not appear to be a significant issue, since only about 30 % reported problems with it.

Chart 4: Writing

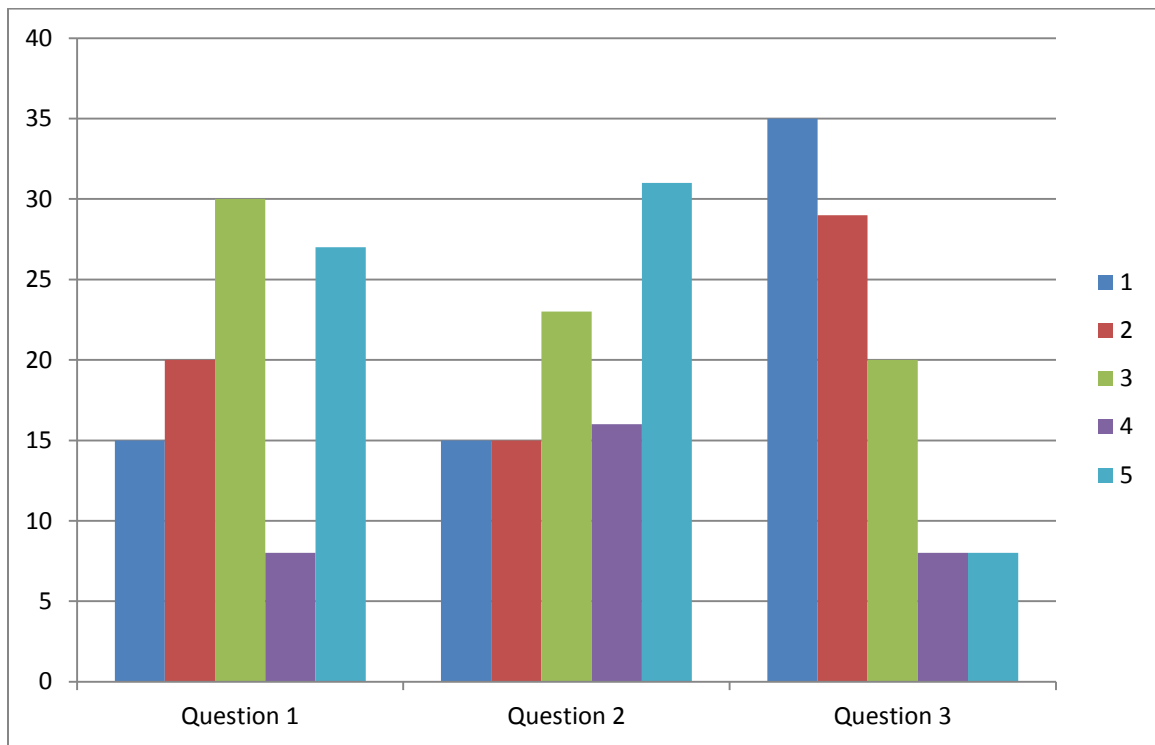


Table 4: Writing

Question 1	I considered the theme of the writing task to be realistic.
Question 2	It was hard for me to write in this context (unreal situation, lack of knowledge of the subject / information).
Question 3	I experienced problems with typing on the computer during the exam.

In Part 3 of the questionnaire, we wanted to ascertain the candidates' feelings after the exam, as well as their overall perception of the exam. Most of the respondents (almost 63 %) reported feeling insecure and doubtful about their performance. Nearly one-third of the respondents (29 %) did not feel well but thought they had passed. Only 8 % of respondents felt satisfied and considered the test to be easy.

One surprising finding concerned the use of practice exams available for student use on the centre's website: 66 % of the respondents reported that they did not use them.

The next question concerned the subjective view on the difficulty of individual skills; all respondents selected at least one option, while 41 % selected two skills as equally the most difficult. Listening comprehension appears to be the most difficult, as 50 % of the respondents have chosen it. Only 9 % fewer respondents identified speaking as the most difficult skill (41 %).

33 % of the candidates selected reading comprehension as a relatively easy language skill, which may have been expected. Most surprisingly, only 16 % of candidates considered writing as the most challenging skill to acquire. This finding is in the sharp contrast to our long-term experience, which suggests that the success rate in this skill has long been the lowest.

Our assumption is that since this examination is a professional one, it should correspond to the everyday working duties of the military personnel. As the preliminary results have shown, practical military life does not require active use of English: only 9 % of the candidates use English often at work, 16 % rarely and 75 % never.

Although the NATO STANAG 6001 examination has been designed for the military, it mostly tests general English. That is why we presumed that the candidates might welcome the prospect of incorporating a greater military context and more military-related vocabulary in the exams. Nonetheless, barely 20 % of respondents stated that they would.

The respondents were also given an opportunity to express themselves freely regarding what they would like to change or improve. Almost 80 % of them have taken advantage of this option and shared their opinions with us. The most frequent commentaries were related to the vocabulary that, in their opinion, was different from that taught in the course. Unfortunately, there is little that can be done about this as the exam is a proficiency test and is not intended to directly measure learning of specific course content, but rather of students' general level of language proficiency. Apart from vocabulary, candidates considered the grammar used, as well as the reading comprehension subtest as difficult. The examination is a bi-level test which might present difficulties to the respondents with lower levels; in fact, they expressed their preference for a mono-level examination. A highly positive response came from three candidates who were impressed by the testing board (speaking part) and generally by the professional approach of testers.

5. Discussion

Although we are fully aware of the fact that the sample can serve only as a rough orientation, the results are nonetheless of some value, though they must be considered cautiously. A summary of the most significant preliminary findings is in order.

As in the subtests of both reading and listening comprehension, around 50 % of candidates believe that some distractors are ambiguous and, hence, that more than one option is correct. That is why the next step we recommend that re-moderating these subtests with the use

of statics as a supporting resource as the next step. If, after this reconsideration, the testers find some items faulty, they should either rewrite or replace them. On the other hand, if testers, as well as statistics, reveal no deficiencies, the tests will not be changed.

Most of the candidates perceived the oral part of the examination on the proper level of difficulty and 60 % of them appreciated the positive and stimulating atmosphere.

Yet the most surprising result of our inquiry was undoubtedly that the candidates considered the writing part of the examination as neither easy nor difficult. This is, however, in sharp contradiction with long-term test results – writing has been for a long time the weakest of the skills. The issue surrounding difficulty in writing is underlined by the teachers' informal feedback stating that writing is the most feared of the subtests.

Another interesting finding surrounds the sample tests which are provided by the Centre on its website. Interestingly, we find out that 66 % of the respondents did use them.

As for the question concerning the most difficult skill, some of the respondents chose more than one and listening comprehension was identified as the most difficult skill by 50 % of the respondents.

6. Scope of Future Research

As the issue of language testing in the Czech military has attracted a lot of attention recently because of its increasingly paramount importance for professional soldiers, our workplace, the Language Centre of the University of Defence, has launched a professional discussion over the issue of testing. Moreover, it has started to question the tests and their validity. Its purpose was clear: to ensure that tests of such importance measure what they are intended to measure. Firstly, exam candidates were addressed and secondly, the teachers were given an opportunity to share their views on the tests. They were given some training concerning testing and on assessing productive skills. At the same time, the conditions were created for exchanging opinions between teachers and testers, workers of the same institution and sharing common goals. Before that, not much cooperation existed between these two groups.

After some informal but quite fruitful discussions, conclusions were drawn. Based on teachers' experience and their informal feedback from students, it became apparent that bi-level examinations may discriminate against candidates whose course was geared towards the lower level. Their test anxiety was reported to be very high and it can be assumed that their feelings of failure contributed to the very low success rate. Apart from tests of receptive skills, the most

pressing issue seems to be the situation in testing writing skills. The instructions are written in English and the language level of instructions is rather high. Thus, they spend a lot of time trying to understand what they should do, which takes from them the time they could spend on writing. Moreover, the second part of writing task is simply too difficult for them to complete, given the time limit and their lower level of proficiency.

This issue has led us to the conclusion that we should consider some changes in test format; or we that should perhaps question the concept of bi-level examination. We have initiated a study in which we are going to administer students (prior to the real exam) a mock writing part of the test with instructions written in both Czech and English (with a time gap of about a week) in order to compare their performance. We wish to clarify whether the language used in the instructions influences the test results or not. If yes, we are prepared to change the test format to create fair conditions for the candidates to perform better.

7. Conclusions

Having gathered the initial results, we have acquired an overview of the situation which enables us to better understand the opinions and impressions of the candidates regarding the language examinations which play such an important role in their careers. This is the first time that feedback has been garnered from the candidates and we consider it to be the first step towards a systematic approach to test validity, as well as launching closer cooperation and communication between teachers and testers. In addition, we expect to broaden our study by getting feedback from language teachers at our workplace. We have already taken some steps towards this endeavour by developing another questionnaire (for teachers) and designing a test familiarization seminar. Ultimately, we hope to construct a better argument for validation, thus creating the conditions for positive washback while ultimately improving the alignment between teaching and testing.

EVALUATION QUESTIONNAIRE

Dear Sir Madam,

We kindly ask you to fill out the following questionnaire regarding your views on the examinations in accordance with STANAG 6001. The survey is anonymous. Please complete it honestly and in

accordance with the actual situation so that the conclusions of this investigation may be considered reliable and valid. The knowledge acquired will be used for the innovation of testing content and process. When filling out the questionnaire, please follow the instructions for the individual questions. If not available, write in the true answer with a cross.

The Centre for Language Preparation

Part I.

The first section consists of identifying information. Responses are very important for statistical purposes.

We guarantee the anonymity of all respondents.

Please provide the following personal information.

1. Please indicate if you are a:

- a. student of University of Defence
 faculty:..... year of study:
- b. student at VŠŠ
- c. soldier
- d. civilian employee
- e. other

2. I just completed the examinations in accordance with STANAG 6001_____

- a. at levels 1-2
- b. at level 2-3

Part II.

Evaluation of the individual parts of the examination

Please evaluate the following statements concerning the individual parts of the exams based upon Standardized NATO Agreement 6001 on a scale of 1-5 in the appropriate box in the table provided below.

1 – strongly agree, 2 – somewhat agree, 3 – neutral, 4 – somewhat disagree, 5 – strongly disagree

Listening Comprehension							
Statement						YES	NO
The tempo (speed) of the language speakers was appropriate.	1	2	3	4	5		
I understood the content of individual test items without any problems.	1	2	3	4	5		
Although I understood the recordings, the correct answer occasionally seemed unclear.	1	2	3	4	5		

Reading Comprehension							
Statement						YES	NO

	I understood the content of individual test items without any problems.	1	2	3	4	5
	Although I understood the content of the individual texts, the correct answer occasionally seemed unclear.	1	2	3	4	5
Speaking						
Statement		YES			NO	
	It was difficult for me to speak in this context (unreal situation, lack of knowledge of the subject / information).	1	2	3	4	5
	The atmosphere during the oral part of the exam was relaxed and motivating.	1	2	3	4	5

Writing						
Statement		YES			NO	
	I considered the theme of the writing task to be realistic.	1	2	3	4	5
	It was hard for me to write in this context (unreal situation, lack of knowledge of the subject / information).	1	2	3	4	5
	I experienced problems with typing on the computer during the exam.	1	2	3	4	5

Part III.

3. **You have just completed your examinations in accordance with STANAG 6001. What are your feelings?**
 - a. I feel satisfied and the test seemed easy to me.
 - b. I feel insecure and doubtful of my performance.
 - c. I don't feel well but probably passed the exam.

4. **I used practice tests on the centre's website in order to prepare for this exam.**
 - a. yes
 - b. no

5. **Which part of the exam did you consider to be the most difficult?**
 - a. listening comprehension
 - b. reading comprehension
 - c. speaking
 - d. writing

6. **How often do you use English to perform work-related duties?**
 daily once a month often rarely never

7. **If you use English at work, which skills do you use? (you may select more than one)**
 - a. listening comprehension
 - b. reading comprehension
 - c. speaking
 - d. writing

8. **As it is a NATO exam, do you believe that the test should be more focused on a military context?**
yes – no

Space for your comments regarding the exam. What would you like to change or improve?

We sincerely thank you for taking the time to fill in this questionnaire.

References

- Alderson, J. C., Clapham, C., & Wall, D. (2006). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alimemaj, Z. (2015). *Language test taking in Albanian schools*. Retrieved from <http://eujournal.org/index.php/esj/article/view/6380>.
- Bachman, L. F., & Palmer, A. S. (2013). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Davidson, F., & Lynch, B. K. (2002). *Testcraft a teacher's guide to writing and using language test specifications*. New Haven: Yale University Press.
- Hughes, A. (2013). *Testing for language teachers*. Cambridge: Cambridge University Press.
- McNamara, T. F. (2008). *Language Testing*. Oxford: Oxford University Press.