

VALIDITY OF HIGH-STAKES MILITARY LANGUAGE TESTS FROM TEACHERS' VIEWPOINT

VALIDITA VOJENSKÝCH JAZYKOVÝCH ZKOUŠEK VYSOKÉ DŮLEŽITOSTI Z POHLEDU UČITELŮ

LUDMILA KOLÁČKOVÁ
MÁRIA ŠIKOLOVÁ

Abstract

The military personnel in NATO (North Atlantic Treaty Organization) countries is commonly tested by high-stakes standardized language tests in accordance with NATO STANAG (Standardization Agreement) 6001. Since passing or failing at the examination can significantly affect the soldiers' careers and personal lives, high validity, reliability and accountability has to be guaranteed. As far as the military in the Czech Republic is concerned, the only body responsible for quality assurance in language testing is the Language Centre of the University of Defence. All language specialists at the Centre are fully aware of the responsibility and permanently devote their efforts to set and maintain the measures to guarantee high validity and reliability of the tests and their results. Nevertheless, one gap that needed to be filled was finding out the teachers' perspective concerning the tests which resulted in launching this survey. The survey was based on a questionnaire distributed to the teachers electronically. The questionnaire covered 11 questions and was answered by 60 respondents. The questions focused on the teachers' opinions on examination in terms of its difficulty, objectivity, as well as problems connected with the preparation for it. The paper will present main results together with potential issues to discuss.

Keyword: language assessment; standardized examination; teachers' opinions; questionnaire; data analysis.

Abstrakt

Komunikační kompetence příslušníků armád v zemích NATO (Severoatlantická aliance) jsou běžně ověřovány pomocí standardizovaných jazykových zkoušek v souladu s NATO STANAG (Standardization Agreement) 6001. Vzhledem k tomu, že výsledek zkoušky může významně ovlivnit nejen kariéru, ale i osobní život vojáků, musí odpovědná instituce zaručovat vysokou validitu a spolehlivost testových výsledků. Pokud jde o situaci v Armádě České republiky, jedinou institucí odpovědnou za zajišťování kvality testování jazyků je Centrum jazykového vzdělávání Univerzity obrany. Všichni odborníci v centru si plně uvědomují svoji odpovědnost a trvale směřují své úsilí k zajištění vysoké validity a spolehlivosti jazykových testů a jejich výsledků. Jedinou mezerou, kterou bylo třeba v jejich aktivitách vyplnit, bylo zjištění pohledu učitelů na vlastnosti používaných testů. To vedlo autorky příspěvku k zahájení průzkumu, založeného na dotazníku, který elektronicky distribuovaly učitelům. Dotazník zahrnoval 11 otázek a odpovědělo na něj 60 respondentů. Otázky byly zaměřeny na názory učitelů na zkoušku z hlediska její obtížnosti, objektivity hodnocení i problémů souvisejících s přípravou na úspěšné složení zkoušky. Předkládaný příspěvek představuje hlavní výsledky tohoto průzkumu spolu s možnými dalšími problémy k diskusi.

Klíčová slova: ověřování komunikační kompetence; standardizovaná zkouška; názory učitelů; dotazník; analýza dat.

Introduction

Nowadays, high-stakes language tests are used for many purposes in different spheres of life. There has been tremendous amount of research in the area of language testing recently and thus, the institutions responsible for test design and administration have a lot of expertise to build their testing systems on. Nevertheless, each testing context has its specifics and so it is up to decision makers how they will apply the results of research in their conditions and manage their testing system.

The current study was conducted in the military context where high-stakes language tests have been used for more than two decades. During this period of time, the tests and testing teams have undergone a lot of changes, although for the whole time, the tests have been based on Standardization Agreement 6001 of NATO countries and their results have been of utmost importance. That is why attention has to be continuously paid to quality assurance of these tests by monitoring their validity and reliability.

The aim of this study was to scrutinize only one aspect falling into the category of validity, and it is namely data gathering and analysis of language teachers' views on these language tests.

Validity

Most of the experts nowadays agree that tests are valid if they accurately measure what they are supposed to measure (Ruch 1924, in Fulcher 2010; Davies et al, 1999; Hughes, 1992; Hennig in Alderson et al, 1995).

While some authors define various types of validity (Davies et al., 1999), the others perceive validity as just one concept. (Bachman, 1997; American Psychological Association, 1985). Winke uses the term "broad validity" which pertains to reliability, concurrent validity, predictive validity, and consequential validity (Winke, 2011).

Many scholars also claim that it is not enough just to state that the test is valid, but that evidence is needed to support this statement (Bachman, Palmer, 1996; Davidson, F., Lynch, B. K. 2002).

Another important idea connected with validity is that it is often not ascribed to a test as its feature, but rather it is related to meaningfulness and appropriateness of the interpretation of the scores (Kyriakides, 2004; Winke, 2011; Sax, 1997 in Kyriakides, 2004; Bachman, 1997; Mac Namarra, 2008, in Fulcher 2010; Messick 1989 in Davidson, F., Lynch, B. K. 2002).

Why teachers' views matter

High-stakes tests are not designed and administered in vacuum. Apart from the fact that they provide the test takers with a score/level, they have an impact on curriculum, teaching methods, teachers and learners. To face the often unlimited power of tests, Shohamy emphasises the need of collaborative and democratic approaches to assessment in which the testers are no longer "know it all"; on the contrary, she sees the testing covering joint effort of not only testers, but other participants of testing process, including teachers (Shohamy, 2008, p. 136).

In authors' opinion, teachers are definitely stakeholders in the educational context. That is why, as Winke states "The teachers' perspectives are therefore valuable pieces of information". As a result, teachers' views have to be taken into consideration when test validity is examined (Bachman, 1990; Hughes, 2003; Messick, 1989, all in Winke, 2011). Although teachers' contribution to validation process seems to be logical and clearly beneficial, in practice, teachers are not routinely involved in validation process (Crocker, 2002; Ryan, 2002; Haertel, 2002; in Winke).

Teachers' opinions on teaching, testing and assessing speaking skills in Finnish upper secondary schools have even become a research topic of a university thesis. The rationale behind the thesis was the fact that at the time of writing the thesis, testing oral skills was not a part of Matriculation Exam, but it was generally perceived as a shortcoming of such a high-stakes exam. Based on the study, the teachers showed positive opinions concerning teaching speaking skills, however, at the same time, they stressed some practical problems to be addressed in teaching

process (too big groups, too little time). That is why they were mostly opposing the inclusion of oral exam into Matriculation examination (Huuskonen and Kähkönen, 2006). Teachers' opinions on tests were scrutinized by different authors with various intentions (Laborda and Álvarez, 2011; Melouk, 2001; Turner, 2006).

Similarly, Kyriakides also considers teachers as important players involved in test validation process claiming that "teachers, as end-users of tests, contribute a distinctive perspective on validity, referred to as inferential validity". He emphasises the importance of investigating the test users' perceptions which he considers to be more important than the perceptions of subject-matter experts (2004). Freire claims that evaluating organizations should learn from educators. He adds that "without this attitude, the evaluators from an external organization will never admit to any gap between their view of reality and reality" (1985, in Shohamy).

Methodology, limitations

To address the population of teachers who, in authors' opinion, should have their say concerning the tests, the authors have decided to construct a questionnaire. As the authors wanted to produce a very respondent-friendly and at the same time not too time-consuming instrument, they have designed a questionnaire consisting of 11 items. Six of them were intended to gather quantitative data, while the rest of them (five) tried to find out personal opinions and attitudes of the teachers (the questionnaire is a part of the attachment).

The questionnaires were distributed via Google Document Management System which enabled to easily address the respondents, as well as to create a user-friendly environment for both respondents and the authors. The questionnaires were distributed from April to May 2020.

Following the trends of the 21st century, it also saved our environment. In addition, the system offered a relatively easy evaluation of the data, as an added value.

Results, interpretation and implications

Out of around 70 teachers teaching English at the Language Centre and thus mostly preparing the students for the examination, 60 teachers have filled up the questionnaires (cca 86 %). Eighty percent of the teachers believe that the content of the courses is designed to lead the students to pass the exam, while twenty per cent partly agree with this statement claiming that the range of the topics is too broad to be covered because of the insufficient number of lessons. Another reason for just partly agreeing was the lack of mock tests for the highest level of proficiency which makes it difficult to illustrate the level of difficulty.

The majority of the respondents (55 %) agree that subtests for individual skills (listening, speaking, reading, writing) are equal in terms of difficulty. The rest of them have presented their disagreement giving the following reasons: they have not been given the opportunity to familiarize with the individual subtests and thus, their opinion may only be based on the students' feedback. Based on this feedback they often deduce that the subtests of reading comprehension at levels 1 and 3 are too difficult if compared with subtests of other skills. Moreover, the texts in reading comprehension subtests are rather long and students would appreciate longer time to complete the test. Yet another respondent thinks that reading subtests require a broad vocabulary; nevertheless, in this respondent's view, productive skills are more difficult to acquire.

What the authors consider to be of profound importance is the way how the teachers see the difficulty of preparing for a particular skill at a particular level. To gather clear results, the respondents were asked to choose just one skill for each level. The table illustrates the results.

In the next item, the respondents were asked to give some details/reasons for their choice of the most difficult skill to prepare the students for on particular levels. For level 1, the most frequently mentioned reasons causing difficulties were as follows: too extensive course content, too little time to train for the test format, different entrance proficiency level, and disproportional range of vocabulary taught from the textbooks and required at the exam. The authors consider of

high significance also the statements concerning the insufficient course length, higher age of students, as well as the low ability to actively use language in productive skills.

	LISTENING COMPREHENSION	SPEAKING	READING COMPREHENSION	WRITING
Level 1	52 %	22 %	17 %	9 %
Level 2	35 %	18 %	10 %	37 %
Level 3	25 %	11 %	9 %	55 %

Table 1 Teachers' point of view: difficulty of preparing for a particular skill at a particular level

Source: Own

As far as L2 exam preparation is concerned, there is one problem similar to problems mentioned in connection with preparation for L1; it is the insufficient vocabulary needed to pass the exam (the range of vocabulary taught from the text books does not seem to be sufficient). Another problem mentioned specifically for this level is dealing with fossilised mistakes. A lot of students of these courses are able to communicate, but they keep making mistakes which consequently may result in failing the exam at this level where certain accuracy is required.

As the reasons for difficulties in preparation for L3 writing, the respondents have mentioned insufficient entrance level in terms of grammar, vocabulary, spelling, and stylistics. Moreover, they claim that the writing tasks that are required at the examination would not be mastered by the students even in their mother tongue. Those who identified receptive skills as the most difficult skills to prepare for claim that there is a lack of sufficient mock tests, as well as appropriate teaching materials.

As far as the perception of test fairness is concerned, almost a half of respondents consider the test to be fair. To specify the reasons for unfairness of the test, the teachers were given the opportunity to mark more than one reason. The most frequently mentioned reason of unfairness was insufficient time limit (42 %). Around a quarter of respondents ascribed the unfairness to the level of education of the candidates – in other words: the tests are too difficult for the candidates because of their insufficient or not fully corresponding education. The same percentage of them sees the reason of unfairness in the professional orientation of the candidates. Almost a third of teachers relates the problem of unfairness to inappropriate subject matter of the tests. The same proportion of them believe that the test measures something else than the language proficiency.

Seventy-one percent of the respondents state that the current system of examinations according to NATO STANAG 6001 has influenced their way of teaching. However, nineteen percent believe the testing system has not affected their teaching and the rest of them, ten percent, do not know. Almost all teachers find the mock speaking exams followed by feedback to be very beneficial/helpful. What surprised the authors, however, was that four teachers are not aware even of the existence of these procedures. All but three teachers see the writing mock tests followed by feedback from methodologists as very beneficial; three teachers do not know that the mock tests at the institution exist. As far as the reading and listening comprehension mock tests are concerned, all teachers find them helpful.

The last two items were open-ended and thus the responses to them offered a broad range of various opinions, suggestions and attitudes which the authors find on the one hand very beneficial, however, on the other hand, difficult to assess, categorize and generalise. The respondents were asked for suggestions in terms of possible improvement of the process of designing mock tests. The item seemed to be very inspiring, since 51 teachers responded. Their responses can be further divided into the areas they referred to. The most numerous part of the teachers prefer the current situation in which some (but not all) individual departments produce mock test items which are further moderated and, as a result, approved, modified or discarded by the methodology team. These respondents also recommended that the methodology team should

continue in this coordinating work. In addition, they suggested further enlargement of the team of teachers producing the items. More seminars/workshops are desired which will offer the teachers more insight into the complex process of designing items. Some of the teachers praised the work of methodology group concerning seminars which they have already organized and which focused on some basics of item writing.

At the same time, there were some critical voices pointing out to the fact that the items were not designed by all of the departments which was implicitly perceived as unfair. Another critical remark referred to the fact that not all departments have a computer lab at their disposal, and so they would appreciate if the mock test of receptive skills were administered not only electronically, but also in pen and paper form.

A teacher stressed that the teachers should have sufficient number of mock tests. Another respondent sees the solution in creating an item bank, with each teacher having the access to it on condition that everyone contributes and everyone shares. Another suggestion was to share the real tests which are not in use anymore. Nevertheless, several teachers admitted they have no idea how to improve the current system.

On the other hand, several others mentioned that testing department should produce mock tests, or they should have more personnel to do it. One respondent suggested that it should be the full responsibility of the methodology group. Other respondents offered various possible technical solutions of storing and administering mock tests of receptive skills. In addition, one teacher would like to share listening tests with other NATO countries which, in their opinion, may contribute to better calibration and equivalence of the tests in particular countries.

In the last item the respondents were encouraged to mention any suggestions, opinions, and/or remarks relevant to language teaching/learning. One respondent claims that the test does not properly reflect the real language needs of professional soldiers in their careers. The MoD should carefully consider which positions really need a certain level of English – otherwise, soldiers learn English being aware of the fact that they will never use it in real life; they might even pass the test, but after a short period of time, the proficiency will go down or disappear completely.

Another teacher expressed her/his worries concerning the soldiers attending intensive level 1 (survival) course. She/he claims that these learners are usually high school graduates and have not had an intention to study anything else, notwithstanding English language. Often, they have no idea of any effective learning styles or strategies. Their thinking is usually very concrete and this fact should be taken into consideration and addressed appropriately in the content of the course and teaching methodology.

Other voices focused again on mock tests articulating the need to have mock tests for receptive skills most urgently on level 3 – otherwise it is rather difficult to prepare the learners for the test well. They also mentioned the fact that the mock tests should not be compiled from non-functional items which, e.g. are on the Centre website, as they are easier than the real ones and thus might mislead the learners' expectations concerning the test difficulty.

There were two more critical opinions referring to Testing Department, one of them stating that the Department should not concentrate on creating items which are as difficult as possible. Moreover, one respondent stated that they feel as if testers did not perceive the teachers as partners, but rather as “thieves” who “want to steal their secret”. The respondent continues that when the teachers express their wish to see some real tests, it is not with the intention to disclose them to the students; what is behind this is their honest effort to grasp the way how the items are designed and what the level of difficulty is. Nevertheless, another teacher praised both methodologists and testers for their cooperation, assistance and willingness to help. One respondent highlighted the fact that the structure of the website of the Centre is a bit confused and does not enable an easy orientation and navigation. A teacher expressed her/himself that she/he has no idea how to do it in a better way.

Discussion

Eighty percent of the teachers agree that the content of the courses is designed to lead the students to pass the exam. At the same time, most of the respondents also claim that the current system of examinations according to NATO STANAG 6001 has influenced their way of teaching. As far as the publications on the issue of “teaching for the exam” are concerned, there are both proponents and opponents of the idea. Opponents of teaching for the exam perceive the meaning of the phrase rather pejorative, as in their opinion it implies narrowing of the curriculum and overemphasising the exam preparation (Berliner, 2011; Reeves, 2015). However, taking into account an enormous number of different language textbooks and courses explicitly claiming that they prepare for different internationally recognised standardized English examinations, the authors do not share the opinion, that “teaching for the exam” is a wrong way to adopt.

That is why, since the courses at the workplace of the authors are all finished by the high-stakes examination according to NATO STANAG 6001, the authors are in agreement with the idea that the courses should prepare for the examination. Having said that, however, they do not agree with narrowing the whole teaching process to training for the examination tasks, but they believe in positive effect of raising awareness of the exam in both teachers and learners and the positive affect of mock examinations followed by professional feedback with learning recommendations.

Teachers expressed their worries concerning level 1 courses, mentioning prevailing characteristics of the learners (high-school graduates, no or little previous experience with learning languages, relatively higher age, mostly concrete and straightforward way of thinking), as well as insufficient length of the courses. The skill they find most difficult to prepare at this proficiency level is listening. As this population and course are very specific, there is not any literature dealing with this combination of problems. The authors consider these ideas significant to approach and to repeatedly recommend to prolong the duration of the course to the MoD, and to brainstorm optimal methodology which would help to overcome the problems mentioned.

From teachers’ responses it is also apparent they consider mock tests of all skills very important and at the same time, they perceive the insufficient number of mock tests of receptive skills negatively, especially in level 3 courses. The responses also suggest that the teachers do not fully understand the way in which the test items are constructed. Consequently, they implicitly might question especially the tests of receptive skills for level 3. As mentioned above, the teachers ask for more mock tests of receptive skills and most of them appreciate the approach introduced recently in which the teachers create items, then the methodologist group moderate them and subsequently the approved items are collated into a complete test and handed over to the head of teaching department. The teachers realize their insufficient skills in item writing and hence appreciate the previous seminars run by methodologists explaining the basics of creating test items and would appreciate more training in this area.

Another suggestion how to address the problem of the lack of mock tests was to share the real tests which are not in use anymore – this is a blatant example of the lack of awareness of the teachers concerning testing – all functional items are being in use permanently, as far as they function well. The results of the questionnaire have nevertheless revealed also some critical opinions. Obviously, not all departments contribute to mock tests with test items, but all want to share the tests. Not all departments have computer labs, so electronic mock tests cannot be used in those workplaces. Although most of the respondents see the teachers’ – methodologists’ – testers’ cooperation as excellent, there was a critical remark concerning testers who are not willing to give teachers at their disposal some real tests to illustrate the real items and their difficulty claiming that test security would be put at stake.

A rather more general critical view has been presented as well stating that the examination does not properly reflect on real language needs of professional soldiers. That seems to be a rather serious claim and possibly justified; nevertheless, this is unfortunately not an issue to be dealt with by the Language Centre.

Conclusion

The main reason for launching this survey was the fact that teachers' views on the standardized language examination were not systematically taken into consideration, even though they are important stakeholders and their opinions may indicate both strengths and weaknesses of high-stakes exams. The survey has revealed some significant views and perceptions of the teachers which in authors' opinion should be addressed by the management of the Language Centre. The most frequently mentioned issues were as follows. The teachers specified the most difficult skills to prepare the students for the exam on all three levels – while at the lowest level the most problematic skill is listening, at intermediate level it is not only listening, but also writing and at the highest level for which the institution prepares, it is writing. These results may help teachers to put more emphasis on the mentioned skills at particular levels. Teachers mostly consider the syllabus to reflect properly the requirements for the examination. They negatively perceive the lack of mock exams and they would appreciate to further develop the system of mock exams design. They generally appreciate the activities conducted by methodology team. They would be glad if they could be shown some of the real tests for receptive skills to be able to better adjust teaching content and methods. Further work would be desirable to complement the findings from the questionnaire by semi-structured interviews with teachers to get a more detailed and in-depth understanding of their view and beliefs.

Literature

- ALDERSON, J. C., CLAPHAM, C., WALL, D. 1995. *Language Test Construction and Evaluation*. Cambridge University Press.
- BACHMAN, L. F. 1997. *Fundamental Considerations in Language Testing*. Oxford University Press.
- BACHMAN, L. F., PALMER, A. S. 1996. *Language Testing in Practice*. Oxford University Press.
- DAVIDSON, F., LYNCH, B. K. 2002. *Testcraft*. Yale: University Press.
- DAVIES, A., BROWN, A., ELDER, C., HILL, K., LUMLEY, T., McNAMARA, T. 1999. *Studies in Language Testing, Dictionary of Language Testing*. Cambridge: Cambridge University Press.
- FULCHER, G. 2010. *Practical Language Testing*. Hodder Education.
- HÖL, D. *Current Problems and Proposed Solutions for Testing Speaking: Opinions of EFL Teachers*. [on-line]. 2020. [accessed 2020-09-09]. Available from: https://www.researchgate.net/publication/328149684_The_Literacy_Trek_27_Current_Problem_s_and_Proposed_Solutions_for_Testing_Speaking_Opinions_of_EFL_Teachers
- HUGHES, A. 1992. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- HUUSKONEN, M. L., KÄHKÖNEN, M. 2006. *Practising, testing and assessing oral skills in Finnish upper secondary schools: teachers' opinions*. [on-line]. [accessed 2020-09-21]. Available from: https://jyx.jyu.fi/bitstream/handle/123456789/7375/1/URN_NBN_fi_jyu-2006384.pdf
- INKE, P. 2011. Evaluating the Validity of a High-Stakes ESL Test: Why Teachers' Perceptions Matter. In: *TESOL Quarterly, Vol. 45, Issue 4, pp. 628 – 660*.
- KYRIAKIDES, L. 2004. *Investigating validity from teachers' perspectives through their engagement in large-scale assessment: The Emergent Literacy Baseline Assessment project*. [on-line]. [accessed 2020-09-09]. Available from: <http://dx.doi.org/10.1080/0969594042000259466>

LABORDA, J. G., ÁLVAREZ, M. F. 2020. *Teachers' opinions towards the integration of oral tasks in the Spanish University Examination*. [on-line]. [accessed 2020-09-09]. Available from: <https://files.eric.ed.gov/fulltext/ED523843.pdf>

MAC NAMARRA, T. 2008. *Language testing*. Oxford: Oxford University Press.

SHOHAMY, E. 2001. *The Power of Tests*. Harlow, England; New York: Longman.

TURNER, C. E. 2006. Professionalism and High-Stakes Tests: Teachers' Perspectives When Dealing With Educational Change Introduced Through Provincial Exams, TESL Canada. In: *Journal Revue E TESL du Canada 55, Vol. 23, Issue 2*.

Kontakt

Mgr. Ludmila Koláčková, Ph. D.
University of Defence
Language Centre
Kounicova 65, 662 10 Brno
Czech Republic
Email: ludmila.kolackova@unob.cz

PhDr. Mária Šikolová, Ph. D.
University of Defence
Language Centre
Kounicova 65, 662 10 Brno
Czech Republic
Email: maria.sikolova@unob.cz