

DOI 10.51558/2490-3647.2021.6.3.265

UDK 811.111:004.738.5

Primljeno: 02. 06. 2021.

Izvorni naučni rad
Original scientific paper

Marijana M. Prodanović, Valentina M. Gavranović

ASSESSING GRAMMAR IN ONLINE ENVIRONMENT: A COMPARISON BETWEEN HIGH AND LOW-STAKES TEST PERFORMANCE

Assessing students' progress and creating objective and reliable tests occupy an important place in the ELT sphere. This paper focuses on the analysis of the answers provided by a group of undergraduate students majoring in English. They responded to two grammar online tests assessing the knowledge of the verb tenses used in context. The analysis includes the comparison of the results obtained via the low-stakes test comprising fill-in type tasks, on the one hand, and the high-stakes test containing multiple-choice tasks on the other hand. This study aims at exploring which tense-related aspects the students find either the most or least demanding and whether there are similarities or differences in the outcomes gained from two different tests and task types. The findings show that the analysed answers exhibit similar characteristics regardless of the format and type of testing. The pedagogical implications arising from this study point to the areas which could be improved in terms of specific uses of tenses in context.

Keywords: low- and high-stakes test; tenses; multiple-choice and fill-in task type; online educational environment

1. INTRODUCTION

Over the course of the previous decades, the concept of digitalisation has undoubtedly brought about a number of changes in all aspects of human life, and education has also been considerably influenced by the immense possibilities the new technologies

offer. The effective use of ICT has been emphasised in applied linguistics, and it seems that it has enveloped all areas of foreign language teaching, learning, and assessment. With the sudden COVID-19 outbreak in 2020, educational systems worldwide faced unprecedented challenges, not only in terms of organising and conducting teaching processes, but also in terms of the process of evaluation and formal testing. The commonly used, traditional mode of face-to-face delivery was impossible to conduct in many institutions, which responded to this newly created situation by using different models of ICT-supported alternative means of assessment. In this regard, the characteristics, options and overall potential of online testing has been explored, exercised and re-examined, thus contributing to creating a more resilient education system that can effectively respond to these abrupt changes and the newly-imposed needs.

This paper reports on the results obtained from two online language tests a group of students took during the first term of the academic year 2020/2021. The first low-stakes test has the characteristics of formative testing, while another test represents the formal part of the exam process. The types of tasks comprising these tests also differ – the first test contains fill-in questions while the latter is characterised by multiple-choice (MC) questions. The main objective of this study is to determine whether there are differences in the answers students provided in these two tests assessing the very same grammatical categories, as well as to identify the most and least challenging aspects for them. It also aims at explaining whether the students made mistakes that relate to similar or different uses of verb tenses which are tested through multiple-choice questions on the one hand, and the fill-in questions on the other.

2. THEORETICAL FRAMEWORK

Test-taking contexts, inevitably characterised by anxiety arousal, become even more challenging once we face situations in which we take tests written in a foreign language (non-native language). As a result, it comes as no surprise that foreign language test-taking scenarios can lead to a number of students feeling anxious and not comfortable enough when they are about to exhibit the language-related knowledge (to list but a few psychological studies addressing the issue, e.g. Bellinger, Decaro & Ralston 2015, Haladyna & Downing 2004, Moran 2016).

Apart from differentiating between facilitative and debilitating anxiety, we can also identify Foreign Language Anxiety (FLA), encompassing a set of “self-perceptions, beliefs, feelings, and behavior related to classroom language learning

arising from the uniqueness of the language learning process” (Horwitz, Horwitz & Cope 1986: 128) which is not solely related to test-taking activities, but also communication apprehension and production pertinent to foreign language phenomenon.

Traditional pen-and-paper tests have been used for the purpose of evaluating students’ knowledge for decades and both language learners and instructors (language practitioners) still rely on them, especially in the context of formal grading. Despite all the recommendations regarding the introduction of continuous assessment in education, including language learning-teaching practices, numerical illustrations of the obtained results are still very prominent in the area (e.g. TOEFL, SAT, IELTS, TOEIC, etc.) and it was believed that even mere suggestion concerning the elimination of the traditional model “would at the very least cause colleagues in language departments to raise eyebrows” (VanPatten, Trego & Hopkins 2015: 661). In a similar vein, high-stakes testing, applied for school accountability, has been on the increase over the course of the previous decades (Gebriil & Eid 2017).

From the array of tests language practitioners have at disposal, we can choose e.g. placement, final achievement, progress, and other tests - some of which do serve the purpose of defining the extent to what our students have acquired the previously taught notions, while other help us obtain a comprehensive insight into the subject matter, with the goal of improving our language programmes and practices (syllabi, curricula, approaches and methods used, etc.).

Testing grammar commonly raises the question of differentiating the knowledge of and about language, i.e. suggests that the first step in the process should be determining whether the recognition of specific patterns/forms or their adequate usage in context is what we test. Back in 1976, Dimitrijević listed some of the frequent, both open and closed-ended task forms, applied in grammar testing – e.g. multiple-choice form, sentence (re)organisation, sentence transformation, question-answer tasks, synonymous sentences test, etc. (1976: 105-108), all of which can still be identified in language classrooms.

Among both language teachers and students, we can identify the advocates of both close and open-ended forms; while some believe that it is only open-ended forms via which real performance can be measured, it is certain that limitless potential characterises the former just like the latter, which results in the fact that hardly any of the language tests one can come across is of monolithic nature (Gergely 2007: 65).

It is a worthy note that e.g. multiple-choice items, in the view of some theorists, may lead to the scores “inflated due to test-wiseness, or the test-taker’s knowledge

about test taking” (Purpura 2005: 129) or may “leave untested a part of the construct” (Gergely 2007: 67).

Students, on the other hand, sometimes tend to believe that e.g. close-ended forms, especially multiple-choice ones, can help them recognise (i.e. recall) the correct answer, compared to open-ended items, which may ask for e.g. paraphrase, completion, transformation, etc. However, despite their ‘close’ nature, such tests may be very demanding and, at the same time, deceiving.

In the newly-imposed educational context, among the dominant guidelines, there is the notion of assessment, along with - teacher presence, design in online learning (Rapanta et. al. 2020: 22-23), computer literacy, the wellbeing of the participants, etc.

Prior to the COVID-19 imposed era, a comparison between pen-and-paper and online language quizzes organised at Michigan State University (MSU), made by Van Patten, Trego & Hopkins (2015), illustrated little difference in the results obtained.

A switch from pen-and-paper, traditional tests to those of electronic nature (organised online), in the sphere of ESP and EAP was illustrated in the study conducted by Prodanović (2020), which highlighted the extent to what both teachers and students adapted to the *new-normal* testing environment - but also pointed to the fact that there still exists enough room for improvement when it comes to productive skills (with specific reference to speaking).

Gavranović and Prodanović (2020) examined the context of low-stakes testing, from the angle of different tasks and their influence on the testing outcomes. The authors confirmed that “different task types – completion, transformation, error recognition, and correction, testing the same grammatical items – verb tenses and articles, can yield different outcomes” (2020: 46), while those related to error correction proved to be the most demanding.

3. RESEARCH METHODOLOGY

The main aim of this research is to determine whether two different test types that focus on the same grammatical categories influence the test-takers’ performance. Its objective is also to explore if there are differences pertaining to the nature of errors the respondents made in the fill-in and multiple-choice questions.

During the first term of the academic year 2020/2021, the students were involved in a four-month Contemporary English language course at the B2 level, containing five sections focusing on all language skills. The course was organised with the

support of online platforms, including some assessment portions as well. At the end of the term, as a part of the lesson, the students took a low-stakes test administered through Quizizz, the online application allowing students to fill in the missing verbs. Another test was administered as the formal type of testing, thus gaining the properties of high-stakes testing. It was conducted via Mtutor platform – the platform which provides objective testing conditions because it allows the test-administrator to monitor the students through the use of cameras, and prevents the test-takers from leaving the test and opening any other window. The high-stakes test comprised the tasks in the form of multiple-choice questions. Both tests were created by the teachers who organised the lessons throughout the term and assessed the progress of the students through a variety of formative activities and tasks, thus having a clear idea of the grammatical concepts and aspects to be tested. The total number of students who took both tests is 50.

The research centers on the analysis and comparison of the results obtained from these two test types. Both tests comprise the same grammatical category – verb tenses – formulated in a similar vein, but with different answer options offered to the students. They contained the questions whose aim was to evaluate students' knowledge of grammar and vocabulary. For this study, the researchers analysed one major segment of these tests – grammatical categories of verbs – *tenses* (present, past, and future), and *aspect* (progressive and perfect). The total number of these questions amounts to 20 in each test. The tasks in both tests are the same – they evaluate students' understanding of verb forms as used in context. The only difference lies in the options the students had when providing the answers – in the low-stakes test they filled in the empty spaces with the verb forms, and in the case of the high-stakes test, they could choose among the four options already provided in multiple-choice questions. In order to meet the criteria of test validity and reliability, the researchers carefully designed questions that address the same or similar usages of verb forms in a broader context. The questions were formulated in such a way so as to comply with the descriptors defined for the B2 level in the CEFR document.

The methods applied in this research are of both quantitative and qualitative nature. The results are first analysed and represented statistically in terms of the number and percentage of correct answers the students provided in both tests. The statistical representation of the results is labeled by either the correct (if all four answers are correct) or the wrong answer (regardless of whether there were only one or more incorrect answers within the question). The further analysis focuses on the nature of the errors made in both tests.

4. TEST RESULTS ANALYSIS AND DISCUSSION

Figure 1 shows the number of students (vertical line) who scored a certain percentage of correct answers (horizontal line) in the low-stakes test.

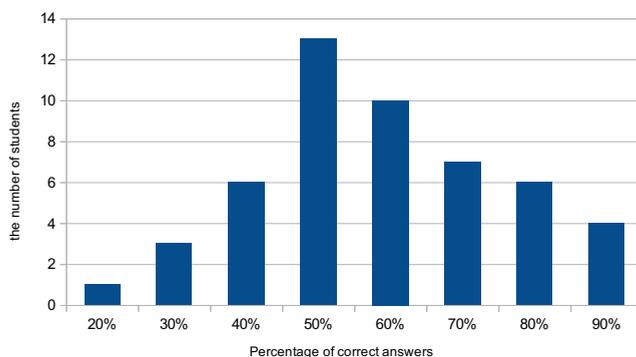


Figure 1. The statistical representation of the low-stakes results

The lowest score of 20% was attained by one student, and no one had the maximum number of points; nonetheless, there were 10 students who had either 80% or 90% of the correct answers in the test. Almost a fourth of the students (13 of them) provided 50% of correct answers in the test. The average score attained on the test was 58.6%.

The statistical analysis of the results obtained from the corpus of 50 high-stakes tests is illustrated in Figure 2. The horizontal line represents the percentage of correct answers, and the vertical line indicates the number of students who scored the given percentage.

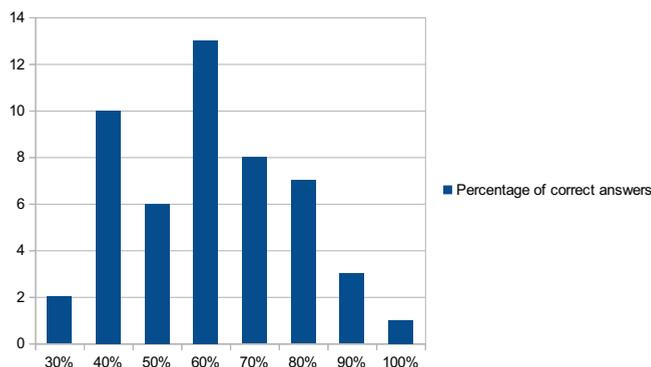


Figure 2. The statistical representation of the high-stakes test results

The results display that the lowest score of 30% was attained by two students and that only one student had the maximum number of points. Almost a fourth of the students (13 of them) provided 60% of correct answers in the test. The average score obtained in the high-stakes test was 60.6%.

The comparison of the statistical data obtained from the analysis of the correct and wrong answers provided by the students indicates that there is insignificant difference between the average score attained in either the low- or high-stakes tests, which amounts to 58.6% and 60.6%, respectively. The majority of students (13 of them) scored 50% in the low-stakes test, whereas the same number of students provided 60% of correct answers in the high-stakes test. Another slight difference is reflected in the number of students who had below 50% of correct answers – 10 students in the low-stakes test, and 12 students in the high-stakes test – the greatest number of whom scored 40%. These numbers show that the type of task – whether it contained fill-in or multiple choice questions – did not result in any statistically significant finding.

The qualitative analysis of the answers the students provided in both test types casts the light on the commonest errors made and the questions these can be attributed to. Additionally, the findings also point to the types of questions/contexts that were the easiest for the respondents.

The mistakes the students often made in the low-stakes test relate to the use of verbs in a broader context. The Examples 1 and 2 illustrate the most demanding tasks the students faced when determining forms used to express future or past time, respectively.

Example 1. Steve has arranged everything concerning his business trip to Barcelona. He _____ (fly) tomorrow and _____ (not be) back until next Monday. By then, his assistant _____ (do) all the paperwork related to the upcoming conference they _____ (prepare) for over three months.

Example 2. When he was a boy, Robert preferred sport to books, and since he _____ (leave) primary school, he _____ (play) a lot of basketball. He _____ (even/consider) to take up sport as profession. However, when he was 18, he discovered the world of science and decided to enroll the university. Robert is now a successful student and he _____ (consider) to continue his master studies abroad.

Only four students provided all correct verb tenses for Example 1. The commonest mistakes the students made relate to the use of The Present Progressive for the planned and arranged future action – the verb *to fly*. They mainly used *will fly* or *is going to fly*, thus failing to make a distinction between the uses of verb tenses which relate to planned, unplanned or arranged future actions. The students also failed to recognise the use of The Future Perfect Tense, and 41 students opted for *will do* instead of *will have done*.

Regarding Example 2, there were five students who performed the task correctly. The most problematic parts in this question related to the use of Past Simple and the Past Perfect tense in the first sentence – the majority of students opted for present and past forms, making combinations of *left* and *has played* or *has left* and *played*. These options could have resulted from the neglect of the broader context wherein the adverb *since* was used for past references.

Apart from these questions which posed the challenge for students in terms of the use of correct tenses, tasks which contained irregular verb forms also yielded many incorrect answers. The example which illustrates this type of task that the students wrote incorrect answers to is as follows:

Example 3. It was New Year's Eve. Tom _____ (flee) home the moment he finished with his work. He was so tired that he did not plan to celebrate it with anyone. When the clock _____ (strike) midnight, he _____ (already/lie) in his bed sleeping soundly.

85% of the students recognised the needed tense, but failed to write the correct Past Simple and Past Participle forms of irregular verbs. Other irregular verbs which were problematic for the students were *cast*, *bid*, and *shed*.

The analysis of the high-stakes test answers shows that the commonest errors the students made relate to the tasks containing verbs used in context with no explicit use of expressions that clearly indicate the temporal relationship between events. The following example, which yielded 42 incorrect answers, illustrates such type:

Example 4. My younger sister studies medicine and she has many years ahead of hard work as a student. She _____ (not be able) to practice full time until she _____ (be) nearly thirty. By the time she _____ (be) 30, she _____ (study) for a total of eight or nine years.

The students could choose one of the following options:

- a. *won't be, is, is, will be studying*
- b. *isn't going to be able, is, will be, will study*
- c. *won't be able, is, is, will have been studying*
- d. *is not going to be, will be, is, will has been studying*

Over 75% of the students circled either a or b option. The possible reason lying behind their choices could be the challenge imposed by the use of The Future Perfect Progressive and the use of The Present Simple with future meaning.

Another type of question which posed a challenge for many students includes verbs which relate to each other throughout a longer context. This type can be illustrated with Example 5, which resulted in wrong answers by over 70% of the students:

Example 5. Last year I decided to go on holiday all by myself. It was the first time I _____ (go) to Morocco. As a matter of fact, I _____ (never be) to Africa. I _____ (feel) so excited when I arrived at Marrakesh airport that I forgot that I was four hours late. My hotel _____ (close) and I had no place to stay.

The students could choose one of the following options:

- a. *had gone, had never been, felt, had closed*
- b. *went, have never been, felt, had closed*
- c. *went, had never been, was feeling, closed*
- d. *had gone, never was, was feeling, closed*

Over 70% of the students circled either b or c option, failing to recognise the use of The Past Perfect with the phrase *It was the first time*.

The type of questions that also posed a problem for the students contained less commonly used irregular verbs, and the following example illustrates such questions:

Example 6. He _____ (lay) the table and _____ (bind) the presents when he noticed he _____ (leave) the card outside the box.

The students could choose one of the following options:

- a. *lay, bound, had left*
- b. *laid, bound, had left*
- c. *lay, bound, left*
- d. *laid, binded, had left*

The results to this question varied, and all options were present in the students' answers, which can be explained by lucky guesses the students made when lacking knowledge related to the exact Past Simple forms.

On the other hand, the easiest questions in the low-stakes test contained verbs used with explicit uses of expressions that clearly indicate the temporal relationship between events in contexts. Example 7 illustrates this type of questions:

Example 7. Whenever Jack _____ (find) a job, there was someone who _____ (suspect) that he _____ (already/be) in prison.

Over 80% of students gave correct answers to this question. There were five more questions of this type that also proved to be correctly filled by over 80% of the respondents. The clear indicators used in this sentence are *whenever* and *already* which helped the students determine the temporal relationships.

Regarding the most successful results attained in the high-stakes test also contained explicitly stated expressions which indicate the tense use. There were 7 questions which were done correctly by the majority of the students. One of the question which belongs to this group is:

Example 8. Well, I _____ (see) my friends today. They _____ (invite) me to go with her to some exhibition. I _____ (think) it _____ at 6.

The students could choose one of the following options:

- a. am seeing, have invited, think, starts*
- b. will see, invited, am thinking, starts*
- c. am going to see, have invited, am thinking, is starting*
- d. will see, invited, think, will start*

Over 85% of the respondents circled the correct option *a*. The reason for such a high percentage of correct answers can lie in the explicit uses of The Present Simple – for the future action associated with timetables (*starts*), and the meaning of the stative verb *to think* (*think*).

The qualitative analysis and the comparison of the answers provided for both tests point out to several conclusions. In the first place, the most demanding tasks in both tests were those which included the application of grammatical knowledge in context which did not contain obvious indicators and temporal markers. In other words, when presented with context which demanded thorough reading and understanding of relations between and among verbs, the students failed to provide correct answers regardless of the format of the question – whether the fill-in or multiple-choice type. Examples 2 and 3 illustrate these relations and show that they demand a proper understanding of the temporal distribution of the verbs – the most problematic issue the students faced was the proper use of aspect. Even though in both low- and high-stakes tests there was a clear clue within the sentences (Examples 1 and 4, *by then* and *By the time she is 30*, respectively), the students still failed to identify the use of The Future Perfect Simple or Progressive. Another problematic aspect the students

faced in both tests relates to the knowledge of less commonly used irregular verb forms – Examples 3 and 6 (*flee, strike, lie, cast, bid, shed, lie, bind*) showing that both fill-in and multiple-choice questions posed a problem for the students, and these examples also yielded many wrong answers.

The students showed a consistent degree of accuracy when dealing with the examples containing straightforward usages of verb tenses – and these can be illustrated by Examples 7 and 8, wherein there are clear clues that explicitly support the grammatical rules. These clues include adverbs such as *whenever, already, today*. The analysis of the low-test results also indicates that the students did not make mistakes in terms of tense formation – there were only few sporadic faulty combinations, such as *will has done, is beeing*, and two instances of the violation of the subject-verb agreement.

5. CONCLUSION

Online language teaching and testing have posed a challenge for both teachers and test developers to create efficient and reliable tests that leave no space for accidents and random results. In the research, we relied on two grammar-based tests – one being the low-stakes test composed of fill-in questions, and the other the high-stakes test comprising only multiple-choice questions. In this paper, we investigated whether these test types that assess the same grammatical categories produce different results.

The analysis and comparison of the results obtained from both tests reveal several conclusions. In the first place, the statistical data related to the average score attained in both tests is rather similar – it amounts to 58.6% in the low-stakes test, and 60.6% in the high-stakes test. The qualitative analysis also yields similar results – the most challenging tasks in both tests relate to the use of tenses in context wherein the temporal relation is less obvious, and demands more careful reading and understanding how actions denoted by verbs relate to each other through time. In some instances, the difficulty to spot these subtleties arise from a highly developed verbal system and aspect as a complex grammatical category in English because of its specific nature and characteristics. Other challenging tasks in both tests relate to less commonly used tenses, such as The Future Perfect and The Future Perfect Progressive. The students also made mistakes when asked to either form or recognise the Past Simple and Past Participle forms of some less commonly used irregular verbs. On the other hand, the students showed a good command of verb tenses in contexts, either shorter or broader, which hold straightforward temporal relations and illustrate

frequent uses. Such examples include the use of The Past Simple for narration in combination with The Past Progressive for description, or with the adverb *while*. The students could also successfully use the Past Perfect in the combination with The Past Simple with the adverb *already*. The use of present forms did not pose a problem when the verbs were used with the adverbs such as *already, since, nowadays, these days*.

The findings of this study show that the outcomes of testing do not depend on the test types, but the way the questions are thought-through and formulated. They also bear pedagogical implications and reveal areas that could be improved in classes. These results also show that it is of vital importance to educate teachers to explore the possibilities of different test types and use them with the most effective results. Nonetheless, due to the scope of this study and the restricted content of the tests, further investigation, including a wider corpus of grammatical tasks, could contribute to a better understanding of the nature and characteristics of testing in foreign language teaching.

BIBLIOGRAPHY

1. Bellinger, David, Marci Decaro, Patricia Ralston (2015), "Mindfulness, anxiety, and high-stakes mathematics performance in the laboratory and classroom", *Consciousness and Cognition*, 37, 123-132.
2. Council of Europe (2018), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Language Policy Division, Strasbourg
3. Dimitrijević, Naum (1976), *Testiranje u nastavi stranih jezika*, Minerva, Beograd
4. Gavranović, Valentina, Marijana Prodanović (2020), "Grammar in Low-stakes, Informal Testing Context – Do Different Task Types Influence Performance?", *Uzdanica*, XVII/2, 33-48.
5. Gebril, Atta, Michael Eid (2017), "Test Preparation Beliefs and Practices in a High-Stakes Context: A Teacher's Perspective", *Language Assessment Quarterly*, 14 (4), 360-379.
6. Gergely, David (2007), "Investing the Performance of Alternative Types of Grammar Items", *Language Testing*, 24 (1), 65-97.
7. Haladyna, Thomas, Steven Downing (2004), "Construct-irrelevant variance in high-stakes testing", *Educational Measurement: Issues and Practice*, 23(1), 17-27.

8. Horwitz, Elaine, Michael Horwitz, Joann Cope (1986), "Foreign Language Classroom Anxiety", *The Modern Language Journal*, 70(2), 125–132.
9. Moran, Tim (2016), "Anxiety and working memory capacity: A meta-analysis and narrative review", *Psychological Bulletin*, 142, 831–864.
10. Prodanović, Marijana (2020), "Testing The Untestable: Humanistic Approaches to Language Testing in The Modern-era", *ICC Journal*, 2 (3), 46-49.
11. Purpura, James (2005), *Assessing Grammar*, CUP, Cambridge
12. Rapanta, Chrysi, Luca Botturi, Peter Goodyear, Lourdes Guardia, Marguerite Koole (2020), "Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity", *Postdigital Science and Education*, 1-23. <https://doi.org/10.1007/s42438-020-00155-y>
13. Van Patten, Bill, Daniel Trego, Walter P. Hopkins (2015), "In-Class vs. Online Testing in University-Level Language Courses: A Research Report", *Foreign Language Annals*, 48 (4), 659–668.

TESTIRANJE GRAMATIKE U ONLAJN OKRUŽENJU: UPOREDNI PRIKAZ REZULTATA TESTOVA VISOKOG I NISKOG RIZIKA

Sažetak

Provera znanja kao i dizajniranje testova oduvek su zauzimali važno mesto u oblasti nastave engleskog jezika. Ovaj rad ilustruje analizu odgovora grupe studenata koji pohađaju studije engleskog jezika, datih u okvirima dva gramatička elektronska (onlajn) testa, u čijem je fokusu upotreba glagolskih vremena u kontekstu. Analiza je obuhvatila poređenje rezultata dobijenih posredstvom testa niskog rizika sačinjenog od zadataka zasnovanih na dopunjavanju informacije, s jedne, odnosno testa visokog rizika, zasnovanog na odabiru tačnog od ponuđenih odgovora, s druge strane. Navedena analiza imala je za cilj ostvarivanje uvida u gramatičke partije koje su studentima bile najviše, odnosno najmanje zahtevne, kao i poređenje srodnosti i razlika u učinku dobijenom posredstvom različitih modela i prirode testova. Rezultati su pokazali, a u vezi sa navedenim, nemali broj sličnosti među odgovorima, bez obzira na prirodu i formu testa.

Ključne reči: testovi visokog i niskog rizika; glagolska vremena; engleski jezik; onlajn okruženje

Adrese autorica
Authors' address

Marijana M. Prodanović
Singidunum University, Belgrade, Serbia
English Language Department
mprodanovic@singidunum.ac.rs

Valentina M. Gavranović
Singidunum University, Belgrade, Serbia
English Language Department
vgavranovic@singidunum.ac.rs