

Teachers in High-stakes Language Tests – Blessing or Curse?

Mária Šikolová, Ludmila Koláčková, Magdalena Veselá

Abstract: The article deals with the involvement of language teachers in high-stakes testing from different perspectives. The theoretical part discusses the problems of language teachers as testers in a broader context, in connection with the importance of language assessment literacy as a part of professional development. In addition, it reviews results of some studies focusing on teachers involved in testing. The main part of the article is devoted to the results of two surveys intended to gather and interpret the teachers' opinions on a standardized language examination used in the framework of NATO, which were conducted among the language teachers of authors' institution, as well as among the language teachers at military language institutions in several foreign countries. Furthermore, the main part presents the results of the questionnaire concerning the opinions and approaches of the teachers towards the item writing training which they received to be able to create test items. The objective of the presented study is to shed light on the teachers' views concerning their participation in testing, since most literature published so far deals predominantly with the principles of language assessment literacy from theoretical viewpoint, not from the viewpoint of the teachers themselves.

1 Introduction

Hand in hand with ever-growing role of the high-stakes tests, the accountability for their results has become an issue. How teachers participate in these tests varies in different contexts – in some institutions, they participate in both test development and rating productive skills, while elsewhere, they “only” teach the learners who then take an examination. However, teachers always need to use some tests for formative assessment in the classroom.

In the context of the Language Training Centre (further the Centre) where the study was conducted, the teachers are primarily responsible for teaching foreign languages, predominantly English, to the Czech military personnel. The language curriculum is built on a military document of high importance – NATO STANAG 6001 (North Atlantic Treaty Standardization Agreement), which was signed by the Czech Republic prior to its joining NATO. This document describes in detail five basic proficiency levels (from level 0 proficiency up to level 5, highly educated native speaker's proficiency), as well as four so-called plus levels in each of the four language skills. Apart from being a basic building block of the curriculum, the document also serves as a solid foundation for test design in NATO member and partner countries.

The quality assurance at the workplace of the authors of this article is of great importance, since the results of the high-stakes test influence the careers of hun-

dreds of soldiers. Nevertheless, the role of the teachers lies predominantly in teaching, bearing in mind the descriptors and requirements for the high-stakes examination. Currently, most of the teachers do not directly participate in the examination. This, in our opinion, however, does not mean that their training in assessment should be neglected.

Standardized high-stakes tests are both developed and administered at the Centre, though not by teachers themselves, but by testing specialists. As mentioned above, the results of the tests have a serious impact on the careers and lives of the test takers, the Czech professional soldiers. That is why we presume that the teachers preparing them for passing these high-stakes tests are important stakeholders who should have their say in testing processes and hence, should have at least some basic knowledge of the principles of testing theory.

The study analyses the results of three different sets of data mapping the Czech teachers' perception of the examination in accordance with NATO STANAG 6001, foreign teachers' perception of the same examination, and the feedback from the Czech teachers who participated in item writing training.

2 Theoretical background

Depending on the context in which teachers need to acquire at least a certain level of testing awareness, they need testing training in different areas, such as item writing, rating speaking, rating writing, or conducting the speaking examination. Nevertheless, all these training needs could be placed under one umbrella – language assessment literacy (further LAL).

The authors dealing with this topic see assessment literacy as a combination of knowledge and skills: “Assessment literacy consists of an individual’s understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (Popham 2011, p. 267; in Glenn Fulcher in Language Assessment Literacy, 2020, p. 8). Fulcher offers a similar understanding of the content of language assessment literacy, emphasizing a bit more practical aspects seeing the assessment literacy in “designing and building tests for their (teachers’) own use, and the institutions for which they work” (Glenn Fulcher in Language Assessment Literacy, 2020, p. 11). Inbar-Lourie defines language assessment literacy from a more general perspective as a complex competence needed for effective testing (Inbar-Lourie, 2017, in Bohn, Tzagari, 2021, p. 222). Davies (2008) sees the pillars of the ability to assess language in the following areas: “knowledge, skills, and principles related to the assessment of language ability” (Davies in Frank, 2018, p. 183).

We fully agree with Vogt and Tzagari that “Training of pre- and in-service teachers constitutes one of the most important aspects in the quality assurance of lan-

guage testing and assessment (LTA)” (Vogt, Tsagari; 2014, p. 374; see also Boyd and Donnarumma, 2018, p. 120). They further claim that teachers of foreign languages come into contact with both standardised tests and common classroom tests, and that is why it is fair to offer them appropriate training in assessment (Vogt, Tsagari; 2014, p. 374).

Bohn and Tsagari argue that teachers should have appropriate curriculum knowledge as well as skills in formative assessment. The authors further conclude that the awareness of test fairness should also be fostered in teacher training, both in university and in-service programmes (Bohn, Tsagari, 2021).

Boyd and David Donnarumma also recommend teachers’ training in assessment; in addition, they point out to the challenge which they see in “the sheer quantity of information to be disseminated in a short period of time and the approach to that training” (2018, p. 105).

Another important point related to teachers and high-stakes tests is the impact of the tests on teaching as such, the content of curriculum, the teaching methods and approaches. In connection with this issue, Vogt and Tsagari (2014, p. 391) emphasize the necessity of teachers’ ability of critical evaluation of the tests (in Frank, 2018, p. 185).

The idea which has attracted our attention and which we have tried to address in our study concerns the problem mentioned by Frank. He claims that the research results in the area of LAL have predominantly offered “a top-down perspective” while the way how the teachers perceive it has not been broadly dealt with in the literature (Frank, 2018, p. 185, see also Vogt, Tsagari; 2014, p. 375).

3 Research goals of the data gathering and results

3.1 Czech Teachers’ Perception of examinations in accordance with NATO STANAG 6001

Teachers at the Language Centre of the University of Defence teach English to both university students and professional soldiers. Their task is mostly to prepare them for passing the standardised examination in accordance with NATO STANAG 6001. That is why the authors wanted to know the teachers’ perception of the examination. Another reason for this survey was to contribute to the validity of the test and test results. A questionnaire was designed and distributed using Google Forms and the results were summarised and evaluated; subsequently, they were presented at the workplace for the management in order to improve the quality assurance. The questionnaires were responded to by 60 teachers, which represents 86% of all addressed respondents from April to May 2020. The questionnaire consisted of 11 questions related to teachers’ opinions on several issues.

Items pertained to the alignment between the syllabus and the examination, the difficulty of the parts of the examination, the difficulty of exam preparation for specific skills at particular proficiency levels, and the fairness of the examination. Additionally, teachers were asked about their personal perception of the wash-back of the examination on their teaching styles and the importance of the mock examination in individual skills. Only the responses to the items relevant to the focus of this paper are further mentioned and commented on.

Around 45% of the teachers do not think that particular parts of the tests are of the same difficulty, although their opinions are based on students' feedback only, as they have not the access to the tests, which is because of threat to test security. The subtests which in their opinion are significantly more difficult than the rest of them are reading comprehension subtests for proficiency levels 1 (survival) and 3 (fully professional).

Test fairness from the viewpoint of teachers was another concept that the authors were interested in. Around a half of them finds the tests fair; the reasons behind the perceived unfairness were identified differently – by 22% respondents it was inadequate time limit; other reasons which were given were connected with the content of the test that, in their opinion, does not correspond with the level and the specialty (military) of students' education (around 17% of the respondents). A relatively low proportion of the respondents (21%) think that the test measures something else than language proficiency level. The test was considered to be unfair because of the level of the education of the students by around 14% of the respondents.

More or less all teachers appreciate an opportunity given to the students to do mock tests in all skills; almost all respondents emphasized the need for more mock tests. Speaking and writing mock exams have in their opinion an added value consisting in feedback provided by methodologists who conduct the mock exam; otherwise, they are accredited testers and they regularly examine.

In the open-ended items the respondents showed their interest in taking part in more workshops to raise awareness in item writing. An urgent need for level 3 reading and listening comprehension mock test was repeatedly mentioned. In addition, one respondent pointed out to the requirements for language levels of professional soldiers which, in his/her opinion have not been set appropriately. Another respondent expressed their opinion that there are some tensions between teachers and testers. Nevertheless, another opinion illustrated good relations and understanding between teachers, testers, and methodologists.

3.2 Foreign Teachers' Perception of examinations IAW STANAG 6001

As soon as the results of the questionnaire were evaluated, the authors realized that they could acquire an added value to the results if they compare them with the results from colleagues from partner workplaces abroad. The same questionnaire was thus distributed to the contacts that the authors have abroad. The questionnaires were responded to by 11 teachers which represents around 50% of all addressed respondents from three countries (Hungary, Croatia, and Poland in June 2020). The questionnaire consisted of 11 questions related to teachers' opinions on several issues. Items pertained to the alignment between the syllabus and the examination, the difficulty of the parts of the examination, the difficulty of exam preparation for specific skills at particular proficiency levels, and the fairness of the examination. Additionally, teachers were asked about their personal perception of the washback of the examination on their teaching styles and the importance of the mock examination in individual skills. Only the responses to the items relevant to the focus of this paper are further mentioned and commented on.

In comparison with the results of the questionnaires distributed among the Czech teachers (45%), more foreign teachers do not think that particular parts of the tests are of the same difficulty (55%). 36% of the respondents considered the subtests of partly the same difficulty. The respondents who do not find the individual subtests of the same difficulty did not specify which subtests they consider to be more or less difficult than the others.

Around 45% of the respondents find the test fair; the main reasons for perceiving the test as unfair were the level of education of the candidates (45%), the content of the test which in their opinion does not correspond with the military profession (18%); measuring something else than the level of language (9%) and insufficient time limit (9%).

In the open-ended questions, one respondent added another reason for the test being unfair towards the students: "Unfair because a range of topics is wide and they are expected to demonstrate their knowledge in terms of issues, accurate grammar and sophisticated vocabulary."

All respondents administer the mock tests in their courses, but they did not comment on their importance.

In the space given to other opinions and suggestions, two respondents highlighted the disparity between the interpretations of the NATO STANAG 6001 descriptors in different countries suggesting that more attention should be paid to the language examination standardization among NATO member states.

3.3 Czech Teachers' Opinions on Item Writing Training

Since the situation at the authors' workplace is very dynamic, an idea was conceived that teachers should start writing mock test items. It had several reasons, one of them ensued from the results of the previous questionnaire (described above) that both teachers and students lack mock tests. Besides, since the Language Centre during the pandemic in the lockdown, before the start of using MS Teams, needed to produce some meaningful work, this idea looked feasible. However, after first attempts of item writing, the specialists who were in charge of moderation test items created by teacher realized that the teachers desperately needed some training. So it was decided to organize seminars for the teachers, delivered via MS Teams. To find out how effective this training was and how the teachers perceived it, a different questionnaire was designed and distributed in April 2022. Sixty-eight teachers of the authors' workplace were addressed and altogether 17 replies were received. The questionnaire consisted of 12 questions designed to uncover whether the respondents are aware of the existence of Language Assessment Literacy; further, the questionnaire searched for opinions concerning mock item development and who should be responsible for it, their motivation to create mock test items, the strengths and challenges of workshops for item writing, as well as their potential interest in and suggestions for future training in the area of mock item writing. Only the responses pertaining to the area of our research interests are further mentioned and commented on.

Almost 70% of the respondents are aware of the concept of Language Assessment Literacy and 90% of them agree that teachers should acquire at least some basic competency in assessment. All of them further clarified their opinion giving different reasons in an open-ended questionnaire item; most of the respondents emphasized the need to be familiar with the examination, its format and assessment criteria to make their teaching more meaningful and effective. Just one of the respondents thinks that the teachers have sufficient knowledge of assessment basics.

Another item asked the respondents if in their opinion mock test items should be written by teachers and more than half of them agrees they should. In the follow-up question they were asked for the reasons why, with a possibility of selecting out of three reasons. More than 60% gave as the reason the lack of mock tests and their necessity; almost 30% think that teachers should know basic principles of item development and about 10% state that teachers need practical experience with item writing.

As far as the item workshops organized at the workplace are concerned, almost 65% find them helpful and around 18% very helpful; the rest of the respondents have either not taken part in them, or they did not find them very useful. More

than 60% of the teachers find the feedback concerning their items very useful or useful.

Seventy per cent of the respondents would appreciate further training in testing; most of them training in assessment of productive skills (more than 80% in speaking assessment, around 75% in writing assessment).

4 Discussion

It is rather complicated to compare the results of the first two surveys concerning the teachers' perception of the standardized examination with the results of other studies, since both populations and examinations are very dissimilar. Nevertheless, some authors have studied if not identical, so at least similar issues, one of them definitely being the perception of broader fairness of the tests. While around a half of the respondents of both Czech and foreign language teachers in the military find the standardized high-stakes test fair, other studies have brought different results. A recent study which was conducted by Scholastic and the Bill and Melinda Gates Foundation concerning teacher perceptions of high-stakes testing and showed the results of a survey of more than 10,000 teachers revealed that and "only 26% thought that the tests were a good reflection of what their students knew" (Rebora, A., 2012, in Gunn et al, 2016, p. 52).

The reasons why the teachers in the Czech and foreign military find the examination unfair were various, such as the level of education of the students, the content of the test, the test measuring something else than the level of language, and insufficient time limit.

Other authors did not search for answers to such concrete questions, however, the teachers in some other studies expressed their opinions that high-stakes tests do not accurately measure what their students have learnt (Gunn et al, 2016, p. 58; Reese et al, 2004, in Gunn et al, p. 53) or they even expressed their "concerns about the validity of high-stakes tests" (Gunn et al, 2016, p. 61). Teachers in our study expressed their worries that the test measures something else than language proficiency as well (around 20% of Czech teachers and less than 10% of foreign teachers).

Seventy per cent of the respondents in the present study expressed their wish in having more training in testing.

In the context of teachers in testing in Spain, more than 60% of the teachers would appreciate training in item writing for reading and listening tests; more than a half of them think they need more training in assessment principles (Cristina Rodrigues, 2018 Teachers in Testing, Spain, in Teacher Involvement in

High-Stakes Language Testing; EOI (Escola Oficial de Idiomas) Official Schools of Languages; p. 140).

More than half of the respondents of our study agree with their involvement in writing mock test items. The teachers at the authors' workplace are not required to develop high-stakes test items, as there is a department specialized in it. Another reasons why they do not develop items for the examinations is their focus on teaching, as well as an issue of test security.

This is in agreement with Kremmel et al. who presents almost the same reasons against teachers' involvement in test tasks development for high-stakes examination: "lack of time (73%), a potential threat to test security (52%)" (Benjamin Kremmel et al., 2018, p. 185, Item Writer Attitudes Towards Involving Teachers in General, in *Teacher Involvement in High-Stakes Language Testing*).

5 Conclusions

Based on the results of our study, with solid arguments from the studies from other countries, authors and contexts, we can state that teachers' involvement in testing is nowadays becoming a must. Yet, the authors are aware of the limitations in the study, especially in terms of the number of returned questionnaires. Another deficiency of the study is that the respondents of the last questionnaire (3) expressed their opinions strictly limited to the examination they prepare the learners for, instead of broader view of LAL.

Nevertheless, LAL as such evolves in various contexts differently and it may be applied in numerous ways. Undoubtedly, from a broader perspective we believe teachers should acquire language assessment literacy, probably according to the conditions and needs of the context in which teachers teach and high-stakes tests are developed and administered.

If teachers are in direct contact with tests and testing, it can not only enrich them, but it can also contribute to the validity of test results.

Although the conditions in which language teaching and testing is conducted vary to a large extent, we still think that research in this field should continue, as it can help to shed light on further aspects affecting teaching and learning languages.

References

- BØHN, H., & TSAGARI, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>
- GIRALDO, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>

- GUNN, J., AL-BATAINEH, A., & ABU AL-RUB, M. (2016). Teachers' perceptions of high-stakes testing. *International Journal of Teaching & Education, IV*(2). <https://doi.org/10.20472/te.2016.4.2.003>
- TSAGARI, D. (2020). *Language assessment literacy: From theory to practice*. Cambridge Scholars Publishing.
- VOGT, K., & TSAGARI, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- XERRI, D., & BRIFFA, P. V. (2018). *Teacher involvement in high-stakes language testing*. Springer.

Authors

Mária Šikolová, PhD., Language Centre of the University of Defence, e-mail: maria.sikolova@unob.cz graduated from the Faculty of Arts of Comenius University in Bratislava, Slovakia, with a focus on translating and interpreting in English and Arabic. After graduating, she started working at the University of Defence, and throughout her career, she has served in various positions, beginning as a lecturer of Arabic and English. Subsequently, she held various managing positions (Deputy Director, Head of the English Department, Director of Studies, Head of the Testing Department, Chief Methodologist). She currently works as a methodologist at the same institution. Apart from both attending and conducting numerous workshops and seminars, she defended her Ph.D. thesis at Pedagogical Faculty of Charles University in 2007. Her primary areas of interest and publication are teaching and testing methodologies, teacher training, and quality assurance.

Ludmila Koláčková, PhD., Language Centre of the University of Defence, e-mail: ludmila.kolackova@unob.cz works at the Language Centre of the University of Defence (Brno, Czech Republic) where she has been a faculty member since 2004 and currently serves as the Head of the 2English Department. Prior to her present assignment, she worked there as a lecturer of English and Czech for Foreigners. Dr. Koláčková completed her PhD. at Jan Evangelista Purkyně University in 2014 and her undergraduate studies at Masaryk University in 1989. Her research interests lie in the area of languages, ranging from theory to design to implementation. She has served on the Association of Teachers of Czech as a Foreign Language (AUCCJ) Executive Committee, and she was the president of the Czech and Slovak Association of Language Centres (CASALC) where she serves as a committee member. She is currently a chief of the Editorial Board of the CASALC Review professional journal.

Magdalena Veselá, Language Centre, 2nd department, e-mail: magdalena.vesela@unob.cz studied French and English at Masaryk University and has been employed as a faculty member at the University of Defence (Brno, Czech Republic) at the Language Centre since 2016. She prepares students for high-stakes exams and is an integral part of a research group. Before her present assignment, she worked as a private teacher and gained skills worldwide. Her interests in the research area cover topics of teaching languages and linguistics, and she is keen on transferring the acquired knowledge into the classroom. She mainly focuses on adult teaching, learning, and preparing her own materials. Another active part of her work life is testing the international military exams, ranging from elementary to advanced level.