

THE IMPORTANCE OF TEST-TAKERS' VIEWS ON LANGUAGE PROFICIENCY EXAMS

A VIZSGÁZÓK NYELVVIZSGÁVAL KAPCSOLATOS VÉLEMÉNYEINEK FONTOSSÁGA

ANIKÓ MÁTYÁS¹

Abstract: This paper aims to show how face validity has been conceptualized in language testing literature and also to reveal what characterizes language testing stakeholders' different approaches and attitudes to this conceptualization. Based on previous research I discuss the impacts of face validity and test-taker related factors on test-takers' perspectives on the language exam and their test performance. This review is not intended to be exhaustive, but merely to illustrate the potential effects of test-takers' perception on the language exam and to make suggestions for improving the quality of tests and teaching methodology and test preparation methodology.

Keywords: *face validity, language exam, test-taker, test-taker related factors*

Absztrakt: Ez a tanulmány arra keresi a választ, hogy a nyelvi tesztelés kutatás, hogyan határozza meg a felszíni validitás fogalmát, valamint azt is feltárja, hogy mi jellemzi a nyelvvizsgáztatásban érdekelt felek különböző megközelítéseit a terminushoz. Korábbi kutatások alapján feltárom, hogy a felszíni validitás és a vizsgázókhoz köthető tényezők milyen hatással vannak a vizsgázók nyelvvizsgával kapcsolatos nézeteire, valamint a nyelvvizsgán nyújtott teljesítményükre. Tanulmányommal nem törekszem teljességre, csupán a nyelvvizsgázók nyelvvizsgával kapcsolatos nézeteinek fontosságát kívánom bemutatni, valamint javaslatokat teszek a nyelvvizsgák minőségének és a tanárok tanításmódszertanának és vizsgafelkészítési módszertanának javítására.

Kulcsszavak: *felszíni validitás, nyelvvizsga, nyelvvizsgázó, nyelvvizsgázóval kapcsolatos faktorok*

1. Introduction

High-stakes language exams that measure the extent to which test-takers use their language ability in target situations have been used to make educational decisions in language learning for the purposes of further education admissions, graduation and promotion at work (Fekete–Csépes 2019). The Hungarian education system recognizes language proficiency through the state accredited language exam

¹ MÁTYÁS ANIKÓ
egyetemi tanársegéd
Modern Filológiai Intézet, Angol Nyelvészeti és Irodalomtudományi Tanszék
Miskolci Egyetem
aniko.matyas@uni-miskolc.hu

certificate, which means a “passport” to further education as a successful exam certificate means plus scores at the university entrance exam for secondary students (Fekete–Csépes 2019: 13). This importance of the language exam resulted in a widely extended language examination system in Hungary, in which test-takers can choose to take language exam at 200 accredited exam sites of 13 state-accredited language exam centers. The number of test-takers who took a language exam in English at B2 level between 2013–2022 in one of the mentioned exam sites surpassed 300,000 (www.nyak.oh.gov.hu).

Not surprisingly, test-takers, language teachers and test developers associate language exams with high-stakes that may induce large consequences (Messick 1996). In recognition of this, it is important to know whether valid decisions are made based on test scores and on the consequences of the test use. Therefore, this article aims to highlight the impacts of test-takers' perception on the language exam and provide recommendations for enhancing test quality, improving teaching methods and test preparation strategies.

2. Face validity – a neglected issue

Language testing has been a subject of debate because test use raises ethical issues in terms of fairness, test development and interpretation of tests (Fulcher 1999; Shohamy 1997). Test validation plays an important role in ensuring appropriate and fair evaluation of the test-takers' ability. Test validity is a basic consideration in language testing, justified by multiple traditional and contemporary frameworks (Bachman 1990; Bachman–Palmer 1996, 2010; Carroll 1980; Lado 1961; Messick 1996) since it represents how accurately a test reflects the test-takers' actual proficiency level (Borsboom et al. 2004).

What characterises validity research is a substantial focus on content and construct validity in terms of test use and the interpretation of test scores (Messick 1995). However, face validity has received only minor attention from testing research owing to its specific nature, as face validity is “the degree to which a test appears to measure the knowledge or abilities it claims to measure, as judged by an untrained observer” (Davies et al 1999: 59). In other words, face validity is the test-takers' unscientific judgment based on subjective views (Bachman 1990; Brown 2004; Henning 1987; Stevenson 1985). Several researchers (Akbari et al. 2015; Davies et al. 1999; Henning 1987; Stevenson 1985) claim that face validity cannot be empirically tested and thus it is not regarded to be scientific. This is presumably the reason why the views of test-takers, language teachers and test-developers on successful performance and qualities of a good test are not similar (Lukácsi 2017).

Differently from researchers' theoretical interpretation of tests, test-takers and language teachers perceive them “in very material ways” (Hamid et al. 2019: 2). To investigate what makes a good language exam, Dávid (2014) asked language teachers to choose one from eleven state-accredited language exams of Hungary and evaluate it. According to Dávid (2014), teachers select based upon their learners' opinion and chose the exams that they found relatively easy with acceptable time

allotted and the fastest results to be received. For test-takers, besides these factors, it is also important to hear what their peers comment on the topics of the exam tasks, the attitude of examiners and the difficulties of the tasks. In this sense, the scientific importance of face validity can be doubted.

However, other researchers do find face validity important, because if the test is “acceptable for users”, the test-takers will be more likely to achieve the best possible results (Alderson et al. 1995: 173). Brown (2000) and Messick (1994) also emphasized the importance of face validity in terms of language testing, as information on behalf of a large population, through test-takers, is gained. Chan and Schmitt (1997), along with Watanabe (2001), revealed a positive relationship between face validity and learning motivation. Further, Watanabe (2001), drawing from interviews held with learners at tertiary level, claimed that test-takers’ perceptions about the difficulty and quality of the exam strongly influence their motivation and also their inclination to learn. Further, Weir (2005) emphasized that face validity can play an important role in maintaining test-takers’ motivation, as a face-valid test may improve their performance. In addition to motivating the learners, a face valid test may also increase the test-takers’ willingness to take the test and the possibility of achieving good scores (Karelitz 2013).

A commonly held belief is that test scores mirror the test-takers’ actual skills, though this assumption may be challenged, if any changes in the tasks or in exam settings occur that the test-taker had not anticipated (Brown 2004). Further, if test-takers do not find the purpose of the test relevant and comprehensible, they may devote less energy to performing well at the test and the scores may not represent their real ability (Bachman 1990). In other words, if dictation is used for measuring writing ability for instance, the test presumably will not be considered face-valid by test-takers as dictation measures listening and spelling skills, while a writing task tests whether the test-taker can organise and express ideas in a clear and coherent way (Davies et al. 1999). Therefore, prior to accepting the test as valid and the tested sample as representative, it should be analysed if the test-takers were motivated to achieve the best scores and were aware of task types and task solving strategies, because test tasks that seem irrelevant may discourage test-takers from spending effort on them (Kane 2004, 2006).

3. The importance of test-taker related factors

Validity is not only a basic consideration in language testing, but also a complex issue that is influenced by different visible or unidentified test-taker related factors that can impact test-takers’ scores and performance. Test-taker related factors are classified by Bachman and Palmer (1996: 64) into the following categories: “personal characteristics, topical knowledge, general level and profile of language ability and predictions about test takers’ potential affective responses to the test and personal characteristics”. As stated by Brown (2004: 117), test-taker related physical and psychological factors such as “a bad night’s rest, illness, an emotional distraction, test anxiety, a memory block” can make the received score deviate from

the true scores. Therefore, he suggests that when designing a test, in addition to the test format and the content, timing and allocation of tasks, test-taker related factors should also be taken into account (Brown 2004). It is essential, as emphasized by (Brown 2000, 2004), because the mentioned factors may change test-takers' performance, so relying solely on test-takers' cognitive abilities is not appropriate (Brown 2000, 2004). Further, as argued by Shohamy (2001), a passionate supporter of fair language testing, test use and test quality can be recognized through the test-takers' feedback. Considering test-taker related factors in test-development is crucial to avoid test bias (Weir 2005).

In their study on self-chosen learning strategies, Sundre and Kitsantas (2004) found a link between test-taker related factors and written exam scores, as the performance of motivated test-takers was better than those of their demotivated peers. Cheng and DeLuca (2011) investigated test-takers' perceptions on validity of four different language proficiency exams and found that in addition to external factors not related to test-takers (such as raters and the location), test-takers' emotions impact their performance. According to the test-takers' perceptions, their emotions were not only influenced by the test invigilators' behaviour but also by their own test preparation and test-taking strategies. The above studies on the interplay between test-taker related factors, their perceptions and test performance showed that test-taker related factors and test-takers' perceptions of face-valid tests are dynamic and also context-sensitive.

4. What makes a face-valid language exam?

In light of the above considerations, it can be concluded that authenticity and reliability are two important components of a face valid language exam (Bachman–Palmer 1996; Brown 2004). According to the framework proposed by Bachman and Palmer (1996), authenticity should be one of the indispensable issues when developing the test as it may be critical, in reference to test-takers' performance, because their perception of authenticity will influence the way in which they approach a certain task and what strategies they will use to carry it out.

Authenticity is a disputed term in language testing because it is regarded as “a matter of perception” (Leung–Lewkowitz 2006: 215). Bachman and Palmer (1996: 23) conceptualized authenticity as “a means for investigating the extent to which score interpretations generalize beyond performance on the test to language use”.

In terms of language exam tasks, Bachman (1991) made a distinction between situational authenticity and interactional authenticity. Situational authenticity refers to “the perceived relevance of the test method characteristics to the features of a specific target language use situation” while interactional authenticity represents the link between the features of the test task and the test-taker's abilities and cognitive processes (Bachman 1991: 90–91). A task is viewed interactionally authentic according to “the degree to which test materials and test conditions succeed in replicating those in the target situation” (McNamara 2000: 131); in other words, if the tasks were to test language functions relevant in real-life

contexts, and the test-takers' strategies for completing the tasks could also be appropriate in non-test situations.

As emphasized by Morrow (1979) and Weir (1990, 2005), authenticity is an important component of communicative testing (Bachman, 1990), which permits the use of texts in the test that originally were not intended for educational purposes. Dávid (2020: 9) finds this approach "a limited, practical conceptualisation of authenticity" as these texts „must have been rewritten to suit the pedagogical purpose". Therefore, according to Dávid (2020), Bachman's (1991) interactional approach is a more profound conceptualisation of authenticity, in which the test developer involves the test-taker and scrutinizes the test-takers' performance to select characteristics of communicative language use and use it as a guidance for test task preparation (Dávid 2020).

However, completely authentic texts cannot be created, as test-takers are well aware that they are being tested and the test cannot imitate reality (Dávid 2020; Davies et al. 1999; Lewkowitz 1997, 2000). Further, tests are taken in a specific time period, at a certain place, which also challenges authenticity (Dávid 2020; Davies et al. 1999; Lewkowitz 1997). Accordingly, it is important that test-takers meet "an authentic testing experience", although it cannot be considered "real-life authenticity or some nontesting situation" (Dávid 2020: 9).

A test that is considered authentic by test-takers may escalate their "concentration and involvement in the target activity more than artificial materials" (Peacock, 1997: 152). Lewkowitz (2000) revealed that several test-takers found authenticity an important supporter of their performance and only realized how important it was when the test tasks were not found to be authentic.

In Chun's (2006) study the test-takers thought that test tasks were unnatural, which increased anxiety and stress that they felt because of the exam. The participants in Fan's (2014) study expressed negative reactions towards the test caused by the speaking tasks that were different from those they would use in real-life communication. Furthermore, to improve the authenticity in test development without abandoning the scientific accuracy in scoring poses a challenge (Chun 2006, 2008). As the participants of the study indicated, since the test was delivered on a computer and phone, the conversation with a machine was found odd. The test-takers were also disappointed as they noted only a limited part of their actual language skills were tested.

In addition to authenticity, keeping up a balance between face validity and reliability is also important, as a reliable test consistently measures what it is supposed to measure under all conditions. In other words, test-takers should receive similar results irrespectively of what version of the test is taken and even if the test is corrected by different raters (Bachman–Palmer 2010; Harrison 1983; Hughes 2003). A reliable test can add to the validity of score interpretations and provides proper information about the test-takers' language skills (Fulcher–Davidson 2007). Along these lines, reliability can be described in two ways, namely the reliability of the scoring and the reliability of the test (Harrison 1983). Similarly to Bachman (1990) and Bachman and Palmer (1996), Brown (2004) also claimed that the

reliability of the test viewed by the test-takers is influenced by several test-taker related factors. Not only “the extent of the sample of material selected for testing, the administration of the test, test instructions” matter in terms of reliability of the test, but also test-taker related factors such as motivation and health issues (Heaton 1975: 162–163).

Bachman and Palmer (1996) emphasized the importance of test-taker related factors more than cognitive variables in terms of the test reliability but also found the former difficult to rate; it was complicated to separate them from the tested language ability. Test-taker related factors can contribute to whether or not a test is reliable and authentic in the eyes of the test-takers and the reverse is also true. As a step towards ensuring that test-takers find the test authentic, Brown (2004: 28) proposed a number of considerations that might be helpful to support the authenticity of the test to be developed including “natural language in text”, “contextualized items”, “meaningful topics”, “thematic organization” of items and “real-world tasks”. In order to maintain reliability of the test Hughes (2003: 44–50) made several suggestions, from “unambiguous items” and “clear and explicit instructions” to making test-takers “familiar with format and testing techniques”, as well as the need to trained scorers and “uniform and non-distracting conditions of administration”.

If the conditions proposed by Brown (2004) and by Hughes (2003) are not met, negative perceptions on authenticity and reliability of the test may appear. Language teachers try to compensate for test-takers' negative perceptions during language exam preparation, where a variety of activities are practiced to review the domain of skills sampled by the language exam (Alderson–Hamp–Lyons 1996). By narrowing the curriculum, teachers can focus on strategies to increase scores and, in the interest of authenticity, to master certain structures such as the passive voice to improve learners' linguistic and communication skills (Vraukó 2022).

Besides improving performance, teachers' aim is also to prevent test-takers from encountering unexpected events at the exam, since face validity will likely be high if test-takers encounter familiar tasks in expected formats that relate to the tasks and topics practiced during preparation and that present test-takers with reasonable challenges.

Not only teachers feel it their duty to improve test-takers' perceptions of the exam, but also the state accredited language exam centers that are competing for test-takers country-wide (www.nyak.oh.gov.hu). The examination centres constantly monitor test-takers' views on the exams and try to adapt accordingly. Due to the strength of the non-experts' perception (Dávid 2014; Lukácsi 2017) the examination centres are determined to ensure the most positive experiences possible for test-takers by providing them with placement tests, mock exam tasks, sample answers as well as feedback from experts to increase the face validity of the exam.

Further, in the last twenty years, exam centers have introduced several alterations in the exam tasks and test administration in Hungary, including but not limited to the followings policies. Test-takers are entitled to use bilingual and/or monolingual dictionaries in monolingual exams. The allocated time of certain tasks has been increased by an average ten minutes in reaction to test-takers' requests. In addition,

speaking tests are taken in pairs and the test-taker can choose the partner prior to the exam. Furthermore, grammatical competence is not assessed in a separate exam task but integrated into sections that measure reading, writing and speaking skills (www.euroexam.org, www.nyest.hu).

5. Steps to be taken for better exams

In light of the above, I find it important to emphasise an urgent need for closer cooperation between test-takers, teachers and test developers in order to avoid the differences in interpretation of test results between test-takers, teachers and test developers (Lukácsi 2017). For this cooperation to occur, the following steps should be implemented:

a) Test-takers need to be informed by teachers that a good test helps the test-taker achieve the best language performance, which means that the scores not only provide information on how the test-takers have performed but also how they would behave in a real language situation. Therefore, the more tasks are involved when a certain skill is being measured, the more knowledge the test-takers can show, regardless of their preference for fewer tasks (Cumming et al. 2005).

b) With this in mind, it is worthwhile for language teachers to provide clear feedback on each step of mock exam tasks that are being practiced in order to teach learners how to manage the allotted time at the exam. Analysing the tasks when giving feedback to the learners can also improve learners' self-assessment skills, promote motivation and develop autonomous learning, as well as nurture commitment in learning the language (Harris 1997).

c) Continual encouragement and support on behalf of teachers can help learners meet the evaluation criteria established in the exam and decrease anxiety and stress caused by the language exam.

d) It is worthwhile for language teachers to use real-life materials when preparing the language course curriculum and also take into consideration what the language learners also find authentic. In this regard, for instance, test developers and teachers can account for the changes that have been introduced and apply new tasks that reflect various forms of digital communication (Harding 2014).

e) Introducing metacognitive strategies to learners – such as planning, monitoring, and adjusting their test preparation – is pivotal to achieving successful test performance. It is also important to show learners how metacognitive strategies can be used to enhance the efficiency of learning.

f) When designing the curriculum, teachers should focus on a needs analysis of test-takers, as well as the incorporation of exam strategies and feedback mechanisms.

g) Frequent feedback should be asked from test-takers and made available to test-developers, either right after taking the exam or in a retrospective form.

h) Clearly, if teachers have proper information about a certain language exam learners have chosen, this will also affect test-takers' perceptions about the exam and raise their awareness of the qualities of a good language exam.

6. Summary

Despite the fact that face validity is underrepresented in testing research and test development, test-takers' unscientific judgments and perceptions of the language exam can provide valuable information for language teachers and test-developers. Along with meeting the criteria above, it is also important to identify the factors that can be attributed to test-takers indicated by Brown (2004) and Bachman and Palmer (1996) – factors that are not part of language ability but either routinely or randomly influence test-takers' performance and the received scores. Having these factors in mind in test development and during teaching could not only help improve test-takers' perceptions of the language exam, but also offer opportunities for teachers to develop their methods and foster learners' willingness to continue studying after taking the exam.

References

- Alderson, Charles. J. – Clapham, Caroline – Wall, Diane 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, Charles. J. – Hamp-Lyons, Liz 1996. TOEFL preparation courses: A study of washback. *Language Testing*, 13/3, 280–297.
- Akbari, Elham – Pilot, Albert – Simons, P. Robert-Jan 2015. Autonomy, competence, and relatedness in foreign language learning through Facebook. *Computers in Human Behavior* 48, 126–134.
- Bachman, Lyle F. 1990. *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, Lyle F. 1991. What does language testing have to offer? *TESOL Quarterly* 25/4, 671–704.
- Bachman, Lyle F. – Palmer, Adrian S. 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, Lyle F. – Palmer, Adrian S. 2010. *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Borsboom, Denny – Mellenbergh, Gideon J. – van Heerden, Jaap 2004. The concept of validity. *Psychological Review* 111/4, 1061–1071.
- Brown, H. Douglas 2000. *Principles of language learning and teaching*. New York, NY: Longman.
- Brown, H. Douglas 2004. *Language assessment: Principles and classroom practices*. Harlow: Pearson Education.
- Carroll, Brendan J. 1980. *Testing communicative performance: An interim study*. Oxford: Pergamon Press.

- Chan, David – Schmitt, Neal 1997. Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology* 82/1, 143–159.
- Cheng, Liying – DeLuca, Cristopher 2011. Voices from test-takers: Further evidence for test validation and test use. *Educational Assessment* 16/2, 104–122.
- Chun, Christian W. 2006. An analysis of a language test for employment: The authenticity of the PhonePass test. *Language Assessment Quarterly* 3/3, 295–306.
- Chun, Christian W. 2008. Comments on “Evaluation of the usefulness of the versant for English test: A response”: The author responds. *Language Assessment Quarterly* 5/2, 168–172.
- Cumming, Alister – Kantor, Robert – Baba, Kyoko – Eouanzoui, Kianre – Erdosy, Usman – James, Mark 2005. *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the New TOEFL*. ETS research report series.
- Dávid Gergely A. 2014. Mitől jó egy nyelvvizsga? *Iskolakultúra* 24/4, 3–25.
- Dávid, Gergely A. 2020. A human face to technology? Digital exams are coming. *WoPaLP* 15, Special Issue.
- Davies, Alan – Brown, Annie – Elder, Cathie – Hill, Kathryn – Lumley, Tom – McNamara, Tim 1999. *Dictionary of language testing*. Cambridge: CUP.
- Euroexam International nyelvvizsgaközpont. *Most még egyszerűbb a nyelvvizsga*. www.euroexam.org/hu/vizsgainfok/most-meg-egyszerubb-a-nyelvvizsga/angol/b2. (Letöltve: 2023. 05. 05.).
- Fan, Jinsong 2014. Chinese test takers’ attitudes towards the versant English test: A mixed-methods approach. *Language Testing in Asia* 4/1, 1–17.
- Fekete Adrienn – Csépes Ildikó 2019. B2-es szintű nyelvvizsga bizonyítvány: útlevelel a diplomás élethez, társadalmi mobilitáshoz. *Iskolakultúra* 28/10–11, 13–24.
- Fulcher, Glenn 1999. Computerizing an English language placement test. *ELT Journal* 53/4, 289–299.
- Fulcher, Glenn – Davidson, Fred 2007. *Language testing and assessment: An advanced resource book*. London: Routledge.
- Hamid, Obaidul M. – Hardy, Ian – Reyes, Vicente 2019. Test-takers’ perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia* 9, Art. 16.
- Harding, Luke 2014. Communicative language testing: Current issues and future research. *Language Assessment Quarterly* 11/2, 186–197.
- Harris, Michael 1997. Self-assessment of language learning in formal settings. *ELT Journal* 51/1, 12–20.

- Harrison, Andrew 1983. *A language testing handbook*. London: Macmillan.
- Heaton, John B. 1975. *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Harlow: Longman.
- Henning, Grant 1987. *A guide to language testing: Development evaluation research*. Rowley, Massachusetts: Newbury House.
- Hughes, Arthur 2003. *Testing for language teachers*. 2nd ed. Cambridge: Cambridge University Press.
- Lewkowitz, Jo. A. 1997. Authenticity for whom? Does authenticity really matter? In: Huhta, Ari – Kohonen, Viljo – Kurki-Suonio, Liisa – Luoma, Sari (eds.): *Current developments and alternatives in language assessment: Proceedings of LTRC 96*. Jyväskylä, Finland: University of Jyväskylä. 165–184.
- Kane, Michael 2004. Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives* 2/3, 135–170.
- Kane, Michael 2006. *Content-related validity evidence in test development*. In: Downing, Steven M. – Haladyna, Thomas. M. (eds.): *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates. 131–153.
- Karelitz, Tzur M. 2013. *Using public opinion to inform the validation of test scores*. Research Report no. 387. Jerusalem: NITE.
- Lado, Robert 1961. *Language testing*. Harlow: Longman.
- Leung, Constant – Lewkowitz, Jo. A. 2006. Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly* 40/1, 211–234.
- Lukácsi, Zoltán 2017. Developing a level-specific checklist for assessing writing. Paper presented at the *14th EALTA Conference*, Ciep, Sèvres, France.
- McNamara, Tim 2000. *Language testing*. Oxford: Oxford University Press.
- Messick, Samuel 1994. The interplay of evidence and consequences in the validity of performance assessment. *Educational Researcher* 23/2, 13–23.
- Messick, Samuel 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50/9, 741–749.
- Messick, Samuel 1996. Validity and washback in language testing. *Language Testing* 13/3, 241–256.
- Morrow, Keith 1979. Communicative language testing: Revolution or evolution? In: Brumfit, Christopher. J. – Keith, Johnson (eds.): *The communicative approach to language teaching*. Oxford: Oxford University Press. 9–25.

- Oktatási Hivatal Nyelvi Akkreditációs Osztály. *Akkreditált nyelvvizsgaközpontok és nyelvvizsgahelyek*. nyak.oh.gov.hu/doc/akk_vizsgakp.asp (Letöltve: 2023. 05. 05.).
- Peacock, Matthew 1997. The effect of authentic materials on the motivation of EFL learners. *ELT Journal* 51/2, 144–156.
- Stevenson, Douglas. K. 1985. Authenticity, validity and a tea party. *Language Testing* 2/1, 41–47.
- Shohamy, Elana 1997. Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing* 14/3, 340–349.
- Shohamy, Elana 2001. *The power of tests: A critical perspective on the use of language tests*. Harlow: Longman.
- Sundre, Donna. L. – Kitsantas, Anastasia 2004. An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology* 29/1, 6–26.
- Vraukó, Tamás 2022. *The English passive construction and its Hungarian equivalents: A comparative analysis*. London: GlobeEdit.
- Watanabe, Yoshinori 2001. Does the university entrance examination motivate learners? A case study of learner interviews. In: Akita Association of English Studies (ed.): *Trans-Equator exchanges: A collection of academic papers in honour of Professor David Ingram*. 100–110.
- Weir, Cyril. J. 1990. *Communicative language testing*. London: Prentice Hall International.
- Weir, Cyril J. 2005. *Language testing and validation: An evidence-based approach*. Palgrave: Macmillan.