

<https://doi.org/10.28925/2617-5266.2023.87>

TESTING AS A LANGUAGE ASSESSMENT TOOL: UKRAINIAN EXPERIENCE

Oksana Vysotska

ORCID ID 0000-0001-9343-9538

PhD in Philology, Associate Professor

Department of the English Language and Communication

Borys Grinchenko Kyiv University,

18/2 Bulvarno-Kudriavska Str, Kyiv, Ukraine, 04053

o.vysotska@kubg.edu.ua

ABSTRACT

There are lots of books devoted to test design techniques for teachers to critically evaluate and reflect on assessment items developed by them and for them. However, the problem is whether teachers, who are “jacks of all trades,” are assessment and test-writing literate, whether they are aware of assessment purposes, they understand and can interpret methods and principles of test item design. The question arises whether, with the help of tests, teachers can ensure valid inferences about test takers’ knowledge and skills, and whether teachers who become item writers and test designers have enough knowledge and skills for this complicated task. The article contributes to helping Ukrainian university teachers reflect on their assessment and evaluation knowledge and skills. Hopefully, the article may raise awareness of the importance of a deeper theoretical and practical expertise in high-quality assessment for language education. Besides, the article may stimulate teachers’ attention and focus on complicated problems around test creation, administration, and analysis.

Keywords: *assessment, knowledge and skills; test; test design; test-task.*

© Oksana Vysotska, 2023

INTRODUCTION

In today’s globalised world, high proficiency in foreign languages is of great demand, especially if the languages are international ones. With a foreign language command, students are able to get education in universities where instruction is not in their native language. They may travel or get employment all over the world. English, the most widely spread of international languages, besides being a native language in a number of countries and a second state language in others, is used as a working language practically across Europe.

Growing demands for high-level users of languages is accompanied by an increasing demand for accountability in language teaching. “Corporations

and businesses are increasingly expecting educational institutions — schools, colleges, and universities — to produce potential employees whose language ability is sufficient for them to function in a multilingual workplace” (Bachman, 2013, p. 2). A bright example of the importance of English is the intention of the Ukrainian government to employ and appoint those officials whose language proficiency is sufficient for using it for communication.

It is understandable that language assessment, in general, and testing, in particular, are also used wider than ever before. There has appeared a separate branch of science which deals with theoretical and methodological foundation and principles of test development and their application in monitoring and evaluating the quality of education. Testing has become an important tool for assessing students’ achievements and the development of their knowledge and skills in any subject, foreign languages included. It is used in order to obtain evidence of language use inside and outside classroom settings, to interpret it, to infer students’ overall language proficiency, and to predict their language performance in real-world situations (Ashraf & Zaki, 2019, p. 164). There are international communities and forums of experts in testing, ILTA (International Language Testing Association) being just one example of them.

It is worth mentioning that the COVID-19 pandemic has launched a digitally-based format of testing and virtual exams in higher education. For example, Morrison & Sepulveda-Escobar (2022) state that online assessment has become pivotal to allow evaluation of the knowledge and competences of teacher students. Mospan et al. (2022) “show a significant increase in EdTech use for online teaching and testing”, particularly for conducting final achievement and progress achievement tests (p. 97). Besides, “the examiners were engaged in designing appropriate distance exam forms, exam materials and choosing a collaboration platform” (p. 98).

Though traditional forms of assessment such as oral and written students’ answers, credit tests and examinations are not done away with because of the development and wide employment of new technology, testing has become an indispensable part of the teaching / learning process. Properly designed tests provide the opportunity to evaluate the quality of education in an objective, fast, and unbiased way.

In this context, **the objectives** of the article — to present testing as a form of language assessment, its principles, stages and elements as well as to reflect on challenges experienced by test writers and test administrators, particularly in Ukrainian education.

METHODOLOGY

The research objectives are reached by applying a literature review and examining the state of the Ukrainian experience of designing and conducting tests in higher education.

RESULTS

Testing as a form of language assessment. It is evident that the employment of tests to assess students' language abilities requires in-depth knowledge on the part of teachers. First and foremost, before designing any test, test developers should identify the purpose for which the test will be created, and a corresponding test specification should be worked out. The Test Specification is the main document on which the test is based. It is the so-called test outline (Brown & Abeywickrama, 2019, p. 63), the rationale that guides the test design. It describes the test content and form, the number and level of test tasks, the assessment system for particular tasks and the test as a whole. The developed test tasks should make it possible to achieve the purpose of the test.

Foreign language tests may be designed for different purposes:

- Diagnostic tests are employed to diagnose language strengths and weaknesses. They present valuable information for teachers to make proper decisions to help their students overcome language problems and make teaching more effective.
- Achievement tests assess students' achievement of specific objectives in the teaching / learning process.
- Placement tests are administered to identify students' level of language acquisition to place them in a particular course appropriate for their level. Mistakes in incorporating students in a group suitable for his / her initial proficiency level may lead to their frustration, boredom, lack of motivation, and failing the course.
- Proficiency tests are developed to measure the general level of language mastery.
- Progress tests help teachers check students' achievements at the end of the lesson, unit, course, or term.

Taking into consideration the time of holding the control, tests may be divided into formative and summative ones. Formative tests are conducted to monitor students' learning and to provide them with ongoing feedback. Formative tests identify areas for improvement for students as well as for teachers with respect to their teaching. Intermediate assessment takes place in the middle of the course to identify how well students are ready for the final examination. Summative assessment tests are administered to measure or sum up what students have learned or can do at the end of the course of instruction.

According to the interpretation of test results, tests may be norm-referenced or criterion-referenced ones. Ukrainian External Independent School-Leaving Examination refers to norm-referenced tests, as its score is interpreted in comparison with the norm — the average score obtained by the whole bulk of school-leavers. The result of the test indicates whether the student has achieved the requirements of the standard — the norm (curriculum, syllabus) and what he or she can do, compared to others. Criterion-referenced tests measure

candidates' performance against a set of criteria which are predetermined and described before the administration of the test.

If one takes into consideration the formal aspect — test structure, the following kinds of test tasks may be distinguished: multiple choice, gap-filling, cloze, true / false / not-given, matching, and open-ended tests.

The books devoted to test design, critically evaluate and reflect on assessment items developed or used by teachers (Alderson et al., 1995; Brown & Abeywickrama, 2010; Brown, 2011; Buck, 2001; Carr, 2011; Green, 2014; Hughes, 2002; McNamara, 2004). They help test item writers create good, quality test tasks. These manuals “walk readers through the genesis, qualities, development, and evaluation of assessments in general” (Giraldo, 2019, p. 130). The main requirements to test may be summarized as a checklist:

- If tests are aimed at assessing candidates' ability to communicate, they are based on communicative principle, not the traditional, linguistic one, i.e., test tasks or items are based on the text which reflects the natural authentic language. On the other hand, grammar and vocabulary tasks used in isolation, in other words, if they are decontextualized, will contradict the principle of assessing communication in the real, authentic language.
- The test is not biased.
- Item stems (questions or statements) do not contain vocabulary unknown to test takers.
- Items are independent and do not overlap or use the same words.
- It is impossible to answer using previous options or the ones that follow.
- Correct options are not based on the universal knowledge or test takers' experience. In this case, the item does not assess the construct.
- Stems are short, options / distractors are of the same length, semantically related, plausible, structurally homogeneous, without redundancies and repetitions.
- Absolute words (“always,” “never,” “only”, etc.) should be avoided.
- As a rule, there is only one unambiguously correct answer.
- Instructions and rubrics are concise and clear to test takers, and test items correspond to the level of their proficiency.

Besides these “nuts and bolts” for writing correct test tasks, it is essential to analyse other qualities of a good test which are to be taken into consideration. Bachman (1996) defined six test qualities of a test. They are *reliability*, *validity*, *authenticity*, *interactiveness*, *impact*, and *practicality*. These characteristics of the test help assess whether the test is properly created and can be used and trusted or not.

Test reliability. Reliability means that the results of the test do not depend on the time and place of holding it, nor on the experts who assess it. It means that if the same test is taken twice by the same examinee at different times, the score will be the same. In statistical terms, reliability is determined with the help of the Reliability Coefficient, Cronbach's alpha (which is calculated

by using a mathematical formula and measured from 0 to 1 scale), test-retest, and inter-rating techniques.

Cohen (1994) revealed the factors affecting test reliability:

Test factors. Assessment criteria not properly developed, low level of item discrimination, invalid construct, ambiguity of instructions which are not clear, improper layout of the test, the test's format unfamiliar to test takers, quality and quantity of items (which should not be too easy or too difficult), a few test tasks. The longer the test is, the more reliable it is. "The more opportunities we give test takers to show what they know, the more accurate and fairer the measurement is likely to be" (Douglas, 2010, p. 4).

Situational factors. Qualities of the physical space, lighting, quality of equipment used, wrong time allotted to the fulfilment of the test.

Individual factors. The physical health and psychological state of test takers, their cognitive abilities and motivation (Cohen, 1994; De Benedetti, 2006).

Validity. Speaking about the test validity, construct and content validities are distinguished.

Construct validity. "Of all the concepts in testing and measurement, it may be argued, validity is the most basic and far-reaching, for without validity, a test, measure or observation and any inferences made from it are meaningless" (Hublely & Zumbo, 1996, p. 207). Besides, one of the major requirements to any language assessment, tests included, is valid interpretations made about test takers' language ability on the basis of assessment results. There should be confidence that the test measures what it is expected to measure, that the interpretations of the test results are justified and meaningful, and that "a given test score can be seen to be an indicator of some abstract theoretical model of language ability" (Bachman, 2013, p. 1-5). The Common European Framework of Reference with its levels of skills descriptors significantly helps understand what language ability is required for this or that particular skill at this or that particular level of proficiency (CEFR, 2001). However, it may happen that, developing tests, teachers trust their intuition, the so-called "gut feeling" which is definitely subjective. If they refer to the CEFR level and skills descriptors, they get a tool that serves "as a reference, a lighthouse to look at for guidance" (Ramirez, 2013, p. 198).

Construct may be defined as the idea or theory, something consisting of different parts is based on. As Bachman defines it, a test construct is an abstract entity, the language ability which the test should measure. Thus, construct validity shows that the test is used for the purpose it has been constructed for. It is possible to say that the construct validity of the test reflects the extent to which the test scores could be interpreted as indicators of the construct or the ability it is intended to measure (Bachman, 2013, p. 1). "Construct validity is about how well a test measures the concept it was designed to evaluate. It presents justification of the interpretation of the test results", in other words, "the extent to which evidence can be found to support the underlying theoretical concept on which the test is based" (Rapi & Miconi, 2014, p. 163).

In order to determine whether the test is valid and to prove that it measures what it is intended to measure, one may resort to the analysis of its descriptive statistics and analyse biserial correlation coefficient, discrimination index, and distribution curve. The other important psychometric characteristics of the test are the mean value (maximum and minimal results) and bias.

Content Validity of the test indicates to which extent the content of a test is sufficiently representative and comprehensive for the purpose of its design (Douglas, 2010; Eckes, & Althaus, 2020).

Authenticity deals with the language presented in the test tasks. It should be authentic, original, live language spoken by native speakers at the present time, today. One of the most difficult and time-consuming tasks in test item creation is to find a suitable text which will meet the requirements of the test specification. It should not be biased — should not violate rights of test takers or discriminate against them on any ground. It should be appropriate for the test-takers level and on the topics included into the curriculum / syllabus.

Interactiveness is about what test-takers bring to the test. It means that the test is designed for a particular group of learners, with their language ability, knowledge of the topic, and the test format.

Practicality is the ease with which the test can be administered and assessed — interpreted and graded (Bachman, 2013, p. 9).

Summing up, the process of test design may be divided into several stages (Hughes, 2002, p. 58-74):

- Stating the problem.
- Writing the test specification.
- Developing test items.
- Trialling on native speakers.
- Trialling on non-native speakers.
- Analysing test results.
- Calibrating scales.
- Validating the test.
- Training the staff involved in test design, administration, and assessing.

Language testing in Ukraine. It is worth mentioning, that Ukraine has been accumulating experience in foreign language testing. The development of the National English Test was initiated by the British Council in Ukraine, Delaso UK Ltd, and the Ukrainian Ministry of Education and Science in 2006. It started with the training course on the theory and practice of test design for item writers, development of the test specification, development of the first tests (to test reading skills) and their trialling. Later the project continued under the auspices of USETI Alliance / American Councils, the Ukrainian Education Quality Assurance Centre, and the Ukrainian Ministry of Education and Science. In 2009, the External Independent Test of foreign language proficiency was introduced. At the beginning, the test comprised the parts to test reading and writing skills plus the use of the language. Then, in 2017, listening was added. Speaking is waiting for its turn.

High-stake tests, including Ukrainian External Independent Test for school-leavers, observe the above procedure. The author of the article with her students at Donetsk National University, Ukraine, participated in pretesting for ESOL, in particular, ILEC and IELTS examinations. This pretesting was important for ESOL since, before the administration of such high-stake tests, it should be trialled on a great number of candidates all over the world. For students, testees, it was useful as they got acquainted with the format and the procedure of these international tests. They also got feedback — information about the obtained score, as the examination score was correlated with the real examination band or CEFR level.

Tests designed at universities are different. A particular example of using testing at Ukrainian higher education institutions is the experience of the author of the article. Tests are widely used by teachers to assess students' language ability, for formative assessment as well as the summative one, at the end of the module or the course. The most suitable examination form was developed, with the examination paper including a test on LMS Moodle and students' direct answers during virtual communication (Mospan et al., 2022, p. 98). The examination format makes it possible to check all the main speech skills plus translation.

Tests are developed by teachers themselves and attached to the e-course on the LMS Moodle. Students have to do three tests every week, a vocabulary test, a grammar test, and a listening one as their self-study assignments. The listening test assesses the language in communication; vocabulary and grammar ones are not based on the communicative principle, but on discrete items. Reading and speaking skills are tested with the help of teacher's assessment or alternative assessment tools in class. Writing is checked in the form of essays. Writing is a time- and effort-consuming skill for both students and teachers who grade students' works, so it is not practised a lot, and not in the classroom. Besides, writing skills are developed with the help of a separate subject, Academic Writing.

University teachers are “jacks of all trades” and as such they do everything for the instruction process. They develop syllabuses, write text books, prepare materials for lessons and examinations, design tests, etc. However, the question arises whether teachers are assessment and test-writing literate, aware of understanding and interpretation of assessment purposes and results, methods and principles of test item design to ensure valid inferences about test takers' knowledge and skills, whether teachers who become item writers and test designers have enough knowledge and skills for this complicated task (Ramirez, 2020, p. 195). Because of that, it sometimes happens that commercial tests are used; sometimes tests are “frankensteined” (Ramirez, 2020, p. 196) — pieces of tests designed commercially or ones taken from textbooks are used. However, these practices do not correspond to the syllabus and the objectives of the course taught, do not take into consideration students' specialization and needs, and, therefore, are not valid and reliable. There is one more thing that

teachers who design tests should be familiar with. It is descriptive statistics. It is a complicated thing and is closely connected with mathematics which is not, as a rule, among favourite subjects for language teachers. For some of them, it is even frightening. In any case, basic statistics connected with test design is necessary for test designers.

Taking into consideration the Ukrainian national school-leaving test, it should be noted that there is a specialized department at the Ukrainian Education Quality Assurance Center which analyses the behaviour of test items for school-leavers after their piloting and pretesting. University electronic tests are also accompanied by graphs presenting the test discrimination curve. Still, it is doubtful that it is paid much attention to and the test is rewritten if the curve of the normal distribution is not “normal”. It does not discriminate students properly in case the Gaussian curve is asymmetrical. If it is shifted right, the test is too easy, and the results are higher than expected. If it moves to the left, the results are lower than expected, and the test is too difficult. The test is considered to be valid, if the asymmetry and excess of the distribution equal zero. Thus, the statistical analysis of the test is based on the deviation of the results from the normal distribution — on the coefficients of asymmetry and excess.

The high-stake school-leaving test is calibrated against the Common European Framework descriptors, and the results are reported on a single scale. The specification, construct and content of tests designed by university teachers are based on the syllabus which, in its turn, is based on the CEFR. There is no wide-scale trialling, mostly the intuition and common sense of teachers who design and administer tests. Then alterations, if they seem necessary, are introduced, some of them after the administration / implementation of the test. Feedback is very important in order to review positive and negative aspects of the test, as it helps improve the test and avoid mistakes when developing future ones. De Benedetti (2006) recommends to concentrate on the following while designing test items and compiling tests: to analyse test items including the test statistics, to design and analyse questionnaires for students and teachers, ask them to fill the questionnaires out after the test administration, and to reflect on the wash-back effect of the test. It also helps if assessment and evaluation are transparent; it raises students' and teachers' awareness of the fairness of the procedure and thus students' motivation (Ashraf & Zaki, 2019, p. 169).

CONCLUSION

The article shows, that despite lot of books on language assessment and evaluation and webinars conducted on these issues, teachers need training in test task design since they enter into the profession without a strong knowledge and skills of assessment (Ashraf & Zaki, 2019). They need greater expertise and diversification of evaluation techniques, as well as the development of constant reflection and critical thinking towards assessment and evaluation (Brown &

Bailey, 2008; Ramirez, 2020). The emphasis on testing is reinforced by the need to investigate teacher assessment concepts in a completely new virtual environment (Morrison et al., 2022). The article may contribute to helping Ukrainian university teachers reflect on assessment and evaluation knowledge and skills and raise awareness of the importance of deeper theoretical and practical expertise in high-quality assessment for language education. It may stimulate teachers' attention to focus on complicated problems around test creation, administration, and analysis.

REFERENCES

- Alderson, C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Ashraf, A. & Zaki, S. (2019). Language assessment in higher education: Challenges and consequences. *Pakistan Journal of Education*, 36(3), 163-185. <https://doi.org/0.30971/pje.v36i3.437>
- Bachman, L.F. (2013). Ongoing challenges in language assessment. *Evaluation, Methodology, and Interdisciplinary Themes*, 3(12), *Interdisciplinary Themes*. <https://doi.org/10.1002/9781118411360.wbcla128>
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practice*. New York, USA: Pearson Longman.
- Brown, J.D. & Bailey, K.M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349–383. <https://doi.org/10.1177/0265532208090157>
- Brown, J. D. (2011). *Testing in language programs: A comprehensive guide to English language assessment*. New York, USA: McGraw Hill.
- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Carr, N. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Cohen, A. (1994). *Assessing Language Ability in the Classroom*. Boston: Heinle and Heinle Publishers.
- CEFRL. (2001). *Common European Framework of Reference for Languages: Learning teaching, assessment*. Cambridge: Cambridge University Press. Council of Europe.
- De Benedetti, K. (2006). Language testing: Some problems and solutions. *Universidad de Guanajuato, México*, 30(1), 25-41. <https://www.mextesol.net/journal/public/files/2bbd95c84cfda1581bedde28ef1333ae.pdf>
- Davies, A. (1990). *Principles of Language Testing (Applied Language Studies)*. Blackwell Pub.
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder-Arnold.
- Eckes, T, & Althaus, H.J. (2020). Language proficiency assessments in higher education admissions. In Oliveri, M.E., & Wendler, C. (Eds). *Higher*

- Education Admission Practices: An International Perspective*, (pp.256-275).
<https://doi.org/10.1017/9781108559607.015>
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. Oxon: Routledge.
- Giraldo, F. (2019). Designing Language Assessments in Context: Theoretical, Technical, and Institutional Considerations. *HOW*, 26(2), 123-143. <https://orcid.org/0000-0001-5221-8245>
- Hubley, A.M. & Zumbo, B.D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology*, 123(3), 207-215. <https://doi.org/10.1080/00221309.1996.9921273>
- Hughes, A. (2002). *Testing for language teachers: Second edition*. Cambridge: Cambridge University Press.
- McNamara, T. (2004). Language testing. In Davies, A. & Elder, C. (Eds.). *The handbook of applied linguistics* (pp. 763-783). Oxford: Blackwell.
- Morrison, A., & Sepulveda-Escobar, P. (2022). Chilean English as a foreign language teacher educators' conceptions and practices of online assessment. *RELC Journal*. <https://doi.org/10.1177/00336882221079059>
- Mospan, N. V., Ognevyuk, V. O., & Sysoieva, S. S. (2022). Emergency higher education digital transformation: Ukraine's response to the COVID-19 pandemic. *Information Technologies and Learning Tools*, 89(3), 90-104. <https://doi.org/10.33407/itlt.v89i3.4827>
- Ramirez, A. (2020). Challenges in the Design and Implementation of an English Placement Test for a Colombian Public University. *Gist Education and Learning Research Journal*, 21, 191-208. <https://doi.org/10.26817/16925777.825>
- Rapi, L. & Miconi, H. (2014). Theoretical and practical problems in designing tests for foreign language programs. *European Scientific Journal*, 10(29), 156-164. <https://doi.org/10.19044/esj.2014.v10n29p%25p>

ТЕСТУВАННЯ ЯК ФОРМА ОЦІНЮВАННЯ МОВИ: ДОСВІД УКРАЇНИ

Оксана Висоцька, кандидат філологічних наук, доцент, Кафедра англійської мови та комунікації, Київський університет імені Бориса Грінченка, Україна 04053, Україна, м. Київ, вул. Бульварно-Кудрявська, 18/2. o.vysotska@kubg.edu.ua

У статті порушуються питання компетентності вчителів, які є «майстрами на всі руки», в оцінюванні та розробленні тестів, їхнього усвідомлення цілей оцінювання, а також здібностей вчителів інтерпретувати методи та принципи розроблення тестових завдань, зокрема для оцінювання англомовних та мовленнєвих компетентностей студентів. У статті, автор також, розмірковує над можливістю вчителів закладів вищої освіти робити за допомогою тестів достовірні висновки

щодо знань і навичок учасників тестування, а також над наявністю у викладачів, які стають авторами завдань і розробниками тестів, достатніх знань і навичок для цього складного завдання. Об'єктом дослідження є аналіз досвіду та практик розроблення та застосування тестів для оцінювання англомовних компетентностей студентів в освітньому процесі у закладах вищої освіти України. Результати та висновки, зроблені автором, можуть сприяти викладачам закладів вищої освіти осмислити специфіку процесу розроблення якісних тестів. Крім того, ця стаття може підвищити обізнаність про важливість глибокої теоретичної та практичної експертизи у високоякісному оцінюванні мовної освіти, а також зосередити увагу вчителів на складних проблемах створення, адміністрування та аналізу тестів для оцінювання англомовних компетентностей студентів у закладах вищої освіти.

Ключові слова: оцінювання; знання та вміння; тест; розроблення тесту; тестові завдання.

Received: 17.10.2023

Accepted: 28.12.2023