

Analiza valjanosti e-procjenjivanja u engleskomu kao jeziku struke

Prosenjak, Bojan

Doctoral thesis / Disertacija

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

<https://doi.org/10.17234/diss.2023.118679>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:962876>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-26**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)





Sveučilište u Zagrebu

Filozofski fakultet

Bojan Prosenjak

**ANALIZA VALJANOSTI
E-PROCJENJIVANJA
U ENGLESKOMU KAO JEZIKU STRUKE**

DOKTORSKI RAD

Mentor: prof. dr. sc. Vesna Bagarić Medve

Zagreb, 2023.



University of Zagreb

Faculty of Humanities and Social Sciences

Bojan Prosenjak

AN ANALYSIS OF THE VALIDITY OF ONLINE ASSESSMENT IN ESP

DOCTORAL THESIS

Supervisor: prof. Vesna Bagarić Medve, PhD

Zagreb, 2023

O MENTORICI

Vesna Bagarić Medve rođena je 1968. u Osijeku. Diplomirala je engleski jezik i književnost i njemački jezik i književnost na Pedagoškom fakultetu Sveučilišta J. J. Strossmayera u Osijeku. Na Filozofskom fakultetu Sveučilišta u Zagrebu stekla je 2001. znanstveni stupanj magistra znanosti, a 2007. i akademski stupanj doktora humanističkih znanosti iz područja filologije. Kao stipendistica DAAD-a i ÖAD-a usavršavala se stručno i znanstveno na sveučilištima u Njemačkoj (Wilhelm-Pieck-Universität u Rostocku, Martin-Luther-Universität u Halleu) i Austriji (Karl-Franzes-Universität u Grazu).

Nakon završetka diplomskog studija radila je kao nastavnica njemačkog i engleskog jezika, a od 2000. zaposlena je pri Odsjeku za njemački jezik i književnost Filozofskog fakulteta u Osijeku. U znanstveno-nastavno zvanje docentice izabrana je 2008., a 2014. u znanstveno-nastavno zvanje izvanredne profesorice. Od 2019. u znanstveno-nastavnom je zvanju redovite profesorice. Od 2009. do 2014. obnašala je funkciju prodekanice za nastavu na Filozofskom fakultetu u Osijeku.

Na diplomskoj i poslijediplomskoj razini predaje kolegije o ovladavanju inim jezicima, ispitivanju i procjenjivanju komunikacijske kompetencije, istraživanju u nastavi stranog jezika, psiholingvističkim metodama istraživanja i glotodidaktičke kolegije.

Njezini primarni znanstveni interesi obuhvaćaju komparativna istraživanja jezične svjesnosti, sastavnica komunikacijske kompetencije s posebnim naglaskom na diskursnoj kompetenciji te jezičnog testiranja u engleskom i njemačkom jeziku. Objavila je dvije znanstvene i jednu stručnu monografiju, tri uredničke knjige te više od 50 znanstvenih i stručnih radova. Izlagala je na 49 međunarodnih znanstvenih skupova u inozemstvu i u Hrvatskoj te sudjelovala kao istraživač u 12 znanstvenih i stručnih domaćih i međunarodnih projekata. Od 2017. do 2020. bila je voditeljica projekta IP-2016-06-5736 „Koherencija pisanoga teksta u inome jeziku: hrvatski, njemački, engleski, francuski i mađarski jezik u usporedbi” Hrvatske zaklade za znanost. Članica je Hrvatskog društva učitelja i profesora njemačkog jezika / *Kroatischer Deutschlehrerverband (KDV)*, Hrvatskog društva za primijenjenu lingvistiku (HDPL), Hrvatskog filološkog društva te udruga European Association for Language Testing and Assessment (EALTA) i EuroSLA. Od 2009. do 2013. bila je zamjenica glavne urednice, a od 2018. do 2020. glavna urednica časopisa Strani jezici.

SAŽETAK

S napretkom tehnologije e-ispiti sve su prisutniji zbog svojih mnogobrojnih prednosti, a njihova primjena postaje uvriježena u e-nastavi. Zbog toga su istraživanja na temu e-ispita te njihovih usporedba s ispitima na papiru sve češća, no istraživanja same valjanosti e-procjenjivanja još su uvijek malobrojna. Stoga je cilj ovog istraživanja analizirati valjanost e-procjenjivanja u engleskomu kao jeziku struke na fakultetu gdje su ispiti često ispiti srednjeg rizika, pa njihovi rezultati moraju biti valjani pokazatelji sposobnosti ispitanika.

U istraživanju je sudjelovalo 122 sudionika, koji su rješavali ispit poslovnog engleskog jezika, koji ispituje sposobnost razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje, i to njegovu e-inačicu te inačicu za na papiru. Također su ispunili četiri strukturirana upitnika, i to Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika, Upitnik o stavovima ispitima općenito, Upitnik o stavovima o računalima, te Upitnik o razini ispitne anksioznosti. Prikupljeni su se podatci zatim obradili primjenom metode potvrje procjenjivanja te postupcima deskriptivne i inferencijalne statistike.

Dobiveni rezultati istraživanja pružaju empirijski dokaz za valjanost e-procjenjivanja i obogaćuju znanstvenu literaturu u području jezičnog procjenjivanja, a s praktične strane doprinose poboljšanju kvalitete procjenjivanja na visokoškolskim ustanovama, ali i u svim školama i centrima gdje se provode razni oblici procjenjivanja. Ovo istraživanje također uvodi nove pojmove u hrvatski jezik u ovom znanstvenom području budući da se radi o prvom istraživanju valjanosti e-procjenjivanja u Hrvatskoj s obzirom na primijenjenu metodologiju istraživanja.

Ključne riječi: e-ispit, e-procjenjivanje, valjanost, potvrjepa procjenjivanja, engleski kao jezik struke (EJS), stavovi o ispitima, stavovi o računalima, ispitna anksioznost

SUMMARY

Theoretical background

The English language is one of the most obvious characteristics of globalization, taught to non-native speakers worldwide, with the aim of helping them achieve communicative competence (cf. Tulasi & Murthy, 2022). English for Specific Purposes (ESP) is only one branch of English as a Foreign Language or English as a Second Language, and it is seen as an approach that puts the learner's needs at the centre of the teaching process (cf. Hutchinson & Waters, 1987; Robinson, 1991; Paltridge & Starfield, 2013). ESP developed at a different speed in different countries and in several stages (cf. Johns, 2013; Upton & Connor, 2013), and today it includes cooperation and collaboration between teachers and experts from the relevant domains, and even team teaching (Dudley-Evans & St. John, 1998).

Language abilities in ESP have to be assessed, and language assessment “involves obtaining evidence to inform inferences about a person's language-related knowledge, skills or abilities” (Green, 2014: 5). Assessment is thus seen in its narrower and its wider sense – if the former, it overlaps with testing (Kramer, 2013), as it includes standardized tests and/or scales/rubrics (Bagarić Medve & Škarica, 2023), portfolios, observations, etc. (McNamara, 2000; Council of Europe, 2001); in the latter, it overlaps with evaluation, as they both include giving judgements on the learner's language knowledge and proficiency, but evaluation goes even further and encompasses the monitoring of the development and progress of learning, as well as the goals and outcomes of teaching programmes (Council of Europe, 2001; Jelaska & Cvikić, 2008).

Technology has been part of language assessment since the first half of the 20th century (cf. Fulcher, 2010; Brooks, 2017), but computers started being used in foreign language teaching in the 1980s, primarily in the USA, the UK and the Netherlands (Dunkel, 1999). Computer-based tests have many advantages in comparison with paper-based tests, but also some potential downsides (cf. Gruba & Corbel, 1997; Chalhoub-Deville, 2001; Roever, 2001; Chapelle & Douglas, 2006; Kramer, 2013; Chapelle & Voss, 2017), and the same is true for online tests or web-based tests, which proliferated with the more widespread use of the internet from the beginning of the 21st century onwards (cf. Dunkel, 1999; Chapelle, Jamieson & Hegelheimer, 2003; García Laborda, 2007; Lim & Kahng, 2012; Lesiak-Bielawska, 2015; Chapelle & Voss, 2016; Isbell & Kremmel, 2021; Li et al., 2021; Turnbull et al., 2021). Though the mode of test administration can be different, all tests have to follow certain development

stages – planning (which includes writing test specifications), writing test items, trialling, validation, and post-testing activities, such as writing test handbooks, training staff and test maintenance (cf. Davies, 1984; McNamara, 2000; Udier & Jelaska, 2008; Fulcher, 2010; Green, 2014; Hughes & Hughes, 2020; Green & Fulcher, 2021).

As every well-designed test needs to undergo the validation process, this means that validity is one of the key qualities of any test. Even though validity is seen by some researchers as a unitary concept, which is also our standpoint, or as comprising its different types (cf. Anastasi, 1976; Messick, 1980, 1988; Angoff, 1988; Cronbach, 1988; Douglas, 2001; Urbina, 2004; Davies & Elder, 2005; Weir, 2005; Fulcher & Davidson, 2007; Milas, 2009), it is never complete, but rather at a certain degree at which the gathered data supports the intended interpretation of the test results (cf. Anastasi, 1976; Angoff, 1988; Messick, 1989; Urbina, 2004; Bachman, 2004; Weir, 2005; Fulcher & Davidson, 2007; Green, 2014; Green & Fulcher, 2021). During the process of validation, it is not enough to claim that a certain assessment is valid, but these claims have to be supported by the corresponding backing with its warrants, and all the possible rebuttals have to be rejected as much as possible, using empirical evidence (cf. Messick, 1980; Davies, 1984; Angoff, 1988; Cronbach, 1988; Messick, 1989; Kane, 1992, 2010; Bachman & Palmer, 1996; Kane, Crooks & Cohen, 1999; McNamara, 2000; Chapelle, Jamieson & Hegelheimer, 2003; Toulmin, 2003; Bachman, 2004, 2005, 2015; Urbina, 2004; Davies & Elder, 2005, 2010; Weir, 2005; Chapelle & Douglas, 2006; McNamara & Roever, 2006; Milas, 2009; Bachman & Palmer, 2010; Shepard, 2016). A conceptual framework that allows for such validation is the Assessment Use Argument (AUA), developed by Bachman (2005, 2015) and Bachman & Palmer (2010), based on Kane's interpretative argument (Kane, 1992, 2001, 2004; Kane, Crooks & Cohen, 1999) and Toulmin's argument structure (Toulmin, 2003). AUA is divided into the Assessment Validity Argument (AVA) and the Assessment Utilization Argument – the former linking test-takers' performance with the interpretations of their results, and the latter linking those interpretations with the possible decisions to be made based on them.

There are numerous possible threats to validity (cf. Chapelle & Douglas, 2006), of which we single out test-takers' attitudes and their level of test anxiety. Attitudes towards language, country, people, tests, teachers, learning, etc. can have an impact on test-takers' performance (cf. Gardner, 1985; Eiser, 1986; Brunfaut & Clapham, 2013; Lasagabaster, 2013), and the level of their test anxiety can have an impact on their performance (cf. Alderson & Wall, 1993; Zeidner, 1998; Cassady & Johnson, 2002; Živčić-Bećirević, 2003; Juretić, 2008;

Erceg Jugović & Lauri Korajlija, 2012). These threats have to be taken into consideration during any validation processes.

AUA was used in several studies to analyse the validity of assessments used in different educational contexts (cf. Llosa, 2008; Wang et al., 2012; Long et al., 2018; Jun, 2021; Park, 2021), which proved beneficial for the present study in terms of comparison of results and the methodology applied. The other groups of selected research that follow were also used to compare the results and to use some of the same instruments, as described later in this summary. The second group of studies compared the validity of paper-based tests with computer-based and online tests (cf. Al-Amri, 2007; Hewson et al., 2007; Khoshsima et al., 2017; Öz & Özturan, 2018; Hewson & Charlton, 2019); the third group investigated whether test-takers' performance is impacted by their attitudes towards tests and computers (cf. Gorsuch, 2004; Fan & Ji, 2014; Dizon, 2016; Hartshorn & McMurry, 2020; Chung & Choi, 2021; Hoang et al., 2021); the fourth group analysed the attitudes of other test users, such as teachers, principals and administrators (cf. Winke, 2011; Abduh, 2021; Ghanbari & Nowroozi, 2021; Yulianto & Mujtahin, 2021; Zhang et al., 2021; Lučev et al., 2022); the last selected group of studies investigated the potential effect of test anxiety on test takers' performance (cf. Cassady & Johnson, 2002; Chapell et al., 2005; Juretić, 2008; Aliakbari & Gheitasi, 2017).

The present study

Owing to technological advancements, online tests are becoming more frequent due to their advantages, and their administration is becoming indispensable in online classes. A lot of research is being published that analyses online tests and compares them to paper-based tests, and there is a great deal of researchers investigating the impact of learners' attitudes and test anxiety on their performance in online tests, but the field of analysing the validity of online assessment applying a systematic and methodological approach is still not fully explored.

In the Croatian context, to our knowledge, there are no extensive studies focusing on the analysis of the validity of online assessment. Therefore, the purpose of our research is to offer empirical evidence in this domain. The main aim of this thesis is to analyse the validity of online assessment in English for Specific Purposes (ESP), and the individual aims are as follows:

- 1) to compare the online test and paper-based test scores;
- 2) to examine the participants' attitudes towards online tests, tests in general and computers;
- 3) to examine the participants' test anxiety level.

The research questions (RQ) guiding us in reaching the abovementioned goals are the following:

RQ1: Do online tests test reading comprehension, comprehension and use of vocabulary and written production abilities in ESP?

RQ2: Are scores in online tests connected with the participants' attitudes towards online tests, tests in general and computers?

RQ3: Are scores in online tests connected with the participants' test anxiety level?

On the basis of the stated aims and previous research, the following hypotheses (H) are formed:

H1A: Online tests test the reading comprehension ability in ESP.

H1B: Online tests test the comprehension and use of vocabulary ability in ESP.

H1C: Online tests test the written production ability in ESP.

H2A: There is no correlation between online test scores and the participants' attitudes towards online tests.

H2B: There is no correlation between online test scores and the participants' attitudes towards tests in general.

H2C: There is no correlation between online test scores and the participants' attitudes towards computers.

H3: There is no correlation between online test scores and the participants' test anxiety level.

Participants, instruments and method

To achieve the abovementioned aims and test the hypotheses, we conducted our study with 122 participants, who were first-year graduate students of Economics at one Croatian university, enrolled in a Business English (BE) course at B1-B2 level according to the CEFRL*. They were taught by ten different teachers, one of whom was also the author of this study and of the test that was used as one of the instruments, both in its online and paper-based versions. The test comprised reading comprehension, comprehension and use of vocabulary, and written production tasks. The participants' performance in the latter was assessed by two ESP experts, who were the two other participants in this study. The other instruments were four structured questionnaires: Attitudes towards online BE tests questionnaire (which was adapted from Fan

* Common European Framework of Reference for Languages (Council of Europe, 2001).

& Ji, 2014), Attitudes towards tests in general questionnaire (which was the translation of the Attitudes Toward Tests Scale, created by Dodeen 2008 in Muñoz, 2017), Attitudes towards computers questionnaire (which was the translation of the Computing Attitudes Study, conducted by Hewson et al., 2007 and Hewson & Charlton, 2019), and Test anxiety questionnaire (which was the translation of the Westside Test Anxiety Scale, created by Driscoll, 2007 in Aliakbari & Gheitasi, 2017).

The data for our study were gathered in 2021 in two stages – the participants first took the online test and completed two questionnaires in Google Forms, and three and a half months later they took the paper-based test and completed the remaining questionnaires on paper. These data were then analysed using a mixed methods approach, starting with the qualitative analysis, which included the Assessment Use Argument (AUA). The quantitative portion of the study comprised the descriptive and inferential statistics – the correlation analysis and the analysis of variance. The aim was to analyse the interaction between the two factors (testing mode, i.e. online and paper-based, and test sections, i.e. reading comprehension, comprehension and use of vocabulary, written production) and to see whether the impact of one factor depends on the levels of the other. Furthermore, we wanted to examine the correlation of variables, i.e. the participants' test results and their responses in the questionnaires.

Ethical principles were adhered to while conducting this study (cf. Shuster, 1997; Steneck, 2006; Resnik & Shamoo, 2011; Kraš & Miličević, 2015; Truog et al., 2015; OUZP, 2016; Cergol, 2021), as well as certain testing rituals (Fulcher, 2010). In addition, a formal request was sent to the Head of the Foreign Languages Department and to the BE teachers, in which they gave their consent for this study, and an informed consent form was sent to the participants, which explained the present study and guaranteed their anonymity and allowed for the possibility to quit the study without any prejudice. Finally, the motivation for the participants to take part in this study was boosted by offering them intellectual and material awards, i.e. opportunities for further practice and gift cards (Cergol, 2021).

Results and discussion

In line with the first aim of this study, the quantitative analysis of the data has shown that the mean value of the participants' scores in the online test was higher than in the paper-based test overall, and in the reading comprehension and the written production section; the opposite is true for the comprehension and use of vocabulary section.

Additionally, the participants' responses in the questionnaires (Table 7) imply that their attitudes towards online BE tests are generally positive ($M = 4.78$, $\min = 2.75$, $\max = 6.00$) and that their attitudes towards tests in general are somewhat positive ($M = 3.2$, $\min = 2.00$, $\max = 4.59$). Next, their attitudes towards computers were grouped according to the factor analysis of this questionnaire, and they show relatively low level of computer anxiety ($M = 2.12$, $\min = 0.96$, $\max = 4.47$) and computer addiction ($M = 2.15$, $\min = 1.05$, $\max = 4.47$). Finally, the participants demonstrated a moderately low level of test anxiety ($M = 3.5$, $\min = 1.00$, $\max = 5.00$).

After the two ESP experts assessed the participants' written production in both online and paper-based tests in three categories (i.e. form, content, language) and in total, a t-test was performed (Table 9) and the ICC was calculated (Table 10) to analyse the interrater reliability. The results have shown that the interrater reliability was excellent or good (ICC values between 0.78 and 0.97) in all variables except in "paper-based test - language", where it was moderate (ICC = 0.73), as explained by Koo & Lee (2016).

Furthermore, to examine hypotheses H1A, H1B and H1C, a two-way repeated measures analysis of variance was conducted and it has shown that the first effect, i.e. the testing mode, is not statistically significant, which means that the online and the paper-based tests test the same participants' abilities, i.e. their reading comprehension, comprehension and use of vocabulary and written production. The other effect, i.e. the test content, was statistically significant, and the interaction between the two main effects can be seen in Picture 22. Further t-tests show that there are differences in the participants' scores in the categories "writing production - total" ($t = 2.996$; $df = 119$; $p < 0.01$) and "comprehension and use of vocabulary" ($t = -2.589$; $df = 119$; $p < 0.05$) – in the former, the participants achieve better results in the online test ($M = 10.837$; $SD = 3.305$), and in the latter, their results are better in the paper-based test ($M = 12.1$; $SD = 2.317$).

To examine hypotheses H2A, H2B, H2C and H3, correlation analyses were conducted (Table 12), and they demonstrate that the participants' attitudes towards online BE tests, tests in general and computers, as well as their level of test anxiety, do not show a statistically significant correlation with any section of the test. The only correlation that has been established was a weak negative correlation between the participants' test anxiety level and the reading comprehension section of the test ($r = -.205$; $p < .05$).

The qualitative analysis applied the part of the Assessment Use Argument which connects the participants' test performance to the interpretation of these results, i.e. the Assessment Validity Argument (AVA), as shown in Figure 22. The first claim in our AVA is

that the participants' online test results are consistent, for which the backing is that the test administration procedures are followed consistently, that the assessment procedures are good and also followed consistently, and that the results are consistent in different testing modes. The warrants for these backings are an expert author of the test, an expert examiner and invigilator, expert BE assessors for one section of the test, the use of Google Forms for automatic scoring of the other two sections of the test, and finally the abovementioned two-way repeated measures analysis of variance. The rebuttals for this first claim were rejected by the rebuttal data, which included the evidence of a quality test design, administration and assessment, as well as the earlier mentioned interrater reliability analysis and the abovementioned correlation analyses.

The second claim in our AVA is that the interpretation that the online test is a valid indicator of the reading comprehension, comprehension and use of vocabulary and written production abilities is meaningful, impartial, generalizable, relevant and sufficient. The backings for this claim are a well-defined construct, clear test specifications, the best possible performance of the participants, impartial test tasks, inoffensive test content, clear scoring criteria, impartial test administration, the online test that corresponds to the target language use, the online test assessment and criteria procedures that correspond to the ESP context, and finally, the scores that are relevant and sufficient indicators of the level of the participants' abilities. The warrants for these backings are the online test which is in accordance with the course outcomes, syllabus and teaching materials, the previously mentioned experts (i.e. the author, examiner, invigilator and assessors) and Google Form, the well-explained scoring criteria that are also written on the test itself, as well as the well-explained procedures and conditions of the test administration. The rebuttals for this second claim were rejected by the rebuttal data, which showed that the course outcomes, syllabus and teaching materials had been used in class and were constantly available to the participants in Google Classroom; that the participants' attitudes towards online BE tests, test in general and computers are not in a statistically significant correlation with any section of the test; that the scoring criteria were clear and explained to the participants; that all participants were able to access the test equally and demonstrate their language proficiency; that the participants' scores were a relevant and sufficient indicator of the level of their abilities. The only data that weakened one of the rebuttals were the results on the test anxiety level questionnaire, because of the previously mentioned weak negative correlation that has been established between the participants' test anxiety level and the reading comprehension section of the test.

Therefore, we can say that hypotheses H1A, H1B and H1C have been confirmed because of the weak positive correlation between the online and paper-based test, and the statistically significant content of the test in all its three sections. Our conclusion that the testing mode has no effect on the participants' performance is in line with the previous studies by Al-Amri (2007), Hewson et al. (2007), Mohammadi & Barzgaran (2012), Öz & Özturan (2018).

Equally, we can say that hypotheses H2A, H2B and H2C have also been confirmed because the participants' attitudes towards online BE tests, tests in general and computers are not statistically correlated with their results in any of the sections of the exam. Our results are in accordance with the studies that show that computer familiarity and attitudes towards computers do not have a significant impact on computer-based tests (Al-Amri, 2007; Khoshshima et al., 2017), on online tests (Hewson et al., 2007; Hewson & Charlton, 2019), and that the computer familiarity level does not have an impact on an online writing test (Mohammadi & Barzgaran, 2012). Furthermore, our results are partly in accordance with those of Fan and Ji (2014), which could be due to the fact that they analysed only the attitudes towards a certain paper-based English test. On the other hand, our results are completely the opposite of Gorsuch's (2004), maybe due to the fact that she analysed only a listening test on a sample of six participants, and to those of Hartshorn and McMurry (2020), who focused on the participants' attitudes towards online English classes due to COVID-19 pandemic.

Hypothesis H3 has mostly been confirmed because the correlation analysis results indicate that the participants' levels of test anxiety are statistically correlated only with the reading comprehension section of the test, as previously mentioned. Therefore, we can say that the participants with a lower test anxiety level achieve better results in this section, which in part corresponds to the studies conducted by Cassady & Johnson (2002), Chapell et al. (2005), Juretić (2008), and Erceg Jugović & Lauri Korajlija (2012),

Conclusion, implications and limitations

The validity of online assessment is still an insufficiently researched area where new insight appears every day. What is important to note is that whatever the reason for its use, online assessment needs to be valid, irrelevant of the fact whether validity is seen as a unitary concept, which is also our standpoint, or as comprising its different types. Validity thus needs to be proven, bearing in mind that it is never complete, but at a certain degree, and in this study the Assessment Use Argument (AUA) was used as one possible validation method, i.e. its Assessment Validity Argument (AVA) part.

Other researchers have also used the AUA in their studies (Chapelle, Jamieson & Hegelheimer, 2003; Llosa, 2008; Wang et al., 2012; Long et al., 2018; Jun, 2021; Park, 2021), which enabled us to compare their methodology with ours, since we have also used it in our research, focusing on its AVA part, in order to analyse the validity of online assessment in English for Specific Purposes (ESP). Our results show that the results that our participants achieved during their performance in an online Business English test are consistent, and that the interpretation of these results – which says that the online test is a valid indicator of the reading comprehension, comprehension and use of vocabulary and written production abilities – is meaningful, impartial, generalizable, relevant and sufficient. Therefore, because the empirical evidence used in our research rejected all possible rebuttals and weakened only one of them, we can conclude that our AVA is strong (cf. Fitzgerald & Clenton, 2010). Taking into consideration the fact that validity “is a matter of degree, not all or none” (Messick, 1989: 33), we can also conclude that our results increase the degree of validity of online ESP tests (cf. Anastasi, 1976; Angoff, 1988; Messick, 1989; Anderson & Banerjee, 2001; Urbina, 2004; Bachman, 2004; Davies & Elder, 2005; Weir, 2005; McNamara & Roever, 2006; Fulcher & Davidson, 2007; Milas, 2009; Green, 2014; Fitzpatrick & Clenton, 2010; Shepard, 2016; Chapelle, 2021; Green & Fulcher, 2021).

Consequently, we can also conclude that the results of this research, which was conducted in the Croatian context, are significant because they are comparable with the results of all the studies previously mentioned in this summary that were conducted in different countries and that used various instruments. This demonstrates the significance of our research, placing it in a global context, to which it contributes with the applied analytical procedures and methods. What is more, being the first systematic research of validity in Croatia that used the AUA, it has also contributed to Croatian terminology by offering translations of some crucial notions in the area of assessment, as is shown in the Glossary (Appendix 7).

Furthermore, the results of the present study can be used by all test users – from curriculum, syllabus and test authors, to teachers and students or test-takers – who have this way been given an argument for using an online assessment mode. This is especially important in situations when paper-based assessment cannot be administered due to a variety of reasons, such as pandemics, wars, etc., but also when online assessment is the preferred mode because of institutional or individual requirements or desires for digitalization of learning, teaching and assessment. Moreover, our results can be applied not only in ESP university courses (cf. Long et al., 2018) but in other tertiary, secondary and primary education institutions, i.e. wherever test users need to validate the applied instrument. This will improve the quality of online

assessment because the results achieved by test-takers will be valid indicators of the abilities that the test in question purports to actually be testing. This can apply to medium-stake assessment, such as the one used in the present study, i.e. one that assesses test-takers' achievement in a certain area, but also to low-stake assessment, i.e. one that is used formatively, providing feedback on their performance (cf. Roeber, 2001). Finally, AVA can be beneficial for teachers, too, because it can point them towards the areas they could improve while assessing the learners' mastery of certain standards (cf. Llosa, 2008).

As far as the limitations of the present study are concerned, the first one is the sample. It included only first-year students, who were also taught by different teachers (one of whom was the author of the present study) and were unevenly distributed in this regard. The second one is the language test – the construct of the BE test was harmonized with the syllabus and with the university's test administration rules, and the researcher was at the same time its author. The third limitation was the assessment of the written production, which was performed by two LSP experts, one of whom was the researcher. The fourth limitation were the other applied instruments, i.e. the four questionnaires. They were used following the decision made after the examination of available scientific literature that was in accordance with the aims of our study. Taking the abovementioned into consideration, the following is suggested as possible further research: include more participants from different years at university and in equal numbers from each respective teacher; include other test users in the research, such as teachers, examiners and administrators, to test their attitudes towards the test that is being validated (Winke, 2011); expand the test so that it includes more test items and/or more sections, such as listening comprehension and spoken production and interaction; use more assessors; use automatic writing evaluation applications if the technological advancements allow, making sure that the assessment is not lead by technology, but that technology only enhances it (Brunfaut, 2023); use more questionnaires that examine the participants' attitudes; compare the syllabi of other courses, not only of foreign languages, and develop an instrument that would test those abilities that are common to all relevant courses; ensure that the author of the test is not simultaneously the validator of that test (cf. Davies & Elder, 2005; Winke, 2011).

Keywords: online test, online assessment, validity, Assessment Use Argument (AUA), English for Specific Purposes (ESP), attitude towards tests, attitude towards computers, test anxiety

SADRŽAJ

1. U V O D.....	1
1.1. SVRHA I CILJEVI RADA.....	1
1.2. STRUKTURA RADA.....	2
2. T E O R I J S K A P O L A Z I Š T A.....	4
2.1. ENGLSKI KAO JEZIK STRUKE.....	4
2.1.1. Definicija engleskoga kao jezika struke	4
2.1.2. Razvoj engleskoga kao jezika struke	6
2.1.3. Engleski kao jezik struke danas.....	8
2.2. PROCJENJIVANJE	10
2.2.1. Definicija procjenjivanja.....	10
2.2.2. Vrste procjenjivanja.....	15
2.2.3. Razvoj jezičnog procjenjivanja	20
2.2.4. Tehnologija u jezičnom procjenjivanju	24
2.2.4.1. <i>Počeci primjene tehnologije u jezičnom procjenjivanju.....</i>	<i>24</i>
2.2.4.2. <i>Ispiti na računalima.....</i>	<i>25</i>
2.2.4.3. <i>Prilagodljivo ispitivanje na računalima</i>	<i>27</i>
2.2.4.4. <i>Automatsko procjenjivanje pisane proizvodnje.....</i>	<i>29</i>
2.3. E-ISPITI.....	32
2.3.1. Definicija, prednosti i nedostaci e-ispita.....	32
2.3.2. Primjena e-ispita.....	35
2.3.3. Izrada e-ispita	37
2.3.4. Svojstva e-ispita.....	41
2.4. VALJANOST.....	43
2.4.1. Definicija valjanosti.....	43
2.4.2. Utvrđivanje valjanosti.....	44
2.4.3. Jedna ili više valjanosti?	46
2.4.3.1. <i>Sadržajna valjanost.....</i>	<i>47</i>
2.4.3.2. <i>Kriterijska valjanost.....</i>	<i>48</i>
2.4.3.3. <i>Konstruktna valjanost</i>	<i>49</i>
2.4.3.4. <i>Ostale valjanosti</i>	<i>51</i>
2.4.3.5. <i>Zaključak.....</i>	<i>52</i>
2.4.4. Valjanost procjenjivanja	53

2.4.5.	Prijetnje valjanosti procjenjivanja	55
2.4.5.1.	<i>Stavovi ispitanika</i>	56
2.4.5.2.	<i>Ispitna anksioznost</i>	58
2.5.	POTKRJEPA PROCJENJIVANJA	61
3.	EMPIRIJSKA POLAZIŠTA	73
3.1.	PRIMJENA METODE POTKRJEPE PROCJENJIVANJA	73
3.2.	JEZIČNI ISPITI NA PAPIRU, NA RAČUNALU I E-ISPITI.....	80
3.3.	STAVOVI I JEZIČNA IZVEDBA NA ISPITU.....	83
3.3.1.	Stavovi ispitanika o ispitima	83
3.3.2.	Stavovi ispitanika o računalima	85
3.3.3.	Stavovi drugih dionika ispita	86
3.4.	ISPITNA ANKSIOZNOST I JEZIČNA IZVEDBA NA ISPITU	90
4.	VALJANOST E-PROCJENJIVANJA U ENGLESKOMU KAO JEZIKU STRUKE	93
4.1.	CILJEVI ISTRAŽIVANJA, ISTRAŽIVAČKA PITANJA I HIPOTEZE	95
4.2.	METODOLOGIJA ISTRAŽIVANJA	96
4.2.1.	Sudionici	96
4.2.2.	Instrumenti	97
4.2.2.1.	<i>Jezični ispit</i>	97
4.2.2.2.	<i>Upitnik o stavovima o e-ispitima PEJ-a</i>	100
4.2.2.3.	<i>Upitnik o stavovima o ispitima općenito</i>	101
4.2.2.4.	<i>Upitnik o stavovima o računalima</i>	101
4.2.2.5.	<i>Upitnik o ispitnoj anksioznosti</i>	102
4.2.3.	Prikupljanje podataka	103
4.2.4.	Analiza podataka	104
4.2.4.1.	<i>Postupci statističke analize</i>	104
4.2.4.2.	<i>Metoda potkrjepe procjenjivanja</i>	105
4.2.5.	Etička načela	107
4.3.	REZULTATI ISTRAŽIVANJA	109
4.3.1.	Preliminarne provjere i statističke analize podataka	109
4.3.1.1.	<i>Analiza i tretman podataka koji nedostaju te univarijatna i multivarijatna odstupanja rezultata</i>	109
4.3.1.2.	<i>Deskriptivni rezultati</i>	110
4.3.1.2.1.	<i>Izvedba sudionika na jezičnom ispit</i>	110

4.3.1.2.2.	Odgovori na Upitniku o stavovima o e-ispitima Poslovnog engleskog jezika	112
4.3.1.2.3.	Odgovori na Upitniku o stavovima o ispitima općenito	114
4.3.1.2.4.	Odgovori na Upitniku o stavovima o računalima	116
4.3.1.2.5.	Odgovori na Upitniku o razini ispitne anksioznosti....	121
4.3.1.3.	<i>Formiranje kompozitnih varijabla, deskriptivni rezultati i normalitet distribucija kompozitnih varijabla.....</i>	<i>123</i>
4.3.2.	Bivarijatne povezanosti varijabla primijenjenih u istraživanju	128
4.3.3.	Testiranje hipoteza.....	131
4.3.4.	Potkrjepa valjanosti e-procjenjivanja u engleskomu kao jeziku struke	133
4.3.4.1.	<i>Prva tvrdnja – rezultati koje su ispitanici ostvarili na e-ispitu dosljedni su.....</i>	<i>136</i>
4.3.4.2.	<i>Druga tvrdnja – tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke smisleno je, nepristrano, uopćivajuće, značajno i dostatno.</i>	<i>139</i>
4.4.	DISKUSIJA REZULTATA ISTRAŽIVANJA	144
4.4.1.	Usporedba rezultata na e-ispitu i ispitima na papiru	145
4.4.2.	Ispitivanje stavova sudionika o e-ispitima, o ispitima općenito i o računalima	148
4.4.3.	Ispitivanje razine ispitne anksioznosti sudionika	151
4.4.4.	Zaključak diskusije	152
4.5.	GLOTODIDAKTIČKE IMPLIKACIJE ISTRAŽIVANJA	153
5.	ZAKLJUČAK.....	155
	POPIS LITERATURE.....	160
	PRILOZI.....	170
PRILOG 1.	Ispit Poslovnog engleskog jezika	170
Prilog 1.1.	Rezultati izvedbe sudionika na jezičnom ispitima	175
Prilog 1.2.	Histogrami procjenjivača A i B na korespondentnim mjerama	179
PRILOG 2.	Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika.....	183
Prilog 2.1.	Izjave iz upitnika o stavovima o ispitima The Fudan English Test (FET).....	184
Prilog 2.2.	Odgovori sudionika na Upitniku o stavovima o e-ispitima PEJ-a.	185

Prilog 2.3.	Matrica komponenata latentne strukture Upitnika o stavovima o e-ispitima PEJ-a (N = 120).....	188
Prilog 2.4.	<i>Scree plot</i> prikaz latentne strukture Upitnika o stavovima o e-ispitima PEJ-a (N = 120).....	189
PRILOG 3.	Upitnik o stavovima o ispitima općenito	190
Prilog 3.1.	Ljestvica stavova o ispitima	191
Prilog 3.2.	Odgovori sudionika na Upitniku o stavovima o ispitima općenito	192
Prilog 3.3.	Matrica komponenata latentne strukture Upitnika o stavovima o ispitima općenito (N = 120)	195
Prilog 3.4.	<i>Scree plot</i> prikaz latentne strukture Upitnika o stavovima o ispitima općenito (N = 120).....	196
PRILOG 4.	Upitnik o stavovima o računalima.....	197
Prilog 4.1.	Istraživanje stavova o računalima.....	199
Prilog 4.2.	Odgovori sudionika na upitniku o stavovima o računalima.....	203
Prilog 4.3.	Matrica komponenata latentne strukture Upitnika o stavovima o računalima (N = 120).....	212
Prilog 4.4.	<i>Scree plot</i> prikaz latentne strukture Upitnika o stavovima o računalima (N = 120).....	213
PRILOG 5.	Upitnik o razini ispitne anksioznosti	214
Prilog 5.1.	Ljestvica ispitne anksioznosti Westside.....	215
Prilog 5.2.	Odgovori sudionika na upitniku o razini ispitne anksioznosti	216
Prilog 5.3.	Matrica komponenata latentne strukture Upitnika o razini ispitne anksioznosti (N = 120).....	219
Prilog 5.4.	<i>Scree plot</i> prikaz latentne strukture Upitnika o razini ispitne anksioznosti (N = 120).....	219
PRILOG 6.	Histogrami distribucija kompozitnih varijabla.....	220
PRILOG 7.	Glosar.....	223
ŽIVOTOPIS.....		229

1. U V O D

1.1. SVRHA I CILJEVI RADA

U situacijama kad se postupci procjenjivanja ispitanikovih jezičnih vještina ne mogu provoditi na tradicionalan način, odnosno ispitima na papiru, alternativa postaje e-procjenjivanje – procjenjivanje putem interneta primjenom e-ispita. U zadnjih nekoliko godina mnoge se sastavnice procesa jezičnog poučavanja, a stoga i jezičnog ispitivanja, procjenjivanja i vrjednovanja, provode putem interneta. Posljedično, stavovi sudionika tih procesa počeli su se istraživati i u hrvatskom i u globalnom kontekstu, a počela su se provoditi i istraživanja razine ispitne anksioznosti kod ispitanika prilikom polaganja ispita na papiru i e-ispita.

Uvid u dostupnu literaturu otkriva nam da područje analize procjenjivanja nije dovoljno istraženo, a posebice manjka uvida u analizu valjanosti procjenjivanja uopće, naročito e-procjenjivanja, i to ne samo engleskoga kao jezika struke, već i općeg engleskog i drugih stranih jezika. Stoga je cilj istraživanja koje se opisuje u ovom radu ispitati i pružiti uvid u valjanost e-procjenjivanja u engleskomu kao jeziku struke, a njegovi su pojedinačni ciljevi:

1. usporediti rezultate na e-ispitu i ispitu na papiru;
2. ispitati stavove sudionika o e-ispitima poslovnog engleskog jezika, o ispitima općenito i o računalima;
3. ispitati razinu ispitne anksioznosti sudionika.

Očekuje se da će rezultati istraživanja valjanosti e-procjenjivanja i primjene e-ispita koji se prikazuju u ovom radu pružiti empirijski dokaz za valjanost primjene e-procjenjivanja, što je iznimno važno u situacijama kad se ne mogu primjenjivati tradicionalni načini procjenjivanja, odnosno ispiti na papiru. Smatra se da će se na ovaj način poboljšati kvaliteta e-procjenjivanja, prvenstveno na visokoškolskim ustanovama na ispitima srednjeg rizika poput završnih semestralnih ispita kojima se prati postignuće studenata u stranom jeziku, ali i u srednjim i osnovnim školama koje primjenjuju e-ispite i koriste se rezultatima koje učenici na njima postižu za procjenjivanje njihovih jezičnih vještina. Nadalje, očekuje se da će se ovim istraživanjem upotpuniti znanstvena literatura na globalnoj razini jer je područje analize valjanosti e-procjenjivanja još uvijek nedovoljno istraženo. Zadnje, ali ne manje važno, u ovom će se radu uvesti određeni izrazi za postojeće pojmove na engleskom jeziku, čime će se obogatiti pojmovlje u hrvatskom jeziku u području procjenjivanja, naročito konceptualnog okvira potkrjepe procjenjivanja.

1.2. STRUKTURA RADA

Rad je podijeljen u pet dijelova – ovaj prvi dio predstavlja uvod, drugi prikazuje teorijska, a treći empirijska polazišta za naše istraživanje, u četvrtom dijelu prikazano je naše istraživanje valjanosti e-ispita u engleskomu kao jeziku struke, dok peti donosi zaključak cijelog rada.

Nakon ovog **prvog dijela** slijedi **drugi dio** u kojemu se daje teorijski okvir našeg istraživanja, a koji je podijeljen na pet poglavlja čiji naslovi proizlaze iz naslova ovog rada – „Analiza valjanosti e-procjenjivanja u engleskomu kao jeziku struke“ – i to počevši od zadnjeg pojma prema prvomu. Stoga se **prvo poglavlje** bavi pojmom engleskoga kao jezika struke, a **drugo poglavlje** problematikom procjenjivanja, i to jezičnog, od definicije, preko njegovih vrsta i razvoja pa do primjene tehnologije u njemu – od prvih ispita na računalima, preko prilagodljivog ispitivanja na računalima, do automatskog vrjednovanja pisane proizvodnje. Nakon toga, u **trećem poglavlju** predstavljaju se e-ispiti – njihova definicija, prednosti i nedostaci, zajedno s njihovom primjenom, izradom i svojstvima. **Četvrto poglavlje** razrađuje pojmove valjanosti, utvrđivanja valjanosti i valjanosti procjenjivanja te ilustrira problematiku definiranja valjanosti budući da se na valjanost može gledati kao na jedinstveni koncept s nekoliko aspekata ili se može smatrati da postoji više vrsta valjanosti. Usto, obrađeni su stavovi o ispitima i ispitna anksioznost kao dvije važne moguće prijetnje valjanosti procjenjivanja. Nadalje, u **petom poglavlju** prikazan je konceptualni okvir potkrjepe procjenjivanja i uvedeno je novo pojmovanje na hrvatskom jeziku.

U **trećem dijelu** ovog rada daje se pregled dosadašnjih istraživanja valjanosti e-ispita, koji započinje **prvim poglavljem**, u kojemu su predstavljena istraživanja koja su se bavila analizom valjanosti e-ispita primjenom metode potkrjepe procjenjivanja. **Drugo poglavlje** predstavlja istraživanja koja su uspoređivala valjanost jezičnih ispita na računalu, na e-ispitima i na papiru. **Treće poglavlje** predstavlja istraživanja koja su proučavala kakav učinak na jezičnu izvedbu na ispitima imaju stavovi ispitanika o ispitima, odnosno o računalima, a predstavljena su i ona koja analiziraju stavove drugih dionika ispita o e-ispitivanju i e-procjenjivanju. **Četvrto poglavlje** predstavlja istraživanja koja su analizirala razinu ispitne anksioznosti ispitanika te njezin učinak na jezičnu izvedbu i na stavove o ispitima.

Četvrti dio donosi prikaz našeg empirijskog istraživanja valjanosti e-ispita u engleskomu kao jeziku struke, a podijeljeno je na četiri poglavlja. Nakon uvoda u ovo istraživanje, **prvo poglavlje** navodi njegove ciljeve, istraživačka pitanja i hipoteze, nakon čega je u **drugom poglavlju** predstavljena metodologija istraživanja. Prvo su prikazani sudionici

istraživanja te instrumenti koji su se u njemu primijenili – nestandardizirani jezični e-ispit poslovnog engleskog jezika, njegova istovjetna inačica na papiru i četiri strukturirana upitnika: Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika, Upitnik o stavovima o ispitima općenito, Upitnik o stavovima o računalima i Upitnik o ispitnoj anksioznosti Westside. Zatim slijede opisi prikupljanja i analize podataka, a na kraju se navode i etička načela koja su se slijedila u istraživanju. U **trećem poglavlju** prikazani su rezultati istraživanja – prvo deskriptivni, nakon kojih slijedi statistička analiza podataka i testiranje hipoteza. U **četvrtom poglavlju** rezultati se raspravljaju u odnosu na postavljene hipoteze. Naposljetku, **peto poglavlje** predstavlja glotodidaktičke implikacije ovog istraživanja.

Posljednji, **peti dio** donosi zaključke našeg istraživanja, zajedno s mogućim smjericama za daljnja istraživanja u području analize valjanosti procjenjivanja. Na kraju rada naveden je **popis literature**, a u **prilozima** nalaze se svi instrumenti primijenjeni u našem istraživanju, rezultati ispitanika na jezičnim ispitima, njihovi odgovori na četirima upitnicima, dodatni prikazi rezultata statističke obrade podataka te glosar pojmova od kojih se neki zahvaljujući ovom istraživanju prvi put uvode u hrvatski jezik.

2. TEORIJSKA POLAZIŠTA

U ovom su poglavlju predstavljani i raspravljani temeljni pojmovi kojima se koristimo u našem radu – počinje se s engleskim kao jezikom struke, zatim se nastavlja s procjenjivanjem i e-ispitima te se naposljetku dolazi do valjanosti i potkrjepe procjenjivanja kao metode analize valjanosti. Ti su pojmovi također uvršteni u glosar u prilogu 7 ovog rada, od kojih se jedni već primjenjuju u hrvatskom jeziku, dok druge uvodimo budući da u njemu ne postoje.

2.1. ENGESKI KAO JEZIK STRUKE

2.1.1. Definicija engleskoga kao jezika struke

Engleski je jezik danas jedan od najvidljivijih obilježja globalizacije i poučava se neizvornim govornicima diljem svijeta kao strani jezik, pri čemu je cilj da u njemu postignu komunikacijsku kompetenciju, što je ujedno i glavna slabost ovog procesa učenja i poučavanja – naime, engleski se jezik vrlo često poučava u okruženju koje mu nije izvorno, u onomu u kojemu nema doticaja s izvornim govornicima i bez izloženosti stvarnim situacijama, a to su dvije glavne sastavnice usvajanja stranog jezika (Tulasi i Murthy, 2022). Poučavanje engleskomu kao stranom jeziku odvija se već stoljećima i dugo se nazivalo jednostavno tako – poučavanjem engleskom jeziku, odnosno *English Language Teaching*, pa se postavlja pitanje kad se javio pojam **engleskoga kao jezika struke**¹ (EJS), odnosno pojam *English for Specific Purposes* (ESP). Prema Hutchinsonu i Watersu (1987), EJS je samo jedna grana engleskoga kao stranog jezika (eng. *English as a Foreign Language*), odnosno engleskoga kao drugog jezika² (eng. *English as a Second Language*), koji su oba glavne grane poučavanja engleskom jeziku, koji u svojim temeljima ima dva ključna koncepta: komunikaciju i učenje.

Istovremeno, EJS dio je šireg koncepta **jezika struke**³ (eng. *Language for Specific Purposes*), čiji počeci sežu u 1964. godinu kad su Halliday, Strevens, i McIntosh objavili svoju

¹ Engleski kao jezik struke još se naziva i engleskim za posebne namjene (EPN), kako navode Kuna i Ferčec (2023).

² Prema Medved Krajnović (2010: 11), pojam „drugi jezik [...] podrazumijeva i strani jezik [...] i treći jezik [...] i bilo koji sljedeći jezik kojime pojedinac nastoji ovladati. U hrvatskoj terminologiji [...] uvriježio se termin ini jezik [...] za sve pojedinčeve jezike osim prvoga usvojenoga.“ Nadalje, prema Kraš i Miličević, (2015: 3): „termin *second language* (drugi jezik) zapravo obuhvaća različite podređene kontekste, odnosno i drugi jezik, koji se usvaja u sredini gdje se govori, i strani jezik, koji se uči izvan takve sredine (obično u učionici)“. U ovom se radu pojmovi drugi jezik i strani jezik rabe dosljedno u skladu s time kako ih razumiju autori koje se navodi.

³ „U hrvatskom se jeziku osim pojma *strani jezik za posebne namjene* [...] još koriste nazivi *jezik struke*, *strani jezik u struci*, *strani jezik struke* i slično“ (Kuna i Ferčec, 2023: 589).

knjigu *The Linguistic Sciences and Language Teaching* (Upton i Connor, 2013). Jezik struke definira se kao područje primijenjene lingvistike koje se usredotočuje na analizu i poučavanje jeziku kako bi se zadovoljile posebne određene jezične potrebe neizvornih govornika jezika, koje mogu biti akademske prirode ili vezane za dotično zanimanje ili posao (ibid.). Ovaj pojam ujedno označava vrstu istraživanja primjene jezika u specifičnim područjima rada i obrazovanja koja se provode kako bi dala lingvističke opise za poučavanje jeziku struke (Truchot, 2013). Naposljetku, valja napomenuti i da se pojmovi EJS-a i jezika struke općenito često međusobno isprepliću jer se većina jezičnih analiza i poučavanja u području jezika struke usredotočuju na engleski, iako su francuski, njemački i ruski također značajni, dok važnost arapskoga, kineskoga i španjolskoga također svakodnevno raste (Upton i Connor, 2013).

Međutim, još uvijek ne znamo što EJS ustvari jest. Kako ističu Hutchinson i Waters (1987) te Paltridge i Starfield (2013), na EJS se ne gleda kao na neki proizvod, nego **pristup** poučavanju engleskomu kao drugom ili stranom jeziku koji ne uključuje određenu vrstu jezika, nastavne materijale ili metodologiju, već koji u središte stavlja učenje jezika i koji se temelji na potrebama učenika, odnosno njihovim razlozima za učenje i na primjeni engleskoga u određenom području kao njihovu cilju (npr. studiranje, sudjelovanje na poslovnim sastancima, rad na recepciji i sl.). Upravo je usredotočenost na **analizu potreba** učenika ono što najviše obilježava EJS, odnosno njegova usmjerenost cilju te osmišljavanje sadržaja i aktivnosti poučavanja koje proizlazi iz te analize kako bi se što je bolje moguće odredilo što učenik treba činiti rabeći engleski jezik kao medij (Robinson, 1991; Paltridge i Starfield, 2013). Istovremeno, ključna pitanja koja se prožimaju kroz poučavanje EJS-u jest kako identificirati te potrebe učenika i sadržaj koji on mora proizvesti, odnosno u kojemu mora sudjelovati, te kako nastavnici znaju da su njihovi učenici to uspješno izveli i što mogu učiniti da im pomognu ako nisu (Paltridge i Starfield, 2013).

Dodatno, EJS kao pristup obilježavaju četiri stalna i dva promjenjiva **obilježja** (Stevens, 1988, u Dudley-Evans i St. John, 1998) – s jedne strane, njegova su stalna obilježja da je različit od općeg engleskog jezika (eng. *General English*), da je osmišljen kako bi udovoljio posebnim potrebama učenika, zatim da se odnosi na određeni kontekst (tj. na teme, discipline, profesiju, aktivnosti) i naposljetku da je usredotočen na jezik koji je tom kontekstu primjeren u sintaksi, vokabularu, diskursu i semantici te na analizu tog diskursa; s druge strane, promjenjiva su mu obilježja da se može ograničiti na određenu vještinu (npr. samo na čitanje) i da se ne može poučavati prema unaprijed određenoj metodologiji.

Naposljetku, EJS valja staviti u odnos s drugim granama i (pod)vrstama poučavanja engleskom jeziku. Kako ističe Brinton (2013), znanstvenici koje smo već spomenuli u ovom

poglavlju – Strevens te Hutchinson i Waters – najistaknutiji su u svojim istraživanjima i definiciji EJS-a, a ono što im je zajedničko, iako je EJS vrlo široko područje koje obuhvaća mnoge grane i podvrste, jest sljedeće:

- EJS se jasno razlikuje od općeg engleskog jezika;
- postoji jasna granica između engleskoga za akademske namjene (eng. *English for Academic Purposes* – EAP) s jedne strane te engleskoga za profesionalne namjene (eng. *English for Occupational Purposes* – EOP)⁴ i engleskoga za strukovne namjene (eng. *English for Vocational Purposes* – EVP) s druge strane;
- engleski za akademske namjene često se izdvaja kao različit od specifičnijih grana, poput engleskoga u znanosti i tehnologiji (eng. *English for Science and Technology* – EST).

Ako se pogledaju podjele EJS-a, primjećuje se da se najčešće dijeli na EAP s jedne strane i na EOP, odnosno EVP s druge strane (Dudley-Evans i St John, 1998; Upton i Connor, 2013). EAP podrazumijeva jezično poučavanje usmjereno na određene potrebe učenika i samog jezika u okruženju u kojemu se poučava, koje je najčešće sekundarno i tercijarno obrazovanje, a naglasak stavlja na akademski jezik, npr. na pisanje članaka ili čitanje tekstova u udžbenicima (Upton i Connor, 2013). Dudley-Evans i St John (1998) EOP dalje granaju na razne podvrste, među ostalima na engleski u znanosti i tehnologiji, engleski u medicini (eng. *English for Medical Purposes* – EMP), pravni engleski ili engleski za pravne namjene (eng. *Legal English* ili *English for Legal Purposes* – ELP) te engleski za poslovne namjene (eng. *English for Business Purposes* – EBP) ili poslovni engleski (eng. *Business English*), za koji ističu da je to grana EJS-a u kojoj se izdaje najviše nastavnih materijala. EVP dalje granaju na njegove dvije podvrste – engleski vezan za određenu struku, odnosno strukovni engleski (eng. *Vocational English*), koji obuhvaća osposobljavanje za određene zanate i profesije, te engleski u pripremi za određenu struku (eng. *Pre-Vocational English*), koji obuhvaća vještine potrebne prilikom pronalaska posla, razgovora za posao i sl.

2.1.2. Razvoj engleskoga kao jezika struke

Prema Hutchinsonu i Watersu (1987), EJS nitko nije osmislio i uveo planski, već su ga iznjedrila **tri čimbenika** opisana niže.

- Po završetku Drugog svjetskog rata napretci u tehnologiji i trgovini stvorili su potrebu za međunarodnim jezikom, čiju je ulogu preuzeo engleski zbog gospodarske moći

⁴ Pojmovi prema Kuni i Ferčec (2023).

Sjedinjenih Američkih Država. Ona je još i više porasla u ranim 70-ima prošlog stoljeća uslijed naftne krize, kad su zapadni novac i iskustvo još većom brzinom pritekali u zemlje bogate naftom, npr. Iran, Kuvajt i Saudijska Arabija, kako navode Dudley-Evans i St. John (1998).

- Došlo je do pomaka u lingvistici s tradicionalnog usredotočivanja znanstvenih istraživanja na opisivanje pravila primjene jezika (tj. gramatike) na načine na koje se jezik uistinu primjenjuje u stvarnoj komunikaciji. Rezultati nekih od njih u 60-ima i 70-ima prošlog stoljeća doveli su do zaključaka da se jezik kojime govorimo i pišemo znatno razlikuje, i to na mnogo načina, te da ovisi o kontekstu (npr. trgovina, građevinarstvo, tehnologija i dr.).
- Razvoj edukacijske psihologije doveo je do naglašavanja važnosti učenika i njihovih stavova prema učenju jer nemaju svi jednake potrebe i interese, što utječe na njihovu motivaciju za učenjem i posljedično na učinkovitost učenja. Shodno tomu, počeli su se osmišljavati specijalizirani nastavni materijali na engleskom jeziku (npr. udžbenici za određeno područje). Upton i Connor (2013) ističu da je upravo ovaj pomak u paradigmi razumijevanja kako učenici uče postao ključnim obilježjem EJS-a jer ne samo da se poučava jeziku koji se primjenjuje u određenom kontekstu, već se uzimaju u obzir potrebe i interesi učenika u tim specifičnim kontekstima.

Hutchinson i Waters (1987) navode da se EJS razvijao različitom brzinom u različitim zemljama u **pet koraka**, koji su opisani niže, pri čemu se svaki naredni javio zbog određenih nedostataka ili nedovoljne iscrpnosti prethodnog.

1. Analiza gramatičkih i jezičnih obilježja u registrima engleskog jezika u 60-ima i 70-ima prošlog stoljeća (npr. inženjerstvo, biologija i dr.), prema čemu su se osmišljavali udžbenici i silabi. Kako navodi Johns (2013), ta su istraživanja bila deskriptivna i uključivala su statističku obradu pojavnosti gramatičkih struktura u pisanim diskursima, poglavito udžbenicima i znanstvenim člancima, zbog čega su imala opisnu valjanost, ali su malo toga objašnjavala.
2. Analiza retorike i diskursa, odnosno nastojanje da se razumije kako se rečenice kombiniraju u diskursu radi proizvodnje značenja, s počecima u 1970-ima, npr. primjena diskursnih oznaka u znanstvenim tekstovima i govorima koja je različita od primjene u ekonomskim i sl. Vjerojatno najvažnije istraživanje, prema Johns (2013), ono je Tarone i sur., objavljeno 1981. godine u prvom izdanju časopisa koji se tada zvao The ESP Journal, a koji je izdalo Američko sveučilište u Washingtonu. U njemu su se autori usredotočili na jedno središnje obilježje znanstvene proze, a to je učestala primjena

pasivnih struktura u znanstvenim člancima, i došli su do rezultata da se oko 25 % glagola u znanstvenim člancima javlja u pasivnoj strukturi. Međutim, Johns (ibid.) također napominje da mnoga istraživanja koja su prethodila ovom nisu uključivala provjeru oblikovanih hipoteza, unatoč postojanju i kontrastivnih analiza diskursa, iako su bila revolucionarna za to doba, i to ponajviše ona tzv. Vašingtonske škole (The Washington School). Na kraju ovog koraka valja naglasiti da su Tarone i sur., kao i znanstvenici Vašingtonske škole, dali još jedan doprinos metodologiji EJS-a, a to je savjetovanje sa stručnjacima iz područja koje je EJS obrađivao (ibid.).

3. Analiza ciljne primjene jezika (eng. *target language use domain*), odnosno analiza potreba učenika u 1970-ima i 1980-ima, pri kojoj se izrađuje iscrpan profil učenikovih potreba u odnosu na svrhu, okruženje i načine komunikacije te jezične vještine, funkcije i strukture. Iako su ove analize bile sistematizirane, i iako su stavile učenikove potrebe u središte procesa osmišljavanja nastavnog programa, bile su prejednostavne.
4. Analiza učenikovih vještina i strategija u 1980-ima, kojoj je polazište da se do značenja diskursa dolazi procesima rasuđivanja i tumačenja koje ne ovisi o površinskim strukturama jezika (npr. pogađanje značenja riječi iz konteksta, primjena srodnih riječi i dr.) i koja ističe da usredotočivanje na određeni registar nije potrebno jer njemu nisu svojstveni ni ovi procesi. Nedostatak ovog koraka jest taj da se težište stavljalo prvenstveno na analizu vještina čitanja i slušanja.
5. Pristup usmjeren na razumijevanje procesa učenja jezika, za razliku od svih prethodnih koji su bili usmjereni na opisivanje primjene jezika, odnosno onoga što govornici čine jezikom, koji se dalje razvija još i danas.

2.1.3. Engleski kao jezik struke danas

Danas se na poučavanje EJS-u gleda kao na zasebnu aktivnost unutar poučavanja engleskom jeziku, a na istraživanja u području EJS-a kao na punopravnu sastavnicu istraživanja u primijenjenoj lingvistici; također se smatra da je EJS ipak razvio vlastitu metodologiju i da svoja istraživanja temelji i na drugim disciplinama osim primijenjene lingvistike, no da je zadržao naglasak na praktičnim ishodima (Dudley-Evans i St John, 1998). S druge strane, valja istaknuti i da istraživanja u području EJS-a također imaju utjecaj i na druge discipline, poput istraživanja pisane proizvodnje u materinskom jeziku, istraživanja retorike u svijetu ekonomista i sl. (ibid.).

Poučavanje EJS-u vođeno je učenikovim vanjskim nejezičnim ciljevima učenja (npr. akademskim, profesionalnim, poslovnim), odnosno primjenom jezika izvan učionice pomoću jezika kao instrumenta, za što mu je potrebna određena jezična kompetencija (Basturkmen, 2010). Učenici su stoga motiviraniji kad ih se poučava EJS-u nego općem engleskomu, upravo zbog usredotočenosti, značajnosti i učinkovitosti tog procesa, budući da se ne obrađuju teme koje su im predobro poznate, koje ih ne zanimaju i sl. (Dudley-Evans i St John, 1998). Zato nastavnik (ili autor silaba) EJS-a mora saznati koje su jezične sposobnosti potrebne učenicima za ostvarivanje ciljeva i osigurati da sadržaj nastave EJS-a to pokriva (Basturkmen, 2010). Poučavanje EJS-u najučinkovitije je kad se u nj uključe stručnjaci iz dotičnog područja, što smo već prethodno spomenuli da se javilo u drugom koraku razvoja EJS-a, a Dudley-Evans i St. John (1998) ističu da se to može postići na tri načina – prvo, savjetovanjem sa stručnjakom kako bi se utvrdila vrsta jezika i primjene jezika kojima učenici moraju ovladati u određenom kontekstu (eng. *cooperation*); drugo, suradnjom stručnjaka koji su izravnije uključeni u planiranje nastave te odabir i izradu materijala (eng. *collaboration*); treće, timskim poučavanjem pri kojemu stručnjaci i nastavnici zajedno rade u učionici (eng. *team teaching*).

Nastavnici EJS-a pred sobom imaju dvije dodatne zadaće za razliku od nastavnika općeg engleskog jezika (Basturkmen, 2010) – prva je da se EJS usredotočuje „ili na područja znanja koja su izvan komunikacijskog repertoara svih obrazovanih izvornih govornika ili na obrađivanje specifičnih vještina kojih su obrazovani izvorni govornici tek prešutno svjesni“ (str. 9), zbog čega nastavnici EJS-a moraju znati kako osmisliti nastavu u konceptualnom okviru kojim još nisu ovladali te razviti sposobnosti analize i opisivanja specifičnih tekstova, što također ističu i Upton i Connor (2013); druga je da će nastava EJS-a često trajati samo određeno vrijeme, ovisno o promjenama potreba i okolnosti. Istovremeno, Dudley-Evans i St John (1998) naglašavaju da nastavnici trebaju biti svjesni dviju stvari – prvo, činjenice da je EJS višedimenzionalno područje te da u svoje poučavanje moraju trebati i željeti uključiti saznanja i rezultate istraživanja iz raznih drugih disciplina; drugo, osjetljivosti na kulturne razlike učenika iz akademskih i poslovnih krugova (npr. neki učenici koji su ujedno i nastavnici neće voljno surađivati u seminarima, nekim će poslovnim ljudima biti neugodno voditi sastanak i sl.).

U budućnosti, kako zaključuju Upton i Connor (2013), poučavanje EJS-u moralo bi slijediti niže opisane smjernice.

- EJS mora nastaviti razvijati perspektive i metode istraživanja koje omogućuju šire, ali i iscrpnije razumijevanje potreba učenika u kontekstu stalno promjenjiva svijeta.

- EJS mora uzeti u obzir odnos između izvornih i neizvornih govornika u kontekstima kad se engleskim služi kao *linguom francom* (npr. poslovne situacije, trgovina i sl.).
- EJS mora proučavati složene kontekstualne, sociološke i ideološke aspekte jezičnog međudjelovanja, koji su često višemodalni te uključuju tekst, govor i druge oblike sadržaja i prijenosa poruka. Za to nije dovoljna samo analiza riječi i diskursa, već su potrebne metode poput promatranja, etnografskih pristupa i analiza međudjelovanja kako bi se razumjela složenost primjene jezika i pragmatičkih strategija potrebnih za učinkovitu komunikaciju u posebnim kontekstima i u posebne svrhe.

2.2. PROCJENJIVANJE

2.2.1. Definicija procjenjivanja

Naslov je ovog rada „Analiza valjanosti e-procjenjivanja u engleskomu kao jeziku struke“, no budući da se u hrvatskom jeziku pojmovi ispitivanje, procjenjivanje i vrjednovanje jezičnog znanja ne primjenjuju dosljedno za označivanje triju različitih pojmova, a nerijetko im se značenja zamjenjuju i isprepliću, potrebno ih je objasniti te ih približiti engleskim istovrijednicama. Usto, dalje u našem radu neće se uvijek izričito navoditi pridjev „jezično“ uz pojmove ispita, ispitivanja, procjenjivanja i vrjednovanja, već se podrazumijeva da se radi o jezičnim ispitima, jezičnom ispitivanju, jezičnom procjenjivanju te jezičnom vrednovanju i u slučajevima bez dotičnog pridjeva.

Jezično procjenjivanje (eng. *language assessment*) definira se na sljedeći način (Green, 2014: 5):

Jezično procjenjivanje uključuje prikupljanje dokaza radi donošenja zaključaka o jezičnomu znanju, vještinama i sposobnostima koje posjeduje određena osoba. Dokazi dolaze od izvedbe⁵ na zadacima koji uključuju primjenu jezika, a zaključci su ono što tumačimo da ta izvedba znači na temelju svojih uvjerenja o prirodi jezika i njegovoj ulozi u životu ispitanika. Zaključcima do kojih dolazimo o ispitanicima [...] koristimo se uglavnom za donošenje odluka [...]

Problematikom ovih pojmova bavile su se i Jelaska i Cvikić (2008) koje su za **procjenjivanje** (eng. *assessment*) zaključile sljedeće:

⁵ Izvedba (eng. *performance*) definira se kao „govoreni ili pisani jezik (ili drugi odgovori) ispitanika koji se proizvodi kao odgovor na ispitne zadatke“ (Chapelle, 2021: 14).

Procjenjivanje — davanje suda o korisnikovome jezičnome znanju i umijeću. Prototipnomu procjenjivanju temelj su standardizirani oblici procjene, poput ispita, usmenoga ili pismenoga. No osim na temelju jezičnoga ispita, procjenjivati se može i na temelju samoga promatranja korisnika u različitim jezičnim djelatnostima. Ne procjenjuje ga se na temelju unaprijednih očekivanja, nego se procjenjuje kako i koliko zna. (str. 124)

Iz gornje se definicije vidi da se na procjenjivanje može gledati u užem i u širem smislu. Prema Bagarić Medve i Škarica (2023), **procjenjivanje u užem smislu** pojam je koji se primjenjuje za procjenu razine znanja i sposobnosti ispitanika pomoću ljestvica za procjenjivanje⁶, a autorice za nj navode sljedeće:

Rezultat toga postupka ne smatra se u punoj mjeri objektivnim jer je takav proces procjenjivanja podložan djelovanju raznih čimbenika (npr. stavova nastavnika, njegova emocionalnog stanja u trenutku procjenjivanja, njegove interpretacije kriterija i opisnika kriterija, vremena i mjesta procjenjivanja itd.) koji mogu utjecati na rezultat procjene, odnosno davanja suda. S druge strane, procjenjivanje pomoću kvalitativnih opisa jezične izvedbe uzima u obzir uvjete stjecanja znanja i sposobnosti te omogućuje nastavniku pružanje jasne, temeljite i individualizirane povratne informacije učeniku. (str. 452)

Razvidno je da se procjenjivanje u užem smislu isprepliće s pojmom **ispitivanja** (eng. *testing*), a Kramer (2013) čak i navodi da pojam procjenjivanja općenito obuhvaća sve metode ispitivanja i procjenjivanja premda se ponekad ispitivanje primjenjuje za formalne ili standardizirane ispite poput TOEFL-a (Test of English as a Foreign Language) i IELTS-a (International English Language Testing System), dok se procjenjivanje primjenjuje za neformalnije metode. Kako ilustrira McNamara (2000), ispitivanje za mnoge podrazumijeva ispitnu sobu s urom na zidu i ispit na papiru koji se rješava olovkom u zadanom vremenu ili sjedenje na stolici ispred sobe u kojoj se provodi usmeni ispit i ponavljanje naučenog izlaganja koje je spremno za recitaciju ispitivačima. Naravno, jezični se ispiti danas više ne provode samo na papiru i uživo, o čemu iscrpnije govore naše potpoglavlje 2.2.4. – Tehnologija u

⁶ Spolsky (2017) ju naziva ljestvicom (eng. *scale*), a u pojmovlju na američkom engleskom jeziku nazivaju se *scoring rubrics* ili *rating scales*, dok se na britanskom engleskom jeziku nazivaju *band/level descriptors* (Fulcher i Davidson, 2007). Usto, pojam *rubric* primjenjuje se češće u sjevernoameričkom kontekstu (Kuiken i Vedder, 2021). Prema Vijeću Europe, „ljestvica za procjenjivanje umijeća [...] određuje niz rastućih razina umijeća [i ona] može obuhvaćati cijeli konceptualni raspon učenikova umijeća ili može samo obuhvaćati raspon umijeća bitan za dotični sektor ili instituciju“ (Council of Europe, 2001, str. 40). Nadalje, kako navode Kuiken i Vedder (2021), opisnici (eng. *descriptors*) podrazumijevaju izjave koje određuju obilježja jezične izvedbe na određenoj razini, a ljestvica za procjenjivanje sadrži jedan ili više opisnika za svako mjerilo bodovanja (eng. *rating criterion*).

jezičnom procjenjivanju i poglavlje 2.3. – E-ispiti. No kakav god da ispit bio, Brown i Abeywickrama (2010) ističu da se **ispit** (eng. *test*) smatra dobro konstruiranim instrumentom ako omogućuje točno mjerenje ispitanikove sposobnosti u određenom području, pri čemu naglašavaju da ispit mjeri izvedbu, a da rezultati podrazumijevaju ispitanikovu sposobnost (eng. *ability*). Taj pojam **mjerenja** (eng. *measurement*) u kontekstu društvenih znanosti definira se kao „proces kvantifikacije ispitanikovih obilježja u skladu s eksplicitnim procedurama“ (Bachman, 1990: 18), a u kontekstu jezičnog procjenjivanja kao „proces kvantificiranja razine jezičnih znanja i sposobnosti učenika prema eksplicitno određenim postupcima ili pravilima, odnosno na iskazivanje jezične razine određenim numeričkim vrijednostima, primjerice brojem bodova u zadacima višestrukog izbora“ (Bagarić Medve i Škarica, 2023: 452). Potrebno je razlikovati kvantitativne od kvalitativnih opisa izvedbe ispitanika – kvantitativni uključuju dodjeljivanje određenih vrijednosti radi rangiranja promatrane izvedbe (npr. brojeve, slova, riječi i dr.), a kvalitativni uključuju pisane opise, usmene povratne obavijesti i određene vrste izvješća (Bachman, 1990; Brown i Abeywickrama, 2010). U svakom slučaju, ispitivanje mora biti ekološki osjetljivo, odnosno udovoljavati potrebama i nastavnika i učenika.

Međutim, procjenjivanje je mnogo više od samog ispitivanja, i tu dolazimo do pogleda na **procjenjivanje u širem smislu**. Zajednički europski referentni okvir za jezike, odnosno ZEROJ (Council of Europe, 2001), pod procjenjivanjem podrazumijeva procjenjivanje jezičnog umijeća⁷ (eng. *language proficiency*) korisnika jezika te napominje da su svi ispiti oblici procjenjivanja, ali da postoje i mnogi drugi oblici koji nisu ispiti, primjerice kontrolni popisi, nastavnikovo promatranje učenika i dr. Daljnji primjeri koje navodi McNamara (2000) jesu da ispitanici mogu izrađivati portfelje, sudjelovati s drugim ispitanicima u razgovorima i raspravama, može ih se promatrati tijekom komunikacijskih aktivnosti u učionici i izvan nje, mogu sami prikupljati dokaze svoje jezične izvedbe te ih se može poticati da sami procijene određene aspekte svojih jezičnih sposobnosti, no procjenu njihovih sposobnosti mogu raditi i drugi, što i dalje spada u ovaj širi smisao procjenjivanja. Naime, ispitivači se koriste podacima prikupljenima prilikom ispitivanja kako bi došli do dokaza učenja, ali ti se dokazi ne javljaju u konkretnu, već apstraktnu obliku tumačenja, i ustvari su stvar prosudbe (ibid.). McNamara i

⁷ Jezično se umijeće definira kao „mjera jezične sposobnosti koja se opisuje kroz promatranja jezične izvedbe. Smješteno je unutar konteksta primjene jezika i često se primjenjuje u svrhe donošenja odluka.“ (Harsch i Malone, 2021: 33). Drugim riječima, jezično se umijeće odnosi na „znanje, kompetencije ili sposobnosti prilikom primjene jezika, neovisno o tomu kako, kad ili u kojim je uvjetima stečeno“ (Bachman, 1990: 16), pri čemu se znanje odnosi na ono što učenik zna ili razumije u jeziku, a kompetencija na to koliko je dobra njegova izvedba, no tomu se mora dodati i njegova sposobnost točna i učinkovita sporazumijevanja s različitim sugovornicima u raznovrsnim okruženjima (Harsch i Malone, 2021).

Roever (2006) dalje ističu da se na procjenjivanje može gledati kao na proces rasuđivanja i prikupljanja dokaza radi donošenja zaključaka o pojedincima. Kroz stalan proces koji uključuje široku paletu metoda i tehnika ti se zaključci donose na temelju kvalitativnih i kvantitativnih pokazatelja jezičnih sposobnosti učenika (Brown i Abeywickrama, 2010). Međutim, ti se zaključci ne odnose samo na ispitanike jer procjenjivanje ne prati samo njihov napredak, već i kvalitetu procesa poučavanja, odnosno mjeri jesu li stvarni ishodi učenja u skladu s planiranim ishodima obrazovnih programa (Ghanbari i Nowroozi, 2021). O ovim procesima prosudaba i zaključivanja iscrpnije govorimo u poglavlju 2.4. – Valjanost.

Iz definicije procjenjivanja Jelaske i Cvikić (2008) na početku ovog potpoglavlja vidljivo je da se pojam procjenjivanja isprepliće s pojmom vrjednovanja u prvom dijelu definicije, dijelu koji je identičan obama pojmovima i koji je podcrtan u donjem navodu, no **vrjednovanje** (eng. *evaluation*) je širi, odnosno nadređeni pojam, za koji zaključuju sljedeće:

Vrjednovanje — davanje suda i o korisnikovome znanju i jezičnome umijeću, ali i o uspješnosti programa koji je završio. Ono može uključivati i samu primjerenosti [sic] načina procjenjivanja u odnosu na okolnosti ili razvijanje njegova znanja, smislenosti toga procjenjivanja i slično. (str. 124)

Bagarić Medve i Škarica (2023) dodaju da se uz vrjednovanje još primjenjuje i pojam evaluacije te da oba pojma označuju davanje suda o znanjima i sposobnostima ispitanika, no da „podrazumijevaju i praćenje razvoja i napretka učenika u učenju jezika u odnosu na ciljeve i ishode učenja određenog nastavnog programa“ (str. 451). Ono što je važno dodati gornjoj definiciji vrjednovanja Jelaske i Cvikić (2008) jest donošenje odluka – koncept koji je ključan pri definiranju vrjednovanja i za njegovo razlikovanje od pojma procjenjivanja ili čak ispitivanja. Bachman (1990) vrjednovanje definira kao „sustavno prikupljanje informacija radi donošenja odluka“ (str. 22), koje nužno ne podrazumijeva ispitivanje, već se provodi kad se rezultati ispita (ili drugih postupaka procjenjivanja) primjenjuju radi donošenja odluka, i to ne samo u nastavnom procesu nego i u širem odgojno-obrazovnom kontekstu (Bachman, 1990; Brown i Abeywickrama, 2010). Svrha je vrjednovanja u ovima dvama kontekstima dvojaka – s jedne strane, odgojno-obrazovne ustanove putem podataka dobivenih vrjednovanjem pokazuju svoju učinkovitost značajnim dionicima i mjere svoju uspješnost u marketinške svrhe, a s druge strane nastavnici pojedinačno vrjednovanjem prate i unaprjeđuju procese učenja i poučavanja (Dobson, 2013).

I ZEROJ ističe da je pojam vrjednovanja širi od pojma procjenjivanja – svako je procjenjivanje oblik vrjednovanja, a u jezičnom programu vrjednuju se mnoge druge stvari osim učenikova umijeća, npr. učinkovitost određenih metoda ili materijala, vrsta i kvaliteta

diskursa koji se proizvodi u programu, zadovoljstvo nastavnika i učenika, učinkovitost poučavanja itd. (Council of Europe, 2001). Nadalje, Pravilnik o izmjenama i dopuni Pravilnika o načinima, postupcima i elementima vrednovanja učenika u osnovnim i srednjim školama (Ministarstvo znanosti i obrazovanja, 2019) vrjednovanje definira na sljedeći način:

Vrednovanje je sustavno prikupljanje podataka u procesu učenja i postignutoj razini ostvarenosti odgojno-obrazovnih ishoda, kompetencijama, znanjima, vještinama, sposobnostima, samostalnosti i odgovornosti prema radu, u skladu s unaprijed definiranim i prihvaćenim metodama i elementima. Vrednovanje obuhvaća tri pristupa vrednovanju: vrednovanje za učenje, vrednovanje kao učenje, vrednovanje naučenog. Vrednovanje za učenje služi unapređivanju i planiranju budućega učenja i poučavanja. Vrednovanje kao učenje podrazumijeva aktivno uključivanje učenika u proces vrednovanja te razvoj učeničkoga autonomnog i samoreguliranog pristupa učenju. Vrednovanje naučenog je ocjenjivanje razine postignuća učenika. Vrednovanje za učenje i vrednovanje kao učenje ne rezultiraju ocjenom, nego kvalitativnom povratnom informacijom. (čl. 2. st. 1)

Izvorni čl. 2. st. 1. Pravilnika o načinima, postupcima i elementima vrednovanja učenika u osnovnoj i srednjoj školi (Ministarstvo znanosti i obrazovanja, 2010) nije uključivao ova tri pristupa vrjednovanju, ali je naveo da su sastavnice vrjednovanja praćenje, provjeravanje i ocjenjivanje. Izmijenjeni ih članak ne navodi kao sastavnice vrjednovanja, već ih samo definira u st. 2., 3. i 4. kako slijedi (Ministarstvo znanosti i obrazovanja, 2019):

(2) Praćenje je sustavno uočavanje i bilježenje zapažanja o postignutoj razini ostvarenosti odgojno-obrazovnih ishoda u svrhu poticanja učenja i provjere postignute razine ostvarenosti odgojno-obrazovnih ishoda i očekivanja definiranih nacionalnim, predmetnim i međupredmetnim kurikuluma, [sic] nastavnim programima te strukovnim i školskim kurikulumima. Uključuje sva tri pristupa vrednovanju: vrednovanje za učenje, vrednovanje kao učenje i vrednovanje naučenog.

(3) Provjeravanje je procjena postignute razine ostvarenosti odgojno-obrazovnih ishoda, kompetencija i očekivanja u nastavnome predmetu ili području i drugim oblicima rada u školi tijekom školske godine.

(4) Ocjenjivanje je pridavanje brojčane ili opisne vrijednosti rezultatima praćenja i provjeravanja učenikovog rada.

Prema čl. 2 Pravilnika o izmjenama i dopuni Pravilnika o načinima, postupcima i elementima vrednovanja učenika u osnovnim i srednjim školama (Ministarstvo znanosti i obrazovanja, 2019), razvidno je da je vrjednovanje definirano kao pojam širi od procjenjivanja, koji u sebi

sadržava postupke od ocjenjivanja (što se podudara s našim prethodno opisanim pojmom ispitivanja, odnosno procjenjivanja u užem smislu), preko praćenja i provjeravanja (što se podudara s našim opisom procjenjivanja u širem smislu jer podrazumijeva donošenje određenog suda odnosno zaključka), do donošenja odluka o učenju i poučavanju (što odgovara gornjoj definiciji vrjednovanja Bachmana te Browna i Abeywickrame).

Nakon što su se proučile gornje definicije i pogledi na procjenjivanje, za potrebe ovog rada odlučeno je da će se primjenjivati pojam procjenjivanja u užem smislu i u širem smislu. Uži će smisao podrazumijevati kriterijsko procjenjivanje zadatka pisane proizvodnje, a širi će se smisao odnositi na metode procjenjivanja valjanosti e-ispita, odnosno na njezinu kvalitativnu i kvantitativnu analizu.

2.2.2. Vrste procjenjivanja

U literaturi i praksi postoje različite podjele i vrste procjenjivanja, a svoj ćemo prikaz početi podjelom procjenjivanja u širem smislu na neformalno i formalno procjenjivanje, koji se još nazivaju neformalnim i formalnim vrjednovanjem⁸. **Neformalno procjenjivanje** (eng. *informal assessment*), koje se također naziva **nestandardiziranim procjenjivanjem** (eng. *non-standardized assessment*), podrazumijeva da procjenjivač ne donosi odluke o izvedbi ispitanika, a u području jezičnog procjenjivanja ono može obuhvaćati neplanirane komentare, odgovore i povratne obavijesti učenicima, primjerice davanje ohrabrujućih komentara o domaćoj zadaći, govorenje riječi ohrabrenja, pozitivno ispravljanje pogrešaka u govoru, davanje savjeta o boljem izgovoru određene riječi, o vođenju bilježaka i dr. (Brown i Abeywickrama, 2010). Ono je sastavni i stalni dio procesa poučavanja određene skupine učenika, nije određeno ni ograničeno vremenom, te kao takvo ima izrazito naglašenu pedagošku notu (Bagarić Medve i Škarica, 2023). Nasuprot tomu, **formalno procjenjivanje** (eng. *formal assessment*), koje se još naziva **standardiziranim procjenjivanjem** (eng. *standardized assessment*), „provodi [se] planski i sustavno, pomoću metrijski provjerenih instrumenata i razrađenih okvira za procjenjivanje i tumačenje rezultata [i] pruža svim sudionicima u procesu vrednovanja [...] uvid u stečena jezična znanja i sposobnosti na kraju određenih obrazovnih razdoblja“ (ibid.: 453). Kako navode Brown i Abeywickrama (2010), iako ono podrazumijeva ispite, ispiti nisu svi oblici formalnog procjenjivanja – u području jezičnog procjenjivanja formalno procjenjivanje obuhvaća i učeničke dnevnike, portfelje,

⁸ Sličnosti i razlike između pojmova procjenjivanje i vrednovanje objašnjene su u prethodnom potpoglavlju (2.2.1. – Definicija procjenjivanja).

sustavno promatranje učestalosti učenikova usmenog sudjelovanja na nastavi i sl. Dakle, autori zaključuju da je ono sustavna, planirana tehnika prikupljanja podataka kako bi nastavnik i učenik dobili procjenu učenikova postignuća.

Druga ovdje izdvojena podjela procjenjivanje dijeli na vanjsko i unutarnje. **Vanjsko procjenjivanje** (eng. *external assessment*) „oblik je procjenjivanja u kojemu ispite, zadatke i pitanja određuje posebna ustanova, i koji se zatim polažu u određenim uvjetima [...] te ih ta posebna ustanova boduje ili procjenjuje“ (Department for Education, 2015). Stoga se često vanjsko procjenjivanje povezuje s pisanim, vremenski ograničenim ispitima, ali to nije nužno uvijek slučaj (Vitello i Williamson, 2017). Ono što obilježava ovaj oblik procjenjivanja jest da se njime „stječe uvid u razinu ostvarenosti odgojno-obrazovnih ciljeva i zadataka određene obrazovne ustanove, obrazovnih ishoda određenih nastavnih predmeta na određenim razinama obrazovanja kao i uvid u čimbenike koji mogu utjecati na njihovo ostvarivanje“ (Bagarić Medve i Škarica, 2023: 453.). S druge strane, **unutarnje procjenjivanje** (eng. *internal assessment*) podrazumijeva da ispite osmišljava, provodi i procjenjuje institucija koju ispitanik pohađa, npr. škola ili radno mjesto (Vitello i Williamson, 2017). Nadalje, svrha je unutarnjeg procjenjivanja stjecanje uvida „u sam proces učenja i poučavanja u određenom nastavnom predmetu te rezultate toga procesa, tj. postignuća učenika u razvoju jezičnih znanja i sposobnosti“ (Bagarić Medve i Škarica, 2023: 453).

Treća podjela koju ovdje prikazujemo jest ona na sumativno i formativno procjenjivanje. **Sumativno procjenjivanje** (eng. *summative assessment*) ispituje učinak ishoda određenog predmeta (Dobson, 2013), odnosno mjeri postignuće kako bi se izvijestilo druge dionike ovog procesa (Woodin i Müller-Hartmann, 2013). Kako navode Bagarić Medve i Škarica (2023: 454):

Sumativno vrednovanje [...] provodi se sa svrhom procjene učeničkih postignuća na kraju određenog nastavnog ciklusa (npr. na kraju polugodišta, na kraju nastavne godine). Obilježje mu je usmjerenost na ishod učenja, a povezuje se najčešće s funkcijom davanja ocjene, selekcijskom funkcijom (npr. na prijemnim ispitima za studij) ili funkcijom usmjeravanja (npr. u vezi s daljnjim učenjem ili školovanjem).

Drugim riječima, cilj je sumativnog procjenjivanja mjeriti ili sažeti ono što je učenik naučio, i uglavnom ne upućuje na njegov budući napredak (Bachman, 1990; Brown i Abeywickrama, 2010). Sumativno se procjenjivanje osvrće unazad kako bi se vidjelo što se naučilo, pa stoga njegova obilježja ima i **procjenjivanje naučenoga** (eng. *assessment of learning*), kako navodi Green (2014), odnosno vrednovanje naučenoga (Ministarstvo znanosti i obrazovanja, 2019). Nadalje, u vezi sa sumativnim procjenjivanjem također su kontinuirano i dijagnostičko

procjenjivanje. **Kontinuirano procjenjivanje** (eng. *continuous assessment*) označava procjenjivanje izvedbe, radova i projekata tijekom obrazovnog razdoblja, provode ga nastavnici, no mogu i učenici, a završna se ocjena odnosi na cijelo razdoblje, semestar ili godinu (Council of Europe, 2001). **Dijagnostičko procjenjivanje** (eng. *diagnostic assessment*) usredotočeno je na dijagnozu ispitanikovih snažnih i slabih strana kako bi se usmjerilo i prilagodilo procese poučavanja i učenja (Xi i Sawaki, 2017), a obično se provodi prije početka tih procesa, što može biti na početku školske godine ili prije početka obrade nove nastavne cjeline, kako navode Bagarić Medve i Škarica (2023).

Nasuprot sumativnom procjenjivanju, **formativno procjenjivanje** (eng. *formative assessment*), osnažuje ili poboljšava predmet koji se procjenjuje (Dobson, 2013), odnosno daje informacije o učenju koje se mogu upotrijebiti za planiranje daljnjeg poučavanja (Bachman, 1990; Woodin i Müller-Hartmann, 2013). Bagarić Medve i Škarica (2023) ističu da se ono provodi redovito tijekom nastavnog procesa te također navode sljedeće:

Svrha mu je prikupiti informacije o napredovanju učenika u postizanju obrazovnih ishoda koje, s jedne strane, nastavniku daju uvid u postojeće stanje u vezi s razinom ovladanosti obrađenog jezičnog sadržaja, a s druge strane pomažu učeniku da promišlja o svojem napredovanju i potencijalnim problemima te načinima kako ih riješiti i unaprijediti svoje učenje [...] [F]ormativno vrednovanje oblikuje, potiče i usmjerava daljnji proces učenja i poučavanja prema krajnjem obrazovnom ishodu [te] nema funkciju davanja ocjene. (str. 454-455)

Drugim riječima, prilikom formativnog procjenjivanja nastavnik isporučuje, a učenik internalizira odgovarajuće povratne obavijesti o izvedbi radi daljnjeg učenja (Brown i Abeywickrama, 2010). Stoga, ono odmah vodi nastavnika i učenika prema onomu što će se napraviti sljedeće, pa je zato u uskoj vezi s **procjenjivanjem za učenje** (eng. *assessment for learning*), kako navodi Green (2014), odnosno vrednovanjem za učenje (Ministarstvo znanosti i obrazovanja, 2019). Nadalje, budući da se formativno procjenjivanje bavi praćenjem, razumijevanjem i unaprjeđivanjem učenja te potiče aktivno sudjelovanje učenika u procesu procjenjivanja, u uskoj je vezi i s **procjenjivanjem kao učenjem** (eng. *assessment as learning*), odnosno vrednovanjem kao učenjem (Bagarić Medve i Škarica, 2023). Također ovdje ističemo i **dinamično procjenjivanje** (eng. *dynamic assessment*), odnosno „pristupe procjenjivanju koji se temelje na sociokulturnoj teoriji u kojoj procjenjivač nastoji procesima procjenjivanja unaprijediti ispitanikovo razumijevanje ili sposobnost samostalnog funkcioniranja“ (Green, 2014: 232). Prilikom njega procjenjivači učenicima koji se uključuju u nastavnu aktivnost pružaju neki oblik potpore, koja može biti unaprijed određena, spontana,

u obliku povratnih obavijesti i dr., a sve s ciljem procjenjivanja njihove izvedbe i promicanja razvoja njihovih sposobnosti koje se procjenjuju (usp. Brown i Abeywickrama, 2010; Fulcher, 2010; Fox, 2017; Poehner i dr., 2017). U uskoj vezi s formativnim procjenjivanjem jesu samoprocjenjivanje i vršnjačko procjenjivanje. **Samoprocjenjivanje** (eng. *self-assessment*) podrazumijeva procjenjivanje prilikom kojega si učenici postavljaju ciljeve, razmišljaju o najboljim načinima učenja, isprobavaju i vrednuju nove tehnike te često razmišljaju o vlastitom napretku prema svojim ciljevima (usp. McNamara, 2000; Council of Europe, 2001; Green, 2014). Prilikom samoprocjenjivanja, koje može biti pojedinačna aktivnost, ali se može provoditi i u skupini, učenici procjenjuju svoju izvedbu u odnosu na određeno mjerilo, što znači da ono gotovo uvijek uključuje određenu kvantitativnu sastavnicu (Falchikov, 2004). Nasuprot samoprocjenjivanju, prilikom **vršnjačkog procjenjivanja** (eng. *peer assessment*) učenici, ponekad i anonimno, procjenjuju izvedbe svojih vršnjaka, a ono može uključivati davanje bodova ili ocjena, no češće podrazumijeva kvalitativne aspekte te stavlja naglasak na povratne obavijesti u odnosu na zadana mjerila, a ne na ocjenu (usp. McNamara, 2000; Council of Europe, 2001; Falchikov, 2004; Green, 2014). I samoprocjenjivanje i vršnjačko procjenjivanje mogu uključivati određene kontrolne popise ili bodovne ljestvice, a nakon što se provedu, i to uz nadgledanje nastavnika, važno je da učenici daju povratne obavijesti o provedenom procjenjivanju i obrazlože svoje odluke te da se bilo kakve moguće nesuglasice odmah riješe (Falchikov, 2004).

Četvrta podjela procjenjivanja ona je na normativno i kriterijsko. Prilikom **normativnog procjenjivanja** (eng. *norm-referenced assessment*) postignuće učenika procjenjuje se u usporedbi s postignućima drugih učenika, dok se prilikom **kriterijskog procjenjivanja** (eng. *criterion-referenced assessment*) postignuće učenika procjenjuje u odnosu na određena mjerila, odnosno kriterije procjenjivanja (usp. Brown i Abeywickrama, 2010; Hughes i Hughes, 2020; Bagarić Medve i Škarica, 2023). Što se tiče jezičnog procjenjivanja, normativno procjenjivanje „ne govori nam izravno što učenik može napraviti u dotičnom jeziku“, a prilikom kriterijskog procjenjivanja „ne saznajemo ništa o tomu kako se izvedba ispitanika uspoređuje s izvedbom drugih. Umjesto toga, saznajemo nešto o onomu što on ili ona ustvari mogu napraviti u jeziku“ (Hughes i Hughes, 2020: 20).

Peta je moguća podjela procjenjivanja na obrazovno procjenjivanje i na procjenjivanje umijeća. **Obrazovno procjenjivanje** (eng. *educational assessment*) ima kao svrhu procijeniti stupanj napretka prema cilju učenja, što je važno u školama i drugim obrazovnim sredinama, uglavnom se temelji na sadržaju koji se usvojio ili koji će se usvojiti, a često ga osmišljavaju i provode nastavnici, pa je stoga fleksibilno i uključuje tehnike promatranja, portfelja,

samoprocjenjivanja te neformalne ispite ili kvizove i formalne ispite koji se polažu u strožim uvjetima (Green, 2014). **Procjenjivanje umijeća** (eng. *proficiency assessment*) ima kao cilj procijeniti je li određena sposobnost ispitanika zadovoljavajuća u odnosu na prethodno određenu potrebu ili standard, a kad se odnosi na jezično procjenjivanje, odnosi se na ono što ispitanik može postići primjenom svojeg jezika u trenutku procjenjivanja, odnosno primjenom jezičnih vještina za obavljanje određenog posla, studiranje u akademskom okruženju i sl. (ibid.). Nadalje, budući da mu težište nije na onomu što je ispitanik naučio ili usvojio, provode ga državne ili međunarodne agencije, a uključuje formalne ispite koji se provode u kontroliranim i standardiziranim uvjetima radi, naprimjer, useljenja u određenu zemlju, zapošljavanja na određenom radnom mjestu, školovanja na određenoj ustanovi, obavljanja određenog zadatka ili zadaće poput rezervacije soba, ispunjavanja obrazaca, čitanja članka, vođenje razgovora i sl. (usp. Brown i Abeywickrama, 2010; Green, 2014; Hughes i Hughes, 2020).

Šesta je podjela procjenjivanja na prognostičko procjenjivanje i na procjenjivanje postignuća. **Prognostičkom procjenjivanju** (eng. *prognostic assessment*) svrha je usredotočiti se na planirani nastavni sadržaj i odlučiti o njegovim prioritetima, a kad se radi o jezičnom procjenjivanju, svrha mu može biti odlučiti kojima je učenicima potreban određeni jezični tečaj, koji će učenici najvjerojatnije uspjeti na određenom tečaju, koji tečaj najbolje odgovara potrebama učenika i sl. (Green, 2014). S druge strane, **procjenjivanje postignuća** (eng. *achievement assessment*) u kontekstu jezičnog procjenjivanja primjenjuje se za mjerenje trenutačnog napretka učenika, primjerice ispitima na kraju tečaja kad se odlučuje o tomu hoće li ga učenik proći ili pasti ili će mu se dodijeliti ocjena; nastavnik procjenjivanjem postignuća dobiva iscrpne dijagnostičke informacije o tomu koje su dijelove obrađenog materijala učenici usvojili, a na kojima će se morati više raditi (usp. McNamara, 2000; Brown i Abeywickrama, 2010; Green, 2014; Hughes i Hughes, 2020).

Od ostalih vrsta procjenjivanja može se izdvojiti i **procjenjivanje izvedbe** (eng. *performance assessment*), pojam koji uvodi i objašnjava McNamara (1996) i za koji navodi da mu je glavno obilježje stvarna ispitanikova izvedba određenih zadataka, a ne apstraktni prikaz znanja, i to često pisanjem odgovora na papir. Dodaje i da je procjenjivanje izvedbe u drugom jeziku složeno kognitivno postignuće u kojemu su objedinjeni mnogi psiholingvistički procesi, a usredotočuje se na kvalitetu izvedbe i na ono što izvedba otkriva o jezičnom znanju pojedinca koji se procjenjuje.

2.2.3. Razvoj jezičnog procjenjivanja

Povijesni se razvoj jezičnog procjenjivanja u užem smislu, dakle jezičnog ispitivanja, može podijeliti na četiri razdoblja opisana niže (usp. Spolsky, 1977, Grotjahn, 1986, Bagarić Medve, 2011; Kramer, 2013; Green, 2014; Bagarić Medve, 2023) iako je Spolsky (2017) kasnije odbacio ovakvu sistematičnu podjelu u nekoliko razdoblja i jednostavno naveo ključne događaje u njegovu razvoju.

1. **Tradicionalno** ili **predznanstveno** razdoblje trajalo je do sredine 20. stoljeća. U početku ovog razdoblja nisu se primjenjivali ispiti, nego se mjerenje temeljilo na subjektivnoj procjeni ispitivača, zbog čega se ispitivanje kritiziralo da nije dovoljno objektivno ni pouzdano (Grotjahn, 1986). Kako su se ispiti počeli primjenjivati, temeljili su se na mjerenju gramatičkog znanja, pisanju sastavaka te sposobnosti razumijevanja i prevođenja (Spolsky, 2017).
2. **Psihometrijsko-strukturalno**, odnosno **moderno** ili **znanstveno** razdoblje počelo je u drugoj polovici 20. stoljeća prema Grotjahnu (1986), iako Kramer (2013) njegov početak smješta već u 1920-e, a obilježava ga utjecaj bihevioralne psihologije i strukturalne lingvistike (Brown i Abeywickrama, 2010). U njemu se jezično ispitivanje institucionaliziralo kao međudisciplinarni pothvat primijenjene lingvistike i psihometrije, no psihometrija je ujedno bila zaslužna za stavljanje društvene dimenzije jezičnog procjenjivanja u drugi plan (McNamara i Roever, 2006). U ovom se razdoblju optimiziraju objektivnost i pouzdanost, koji su nužni preduvjeti valjanosti, naglašavaju se statistički argumenti, a jezično se znanje mjerilo standardnim, objektivnim ispitima, najčešće s pitanjima višestrukog izbora koja su se odnosila na gramatiku i vokabular te prijevode riječi, rečenica i kratkih odlomaka (Grotjahn, 1986; Brown i Abeywickrama, 2010; Green, 2014), i u kojima su se „autori ispita usredotočili na individualne jezične elemente poput fonoloških, gramatičkih i leksičkih, ponajviše stoga što se vjerovalo da je znanje pojedinih elemenata jednako ukupnomu znanju jezika“ (Bagarić Medve, 2011: 320). Ovo se razdoblje još naziva razdobljem industrijalizacije i centralizacije jezičnog ispitivanja velikih razmjera, koje je svoja sjedišta imalo na sveučilištima Princeton i Cambridge, a u ovom je razdoblju također započela i računalizacija procesa primjene i bodovanja jezičnog ispitivanja (Spolsky, 2017), o čemu više govori sljedeće potpoglavlje 2.2.4. – Tehnologija u jezičnom procjenjivanju.
3. **Psiholingvističko-sociolingvističko**, odnosno **integrativno-sociolingvističko** ili **poslijemoderno** razdoblje počelo je u 1970-ima. Njegov je glavni cilj osigurati valjanost

te istražiti odnos kompetencije (eng. *competence*) i izvedbe na jezičnim ispitima (Grotjahn, 1986), a zbog „međuzavisnosti komponentijalnih dijelova jezika i svakog posebnog dijela s lingvističkim i nelingvističkim kontekstom“ (Bagarić Medve, 2011: 320) u ovom se razdoblju počinju primjenjivati poglavito diktati i *cloze*-ispiti⁹ budući da su se tada smatrali najboljim primjerima integrativnih ispita (Brown i Abeywickrama, 2010; Kramer, 2013), što je u skladu sa „snažnom inačicom hipoteze jedinstvene jezične kompetencije¹⁰“ (Oller, 1980: 141), koju je Oller predstavio, no kasnije i odbacio jer je na prejednostavan način gledala na jezično umijeće. Grotjahn (1986) za ovo razdoblje dalje naglašava da ispiti ne bi više trebali mjeriti znanje jezičnih pravila, već produktivnu primjenu jezika u funkcionalnim situacijama, a jezična se sposobnost treba mjeriti integrativnim ili holističkim postupcima, primjenom statističkih analiza, poput faktorske analize ili teorije odgovora na zadatke (eng. *item response theory*), koja je iscrpnije pojašnjena niže u potpoglavlju 2.2.4.3. – Prilagodljivo ispitivanje na računalu. Na taj bi se način bolje razumjela valjanost ispita, pri čemu je ipak važno istaknuti da se oslanjanjem prvenstveno na statističke argumente neće dobiti pravi uvid u ono što jezični ispit mjeri (ibid.). Stoga se ovo razdoblje smatra i početkom dubljih istraživanja valjanosti ispita, odnosno njegova svojstva na koje su razni dionici procesa jezičnog ispitivanja gledali drugačije (Spolsky, 2017), a o kojemu iscrpnije govorimo u poglavlju 2.4. – Valjanost. Naposljetku, valja napomenuti da se „krajem 1970-ih naglasak prenio na ispitivanje jezičnoga ponašanja ispitanika u stvarnim ili približno stvarnim situacijama jezične uporabe“ (Bagarić Medve, 2011: 320), odnosno na komunikacijske vještine ispitanika i na njihovu jezičnu izvedbu u autentičnoj situaciji, pa se može reći da je došlo do kraja trećeg razdoblja u razvoju jezičnog ispitivanja i da je započelo četvrto.

4. **Komunikacijsko** razdoblje počelo je krajem dvadesetog stoljeća, kad se naglasak stavio na autentične zadatke u smislu sadržaja i svrhe te na valjanost umjesto na pouzdanost (Kramer, 2013; Green, 2014). Počeli su se primjenjivati ispiti koji su mjerili kvalitetu, a ne kvantitetu jezične izvedbe ispitanika, ispiti čija je svrha bila, među ostalim, „potaknuti izradu komunikacijskih jezičnih zadataka u kojima bi se performansa ispitanika

⁹ Kako navode Brown i Abeywickrama (2010), *cloze*-ispiti, odnosno *cloze*-zadatci podrazumijevaju tekst u kojemu su izostavljene riječi, a ispitanik mora ponuditi riječ koja odgovara praznini koja je na mjestu te izostavljene riječi. Bagarić Medve (2023) za *cloze*-zadatak, među ostalim, navodi da se u tekstu „izostavlja svaka n-ta riječ, obično svaka peta riječ, ali je moguće izostaviti i svaku šestu ili svaku sedmu riječ. Što je više riječi između praznina u tekstu, to je zadatak lakši.“ (str. 477)

¹⁰ Kako navodi McNamara (2000), izvedba ispitanika na raznim ispitima ovisi o istoj temeljnoj kompetenciji ispitanika – sposobnosti da objedini gramatičko, leksičko, kontekstualno i pragmatično znanje u izvedbu na ispitu – i naziva se hipotezom jedinstvene jezične kompetencije (eng. *unitary competence hypothesis* ili *unitary factor hypothesis*).

podudarala s jezičnom uporabom u stvarnoj komunikacijskoj situaciji“ (Bagarić Medve, 2011: 321). U skladu s ovim pogledom na jezično ispitivanje jest i prethodno spomenuto Ollerovo (1980) napuštanje njegove inačice hipoteze jedinstvene jezične kompetencije koju je zamijenio složenijom, komunikacijskom kompetencijom. Ovo je razdoblje, dakle, vrijeme u kojemu su se počeli primjenjivati autentični ispiti u kojima su se morale podudarati izvedba ispitanika i stvarna primjena jezika (Brown i Abeywickrama, 2010; Spolsky, 2017). Što se tiče zadataka na tim ispitima, Bagarić Medve (2011) kaže sljedeće:

Danas takvi zadatci najvećim dijelom odgovaraju poluotvorenim i otvorenim tipovima zadataka (npr. prošireni odgovor, intervju, rasprava, pričanje priče, vođeni sastavak, odgovor na pismo, esej i sl.) kojima se ispituje i procjenjuje usmena i pisana jezična proizvodnja ispitanika, često u kombinaciji sa slušanjem ili čitanjem. (str. 321)

Kako navodi Spolsky (2017), u ovom je razdoblju došlo do šire primjene ljestvica za procjenjivanje, od kojih je danas možda najpoznatija Zajednički europski referentni okvir za jezike (Council of Europe, 2001), a koje dovode do novih problema jer se na jezično umijeće na ovaj način gleda kao na svojstvo koje se može mjeriti na ljestvici, a ne kao na skup djelomično povezanih sposobnosti pri izvedbi raznih jezičnih funkcija, zbog čega se one trebaju prihvatiti konsenzusom umjesto da se njihova valjanost utvrdi u teoriji ili praksi. Nadalje, Green (2014) u ovo razdoblje uključuje formativno procjenjivanje, odnosno naglasak na proizvod učenja, te procjenjivanje za učenje, odnosno naglasak na proces učenja, a kao primjere navodi konferencije za nastavnike, procjenjivanje portfelja, samoprocjenjivanje, vršnjačko procjenjivanje i dr. Ono što još obilježava ovo razdoblje jest širenje sadržaja kako bi uključio sociolingvističke aspekte jezika, odnosno kako bi se jezično ispitivanje prilagodilo svojem društvenom kontekstu i etičkoj primjeni, primjerice kako bi omogućilo inteligentnoj djeci slabijeg društvenoimovinskog statusa napredak u društvu, kako se jezično ispitivanje ne bi diskriminatorno primjenjivalo pri identifikaciji imigranata ili tražitelja azila i dr. (Spolsky, 2017). Konačno, trenutačno se u procesu jezičnog ispitivanja sve više uzima u obzir višejezičnost ispitanika i položaj engleskog jezika kao *lingue francae* te se napušta izvorni govornik kao cilj kojemu se teži, pa se razmatraju izrade tomu prilagođenih ispita (ibid.).

Danas se u području jezičnog procjenjivanja mogu primijetiti tri važne promjene koje u njemu potiču sve više istraživanja (Chapelle i Voss, 2021) i koje su opisane niže.

- Uloge za primjenu jezičnog procjenjivanja sve su brojnije i raznovrsnije te stoga zahtijevaju pristupe istraživanjima valjanosti dotičnog ispita koji uzimaju u obzir njegovu namjenu, o čemu se iscrpnije govori u poglavlju 2.4. – Valjanost. Naprimjer,

procjenjivanje kojemu je namjena nastavnicima i učenicima pružiti povratne obavijesti o učenju na određenom jezičnom tečaju neće slijediti jednaka mjerila kao ono kojemu je namjena odrediti razinu ispitanikovih jezičnih sposobnosti kako bi se donijela odluka o njihovoj spremnosti za određeni posao.

- Jezično procjenjivanje u mnogim slučajevima zahtijeva primjenu tehnologije za izradu, provedbu i bodovanje ispita, ali i za objedinjavanje procjenjivanja s nastavnim materijalima dostupnima na internetu, s informatičkim sustavima za upravljanje procesima učenja i poučavanja, s alatima za prikupljanje i sintezu podataka te s programima za izvješćivanje o bodovima i profilima ispitanika. Primjena tehnologije u jezičnom procjenjivanju iscrpnije je predstavljena u sljedećem potpoglavlju 2.2.4. – Tehnologija u jezičnom procjenjivanju). Ovdje valja samo istaknuti da ispitne čestice koje računalo predstavlja i boduje imaju posljedice na značenje bodova ostvarenih na tom ispitu, a podatci o izvedbi ispitanika prikupljeni u digitalnom obliku putem računala i/ili interneta otvaraju nove mogućnosti za njezino istraživanje. Naprimjer, ako ispit od ispitanika zahtijeva da napišu tekst na računalu, ili ako trebaju izgovoriti određeni tekst u mikrofonski, gledajući u ekran, to će od njih zahtijevati da primijene sposobnosti drugačije od onih koje su potrebne da isti tekst napišu olovkom na papiru, odnosno da razgovaraju sa sugovornikom gledajući ga u oči.
- Broj i raznovrsnost sudionika jezičnog procjenjivanja raste ne samo zbog prethodno spomenutih novih tehnologija, već i zbog demografskih promjena, veće mobilnosti ispitanika, drugačijih ekonomskih sustava i političkih trendova. Naime, u mnogim zemljama povećava se stupanj odgovornosti škola u procesima jezičnog ispitivanja jer na taj način mogu, primjerice, identificirati učenike kojima je potrebna dodatna pomoć ili pak mogu izdavati potvrde koje jamče da su ispitanici uspješno završili dotične programe. Posljedično, tamo gdje su postavljeni određeni jezični uvjeti, postoje i jezični ispiti, a tamo gdje postoje jezični ispiti, postoje i ispitni centri koji ih izrađuju. Na taj se način povećava broj i spektar dionika koji su uključeni u procese izrade ispita i oglašavanje novih proizvoda za rastući broj njihovih namjena.

2.2.4. Tehnologija u jezičnom procjenjivanju

2.2.4.1. Početci primjene tehnologije u jezičnom procjenjivanju

Tehnologija se često poistovjećuje s učinkovitosti, pa se u skladu s time u primijenjenoj lingvistici tehnologija u jezičnom procjenjivanju odnosi poglavito na način na koji unaprjeđuje procese ispitivanja (Chapelle i Voss, 2017). Iako pojam tehnologija može uključivati širok raspon uređaja, poput opreme za snimanje, statističkih programa, baza podataka, programa za prepoznavanje jezika i dr., ipak je primjena računalne tehnologije za provedbu ispita i obradu ispitanikovih odgovora ona koja ima najizravniji učinak na ispitanike i obrazovne programe (ibid.). Upravo ta primjena računalne tehnologije u jezičnom procjenjivanju naziva se **računalno potpomognutim jezičnim procjenjivanjem** (eng. *computer-assisted language assessment*) ili **računalno potpomognutim jezičnim ispitivanjem** (eng. *computer-assisted language testing*), kako ih definiraju Chapelle i Voss (ibid.). Također, budući da su učenje i poučavanje jezika povezani s jezičnim ispitivanjem i procjenjivanjem, a tehnologije koje se primjenjuju u jezičnom učenju i poučavanju također su prisutne u jezičnom procjenjivanju i ispitivanju, Garret (2009) definira i pojam **računalno potpomognuto učenje jezika** (eng. *computer-assisted language learning*), koji se odnosi na dinamičan odnos u kojemu su tehnologija, teorija i pedagogija neodvojivo isprepleteni.

Iako začetci računalno potpomognutog jezičnog procjenjivanja sežu u razdoblje Drugog svjetskog rata, zanimljivo je istaknuti primjer stroja koji je IBM izumio netom prije početka rata, 1938. godine, pod nazivom IBM 85 – uređaj za bodovanje zadataka višestrukog izbora, koji bi odaslao elektronički signal kad bi prepoznao oznaku napravljenu olovkom na listu papiru koji bi se stavio u stroj, a kojih je po satu mogao bodovati između 800 i 1000, od kojih je svaki mogao sadržavati do 150 ispitnih čestica (Fulcher, 2010). Također je važno napomenuti da se prije 1940-ih godina jezično procjenjivanje prvenstveno usredotočivalo na procjenjivanje umijeća čitanja, i to u učionici, dok su Drugi svjetski rat i uključenost SAD-a u nj pomaknuli težište na slušanje i govorenje, što je 1949. godine dovelo do prvog standardiziranog ispita jezičnog umijeća čitanja, slušanja, pisanja i gramatike na 25 jezika pod nazivom Army Language Tests (Brooks, 2017). Ministarstvo obrane SAD-a, uvidjevši moguće koristi računala u ispitivanju, u 60-ima i 70-ima godinama prošlog stoljeća počelo je znatno ulagati u teorijska istraživanja u ovom području; međutim, u to su vrijeme s praktične strane troškovi bili značajni zbog skupih i velikih računala tako da su se tek u 1980-ima računala počela intenzivnije primjenjivati u praksi, i to u području poučavanja stranim jezicima na

jezičnim institutima i sveučilištima prvenstveno u SAD-u, UK-u i Nizozemskoj (Dunkel, 1999).

2.2.4.2. *Ispiti na računalima*

Ispiti koji su se počeli provoditi na računalnim terminalima ili na osobnim računalima nazvali su se **ispitima na računalima** (eng. *computerized tests* ili *computer-based tests*) (usp. Dunkel, 1999; Roever, 2001). Njihova je svrha bila učiniti procjenjivanje učinkovitijim, što je ujedno i njihova najveća prednost u odnosu na tradicionalne ispite na papiru; od ostalih **prednosti** valja izdvojiti da se ispitima na računalima omogućila primjena ispita bilo kad i bilo gdje na svijetu, prilagodljivija i individualiziranija primjena ispita, nove vrste zadataka, automatsko ispravljanje zadataka, krojenje težine zadataka prema ispitanikovim rezultatima, bolje praćenje rezultata ispitanika, istovremeno slanje povratnih obavijesti te povećana razina sigurnosti ispita (usp. Gruba i Corbel, 1997; Chalhoub-Deville, 2001; Roever, 2001; Chapelle i Douglas, 2006). Ovim prednostima valja dodati i mogućnost primjene tehnologije radi osmišljavanja načina na koji će još više moći ispitati i još bolje izmjeriti ključne aspekte konstrukta¹¹, a ona će im također pomoći da bolje razumiju kako se ti aspekti konstrukta razvijaju kroz različite razine ispitanikovih jezičnih sposobnosti (Chalhoub-Deville, 2001). Sljedeća je prednost primjena metode praćenja pokreta očiju (eng. *eye-tracking method*)¹² koja se pokazala učinkovitom pri istraživanju međudjelovanja ispitanika i ispitnih zadataka (Chapelle i Voss, 2017). Kao prednost također se može navesti primjena videokamera zahvaljujući kojima se i zadatci govorne proizvodnje mogu uključiti u ispite na računalima, a Massachusetts Institute of Technology razvio je program koji u simuliranu razgovoru, uz prepoznavanje govora i prozodije, analizira izraze lica, kimanje glavom i osmijehe, a

¹¹ Apstraktne imenice koje se ne mogu promatrati, ali koje su ljudima važne (npr. ljubav, dobrota, zlo i sl.) intuitivno razumijemo kao koncepte, a jednak je slučaj i u području jezičnog procjenjivanja (npr. tečnost, razumljivost, točnost i sl.) – ovi pojmovi prestaju biti koncepti i postaju konstrukti kad ih se definira na način koji omogućuje njihovo mjerenje i njihove odnose s drugim konstruktima koji su od njih različiti (Fulcher i Davidson, 2007; Fulcher, 2010; Green i Fulcher, 2021). Konstrukt se stoga definira kao „[...] smisljeno tumačenje ili objašnjenje dosljednosti izvedbe. Konstrukti mogu biti određeni u smislu stvarnih pojava ili struktura koje su osmislili autori ispita ili kao kombinacija toga dvojega“ (Chapelle, 2021: 14). Konstrukt se odnosi na aspekte znanja ili vještina koji se mjere, a kako bi se konstrukt odredio, treba biti jasno od čega se sastoji znanje o jeziku i kako se tim znanjem koristi u primjeni jezika, odnosno jezičnoj izvedbi (McNamara, 2000).

¹² Začetci ove metode sežu još u 1879. godinu, no s prvim se istraživanjima počinje na početku 20. stoljeća, dok u posljednjoj trećini 20. stoljeća nastupa novi val istraživanja zahvaljujući razvoju novih metodoloških pristupa i kognitivnih teorija obrade; danas se istraživanja koja se temelje na ovoj metodi primjenjuju u fonološkoj, ortografskoj, morfološkoj i sintaktičkoj obradi, obradi slogova i leksičkoj dvosmislenosti (Matić, 2017).

trenutačno se razvija i tehnologija prepoznavanja ljudskih emocija i raspoloženja, no ona još nije integrirana u jezično ispitivanje (ibid.). Još je jedna prednost ispita na računalima ta da, ako su dobro osmišljeni, neće samo razvrstati ispitanike prema njihovim rezultatima, već će stvoriti odnos između različitih obilježja ispitnih zadataka (jezičnih, kognitivnih, situacijskih i dr.), izvedbe ispitanika i zaključaka o ispitanikovim jezičnim sposobnostima budući da će ovakav pristup dovesti do bogatijeg i značajnijeg prikaza ispitanikovih jezičnih sposobnosti (Chalhoub-Deville, 2001).

Kao što smo spomenuli na kraju prethodnog odlomka, ispiti na računalima i njihova provedba moraju biti dobro osmišljeni kako bi imali navedene prednosti u odnosu na ispite na papiru i provedbu na taj tradicionalan način. Međutim, imaju i određene **nedostatke**, poput ograničenja primjene tipova zadataka koji se mogu bodovati automatski (Chapelle i Douglas, 2006), različite razine ovladanosti računalnim vještinama ispitanika, visokih troškova osnivanja i održavanja računalnih centara te mogućnosti tehničkih problema (Roever, 2001). Nadalje, Chapelle i Douglas (2006) pitaju se koliko primjena multimedijских zadataka ustvari smanjuje valjanost tumačenja i primjena rezultata procjenjivanja jer će gledanje videa možda utjecati na izvedbu ispitanika na procjenjivanju njegove razine vještine slušanja, dok Chapelle i Voss (2017) navode da su zahvaljujući tehnologiji ispitanicama u videima dostupni i vizualni znakovi, koji su bitni za tumačenje značenja govorne poruke, što opet postavlja nova pitanja koja valja istražiti, a tiču se zadataka koji u sebi objedinjuju više vještina (Kramer, 2013).

Još jedan od mogućih nedostataka ispita na računalima može se javiti u zadacima pisane proizvodnje, primjerice u slučajevima kad ispitivač i računalo možda neće na jednak način tumačiti bodove na zadatku pisanja (Chapelle i Douglas, 2006). Usto, iako je učinjen značajan pomak u istraživanju obrade prirodnog jezika (eng. *natural language processing*)¹³, i to više u pisanima nego govorenima odgovorima ispitanika (Chapelle i Voss, 2017), strojna je obrada prirodnog jezika vrlo složen problem, kako navodi Glavaš (2011), jer „prirodni je jezik inherentno nepogodan za automatsku (strojnu) analizu zbog svoje kompleksnosti, nepreciznosti i velikog broja prisutnih višeznačnosti“ te su potrebna daljnja istraživanja, od kojih su mnoga i napravljena u proteklom desetljeću, ali ovo je područje izvan opsega našeg rada.

¹³ Kako definira Šuman (2021), obrada prirodnog jezika „dio je područja umjetne inteligencije koje je povezano s lingvistikom [i koje] istražuje načine uporabe računala za obradu ili razumijevanje ljudskih – prirodnih jezika. Upotrebljava se za pretvaranje ili prevođenje podataka s prirodnog jezika na računalo razumljiv jezik – strojno razumljiv format“ (str. 373). Dalje navodi da su uobičajene primjene obrade prirodnog jezika raspoznavanje govora, leksička i semantička analiza jezika, strojno prevođenje, automatsko sažimanje, analiza mišljenja, dohvaćanje informacija, odgovaranje na pitanja i dr.

Naposljetku, Sawaki (2001) u svojem preglednom članku o usporedbi ispita čitanja na drugom jeziku na računalu i na papiru zaključuje da je teško doći do općenitih zaključaka zbog raznolikih obilježja ispitanika, ispitnih zadataka, uvjeta provedbe ispita, računalnih vještina i drugih čimbenika. Međutim, zaključuje da način primjene ispita može imati učinak na promjenu vrste zadatka, na odluku koja se donosi na temelju rezultata na ispitu, na vrijeme koje je potrebno za rješavanje ispita te na osjećaje ispitanika, a ne nužno na njihove rezultate. K tomu valja dodati i napomenu Chapelle i Vossa (2017) da mladi učenici danas odrastaju uz računalnu tehnologiju te da se računalna i jezična pismenost razvijaju usporedno zahvaljujući elektronskim uređajima koji su u svakodnevnoj primjeni; stoga neki ispitni centri čak nude kandidatima mogućnost odabira načina polaganja ispita – na računalu ili na papiru.

2.2.4.3. Prilagodljivo ispitivanje na računalima

Način ispitivanja koji se prvi počeo primjenjivati na računalima u 1960-ima i 1970-ima bio je **prilagodljivo ispitivanje na računalima** (eng. *computer-adaptive testing*), što znači da se ispit svakom ispitaniku pojedinačno i interaktivno kroji tijekom samog polaganja (usp. Dunkel, 1999; Fulcher, 2010; Chapelle i Voss, 2016, 2017), a „mogućnosti obrade koje imaju računala omogućuju razvoj ispita koji nadilaze granice konvencionalnih oblika“ (Gruba i Corbel, 1997: 141). Primjena prilagodljivih ispita na računalima u području procjenjivanja stranih jezika započela je u 1980-ima ispitom S-CAPE, na kojemu su ispitanici odabirali točan odgovor među ponuđenima radi razmještaja na prikladan tečaj španjolskog jezika Sveučilišta Brigham Young, nakon čega se u sklopu povezanog projekta razvio i ispit Computest ESL za ispitivanje vještina čitanja i slušanja na visokoškolskim ustanovama, a potom su uslijedili i drugi ispiti, kao što su ToPE (Test of English Proficiency), LUCAS (Lancaster University Computer-Based Assessment Systems), E-CAPE (nasljednik S-CAPEA, samo ovaj put za ispitivanje engleskog jezika), Enhanced TOEFL i dr. (ibid.: 142-145).

U 1980-ima mikroročunala su počela biti dostupna mnogim primijenjenim lingvistima koji su u svojim istraživanjima proučavali učinkovitost ovog načina ispitivanja (Chapelle i Voss, 2017), koji se isprva odnosio na određenu vrstu algoritma koji se oslanja na psihometrijsku metodu nazvanu **teorija odgovora na zadatke** radi kontrole prilagodljivosti koja se temelji na izvedbi ispitanika na svakom zadatku na ispitu (Chapelle i Voss, 2016). Kako objašnjava Dunkel (1999), u ovom načinu ispitivanja računalo odabire ispitne zadatke prema procijenjenoj razini ispitanikovih jezičnih vještina – prvi je zadatak srednje težine, a ako ispitanik na nj odgovori točno, sljedeći je teži, no ako odgovori netočno, sljedeći je lakši i tako

dalje. Ona također naglašava da se ovaj način ispitivanja razvio upravo kako bi uklonio dugotrajne i neučinkovite ispite koji ispitanicima s visokom razinom jezičnih sposobnosti postavljaju prelagana pitanja, a preteška onima niže razine jezičnih sposobnosti. Međutim, iako se ovim načinom ispitivanja gotovo u potpunosti otklanja mogućnost da dvoje ispitanika dobije ispit s jednakim ispitnim česticama, to se može ostvariti samo pod uvjetom da je baza ispitnih čestica dovoljno velika, što predstavlja najveći izazov većini nastavnika i institucija (Fulcher, 2010).

Što se tiče **prednosti** prilagodljivog ispitivanja na računalu koje se temelji na teoriji odgovora na zadatke u odnosu na ispitivanje na papiru, Chapelle i Voss (2016) navode tri najznačajnije, a to su kraće vrijeme ispitivanja, manji broj ispitnih zadataka te točnija procjena pouzdanosti ispita. Dunkel (1999) navodi još neke prednosti, kao naprimjer da svaki ispitanik može rješavati ispit svojim ritmom, da se svakom ispitaniku daju zadatci primjereni njegovoj razini jezičnih sposobnosti, da ispitanici mogu dobiti istovremene povratne obavijesti o točnosti svojih odgovora, da multimedijske sastavnice mogu prikazati autentične situacije primjene jezika te da je sigurnost ispita veća jer se određeni zadatci uzimaju iz bogate baze zadataka, pa nije moguće zapamtiti sve zadatke koji se u njoj nalaze. Chapelle i Voss (2016) dodaju da se vrijeme koje ispitanici provedu na svakom zadatku bilježi i služi kao pokazatelj automatizma ili tečnosti ispitanikove učinkovitosti, što znači da se vrijeme ispitivanja primjenjuje učinkovito jer ispitanici ne moraju odgovarati na za njih prelagana niti preteška pitanja. Također ističu da je na ovaj način sigurnost ispitivanja pojačana jer svaki ispitanik polaže drugačiju inačicu ispita, čime nestaju mogućnosti tradicionalnih načina varanja na ispitima. Kao još jednu prednost valja dodati i mogućnost dinamičnog procjenjivanja na računalu, prema zaključcima Teo (2012), koja je istraživala usvajanje metakognitivnih strategija čitanja putem računala. U njemu je objedinila davanje uputa i procjenjivanje tako da je program kojime su se ispitanici koristili stvorio prilike za međudjelovanje i najpovoljniju količinu povratnih obavijesti kao konstruktivna posredovanja u formativnom vrjednovanju koje je služilo kao potpora usvajanju vještina razumijevanja čitanjem.

Prilagodljivo ispitivanje na računalima ima i svoje **nedostatke** zato što se zadatci trebaju pilotirati na veliku broju ispitanika kako bi se osiguralo da su stabilni i da mjere ono što se namjerava mjeriti te zato što se baze zadataka za ispite visokog rizika trebaju redovito osvježivati kako bi se osigurala sigurnost ispita (Chapelle i Douglas, 2006). Međutim, kao što navodi Chalhoub-Deville (2001), tehnologija ne bi trebala samo olakšati provedbu ispita, već potaknuti autore ispita na načine razmišljanja drugačije od onih prilikom izrade ispita na papiru, što predstavlja određeni novitet u ustaljenom načinu gledanja na procjenjivanje. Naime,

poenta nije samo uspoređivati ispite na računalima i na papiru kako bi se uvidjelo koji je učinkovitiji, već „potaknuti te 'remetilačke' načine primjene tehnologije, koji procjenjivanje konceptualiziraju i primjenjuju na raznolike inovativne načine“ (ibid.: 97). Jedna je od tih inovacija prethodno opisano prilagodljivo ispitivanje na računalima, a druga automatsko vrjednovanje pisanje proizvodnje, koje je predstavljeno u narednom potpoglavlju.

2.2.4.4. Automatsko procjenjivanje pisane proizvodnje

Istovremeno s prilagodljivim ispitivanjem na računalima počela je i šira primjena **automatskog procjenjivanja pisane proizvodnje** (eng. *automated writing evaluation*) radi daljnjeg poboljšanja učinkovitosti u jezičnom poučavanju i ispitivanju putem programa za automatsko i visoko kvalitetno procjenjivanje eseja (Chapelle i Voss, 2016). Razvoj ovih sustava koji automatski boduju i šalju povratne obavijesti radi procjenjivanja i učenja jezika iznimno je složen pothvat koji zahtijeva veliku razinu stručnosti u raznim disciplinama, od kojih su neke usvajanje prvog ili drugog jezika, jezično procjenjivanje, računalna lingvistika (Xi, 2010), pa i umjetna inteligencija. Svakodnevni napretci u obradi prirodnog jezika te tehnologijama prepoznavanja i obrade govora znače da automatsko procjenjivanje nije više ograničeno na bodovanje zadataka u kojima ispitanici samo odabiru točan odgovor (Kramer, 2013).

Što se tiče prednosti sustava za automatsko procjenjivanje pisane proizvodnje, najveća je svakako rasterećenje nastavnika koji na ovaj način imaju više vremena za druge aktivnosti koje bi inače morali utrošiti na ocjenjivanje (Fulcher, 2010). Usto, također valja istaknuti pružanje automatskih povratnih obavijesti ispitanicima o greškama u njihovim esejima (eng. *error feedback*) i o savjetima vezanima za primjenu diskursnih oznaka radi bolje organizacije i razrade ideja (eng. *discourse feedback*), kako ističu Liu i Kunnan (2016). Međutim, ova je inovacija imala i svojih nedostataka. Kao što napominje Godwin-Jones (2008), mnogi primijenjeni programi nisu bili dovoljno sofisticirani jer su procjenjivali samo neka površna obilježja pisane proizvodnje, poput pravopisa te primjene vokabulara i gramatike na razini rečenice, ne uzimajući u obzir ostale sastavnice, pa stoga ni povratne obavijesti ispitanicima nisu bile od velike koristi, a najmanje neizvornim govornicima engleskog jezika. Stoga nastavnici u nekim slučajevima trebaju dati i svoje povratne obavijesti o greškama ispitanika koje često krivo uoče ili uopće ne (Liu i Kunnan, 2016).

Mnoga su istraživanja sustava automatskog procjenjivanja pisane proizvodnje provedena kako bi se analizirala korelacija između ljudskih procjenjivača i automatskog

bodovanja te kako bi se međusobno usporedilo procjenjivanje u ovim sustavima (usp. Cotos, 2013; Hoang i Kunnan, 2016; Liu i Kunnan, 2016), od kojih niže predstavljamo tri najznačajnija, ali kojima se mogu dodati i sveprisutni Googleovi dokumenti, Microsoftov Word te Grammarly i njegove alternative kao što su ProWritingAid, Readable, Wordtune, Ginger, Whitesmoke i dr.

Prvi sustav automatskog procjenjivanja pisane proizvodnje bio je IntelliMetric[®], koji je na tržište stavio Vantage Learning 1997. godine, te koji je bio prvi robot za bodovanje eseja koji je pokretala umjetna inteligencija i koji je dosegao razinu ljudskih procjenjivača te do danas bodovao preko milijardu eseja, a primjenjuje se, među ostalim, i pri bodovanju eseja na računalu na dijelu analitičke pisane proizvodnje ispita GMAT – Graduate Management Admission Test (Vantage Labs, 2023). Usporedno s njim, Vantage Learning osmislio je i MyAccess![®] – jedan od prvih prilagodljivih formativnih alata za pisanje u oblaku, u kojemu tzv. Intelligent Agent[™] šalje povratne obavijesti radi unaprjeđenja kvalitete pisanja ispitanika, odnosno koji se koristi IntelliMetricovim[®] sustavom bodovanja (Vantage Learning, 2016). Chen i Cheng (2008) analizirali su povratne obavijesti koje je MyAccess![®] slao ispitanicima i utvrdili su da su ispitanici najviše voljeli kombinaciju automatskog bodovanja koncepta u prvim koracima, a ljudske povratne obavijesti u kasnijim koracima pisane proizvodnje, što je pokazalo da bi ova kombinacija mogla biti najbolja potpora vrjednovanju učenja i vrjednovanju za učenje, koji su predstavljeni u potpoglavlju 2.2.1. – Definicija procjenjivanja. Naglašavaju da iako tehnologija automatskog procjenjivanja pisane proizvodnje ima mnoga ograničenja, treba se raditi na razvoju učinkovitijih pedagoških praksa koje će povećati koristi i smanjiti probleme koje ta tehnologija donosi učenicima pri usvajanju vještina pisanja.

Drugi poznati sustav automatskog procjenjivanja pisane proizvodnje razvio je ETS (Educational Testing Service) pod nazivom Criterion[®], a koji pokreće e-rater[®]. Kad ispitanik pročita zadatak, odnosno temu, i napiše esej u Criterionu[®], dobit će rezultat koji odgovara procjeni ljudskog procjenjivača na temelju mjerila bodovanja za dotičnu temu, kao i iscrpne povratne obavijesti o mogućim greškama; ovaj se sustav temelji na obradi prirodnog jezika kako bi se iz eseja izvukla određena obilježja i kako bi se statistički predvidjelo na koji bi način ljudski procjenjivač holistički bodovao taj esej (ETS, 2023). E-rater[®] se, među ostalim, primjenjuje na inačici ispita TOEFL koja se polaže putem interneta u zadatku samostalnog pisanja eseja (ibid.). Lim i Kahng (2012) analizirali su program Criterion[®], koji je mogao utvrditi greške u pisanom tekstu, a koji je imao i mogućnost vršnjačkog procjenjivanja. Također je ispitanicima mogao slati povratne obavijesti i dodatne zadatke za vježbu, a

nastavnicima izrađivati izvješća o ispitanikovu radu, što pokazuje kako i razvoj računalnih programa može staviti težište na učenje, a ne samo ispitivanje.

Treći sustav internetskog alata za automatsko bodovanje kvalitete eseja koji se šalje elektroničkim putem jest Intelligent Essay Assessor™, koji je na tržište stavio Pearson, te koji automatski procjenjuje značenje teksta, a ne samo gramatiku, stil i sl.; pokreće ga Knowledge Analysis Technologies™, koji se temelji na latentnoj semantičkoj analizi i metodama rasprostranjenima u područjima automatskog prepoznavanja govora, računalne lingvistike i drugih oblika statističke umjetne inteligencije (Pearson, 2023). Drugi, aktualniji Pearsonov sustav jest WriteToLearn®, koji sadrži knjižnicu od trenutačno preko 650 zadataka/tema raznih žanrova i stilova, a nudi i mogućnost unosa vlastitog zadatka; eseji se boduju analitički i holistički, a greške se ne označuju u eseju, već se ispitanicima ukratko objasni koji su problem imali i nude im se prijedlozi za ispravak grešaka (Liu i Kunnan, 2016).

Prema Fulcheru (2010), utvrđivanju valjanosti sustava automatskog bodovanja (eng. *automatic scoring*) prvenstveno se pristupalo korelacijom bodova koje je dodjeljivalo računalo i onih koje je dodjeljivao čovjek, pri čemu su se bodovi potonjega tretirali kao „zlatni standard“, a pretpostavljalo se da „ako su bodovi računala i čovjeka na visokoj razini korelacije, ispit mjeri sličan konstrukt“ (str. 216). Proces utvrđivanja valjanosti iscrpnije je predstavljan u potpoglavlju 2.4.2. – Utvrđivanje valjanosti. Fulcher dalje skreće pozornost na dvije stvari: prvo, potrebno je još dokaza kako bi se pokazalo da je uzrok korelacije osjetljivost na jednak konstrukt, što ima učinak na varijacije obiju mjera; drugo, upitno je može li računalo u proces bodovanja unijeti jednako razumijevanje teksta i nijansa u značenju kao što to može čovjek. Zbog tih dvaju razloga, ali i zbog činjenice da mnoga računala ne pružaju ni približno slične povratne obavijesti učenicima kao nastavnici, iako se učenici vole koristiti automatskim bodovanjem i možda od njega imaju određene koristi zbog toga što predstavlja određeni novitet, mnogi su nastavnici i dalje skeptični u vezi s vrijednosti značenja bodova na ispitima na kojima izvedbu ispitanika boduju računala (ibid.). Ipak, računalne tehnologije svakodnevno napreduju i sveprisutne su u jezičnom učenju, poučavanju, procjenjivanju i vrjednovanju, koje obogaćuju mnogobrojnim izvorima i čiju učinkovitost itekako unaprjeđuju, uz uvjet da se mogućnostima računala koristi primjereno i odgovorno (Xi, 2010).

2.3. E-ISPITI

2.3.1. Definicija, prednosti i nedostaci e-ispita

Do devedesetih godina prošlog stoljeća primijenjeni su lingvisti iskazivali svoju zabrinutost u vezi s provedbom ispita visokog rizika na računalima jer su se bojali da će negativne posljedice prevagnuti u odnosu na prednosti ovog načina ispitivanja, ali „tehnologija koja je utjecala na jezično procjenjivanje nije čekala na odobrenje i podršku primijenjenih lingvista. Do sredine 1990-ih počeli su se razvijati mnogi programi za ispitivanje i računalno potpomognuti jezični ispiti“ (Chapelle i Voss, 2017: 152). S pojavom interneta uvelike je porastao potencijal inovacija za primjenu u području jezičnog ispitivanja i procjenjivanja jer je internet postao dostupniji autorima ispita, nastavnicima i učenicima. Te inovacije podrazumijevaju način razmišljanja o jezičnom procjenjivanju – autori ispita, nastavnici i učenici tehnologiju vide kao sredstvo za poboljšanje metoda i povećanje mogućih primjena jezičnog procjenjivanja (Chapelle i Douglas, 2006). Prema Chapelle i Vossu (2016), uspješna inovacija također može dovesti do novih spoznaja o međudjelovanju tehnologije i procjenjivanja, a primjere toga naveli smo u potpoglavlju 2.2.4.4. – Automatsko vrjednovanje pisane proizvodnje. Oni dalje naglašavaju da inovacije u jezičnom procjenjivanju trebaju ispitanicima dati priliku učiti ne samo iz rezultata na ispitu, već i iz samog postupka polaganja ispita.

Kako se širila primjena interneta na početku 21. stoljeća, raslo je zanimanje za **e-ispitivanje** (eng. *web-based testing*) i **e-ispite** (eng. *web-based tests* ili *Internet-based tests*), odnosno jezične ispite na računalu koji se isporučuju putem interneta (Roever, 2001; García Laborda, 2007)¹⁴. Drugim riječima, e-ispiti od ispita na računalu razlikuju se jedino po tomu što se koriste internetom kao sredstvom isporučivanja. Roever (2001) razlikuje dvije vrste e-ispita – napredne, odnosno *high-tech*, i jednostavne, odnosno *low-tech*. Prvi se koriste serverom koji odabire ispitne zadatke putem prilagodljivih algoritama ili programom baze podataka koji prikuplja i analizira odgovore ispitanika, što iziskuje velike troškove i zahtijeva visoku razinu računalnih vještina autora ispita ili angažiranje računalnih stručnjaka, odnosno programera za izradu i održavanje ispita. Drugi su lakši za izradu jer ne zahtijevaju visoku

¹⁴ U svojem smo se radu opredijelili za naziv sa sastavnicom *e-*, koja znači *elektronički* iako se kao istovrijednica javlja i naziv u kojem je prva sastavnica riječ *online*. Naime, Halonja i Mihaljević (2012) navode da se razlika između tih dviju sastavnica gubi kad su u dvorječnoj svezi, a strani se pridjev *online* treba isključiti jer se ne uklapa u hrvatski jezični sustav.

razinu računalnih vještina autora ispita niti angažiranje računalnih stručnjaka, odnosno programera, te nisu skupi za izradu i provedbu.

U skladu s gore navedenom definicijom e-ispita, jasno je da s tradicionalnim ispitima na računalu dijele mnoge njihove prednosti i nedostatke, od kojih smo neke već spomenuli u potpoglavlju 2.2.4.2. – Ispiti na računalima. Dodatne **prednosti** e-ispita koje ćemo ovdje istaknuti jesu ušteda vremena i novca zbog brzine ispravljanja i jednostavnosti izrade zadataka, veća učinkovitost ispitivanja, veća točnost ispitivanja, veća sigurnost ispitivanja, više mogućnosti te brže i iscrpnije davanje povratnih obavijesti i izvješća o izvedbi i rezultatima ispitanika, bolja prilagodba ispita ispitanicima, veća dostupnost ispita, polaganje ispita kad to ispitanicima odgovara, mogućnost učenja i procjenjivanja na daljinu, više prilika za učenje te primjena multimedijских sastavnica koje doprinose autentičnosti procjenjivanja (usp. Dunkel, 1999; Roever, 2001; Chapelle, Jamieson i Hegelheimer, 2003; García Laborda, 2007; Lim i Kahng, 2012; Lesiak-Bielawska, 2015; Chapelle i Voss, 2016; Isbell i Kremmel, 2021). Još je jedna njihova prednost da se, uz primjenu u redovnoj nastavi, mogu primijeniti i u nastavi na daljinu iz krajnje nužde (eng. *emergency remote teaching*), kao što to može biti slučaj uslijed pandemija (Lučev i dr. 2022). Chapelle i Voss (2017) kao dodatnu prednost navode da neki e-ispiti omogućuju ispitanicima služiti se pomoćnim alatima prilikom polaganja, poput pristupa e-rječniku u zadacima razumijevanja čitanjem, što jest prednost e-ispita, ali se u tom slučaju mora pažljivo odrediti konstrukt dotičnog ispita jer taj konstrukt tada mora uključivati određene strateške sposobnosti za uspješno razumijevanje čitanjem u raznim kontekstima (konstrukt je iscrpnije opisan u bilješci 11).

S druge strane, neki od njihovih **nedostataka** jesu ugrožena sigurnost ispita (varanje, plagiranje te osjetljivost osobnih podataka), pojava tehničkih problema, obilježja i oblik e-ispita koji su drugačiji od onih na papiru, ograničene računalne vještine ispitanika, odnosno ispitivača, nedostatak uživanja ispitanika prilikom polaganja ispita, nedostatno tumačenje

rezultata na ispitu ispitanicima, negativan povratni učinak¹⁵ na ispitanike/učenike i na ispitivače/nastavnike te izostanak institucijske potpore i usavršavanja ispitivača/nastavnika i ispitanika/učenika (usp. Dunkel, 1999; Roever, 2001; García Laborda, 2007; Isbell i Kremmel, 2021; Li i sur., 2021; Turnbull i sur., 2021). Ovdje valja napomenuti, kao što naglašavaju Chapelle i Voss (2016), da iako će se danas rijetko naći čitatelj kojemu je strano čitati s ekrana i da će ispitanicima možda biti stranije čitati s papira nego s ekrana, ne smije se zanemariti činjenica da mnogi nastavnici još uvijek osjećaju anksioznost tijekom priprema učenika za e-ispite visokog rizika jer smatraju da će njihovi učenici biti u nepovoljnijem položaju budući da neće moći pokazati sve svoje jezične sposobnosti koje bi mogli na ispitu na papiru (Chapelle i Douglas, 2006). Još jedan mogući nedostatak „može biti nedovoljna pripremljenost ili obučenost nastavnika za nove metode koje podrazumijevaju elektronički mediji, ili nespremnost studenata da u njima sudjeluju“ (Lučev i dr., 2022).

Chapelle i Voss (2017) dodaju kako nedostatak primjerene sigurnosti predstavlja problem pri ispitivanju putem mobilnih platforma, ali budući da mobilni uređaji s multimedijским mogućnostima i pristupom internetu postaju uobičajeni, radi se na razvoju jeftinih jezičnih ispita visokog rizika velikih razmjera, no ti su uređaji trenutačno prilagođeniji zadacima govorenja i slušanja zbog malenih ekrana i tipkovnice. Nadalje ističu da programi za jezično ispitivanje moraju redovito ažurirati softver, hardver i tehničko znanje svojih zaposlenika, što je lakše većim ispitnim centrima, od kojih neki taj trošak prebacuju na krajnje korisnike ispita – same ispitanike – dok se manji centri i ispitne kompanije, izdavači i škole moraju oslanjati na strateška partnerstva radi povezivanja iskustva, ograničenih resursa i tehnologija.

¹⁵ Povratni učinak (eng. *washback* ili *backwash*) prema Bachmanu i Palmeru (1996) definira se kao izravan učinak ispitivanja na pojedince (ispitanike, korisnike ispita i donositelje odluka, među kojima su i nastavnici, ali i šire – razredne i poslovne kolege ispitanika, buduće poslodavce ispitanika i dr.) te na obrazovni sustav i društvo u cjelini, posebice kod ispita visokog rizika. Hughes i Hughes (2020) ovaj učinak na učenje i poučavanje dijele na štetan i pozitivan – štetan je, primjerice, ako se ispitanici na pripremnom tečaju za prijem na sveučilište pripremaju za ispit koji će ispitivati pisanu proizvodnju, a sam ispit ne ispituje tu sposobnost izravno, nego, naprimjer, pitanjima višestrukog izbora; pozitivan je, primjerice, ako se ispit izrađuje u skladu s analizom potreba ispitanika na prvoj godini fakulteta i ako uključuje zadatke slične onima koji će im biti zadani na samoj nastavi te se u skladu s tim izrađuje nastavni silab, odabiru odgovarajući udžbenici, prilagođava se nastava i dr. McNamara i Roever (2006) ističu razliku između povratnog učinka, koji vide kao učinak ispita na poučavanje i učenje koje prethodi ispitivanju, i utjecajnosti, pod kojom podrazumijevaju učinke izvan neposredne situacije učenja jezika (v. potpoglavlje 2.3.4. – Svojstva e-ispita), no napominju da oboje može uključivati planirane i neplanirane posljedice, od kojih potonje mogu biti pozitivne ili negativne. Naposljetku, valja navesti hipotezu povratnog učinka (eng. *the washback hypothesis*), koja pretpostavlja da nastavnici i učenici zbog ispita rade ono što inače ne bi – ona navodi da ispit utječe na poučavanje i učenje (odnosno na to što, kako, kojom brzinom i redosljedom, na kojem stupnju i do koje dubine nastavnici poučavaju, a učenici uče), na stavove o sadržaju, metodi itd. poučavanja i učenja, da ispit koji ima važne posljedice ima povratni učinak i obratno, te da ispit ima povratni učinak na neke učenike i nastavnike, ali ne na druge (Alderson i Wall, 1993).

Također se postavlja pitanje jesu li e-ispiti primjereni za sve vrste jezičnih ispita. Roever (2001) smatra da su ispiti niskog rizika najprimjereniji za e-ispite jer je svrha ispita niskog rizika dati povratne obavijesti ispitanicima o postignutim rezultatima, pa ispitanici na njima nemaju potrebu varati. Na ispitima srednjeg rizika, poput prijemnih, semestralnih ili završnih ispita, ispitanici imaju poticaj varati, pa se za njih e-ispiti ne bi smjeli primjenjivati osim ako ih se ne nadgleda na primjeren način. Primjena e-ispita visokog rizika, poput nekih prijemnih ispita, ispita za određene certifikate ili dobivanje državljanstva, dolazi u obzir jedino ako su u nju uključeni računalni stručnjaci koji mogu jamčiti da neće doći do tehničkih i sigurnosnih problema.

2.3.2. Primjena e-ispita

Kako su se razvijale mogućnosti računala i interneta te kako su u učionice dolazili učenici sa sve većom razinom računalnih vještina i vještina snalaženja na internetu, nastavnici su trebali osmisliti nove načine na koje učinkovito primjenjivati računala i internet u jezičnom učenju, poučavanju, ispitivanju i procjenjivanju. Najveći izazov na početku uvođenja prvih e-ispita nije bio kako izraditi ispit ili koji alat primijeniti kojime bi se brže i učinkovitije provelo ispitivanje u odnosu na ispitivanje koje provode nastavnici, već kako osmisliti zadatke namijenjene za provedbu pomoću računala, odnosno interneta, te proučiti kako rezultati na e-ispitima i njihovi ispravci dovode do učenja (Chapelle i Douglas, 2006).

Što se tiče potonjega, jedan od prvih projekata – DIALANG, razvijen netom nakon adventa interneta, a koji je financirala Europska unija, imao je namjenu dati dijagnostičke informacije učenicima kako bi unaprijedili svoje razumijevanje učenja jezika, a ujedno je uključio i obilježja samoprocjenjivanja (Alderson i Banerjee, 2001), dok je ispit pod nazivom Longman English Interactive, razvijen tek nešto kasnije, učenicima davao povratne obavijesti nakon svake cjeline o tomu koliko su dobro naučili obrađeno gradivo (Chapelle i Voss, 2017). Ovakva primjena tehnologije u ispitima niskog rizika, kojom se procjenjivanje učenicima učinilo dostupnijim, počela je mijenjati dinamiku u odnosu ispitanika i ispita – učenici su mogli saznati kako napreduju u učenju, što trebaju ponoviti, je li razina njihova samopouzdanja u vezi s usvojenim znanjem opravdana i dr. – i nastavila se razvijati do današnjih dana, kad su dostupne brojne aplikacije koje prate učenje jezika ili uvježbavaju kandidate za standardizirane jezične ispite poput TOEFL-a i IELTS-a (ibid.).

Ispiti visokog rizika, s druge strane, još se uvijek u većini slučajeva provode uživo, a ne (isključivo) putem interneta, i to na daljinu. Početkom ovog desetljeća, jedni se ispiti nisu

mogli provoditi uživo zbog pandemije (npr. međunarodni ispiti francuskoga – DALF, DELF, Ev@lang, TCF1; njemačkoga – TestDaF, Goethe, ÖSD, ÖIF2; korejskoga – TOPIK3; španjolskoga – DELE4; engleskoga – CAEL, CELPIP5, PTE6, Aptis7, Password8 te većina Cambridgevih ispita), unatoč tomu što su neki od njih imali spremne digitalne inačice, drugi su ispitanicima ponudili samo ograničene mogućnosti polaganja putem interneta (npr. ispit umijeća kineskog jezika Konfucijeva instituta), dok su se treći uspješno brzo prilagodili i ponudili takvu mogućnost (Isbell i Kremmel, 2020). Od potonjih izdvajaju se sljedeći ispiti (ibid.):

- Ispiti Američkog odbora za poučavanja stranim jezicima (ACTFL), koji su svoje početke imali još u 1990-ima, kad su se usmeni ispiti provodili telefonski, a zatim i putem interneta od 2006., a primjenjuju se na američkim fakultetima u nastavi stranih jezika, za prijem na sveučilišta, pri zapošljavanju nastavnika stranih jezika, pri napredovanju na poslu, za vrjednovanje nastavnih programa na stranim jezicima i dr.
- Educational Testing Service ponudio je TOEFL iBT Special Home Edition kao posebnu inačicu svojeg ispita TOEFL iBT (Internet-based Test), koji je dostupan još od 2005. godine, ali se mogao polagati samo u ispitnim centrima, s nadzirateljem i samo na određene datume. Ovaj se ispit primjenjuje za potrebe studiranja na programima na engleskom jeziku, odnosno prijema na visoka učilišta u SAD-u i 150-ak drugih zemalja.
- IELTS Indicator, koji se primjenjuje od travnja 2020. godine i koji bi trebao biti u primjeni samo dok su ispitni centri uživo zatvoreni zbog pandemije, i čije rezultate prihvaćaju neka visoka učilišta diljem svijeta za prijem studenata.
- Pearsonovi ispiti Versant, koji nude ispite pisanja, govorenja, četiriju vještina, jezika struke i dr., pri čemu se i ispiti govorenja automatski strojno boduju (i to trenutačno ispiti engleskoga, francuskoga, španjolskoga, nizozemskoga i arapskoga). Ispiti Versant primjenjuju se u svrhe zapošljavanja i prijema na razne nastavne programe na stranom jeziku.
- Ispiti LanguageCert, koje regulira Ured za regulaciju kvalifikacija i ispitivanja Ujedinjene Kraljevine (Ofqual), i koji su dostupni još i otprije pandemije, i koji nude ispite engleskog, španjolskog i turskog jezika, od kojih su putem interneta dostupni samo neki ispiti prvih dvaju jezika radi potreba studiranja i zapošljavanja.
- Duolingov ispit engleskog jezika, čije rezultate mnoga visoka učilišta uzimaju u obzir prilikom donošenja odluka o prijemu studenata, iako ispit ne tvrdi izričito da mjeri umijeće u akademskom kontekstu.

- Ispit TEF Express (Test d'évaluation de français), koji nudi Pariška trgovinska i industrijska komora, članica Udruge europskih jezičnih ispitnih centara (ALTE), za prijem na frankofona kanadska visoka učilišta.

Važno je i istraživati učinke ispitivanja putem interneta na daljinu kako bi se unaprijedile daljnje prakse u ovom području, a ispitni bi centri svakako trebali dalje unaprjeđivati mogućnosti ispitivanja putem interneta, i to ne samo zbog mogućih nepredvidivih situacija u budućnosti koje će onemogućiti ispitivanje uživo u njihovim prostorima, već i zbog pravedna pristupa ispitu osoba koje žive u izoliranim sredinama, osoba s teškoćama u kretanju, osoba nižeg ekonomskog statusa zahvaljujući manjim troškovima održavanja i polaganja ispita na daljinu, zbog mogućnosti polaganja ispitivanja samo onih vještina koje su ispitanicima potrebne, zbog sve veće fleksibilnosti koju ispitanici traže i koja je nužna radi, među ostalim, studiranja, zapošljavanja, selidba i sl., što podrazumijeva i širu primjenu ispita putem mobilnih uređaja, a ne samo računala (Isbell i Kremmel, 2020).

2.3.3. Izrada e-ispita

Prema mišljenju Daviesa (1984) e-ispiti, kao i svi drugi jezični ispiti, ustvari su „psihološki testovi sa stručnim sadržajem, pa autori jezičnih ispita moraju naučiti stvari vezane za ispitivanje, a ne jezik“ (str. 50). Povlačeći paralelu između jezičnog ispitivanja i psihologijskog testiranja, jezične ispite postignuća, jezičnog umijeća i dijagnostičke ispite¹⁶ uspoređuje s kognitivnim testovima, a jezične ispite sposobnosti¹⁷ (eng. *aptitude tests*) s nekognitivnima, no dodaje da ta razlika nije sasvim jasna jer niti razlika između osobnosti i znanja nije jasna. Davies (ibid.) dalje napominje da su jezično i psihologijsko testiranje u potpunosti usporedivi kad se pogledaju koraci u njihovoj izradi, koji su prikazani niže i o kojima iscrpnije govori niz autora koji su na razne no međusobno slične načine organizirali te korake (usp. Davies, 1984; McNamara, 2000; Udier i Jelaska, 2008; Fulcher, 2010; Green,

¹⁶ Pojmovi prema Bagarić Medve (2011): ispiti postignuća (eng. *achievement tests*), ispiti jezičnog umijeća (eng. *proficiency tests*) i dijagnostički ispiti (eng. *diagnostic tests*). Ispiti postignuća još se nazivaju ispitima vrsnoće (usp. Jelaska i Cvikić, 2008), no u ovom smo se radu odlučili primjenjivati pojam ispiti umijeća. Prema Bagarić Medve (2023), svrha ispita jezičnog umijeća jest „procjena općih jezičnih znanja i sposobnosti na kraju određenog obrazovnog ciklusa zbog čega se često smatra pretpostavkom za nastavak obrazovanja“ (str. 464), svrha ispita postignuća jest „procjena količine i kakvoće usvojenosti obrađenih sadržaja u određenom predmetu, odnosno ovladanosti obrazovnim ishodima u tom predmetu nakon ograničenog razdoblja (npr. na kraju nastavne jedinice, cjeline ili na kraju nastavnog programa)“ (str. 464-465), a svrha je dijagnostičkih ispita „otkrivanje aspekata jezičnog znanja i sposobnosti učenika koji su nedostavno usvojeni“ (str. 465).

¹⁷ Sposobnost učenja jezika (eng. *language learning aptitude*) definira se kao „komplet kognitivnih i opažajnih sposobnosti koje olakšavaju procese učenja jezika; određeno svojstvo za učenje stranih ili drugih jezika (J2) brzo, uspješno i s lakoćom“ (Roehr-Brackin, 2021: 147).

2014; Hughes i Hughes, 2020; Green i Fulcher, 2021). Valja istaknuti i da jasan plan za osmišljavanje, izradu i primjenu ispita poboljšava kvalitetu ispitivanja, a ako se proces ispitivanja na ispravan način osmisli, provede i dokumentira, podložan je provjerama i kontrolama, što je preduvjet za postizanje povjerenja u njegove rezultate (Green i Fulcher, 2021). Koraci su izrade (standardiziranog) ispita predstavljeni niže.

- **Planiranje ili osmišljavanje.** Ono uključuje definiranje svrhe ispita, određivanje dionika ispita (autora, ispitanika, drugih korisnika), utvrđivanje sadržaja, izgleda i trajanja ispita, postavljanje zadataka, pitanja i uputa, odlučivanje o vrsti zadataka, planiranje očekivanih odgovora ispitanika, razradu bodovanja i procjenjivanja, a odabir ispitnih čestica „obavlja se po načelu karakterističnosti, učestalosti i važnosti u stvarnomu životu [...] te prema određenomu stupnju poznavanja jezika koji se provjerava“ (Udier i Jelaska, 2008: 248). Ove prethodno navedene sastavnice koraka osmišljavanja ispita uključuju se u ispitnu specifikaciju¹⁸, koju pregledavaju stručnjaci dotične struke, u našem slučaju jezični stručnjaci, a eventualno i drugi dionici procesa ispitivanja, kako bi vidjeli jesu li ispiti značajni i prikladni te uključuju li sva važna područja određene djelatnosti (usp. Bagarić Medve, 2023). Nakon toga, ispitna se specifikacija može doraditi, a u skladu s time i ispitne čestice.
- **Izrada ispitnih čestica.** Ispitne se čestice odabiru prema sadržaju ispita, a izrađuju se slijedeći ispitnu specifikaciju, i to barem trostruko više nego što će ih naposljetku biti potrebno. Barem dvoje kolega od kojih ni jedan nije autor ispita moderiraju čestice – pronalaze njihove slabosti, neke čestice poboljšavaju, a neke izbacuju.
- **Predispitivanje.** Prema Daviesu (1984), prva radna inačica ispita provodi se na malenoj skupini zainteresiranih ljudi radi introspekcije o učinku ispita i micanja

¹⁸ Ispitna specifikacija (eng. *specifications*) ključni je dokument koji navodi standarde i smjernice o tomu kako proizvesti, provesti i ponoviti procjenjivanje; u njima je predstavljen plan za prikupljanje uočljivih dokaza na temelju kojih će se doći do zaključaka o tomu koliko znanja, vještina ili sposobnosti ispitanik posjeduje, povezujući teorijsko razumijevanje jezika s odlukama koje će korisnici ispita donijeti o ispitanicima (Green, 2014). Drugim riječima, ispitna specifikacija podrazumijeva upute za sastavljanje ispita koje su napisane kao da će po njima postupati netko tko nije autor ispita, a čija je uloga osigurati jasnoću prilikom donošenja odluka o izradi ispita i omogućiti izradu novih inačica u budućnosti (McNamara, 2000). Ona uključuje informacije o svrsi ispita, sadržaju ispita, ispitanicima kojima je namijenjen, materijalima za ispitivanje i njihovim izvorima ako su autentični, obliku ispita, dužini i strukturi svakog dijela ispita, jeziku uputa, trajanju i načinu provedbe ispita, vrsti zadataka, kriterijskim razinama izvedbe ispitanika, načinu odgovaranja na pitanja i rješavanju zadataka, ljestvicama odnosno postuptcima bodovanja izvedbe ispitanika i dr. (usp. McNamara, 2000; Udier i Jelaska, 2008; Fulcher, 2010; Hughes i Hughes, 2020). Hughes i Hughes (2020) napominju da temeljne sastavnice ispitne specifikacije moraju uvijek biti zastupljene u svakoj inačici ispita, no ipak se mogu mijenjati nakon pilotiranja ili uslijed promijenjenih okolnosti provedbe ispita, a i zato što su neke možda bile nepoznate u trenutku kad su se određivale. Također naglašavaju da se sadržaj ispita često određuje prema onomu što je lagano, a ne ono što je važno ispitati, pa je nužno odrediti cjelovitu ispitnu specifikaciju i osigurati da se u njoj odražava sadržaj ispita kako bi se izbjegla ta opasnost.

nezadovoljavajućih čestica, a nakon toga dvije ili više probnih radnih inačica provode se na veliku uzorku osoba jednakih onima koje će biti ispitanici radi provjere provedbe i analize čestica. Prema Hughesu i Hughesu (2020), prva se inačica ne provodi u ispitnim uvjetima i daje se stručnim govornicima¹⁹, barem njima dvadeset ili više, koji nisu jezični stručnjaci niti stručnjaci u području jezičnog ispitivanja, ali su slični namijenjenima ispitanicima po dobi, obrazovanju i dr., nakon čega se problematične čestice mijenjaju ili odbacuju; druga se inačica ispita daje nestručnim govornicima koji su slični namijenjenim ispitanicima, a ona se provodi u ispitnim uvjetima. U svakom slučaju, cilj je predispitivanja dobiti podatke iz kojih se vidi daju li dotični ispiti dobre rezultate. U koraku predispitivanja ispitanicima se daje i upitnik u kojemu „ocjenjuju stupanj složenosti pitanja i zadataka te jasnoću uputa za rješavanje zadataka. Traži se i njihovo mišljenje o zadacima i materijalima u ispitu.“ (Udier i Jelaska, 2008). Nakon statističke i kvalitativne analize rezultata predispitivanja rade se određene promjene prema potrebi, a jezični stručnjaci podešavaju bodovne ljestvice u slučaju zadataka koji ispituju vještinu pisanja i/ili govorenja.

- **Utvrđivanje valjanosti.** Ovaj je postupak iscrpnije opisan niže u poglavlju 2.4.2. – Utvrđivanje valjanosti, no ovdje ćemo samo istaknuti da ga Davies (1984) naziva probom završne inačice ispita radi prikupljanja dokaza njegove praktične korisnosti (eng. *test usefulness*) ili valjanosti te radi prikupljanja norma, odnosno načina na koji će se procjenjivati značajnosti rezultata ispita. Usto, Hughes i Hughes (2020) ističu kako je utvrđivanje valjanosti ključno za ispite visokog rizika ili za javno dostupne ispite, dok za ispite niskog rizika koji se primjenjuju unutar određene institucije ono nije potrebno, iako je neslužbeno utvrđivanje valjanosti manjeg opsega poželjno za ispite koji će se primjenjivati više puta tijekom određenog vremenskog razdoblja.
- **Zadnji koraci** jesu pisanje priručnika za dionike ispita, obuka osoblja koje će sudjelovati u procesu jezičnog ispitivanja (nadziratelje, usmene ispitivače, procjenjivače, bodovatelje računalne tehničare, programere i dr.) te redovita statistička i kvalitativna analiza ako se ispit provodi duže vrijeme kako bi se identificirali i otklonili mogući problemi, što nekad znači i izradu novih inačica ispita.

Od iznimne je važnosti izbjeći nesklad između namjera autora ispita u odnosu na sadržaj ispita i stručnih prosudaba o tomu što određene čestice mjere, a uzrok tomu može biti

¹⁹ Hughes i Hughes (2020) koriste se izrazom stručnog govornika (eng. *expert speaker*), a ne izvornog govornika (eng. *native speaker*) zbog, kako ističu, „sveopćeg odbacivanja pojma 'izvornog govornika', koji se odnosi na osobu najviše razine jezičnog umijeća“ (str. 63).

loša izrada ispitne specifikacije i/ili čestica ili ozbiljniji problemi povezanosti ispitne specifikacije i valjanosti ispita (Alderson, 1988). Ispitna specifikacija ispita (prethodno iscrpnije predstavljena u bilješci 18) inače ima dvojaku ulogu – pružiti izjavu o tomu što ispit namjerava mjeriti, odnosno što je njegov konstrukt (prethodno iscrpnije predstavljen u bilješci 11), te pružiti smjernice autorima ispita o tomu kako će se ispit izraditi i kako se može jamčiti podudarnost novih inačica (ibid.). Naime, prema Američkoj udruzi za istraživanje u obrazovanju (The American Educational Research Association, dalje u tekstu: AERA), Američkoj psihološkoj udruzi (The American Psychological Association, dalje u tekstu: APA) i Nacionalnom vijeću za mjerenja u obrazovanju (The National Council on Measurement in Education, dalje u tekstu: NCME), dvije ili više inačica ispita koji se smatraju međusobno zamjenjivima nazivaju se ispitima zamjenjiva oblika (eng. *alternate form tests*), a smatraju se takvima kad mjere jednak konstrukt, kad su namijenjeni jednakoj svrsi i kad se provode u jednakim uvjetima (Urbina, 2004), pa bi stoga svi ispitanici trebali imati jednaku mogućnost pokazati svoju razinu izvedbe neovisno o obliku ispita koji polažu (Bae i Lee, 2010). Generički naziv „zamjenjiv oblik“ uključuje tri kategorije – usporediv oblik (eng. *comparable form*), istovrijedan oblik (eng. *equivalent form*) i podudaran oblik (eng. *parallel form*), od kojih je potonji najstroži budući da ispiti koji su podudarni, uz gore nabrojana obilježja zamjenjivih ispita, moraju imati određenu vrstu statističke istovrijednosti (Urbina, 2004; Bae i Lee, 2010).

Što se tiče osmišljavanja zadataka za ispite na računalima, a naročito za e-ispite, za njih je potrebno osmisliti nove tipove zadataka koji se razlikuju od onih za ispite na papiru (García Laborda, 2007). Od početka primjene e-ispita razvili su se razni programski alati i materijali kojima se nastavnici i autori ispita mogu koristiti za izradu zadataka i interaktivnih e-ispita u koje se može dodati multimedijски sadržaj, automatsko ispravljanje, slanje povratnih obavijesti, pohrana rezultata ispitanika i sl., primjerice Test Pilot, Hot Potatoes, Questionmark, Blackboard, Moodle, Google Forms i dr. (usp. Polio, 2001; Winke i MacGregor, 2001; Chapelle i Douglas, 2006; Kessler, 2013; Lesiak-Bielawska, 2015). Ono što je važno istaknuti jest da programi za izradu ispita ne trebaju od autora ispita zahtijevati da budu računalni programeri, već im omogućiti raznovrsnost interaktivnih načina procjenjivanja putem računala s ugrađenim multimedijским alatima, automatskim bodovanjem, mogućnostima slanja povratnih obavijesti i pohranom rezultata ispitanika (Kessler, 2013). Chapelle i Voss (2016) još dodaju da mnogi programi omogućuju predloške za tradicionalne tipove zadataka, poput pitanja višestrukog izbora, pridruživanja, dopunjavanja i dr., te pružaju mogućnost različita oblikovanja ispita. Međutim, kojim se god programom koristilo za izradu ispita, moraju se uzeti u obzir njihova svojstva, koja su opisana u sljedećem potpoglavlju.

2.3.4. Svojstva e-ispita

Pri izradi svakog ispita, pa tako i jezičnih e-ispita, valja se voditi njegovim temeljnim svojstvom, a to je **korisnost**, odnosno primjena za koju je dotični ispit namijenjen (Bachman i Palmer, 1996; Shepard, 2016). Međutim, sama činjenica da se ispitom koristi, što znači da se primjenjuje u određenu svrhu, ne znači da je koristan – koristan je tek ako ovo krovno svojstvo u sebi objedinjuje uključuje šest sljedećih svojstava: pouzdanost, valjanost konstrukta, prirodnost, poticajnost, utjecajnost i izvedivost (Bachman i Palmer, 1996). Valjanost je iscrpnije obrađena u poglavlju 2.4. – Valjanost, a ostala su svojstva kraće predstavljena u sljedećim odlomcima.

Pouzdanost (eng. *reliability*) podrazumijeva „da višestruka mjerenja uvijek daju jednake ili gotovo jednake rezultate“ (Kraš i Miličević, 2015: 38). Drugim riječima, odnosi se na dosljednost rezultata na ispitu na različitim aspektima ispita, što znači da bi korisnik ispita trebao imati neko jamstvo da se isti rezultati mogu ponoviti u slučaju da isti ispitanici polažu isti ispit u sličnim okolnostima (Fulcher i Davidson, 2007). Pouzdanost se može definirati i na sljedeći način (Chapelle i Lee, 2021: 22): „Dosljednost bodova ostvarenih na raznim ispitnim zadacima, oblicima ispita, prigodama ispitivanja i bodovanjima. Kao i konstrukti, pouzdanost se može definirati i istraživati na mnoge različite načine, pa je potrebno navesti posebno značenje pouzdanosti za dotično istraživanje.“ Naposljetku, pouzdanost se može podijeliti na vanjsku pouzdanost, što znači da drugi istraživači mogu ponoviti dotično istraživanje, i na unutarnju pouzdanost, što znači da drugi istraživači trebaju dobiti isti rezultat kad ponovo analiziraju podatke dotičnog istraživanja (Kraš i Miličević, 2015).

Prirodnost ili autentičnost (eng. *authenticity*) predstavlja odnos između obilježja ispitnih zadataka i obilježja zadataka u stvarnom svijetu (Fulcher i Davidson, 2007). Udier i Jelaska (2008) za prirodnost navode sljedeće. „[...] stupanj je oponašanja, s obzirom na uvjete, stvarnih životnih okolnosti ispitanika. Prirodnost se u ispitivanju očituje kao usklađenost između ispitanikovih obilježja i uporabnih obilježja ispita. [...] obilježje prirodnosti ima ispit čiji se rezultati mogu povezati s ispitanikovom ovladanošću jezikom“ (str. 250). Prema Chapelle i Lee (2021), prirodnost jezičnog ispita predstavlja „podudarnost jezika i jezičnih zahtjeva u ispitnim zadacima i onih u kontekstu primjene jezika. Prirodnost se često podrazumijeva kao pozitivno svojstvo jezičnog ispita zbog povezanosti s tumačenjem konstrukta ispita“ (str. 22). Drugim riječima, prirodnost predstavlja „jamstvo da ispitni zadatci odgovaraju jezičnim zadacima izvan samog ispita i da se rezultat ostvaren na ispitu može

proširiti na rezultat i njegovo tumačenje u ciljnoj primjeni jezika“ (Bachman, 2005: 26), o čemu iscrpnije govorimo u metodi potkrjepe procjenjivanja u poglavlju 2.5.

Poticajnost (eng. *interactiveness*) podrazumijeva stupanj do kojega su pojedinačna obilježja ispitanika, poput jezične sposobnosti, pozadinskog znanja, motivacije i dr., uključena prilikom polaganja ispita (Fulcher i Davidson, 2007). Udier i Jelaska (2008) za poticajnost navode da je ona „veza između rezultata i odluka koje donosimo tumačeći ih“ (str. 250). Također, na poticajnost se može gledati kao na „jamstvo da su procesi koje ispitni zadatak potiče jednaki onima koji su određeni konstruktom“ (Bachman, 2005: 26).

Utjecajnost (eng. *impact*) definira se kao „stupanj kojim ispit potiče ispitanikove sposobnosti koje želimo mjeriti“ (Udier i Jelaska, 2008: 250). Prema Bachmanu (2005: 26), „svojstvo utjecajnosti [...] uključuje povratni učinak, odnosno učinak na poučavanje, kao i načine na koje primjena ispita utječe na pojedince, institucije i društvo u cjelini“, a pod ovo svojstvo dodaje još nepristranost (tj. utjecaj na to jesu li sve društvene skupine uključene u procese ispitivanja, kakva im je izvedba, kako se procjenjuju i kako se primjenjuju rezultati njihove izvedbe) te jednak pristup ispitu svim ispitanicima, neovisno o čimbenicima vezanima za obrazovanje, financijski status, lokaciji, opremi, administraciji i dr. Fulcher i Davidson (2007) također napominju da se istraživanja utjecajnosti ispita preklapaju s pojmom njegova povratnog učinka, koji prelazi granice učionice te se odnosi na cijelu školu, obrazovni sustav pa i državu. Međutim, Wall (1997) utjecaj definira kao „bio koji učinak koji ispit može imati na pojedince, politike i prakse, u učionici, školi, obrazovnom sustavu ili cijelom društvu“ (str. 291), a povratni se učinak ponekad „upotrebljava kao sinonim za 'utjecaj', ali češće se upotrebljava za učinke ispita na poučavanje i učenje“ (ibid.).

Izvedivost ili **praktičnost** (eng. *feasibility* ili *practicality*; usp. McNamara, 2000; Council of Europe, 2001) bavi se provedbom ispita, a ne značenjem ispitnih rezultata (Fulcher i Davidson, 2007). Prema Udier i Jelaski (2008: 251):

Da bi ispiti bili izvedivi, mnogobrojnost različitih (mogućih) vrsta provjera mora se svesti na ograničen broj i vrstu kriterija uslijed vremenskih, novčanih i drugih ograničenja. Drugim riječima, ako je ispit i valjan i pouzdan, može biti sasvim neizvediv u zadanim okolnostima.

McNamara i Roeber (2006) napominju da postoji „značajna napetost između osmišljavanja autentičnih ispitnih zadataka i praktičnosti“ budući da bi „bilo idealno da se simuliraju situacije iz stvarnog svijeta te da izvedbu ispitanika procjenjuju ljudski procjenjivači, što je vrlo skupo“ (str. 54).

Iako krovno svojstvo korisnosti ispita u sebi objedinjuje šest drugih svojstava, Bachman i Palmer (1996) ističu kako se dvama od njih, pouzdanosti i valjanosti, ipak treba prvenstveno voditi pri izradi svakog ispita jer ona daju potporu za primjenu rezultata na ispitu kao temelj za donošenje određenih zaključaka ili odluka. Davies još ističe sljedeće: „Na pouzdanost se tradicionalno gledalo kao na zasebno svojstvo, usporedno i mlađe od valjanosti. No može se smatrati i sastavnicom valjanosti“ (Davies, 2011: 38). Ono što smatra neobičnim jest da je odnos između pouzdanosti i valjanosti jednosmjernan – što je veća pouzdanost ispita, to je veća mogućnost njegove valjanosti, a ako se može dokazati dobra valjanost nekog ispita, njegova se pouzdanost može pretpostaviti i postaje sporedna (ibid.). Naposljetku, Davies (1984) čak napominje da je suvišno spominjati valjanost ispita kad se govori o izradi ispita jer ju izrada ispita podrazumijeva, odnosno sadrži ju u sebi kao svoj rutinski dio, što se vidi i u definiciji ispita koju su dali McNamara i Roeber (2006), a koja glasi da je „ispit postupak prikupljanja dokaza radi potpore odlukama koje se donose i radi pažljiva tumačenja tih dokaza“ (str. 17). U sljedećem se poglavlju iscrpnije obrađuju pojmovi valjanosti i utvrđivanja valjanosti te se prikazuje njihova važnost u području jezičnog procjenjivanja, a posebice e-procjenjivanja.

2.4. VALJANOST

2.4.1. Definicija valjanosti

Klasična definicija **valjanosti** iz 1921. godine, a koju je dala Nacionalna udruga voditelja istraživanja u obrazovanju (National Association of the Directors of Educational Research), jest da je ona opseg u kojem ispit mjeri ono što tvrdi da mjeri; drugim riječima, valjanost ispita odnosi se na ono što ispit mjeri i kako dobro to radi (Urbina, 2004). I Milas (2009) valjanost u najširem smislu vezuje uz pitanje „[...] jesmo li doista mjerili ono što smo namjeravali mjeriti?“ (str. 113). Davies i Elder (2005) ovom starom, deduktivnom pogledu na valjanost stavljaju u opreku novi, induktivni pogled – što je to što ispit mjeri? Prema njima, stari pogled „misli da zna“, dok novi ne – prema njemu, valjanost zanima što ispit predstavlja, što on znači, do kojih se zaključaka može doći na temelju rezultata ostvarenih na njemu. No još je 1980. godine Messick valjanost definirao kao ukupan stupanj opravdanosti tumačenja i primjene ispita, odnosno kao vrijednosnu prosudbu dostatnosti i primjerenosti tumačenja rezultata ispita.

Stoga se na valjanost ne treba gledati kao na opće svojstvo ispita niti pretpostavljati da može biti potpuna, već ju treba gledati u smislu njezina stupnja na kojemu prikupljeni podatci podupiru namijenjeno tumačenje rezultata ispita za predloženu svrhu; ona je uvijek svojstvena toj određenoj svrsi, odnosno namijenjenoj primjeni (eng. *intended use*) ili tumačenju (usp. Anastasi, 1976; Angoff, 1988; Messick, 1989; Urbina, 2004; Bachman, 2004; Weir, 2005; Fulcher i Davidson, 2007; Green, 2014; Green i Fulcher, 2021). I Chapelle (2021) navodi da se valjanost odnosi na rasuđivanje koje uključuje određene tvrdnje o tumačenjima i primjenama ispita, te da ključna pitanja u vezi s valjanosti leže u konceptima koji su potrebni za uspješno rasuđivanje, od kojih je najosnovniji zaključivanje. McNamara i Roever (2006) dodaju i da valjanost nije neko matematičko svojstvo, kao što je to primjerice pouzdanost, već da je ona stvar prosuđivanja. Za valjanost Milas (2009) također ističe da se ne odnosi „toliko na sam ispit, koliko na *tumačenje* što se temelji na njegovoj specifičnoj primjeni“ (str. 373). Stoga Bachman (2004) valjanost podrazumijeva kao kvalitetu tumačenja i primjene rezultata procjenjivanja, a ne kao kvalitetu samog ispita ili samih rezultata.

Nadalje, Cronbach (1988) naglašava da se na valjanost ne smije gledati kao na izricanje presude palcem prema gore ili dolje, već da se potvrđivanje tumačenja određenog ispita mora potkrijepiti dokazima. Fitzpatrick i Clenton (2010) također tvrde da je utvrđivanje valjanosti, koje je iscrpnije predstavljeno u sljedećem potpoglavlju, neprekidan i stupnjevit proces te naglašavaju kako valjanost uvelike, a možda i najviše, ovisi o primjeni ispita, dok i Davies i Elder (2005) napominju da nijedan ispit nije posve valjan, već da se razvija nizom zaključaka i tumačenja. Naposljetku, Shepard (2016) ističe da najnovija definicija valjanosti APA-e, AERA-e i NCME-a u Standardima za obrazovne ispite i psihološke testove (Standards for Educational & Psychological Tests) iz 2014. godine definira valjanost kao stupanj do kojega dokazi i teorija podupiru tumačenja rezultata ispita koja proizlaze iz namijenjenih primjena ispita. Međutim, ona naglašava i da valjanost mora uključivati i naredne odluke koje se donose i radnje koje se poduzimaju na temelju tumačenja rezultata određenog ispita.

2.4.2. Utvrđivanje valjanosti

Messick (1989) je među prvima definirao **utvrđivanje valjanosti** (eng. *validation*) kao objedinjenu prosudbu stupnja do kojeg empirijski dokazi i teorijska podloga podupiru dostatnost i primjerenost tumačenja i radnja koje se temelje na rezultatima ispita ili drugog načina procjenjivanja. Dvadesetak godina kasnije, McNamara (2000) je tu definiciju doradio navevši da se utvrđivanje valjanosti može vidjeti kao poduzimanje mjera kojima se istražuju

postupci kojima se dolazi do prosudaba o određenom ispitu, a da je njegova svrha omogućiti obranjivost i pravednost tumačenja koja se temelje na ispitanikovo izvedbi na ispitu. Stoga naglašava da utvrđivanje valjanosti uključuje razmišljanje o logici ispita, naročito o tomu kako je osmišljen i koje su mu namjere, i da također uključuje razmatranje empirijskih dokaza, odnosno konkretnih činjenica do kojih se dolazi pilotiranjem ispita i operativnim provedbama. I drugi znanstvenici utvrđivanje valjanosti opisuju kao proces prikupljanja dokaza koji doprinose valjanosti ispita (usp. Urbina, 2004; Davies i Elder, 2005; Weir, 2005; Fulcher i Davidson, 2007; Green, 2014), a Llosa (2008) ističe da ono zahtijeva istraživački program koji se oslanja na višestruke izvore dokaza kako bi dionicima opravdao tumačenja i primjene koje se temelje na određenom procjenjivanju. Naposljetku, i sami Standardi za obrazovne ispite i psihološke testove u svojoj najnovijoj inačici iz 2014. godine napominju da proces utvrđivanja valjanosti uključuje prikupljanje značajnih dokaza radi pružanja čvrstih znanstvenih temelja za predloženo tumačenje rezultata te da se ne vrjednuje sam ispit, već tumačenja rezultata ispita za njegovu predloženu primjenu (Shepard, 2016).

Kako bi se utvrđivanje valjanosti ispita bolje razumjelo, Urbina (2004) ga je usporedila s testiranjem hipoteza, a McNamara (2000) i Davies i Elder (2010) s obranom na sudskom ročištu, naravno, uz određene razlike između tih dvaju koncepata. Naime, kao na sudu, nakon što policija obavi istragu i na temelju raspoloživih dokaza donese zaključak da je počinjeno kazneno djelo, počinitelja se optuži, a nakon toga taj se njihov zaključak preispituje neovisnim postupkom, što radi sudac i/ili porota (McNamara, 2000). Međutim, valja dodati da za razliku od sudskog postupka, postupak utvrđivanja valjanosti nikad ne završava – uvijek se može doći do novih dokaza i na taj način poboljšati kvalitetu određenog procjenjivanja (Anderson i Banerjee, 2001; Green, 2014). Još jedna stvar koja je istovjetna sudskom ročištu i utvrđivanju valjanosti ispita jesu dva koraka obaju postupaka – prvi je korak rezultat ostvaren na ispitu na temelju kojega se zaključivanjem dolazi do tvrdnje o sposobnosti ispitanika, a drugi je korak istraživanje razumnosti te tvrdnje (Davies i Elder, 2010). Važnost drugog koraka uvelike ovisi o posljedicama prvog, pa što su veći ulozi, to je veća potreba za utvrđivanjem valjanosti, kao što je to slučaj kod ispita visokog rizika (McNamara, 2000). Stoga je vrlo rijetko slučaj da postupak utvrđivanja valjanosti ispita pokreće sam autor ispita jer on može imati određeni subjektivan interes za ishod tog postupka te može nevoljko tražiti dokaze protiv namijenjenih tumačenja dotičnog ispita (Davies i Elder, 2005).

Međutim, autor nije jedini niti najvažniji dionik u postupku utvrđivanja valjanosti. Kako navodi Angoff (1988), prilikom utvrđivanja valjanosti ne utvrđuje se valjanost samog ispita niti rezultata koji se na njemu postižu, već se utvrđuju tumačenja i zaključci do kojih

korisnik ispita dolazi na temelju tih rezultata te odluke i radnje koje su posljedica dotičnih zaključaka. Na taj način valjanost postaje odgovornost svih dionika – od autora ispita do korisnika ispita, a osobe koje provode testiranje i one koje se koriste rezultatima odgovorne su za prikupljanje dokaza valjanosti ispitivanja, odnosno valjanosti tumačenja rezultata ostvarenih na ispitivanju (usp. Messick, 1980; Angoff, 1988; McNamara, 2000; Urbina, 2004). Urbina (2004) dalje ističe da doprinos dokazivanju rezultata ispita može dati bilo koje sustavno istraživanje koje podupire ili obogaćuje njihovo značenje, neovisno o tomu tko ga provodi i kad ga provodi, te napominje da sve dok postoje znanstveni dokazi predloženoj primjeni rezultata ispita, korisnici ispita mogu slobodno primijeniti te rezultate u svoje svrhe, neovisno o tomu jesu li ih predvidjeli autori ispita (ibid.).

Nadalje, Shepard (2016) navodi da se prilikom utvrđivanja valjanosti treba voditi namijenjenom primjenom ispita, koju ona tumači kao ishode ispita. Međutim, primjena ispita može imati razna značenja – od procjene razine sposobnosti, preko davanja ocjena, do poticanja učenika da uče određeno gradivo (Chapelle, Jamieson i Hegelheimer, 2003). Stoga se prema Read i Chapelle (2001) svrha ispita sastoji od triju dijelova – prvo, od zaključaka o sposobnosti ispitanika i naknadnom potencijalu za izvedbu; drugo, od primjene rezultata ispita, odnosno od praktičnih ishoda rezultata ispita; treće, od učinka koji ispit namjerava imati izvan određenog mjerenja, odnosno od namijenjenih učinaka ispita na njegove dionike (učenike, nastavnike, tečajeve, programe, pa čak i društvo u cjelini).

2.4.3. Jedna ili više valjanosti?

Kako navode Anastasi (1976) i Urbina (2004), APA je 1954. godine objavila Tehničke preporuke za psihološke testove i dijagnostičke tehnike (Technical Recommendations for Psychological Tests and Diagnostic Techniques) u kojima su navedene četiri podvrste valjanosti – sadržajna valjanost (eng. *content validity*), prediktivna valjanost (eng. *predictive validity*), usporedna valjanost (eng. *concurrent validity*) i konstruktna valjanost (eng. *construct validity*)²⁰. Međutim, dvadeset godina kasnije APA je u Standardima za obrazovne ispite i psihološke testove objedinila prediktivnu i usporednu u kriterijsku valjanost te je ukazala na to da su sadržajna, kriterijska i konstruktna valjanost zapravo aspekti, a ne vrste valjanosti. Usto, i Milas (2009) ističe da te tri valjanosti „tek združene mogu ponuditi cjelovitu sliku o ukupnoj upotrebljivosti testa za predviđenu namjenu“ (str. 374), a Cronbach

²⁰ Pojmovi na hrvatskom jeziku prema Milasu (2009).

je još 1988. godine napomenuo da je dotadašnji „tridesetogodišnji pogled na tri vrste valjanosti, odvojenih, ali možda jednakih, pogled koji je vrijeme pregazilo“ (str. 4).

Stoga i naše istraživanje valjanost ne vidi kao skup više različitih (pod)vrsta, nego kao jedinstveni koncept koji istovremeno objedinjuje više aspekata, koji prema Douglasu (2001) čine dijelove „mozaika valjanosti“ (str. 447), a koji su zasebno iscrpnije predstavljeni u potpoglavljima niže. Prema Fulcheru i Davidsonu (2007), „u počecima istraživanja valjanosti, valjanost se dijelila na tri 'vrste' na koje se tada u pravilu gledalo zasebno“ (str. 4), a danas se na njih gleda kao na aspekte valjanosti, a ta su tri aspekta najistaknutija i najviše istraživana:

- sadržajna valjanost, koja se još naziva i kontekstna valjanost (eng. *context validity*), a u sklopu koje se nalazi i pojavna valjanost (eng. *face validity*);
- kriterijska valjanost, koja objedinjuje prediktivnu valjanost i usporednu valjanost, a koja se dijeli na dva tipa – dijagnostičku i prognostičku valjanost, koja se još naziva i pragmatičkom valjanosti (eng. *pragmatic validity*); u sklopu kriterijske valjanosti također se spominju vanjska, populacijska, ekološka i sintetička valjanost;
- konstruktna valjanost, koja obuhvaća konvergentnu valjanost (eng. *convergent validity*) i divergentnu valjanost (eng. *discriminant validity*) te unutarnju valjanost (eng. *intrinsic validity*).

Druga četiri značajnija aspekta valjanosti jesu:

- pozitivistička valjanost (eng. *positivist validity*),
- interpretivistička valjanost (eng. *interpretivist validity*),
- eksperimentalna valjanost (eng. *experimental validity*),
- posljedična valjanost (eng. *consequential validity*).

2.4.3.1. Sadržajna valjanost

Prvi od aspekata valjanosti jest **sadržajna valjanost**, odnosno sustavna analiza ispita kako bi se utvrdilo pokriva li taj ispit reprezentativan uzorak predmeta mjerenja (Anastasi, 1976; Fulcher i Davidson, 2007; Green, 2014). Milas (2009) navodi da njezini pokazatelji „kazuju koliko je neki test svojim sadržajem – zadacima, pitanjima ili drugim tipovima elementarnih jedinica, *tematski prilagođen području* koje bi trebao pokrivati“ (str. 374). Weir (2005) predlaže da bi bolji naziv za ovaj aspekt valjanosti bio **kontekstna valjanost** jer je primjereniji za društvenu dimenziju primjene jezika i prikladniji za socijalno-kognitivni pristup ispitivanju. Utvrđivanje ovog aspekta valjanosti primjenjuje se ponajviše kod ispita postignuća, koji mjere koliko je dobro ispitanik ovladao određenom vještinom ili gradivom, pa je zato

važno da se njihov sadržaj utvrdi dosta široko kako bi uključio glavne ishode te primjenu načela i tumačenje podataka kao i činjenično znanje (Anastasi, 1976). Naprimjer, sadržajna je valjanost utkana u određeni obrazovni ispit odabirom primjerenih čestica, koje se pripremaju nakon što se proučio predmetni silab i udžbenici te nakon savjetovanja s predmetnim stručnjacima; nakon toga slijedi izrada ispitne specifikacije za autore ispita (Anastasi, 1976). Milas (2009) također naglašava da valja pribaviti mišljenje stručnjaka kako bi se utvrdila sadržajna valjanost, „podjednako o sadržajnom slaganju testa i intencijski zahvaćene psihičke osobine, kao i reprezentativnosti uzorka čestica za ukupno područje koje bi trebalo obuhvaćati“ (str. 374). U sklopu sadržajne valjanosti spominje se i pojam **pojavne valjanosti**, s kojom se ne smije poistovjećivati, a prema kojoj je važno da se ispit izrađuje tako da jezik i kontekst ispitnih čestica izgledaju valjani i prihvatljivi korisnicima tog ispita – ispitanicima, administraciji, nestručnim promatračima i općoj javnosti (Anastasi, 1976; Angoff, 1988; McNamara, 2000). Naime, mišljenje laika, primjerice ispitanika, roditelja, poslodavaca i dr., jednako je važno kao i ono stručnjaka budući da su i jedni i drugi dionici procjenjivanja (Davies i Elder, 2005; Green, 2014).

2.4.3.2. Kriterijska valjanost

Drugi je aspekt valjanosti **kriterijska valjanost** (eng. *criterion-related validity*), koja pokazuje koliko je ispit učinkovit u predviđanju ponašanja ispitanika u određenim situacijama (Anastasi, 1976). Drugim riječima, ona podrazumijeva odnos između rezultata ispitanika i nekog drugog pokazatelja istih znanja, vještina ili sposobnosti ispitanika, primjerice nastavnikova mišljenja, rezultata na drugom valjanom procjenjivanju (Fulcher i Davidson, 2007; Green, 2014). Milas (2009) napominje da će povezanost primijenjenog ispita i rezultata na drugoj varijabli utvrditi kriterijsku valjanost, a „ako je korelacija visoka, to znači da na temelju testa možemo razmjerno dobro predvidjeti kakav će rezultat neka osoba postići u kriterijskoj aktivnosti“ (str. 375), odnosno u aktivnosti na koju upućuje kriterij. Kriterijska valjanost objedinjuje prediktivnu valjanost i usporednu valjanost. **Prediktivna valjanost** obično se opisuje kao čisto empirijska, odnosno kao usporedba predviđanja s kasnijim događajima (Cronbach, 1988), odnosno kao korelacija između rezultata ispita i kasnijeg ponašanja koje je ispit imao namjenu predvidjeti (Angoff, 1988; McNamara, 1996; Weir, 2005; Fulcher i Davidson, 2007; Green, 2014). S druge strane, **usporedna valjanost** podrazumijeva uspoređivanje rezultata procjenjivanja čija se valjanost utvrđuje s drugim pokazateljima ispitanikova trenutačnog stanja sposobnosti; primjerice, rezultati bodova studenata na ispitu

moгу se usporediti s njihovim općim prosjekom u trenutku ispitivanja, ocjene jednog nastavnika mogu se usporediti s ocjenama drugih i sl. (Anastasi, 1976; Fulcher i Davidson, 2007; Green, 2014). Ona je u mnogim slučajevima zamjena za prediktivnu valjanost jer često nije praktično provoditi utvrđivanje valjanosti u vremenskom razdoblju koje je potrebno za prediktivnu valjanost, pa se ispiti provode na skupini za koju podatci o kriteriju već postoje

Postoje dva tipa kriterijske valjanosti – dijagnostička valjanost i prognostička valjanost koja se još naziva i pragmatičkom valjanosti. **Dijagnostičkom valjanosti** utvrđujemo koji je predmet mjerenja određenog ispita, odnosno što on mjeri (Dizdar, 2006), a odnosi se na „situacije u kojima su test i kriterijska varijabla izmjereni istovremeno ili u vrlo kratkom vremenskom razmaku“ (Milas, 2009: 375). S druge strane, **prognostičkom valjanosti**, odnosno **pragmatičkom valjanosti**, na temelju rezultata određenog ispita pokazujemo „koliko uspješno, odnosno s kolikom sigurnošću možemo predvidjeti uspjeh u nekoj praktičnoj aktivnosti“ (Dizdar, 2006: 311). Prognostička/pragmatička valjanost istražuje „povezanost između testovnog rezultata i kriterijske varijable značajno odmaknute u vremenu“ (Milas, 2009: 375), a ona podrazumijeva i odluku o tomu je li potkrjepa dovoljna za potporu namijenjenoj primjeni ispita (Fulcher i Davidson, 2007; Fitzpatrick i Clenton, 2010).

Nadalje, u sklopu aspekta kriterijske valjanosti spominje se i **vanjska valjanost**, odnosno vjerojatnost kojom se rezultati mogu uopćiti na druge ispitanike (**populacijska valjanost**), na druge okoline (**ekološka valjanost**) i na druge načine operacionalizacije konceptualnih varijabla, tj. na primjenu drugih mjernih instrumenata (Kamenov, 2017). Milas (2009) za vanjsku valjanost u užem smislu kaže da nam ona govori „o mogućnosti uopćavanja uzročnih zaključaka“ (str. 119) te da odgovara na pitanje: „Pod uvjetom da je potvrđena uzročna veza odgovarajućih konstrukata, koliko ih se može uopćiti?“ (str. 120).

U sklopu kriterijske valjanosti također se spominje i **sintetička valjanost**, i to u području psihologije u odjelu ljudskih potencijala, a ona podrazumijeva zaključivanje o valjanosti u određenoj situaciji na temelju sustavne analize sastavnica radnog mjesta, utvrđivanje valjanosti ispita za te sastavnice te kombinaciju tih valjanosti sastavnica u jednu cjelinu (Anastasi, 1976).

2.4.3.3. Konstruktna valjanost

Treći aspekt jest **konstruktna valjanost**, odnosno opseg u kojemu ispit mjeri teorijski konstrukt ili obilježje, a zahtijeva postupno prikupljanje informacija iz različitih izvora (Anastasi, 1976). Drugim riječima, ovaj se aspekt valjanosti odnosi na „povezanost rezultata

dobivenih na instrumentu za koji se utvrđuje konstruktna valjanost s rezultatima dobivenim na drugim instrumentima“ (Milošević, 2010: 21). Milas (2009) napominje da je ona „stupanj u kojem je promjene zavisne varijable moguće pripisati izučavanom teorijskom konstrukt“ (str. 120), odnosno „stupanj u kojem neki testovni rezultat upućuje upravo na određenu psihičku osobinu“ (str. 375), te da odgovara na pitanje: „Pod uvjetom da je potvrđena uzročna veza između nezavisne i zavisne varijable, što ona u teorijskom smislu zapravo predstavlja?“ (str. 120).

Konstruktna valjanost obuhvaća konvergentnu valjanost i divergentnu valjanost. **Konvergentna valjanost** postoji u slučaju visoke korelacije ispita s drugim varijablama s kojima bi trebao teoretski korelirati (Anastasi, 1976), odnosno u slučaju stupnja „povezanosti među varijablama za koje se pretpostavlja da imaju isti predmet mjerenja (Milošević, 2010: 21). Drugim riječima, konvergentna valjanost postoji ako su rezultati jednog procjenjivanja slični rezultatima drugog procjenjivanja koje sadrži drugačije vrste ispitnih zadataka, ali mjeri isto znanje, vještine ili sposobnosti (Green, 2014). Nasuprot tomu, **divergentna valjanost** postoji kad nema značajne korelacije između ispita i varijabla od kojih bi se trebao razlikovati (Anastasi, 1976), odnosno u slučaju odsutnosti „povezanosti među varijablama koje mjere konstrukte koje [sic] se smatraju neovisnima“ (Milošević, 2010: 21), odnosno u slučaju korelacije između različitih konstrukata koje mjeri ista metoda (Angoff, 1988). Drugim riječima, divergentna valjanost postoji ako rezultati na drugom procjenjivanju koje sadrži jednake vrste ispitnih zadataka kao i prvo procjenjivanje, ali procjenjuje druga znanja, vještine ili sposobnosti, rangira ispitanike drugačijim redoslijedom (Green, 2014). Ovdje možemo usporediti i Milasova (2009) objašnjenja ovih dvaju aspekata valjanosti – konvergentnu valjanost objašnjava na sljedeći način: „Korelacija između mjera istih konstrukata dobivenih na temelju različitih metoda [...] znači konvergentnu valjanost“ (str. 376), a divergentnu valjanost na ovaj način: „Korelacije između mjera različitih konstrukata mjerenih istom metodom [...] znače varijancu metode i poželjno je da budu niske jer odražavaju divergentnu valjanost, kao i one koje ne dijele ni metodu ni konstrukt“ (str. 376).

Nadalje, konstruktna valjanost u sebi obuhvaća i **unutarnju valjanost**, u kojoj se teoretizira postojanje sličnih (ili, u alternativnom slučaju, različitih) konstrukata i ispituju podatci s nekoliko mjerenja kako bi se istražili očekivani obrasci korelacija (Angoff, 1988), odnosno vjerojatnost kojom dobiveni rezultat možemo pripisati djelovanju varijabla koje se manipuliraju ili kojima se koristi za predviđanje tog ishoda (Kamenov, 2017). Milas (2009) unutarnju valjanost definira kao „stupanj u kojem je pojedinim eksperimentalnim nacrtom promjene izučavanog ponašanja moguće pripisati manipulaciji nezavisnom varijablom“ (str.

131) te navodi da odgovara na pitanje: „Pod uvjetom da uopće postoji, može li se povezanost između nezavisne i zavisne varijable protumačiti kao uzročno-posljedična?“ (str. 120). Dodaje i da ona „predstavlja 'otpornost' nacrtu na greške pristranosti (sustavnih vanjskih utjecanja na zavisnu varijablu)“ (str. 121). Također, valja napomenuti da je za konstruktivnu valjanost od značajne važnosti faktorska analiza – statistički postupak utvrđivanja određenih psiholoških obilježja (Anastasi, 1976).

2.4.3.4. *Ostale valjanosti*

Valjanost se općenito još može podijeliti na **pozitivističku valjanost**, koja se koristi korelacijskim informacijama između ispita i određenog kriterija, i na **interpretivističku valjanost**, koja se pita što je to što se ispituje i ne oslanja se na pouzdanost kao pozitivistička, već na konsenzus među brojnim dionicima i izvorima dokaza (Davies i Elder, 2005).

Nadalje, Milas (2009) objašnjava i **eksperimentalnu valjanost**, odnosno „stupanj u kojem je neki eksperimentalni nacrt primjeren logički okvir za izvođenje odgovarajućih zaključaka“ (str. 119), te navodi da ima četiri tipa – unutarnju valjanost, vanjsku valjanost, konstruktivnu valjanost i valjanost statističkog zaključka. Za potonju napominje da predstavlja „stupanj u kojem je statističkim postupcima moguće dokazati povezanost (kovarijaciju) nezavisne i zavisne varijable“ (str. 120) te da to dokazivanje „proizlazi iz primijenjenog nacrtu i korištene metodologije“ (str. 121). Valjanost statističkog zaključka također je odraz „neosjetljivosti na nesustavne utjecaje koje se primjerenom brigom oko eksperimentalnog postupka, uzorka i statističke analize, nastoji u najvećoj mjeri isključiti iz konačnih analiza“ (ibid.: 120).

Naposljetku, u literaturi se spominje i **posljedična valjanost**, koju je uveo Messick (1989), a koja nastoji odgovoriti ne samo na pitanje podupiru li moguće ili stvarne društvene posljedice tumačenja i primjene ispita namijenjenu svrhu ispitivanja, već i na pitanje jesu li te posljedice istovremeno u skladu s drugim društvenim vrijednostima – stoga prikladnost, smislenost i korisnost tumačenja na temelju rezultata ispita ovise i o društvenim posljedicama ispitivanja. Drugim riječima, posljedična valjanost podrazumijeva okupljanje svih postojećih izvora valjanosti pod okrilje konstruktivne valjanosti i dodavanje značajnosti, korisnosti te vrijednosnih i društvenih posljedica za utvrđene sadržajne, usporedne, prediktivne i konstruktivne pristupe (Davies i Elder, 2005). McNamara (2000) posljedičnu valjanost opisuje kao promjene do kojih dolazi zbog uvođenja ispita, a koje imaju učinak na predmet mjerenja ispita u toj mjeri da dovode u pitanje pravednost zaključaka o ispitanicima. Stoga se mora imati

na umu da je povratni učinak određenoga ispita, predstavljen u bilješci 15 u potpoglavlju 2.3.1., dio njegove posljedične valjanosti (Fulcher i Davidson, 2007). Kao primjer posljedične valjanosti McNamara (2000) navodi pripremu za određeni ispit koju si mogu priuštiti samo neki ispitanici, pa nešto što je trebalo biti reforma tog ispita ustvari dovodi do neželjene posljedice da se ispitanici ne mogu razlikovati prema njihovim sposobnostima koje taj ispit ima namjeru mjeriti. I Fulcher (2010) ističe da ispiti imaju posljedičnu valjanost ako se njihovom primjenom promiče društvena pravednost i jača meritokracija. Međutim, posljedična valjanost „zahtijeva daljnja istraživanja o onomu što određeni dionici rade s informacijama koje dobiju o ispitu te o tomu dovode li njihova tumačenja do bolje dijagnoze, boljih odluka i čvrste teorije“ (Read i Chapelle, 2001 u Davies i Elder, 2005: 803).

2.4.3.5. Zaključak

Anastasi je još 1976. godine u svojoj knjizi napomenula da je konstruktna valjanost sveobuhvatan koncept koji uključuje ostale vrste. No iako ih je nazvala vrstama, a ne aspektima, već se tada na valjanost počelo gledati kao na jedinstven koncept, što je u skladu i s Messickovim (1980) objašnjenjem da je konstruktna valjanost jedinstven koncept valjanosti koji objedinjuje razmatranja o kriterijima i sadržaju u zajednički okvir za ispitivanje racionalnih hipoteza o teoretski bitnim odnosima. Messick (ibid.) nastavlja da ono što omogućuje ovo objedinjavanje jest smislenost ili mogućnost tumačenja rezultata ispita, što i predstavlja cilj procesa utvrđivanja valjanosti. Nadalje, McNamara (1996) također napominje da je utvrđivanje konstruktne valjanosti preduvjet za bilo kakvu izradu ispita. I Cronbach (1988) je utvrdio da je ideja o trima vrstama odvojenih, ali sličnih valjanosti zastarjela te da valjanost konstrukta obuhvaća cjelokupnu valjanost i da te tri vrste valjanosti ustvari predstavljaju samo tri aspekta jednog psihometrijskog fenomena (Angoff, 1988). Usto, kako napominje Messick (1980), različite vrste tumačenja rezultata ispita ne zahtijevaju različite vrste valjanosti nego različite vrste dokaza, tj. podatke, odnosno činjenice i argumente koji te činjenice „cementiraju u opravdanje tumačenja rezultata ispita“ (str. 1014). Stoga na valjanost konstrukta treba gledati kao na jedinstven, ali višedimenzionalan koncept kojemu čimbenici koji se odnose na sadržaj i mjerila daju određeni doprinos. Bit ovog jedinstvenog koncepta valjanosti jest neodvojivost primjerenosti, smislenosti i korisnosti zaključaka koji se temelje na rezultatima, čija ujedinjujuća snaga predstavlja empirijski utemeljeno tumačenje konstrukta (Messick, 1988).

Davies i Elder (2005) još napominju kako je trenutačno moderno govoriti ne o višestrukim valjanostima, već o višestrukim izvorima dokaza valjanosti. I Weir (2005) ističe da su potrebne različite vrste dokaza kako bi poduprle bilo koju tvrdnju za valjanost rezultata ispita te da ne postoji jedna valjanost koja je nadređena drugoj, već da valjanost predstavlja međusobno nadopunjujuć pojam. Usto, valja dodati i da se u kasnijim izmjenama i dopunama Standarda za obrazovne ispite i psihološke testove utvrdilo da koja se god klasifikacija koncepata valjanosti primjenjivala, ona ne treba biti povezana s ispitima, već s vrstama dokaza koji su predočeni za moguća tumačenja rezultata ispita u određenu svrhu (Urbina, 2004).

Ovdje svakako valja još dodati i objašnjenje koje su dali Anderson i Banerjee (2002), a koje je u skladu s Messickovim jedinstvenim konceptom valjanosti. Naime, u slučaju da dođe do promjenjivih rezultata na ispitu, to neće naštetiti valjanosti dotičnog ispita jer je narušena njegova pouzdanost, već je bitno „razumjeti je li do tih promjena došlo zbog pogreške – koje se tradicionalno poistovjećuju s izvorima manjka pouzdanosti – ili zbog konstrukata koji se ne trebaju mjeriti“ (str. 102). Zaključuju kazavši sljedeće: „Stoga je razlika između pouzdanosti i valjanosti nebitna u ovom jedinstvenom pogledu na valjanost. Ono što je važno jest objasniti izvore promjenjivosti.“ (ibid.)

U svakom slučaju, prema Daviesu i Elder (2005: 809) „valjanost nije lijepo spremljena u određeni ispit, već leži u značenjima koja se pridaju rezultatima tog ispita. Ta značenja nisu statična nego dinamična i mijenjaju se usporedno sa stavovima i praksama u dotičnom kontekstu primjene ispita.“ Ipak, autori dodaju i kako trenutačna istraživanja nisu dostatna da se svi aspekti valjanosti stave pod jedinstven, zajednički krov kao što to predlaže Messick (1980; 1988). Međutim, jedina istina o valjanosti koja nam preostaje ona je koju u svojim umovima stvaraju dionici određenog ispita pa i oni kojima je zadaća utvrđivanje njegove valjanosti (Davies i Elder, 2005).

2.4.4. Valjanost procjenjivanja

Kao što smo prethodno objasnili u potpoglavlju 2.3.3. – Izrada e-ispita, prvi je korak u bilo kojem postupku procjenjivanja odrediti konstrukt, odnosno ono što se ispituje, kako bi rezultati na ispitu omogućili donošenje zaključaka o, primjerice, ispitanikovim ukupnim jezičnim vještinama govorenja, čitanja, slušanja i pisanja u stranom jeziku (Bachman i Palmer, 1996) ili o uspjehu ispitanika u primjeni stranog jezika u nekom prethodno određenom području, primjerice ostavljanju poruke poslovnom partneru, pisanju sažetka ili iznošenju završnog govora u sudnici (McNamara, 1996). Takav je slučaj i kod jezičnih ispita, neovisno

o tomu provode li se putem interneta ili na papiru. Jezični se ispiti stoga daju kako bi se dobio rezultat koji se tumači kao pokazatelj onoga što ispitanik zna ili može na dotičnom jeziku, a tim se tumačenjima koristi kao temeljem za donošenje odluke o primjeni tog jezičnog ispita (Bachman, 2004).

Kako bi se opravdala određena tumačenja rezultata na ispitu, treba se pružiti dokaz toga da rezultati na ispitu odražavaju područje, odnosno područja jezične sposobnosti ispitanika koja se žele mjeriti, a ta tumačenja o jezičnoj sposobnosti trebaju biti dovoljno općenita kako bi se primijenila i na određeno područje ciljne primjene jezika, a ne samo na kontekst ispita (Bachman i Palmer, 1996). Stoga je utvrđivanje valjanosti konstrukta (eng. *construct validation*) stalan postupak dokazivanja da je određeno tumačenje rezultata na ispitu opravdano, a koji uključuje logičku potporu određenom tumačenju s predstavljenim empirijskim dokazima koji opravdavaju to tumačenje (Bachman i Palmer, 1996; Bachman, 2004). Ovo je u skladu i s Messickovom (1989) definicijom utvrđivanja valjanosti, koje zahtijeva utvrđivanje značajne teorije i empirijskih dokaza te kombinaciju toga dvojega na način kojime će ta valjanost potkrijepiti svrhu dotičnog ispita (Chapelle, Jamieson i Hegelheimer, 2003). I Angoff (1988) navodi da se prilikom utvrđivanja valjanosti konstrukta istražuje konstrukt, što dovodi do stanja stalnog međudjelovanja rezultata ostvarenih na ispitu i teorije u podlozi konstrukta; na taj način teorijski koncept konstrukta koji se ispituje određuje prirodu podataka koji se prikupljaju radi utvrđivanja valjanosti rezultata na ispitu i koji se primjenjuju pri tumačenju rezultata ispitivanja – stoga se podacima do kojih se dolazi provedbom ispita potvrđuje, opovrgava ili izmjenjuje sama teorija.

Što se tiče usporedbe valjanosti ispita na papiru i ispita koji se polažu putem interneta, odnosno na računalima i drugim uređajima, valja imati na umu da se pri prijenosu konvencionalnog ispita na računalo, u teoriji, prenosi i njegova valjanost jer ako su ispiti istovrijedni, tada su im istovrijedne i valjanosti (Green, 1988). Stoga se ni utvrđivanje valjanosti ne razlikuje između procjenjivanja na papiru i e-procjenjivanja. U svakom slučaju, potrebno je uzeti u obzir moguće probleme do kojih dolazi prijenosom ispita na papiru na računalo te probleme utvrđivanja valjanosti novih mjerenja ponašanja ispitanika (ibid.), odnosno moguće probleme utvrđivanja valjanosti do kojih dolazi s razvojem e-ispita, neovisno o tomu jesu li oni istovjetne inačice ispitima na papiru ili ispiti koji se izrađuju zasebno. Kako ističu Isbell i Kremmel (2020: 614), „u određenim aspektima e-ispiti nisu i ne mogu se smatrati jednostavno ispitima na papiru koji se isporučuju digitalnim putem, niti bi trebali biti takvi“, a zadatci koji se isporučuju digitalnim putem trebali bi iskoristiti važnost primjene jezika u

stvarnom svijetu (Oh, 2020). Ovi prethodno najavljeni problemi predstavljeni su u sljedećem potpoglavlju.

2.4.5. Prijetnje valjanosti procjenjivanja

Prema Messicku (1989), dvije su osnovne prijetnje valjanosti tumačenja rezultata ispita – beznačajnost konstrukta (eng. *construct irrelevance*), odnosno one sastavnice koje su uključene u procjenjivanje, a koje nisu povezane sa znanjem, vještinama i sposobnostima koje se namjeravaju mjeriti, te nedovoljna zastupljenost konstrukta (eng. *construct under-representation*), odnosno neuključivanje onih sastavnica koje su bitne za znanje, vještine i sposobnosti koje se namjeravaju mjeriti u procjenjivanju. Međutim, ako se usredotočimo na valjanost one vrste ispita kojima se bavimo u svojem istraživanju – e-ispita – mogu se izdvojiti sljedeći problemi koji mogu imati učinak na njihovu valjanost: nesvjesno propuštanje čestica, nemogućnost povrata na već odgovorenu česticu, ograničena veličina ekrana, ograničena grafička svojstva određenih uređaja, brzina odgovaranja, vremensko ograničenje i dr. (Green, 1988). Također, neki su ispitni centri doveli u upitnost valjanost e-ispita koji se polažu kod kuće zbog mogućnosti da se do rezultata došlo na potpuno ili djelomično nedopušten način, a polaganje e-ispita kod kuće stvara više prilika i povećava razinu iskušenja za varanjem (Isbell i Kremmel, 2020; Li i sur., 2021). Li i sur. (2021) navode neke od alata za e-nadziranje ispita, kao što su TOP HAT, Examity i Proctortrack, koje nadziru ispitanike putem kamera i snimanja zaslona, zahtijevaju pisanje ispita na čitavom zaslonu, onemogućuju dijeljenje sadržaja i dr., ali koji otvaraju pitanja zadiranja u privatnost ispitanika i nepovjerenje u njihovu poštenost. Stoga postoje i programi, kao što je OpenProctor, koji putem bihevioralne biometrije koja iz podataka koje ispitanik unosi pisanjem utvrđuje njegov stil pisanja i valjanost ispitanika kao autora (ibid.). Uz ove moderne načine, postoje i tradicionalniji, poput nasumična poretka ispitnih čestica, njihova prikazivanja u određenom vremenskom roku te odabir čestica iz velike baze čestica (ibid.).

Nadalje, neki e-ispiti favoriziraju ispitne zadatke koji su pogodni automatskom strojnom bodovanju i uključuju sadržaj općeg jezika koji nije u skladu s, primjerice, akademskim registrom ili određenim komunikacijskim potrebama (Li i sur., 2021). Iako su rezultati na, primjerice, Duolingovu ispitu i Versantu (v. potpoglavlje 2.3.2. – Primjena e-ispita) ukazali na umjerenu do visoku korelaciju s ljudskim procjenjivačima za ispite sa zadacima jezične proizvodnje i s akademski usmjerenim ispitima poput IELTS-a i ispita TOEFL iBT11, to ne znači da su bodovi jednostavno međusobno zamjenjivi, pa se odluke o

primjeni rezultata ispita moraju donositi kritički i s dodatnom potporom, posebice za ispitanike koji su tik iznad i ispod bodovnih pragova, odnosno unutar standardne greške u mjerenju, primjerice u uzimajući u obzir ispitanikov profil, uz ljudsko procjenjivanje govorne i/ili pisane proizvodnje, putem intervjua i dr. (ibid.).

Chapelle i Douglas (2006) prijetnje valjanosti e-procjenjivanja dijele u šest skupina: prvo, različita izvedba na ispitu na računalo u odnosu na druge načine; drugo, nove vrste zadataka osmišljenih za ispite na računalima; treće, ograničenja zbog prilagodljiva odabira zadataka koji može biti neprimjeren ispitaniku i uzrokovati anksioznost; četvrto, netočno automatsko bodovanje odgovora koje ne uzima u obzir svojstva odgovora koja su bitna za konstrukt koji ispit mjeri; peto, kompromitirana sigurnost ispita; šesto, negativne posljedice za učenike, učenje, nastavu i društvo. Napominju da je ovo ipak skeptičan pogled na primjenu tehnologije u jezičnom procjenjivanju koja narušava ustaljene prakse umjesto da nudi inovacije kojima će na revolucionaran način doprinijeti poboljšanju korisnosti procjenjivanja. U svakom slučaju, za naše je istraživanje važno izdvojiti prvu i treću moguću prijetnju koje su naveli – moguć učinak načina provedbe ispita na izvedbu ispitanika, odnosno analizu razlike između njihove izvedbe na e-ispitima i ispitima na papiru, te moguć učinak ispitne anksioznosti na njihovu izvedbu. Njima dodajemo i moguć učinak stavova ispitanika. Stavovi ispitanika i ispitna anksioznost iscrpnije su predstavljeni u sljedećim dvama potpoglavljima.

2.4.5.1. Stavovi ispitanika

Kao što smo prethodno naveli u ovom potpoglavlju o prijetnjama valjanosti procjenjivanja, jedan od čimbenika koji može imati učinak na valjanost procjenjivanja jesu stavovi ispitanika – od onih o samom ispitu do onih o cijelom procesu procjenjivanja. Prvo valja definirati stavove pa stoga počinjemo Allportovom (1971) definicijom stava kao „mentalnog i neuralnog stanja pripravnosti, organiziranog kroz iskustvo, koje izravno ili dinamički utječe na reakciju pojedinca na sve objekte i situacije s kojima je povezan“ (str. 13). Novija definicija navodi da je stav stečena, relativno trajna i stabilna struktura pozitivnih ili negativnih emocija, vrjednovanja i ponašanja prema nekom objektu (osobi, skupini, pojavi, ideji), a čine ga kognitivna sastavnica, tj. znanja i spoznaje o objektu stava, emocionalna sastavnica, tj. čuvstva odnosno osjećaji prema objektu stava, te konativna ili akcijska sastavnica, tj. spremnost na djelovanje odnosno ponašanje prema objektu stava (Eiser, 1986; Petz, 2005; Hrvatska enciklopedija, 2021). Ovim definicijama dodajemo i objašnjene Brunfaut i Clapham (2013) da stavovi kao preduvjeti ili spremnost za ponašanje mogu postojati u raznim

oblicima, od najskrivenijih tragova zaboravljenih navika do impulsa koji odmah izaziva određeno djelovanje.

Usto, sljedeće se pretpostavke podrazumijevaju prilikom primjene pojma stavova: stavovi su subjektivni doživljaji određenog problema ili objekta, uključuju dimenzije vrjednovanja i prosudbe, mogu se izraziti jezikom, u načelu su razumljivi, komuniciraju se drugima, različiti se pojedinci mogu slagati i ne slagati u vezi oko svojih stavova, ljudi različitih stavova prema određenom objektu razlikovat će se prema tomu što smatraju točnim ili netočnim u vezi s dotičnim objektom te, naposljetku, stavovi se predvidljivo odnose na društveno ponašanje (Eiser, 1986). Nadalje, nisu urođeni, već ih se usvaja i može naučiti putem izravnih ili neizravnih društvenih međudjelovanja, zahvaljujući čimbenicima utjecaja poput obitelji, posla, religije, medija, prijatelja i obrazovanja, a kroz učenje stavovi se mijenjaju i razvijaju (Lasagabaster, 2013).

Što se tiče povezanosti stavova s jezičnim učenjem, Brunfaut i Clapham, (2013) dolaze do zaključka da empirijska istraživanja dovode do dvaju gledišta, odnosno do dviju hipoteza – rezultatska hipoteza temelji se na pretpostavci da iskustvo uspjeha ima učinak na stavove o jeziku, zemlji i ljudima, dok se motivacijska hipoteza temelji na uvjerenju da stavovi, kao konstrukti koji su stabilni i slični motivima, odlučuju o načinu na koji će doći do uspješna učenja. Autori dalje ističu da u socijalnoj psihologiji postoje dva temeljna stajališta o stavovima – prvi je bihevioristički pristup, koji stavove vidi kao „odnose između podražaja i reakcije koji su naučeni i koji odgovaraju promatranom ponašanju pojedinca“ (str. 61-62); drugi se odnosi na „intervenirajuću varijablu između podražaja i reakcije, na skriven unutarnji psihološki proces koji se ne može promatrati i koji utječe na ponašanje u obliku spremnosti na djelovanje“ (str. 62).

Nadalje, Gardner (1985) je predstavio tri klasifikacije stavova – prva je dimenzija određenosti i općenitosti (npr. stav prema učenju određenog jezika, koji opisuje određenu aktivnost, te zanimanje za strane jezike općenito, koje je opći konstrukt i koje ne opisuje neku određenu aktivnost, već ih može uključivati mnoge, poput učenja, slušanja i govorenja stranih jezika); druga je klasifikacija u odnosu na njihovu značajnost za postignuće u stranom jeziku (npr. korelacija između varijable stava i varijable postignuća); treća je klasifikacija na obrazovne i društvene stavove (obrazovni su, primjerice, stavovi o nastavniku, tečaju, učenju stranog jezika itd., a društveni uključuju stavove koji se usredotočuju na društvene posljedice usvajanja stranog jezika, poput stavova o zajednici u kojoj se govori određeni strani jezik). Naposljetku, Lasagabaster (2013) ističe kako su stavovi o jezicima povezani s motivacijom za učenjem jezika te da istraživanja, koja su započela u 1950-ima i još uvijek traju, ističu

povezanost stavova i motivacije te njihovu važnost u teorijama i modelima usvajanja drugog jezika (usp. Aliakbari i Gheitasi, 2017).

2.4.5.2. Ispitna anksioznost

Iako se pri izradi ispita u obzir uzimaju sva njegova svojstva, i iako se utvrdi da je on valjan, još je jedan čimbenik koji se treba uzeti u razmatranje pri donošenju odluka o primjeni dotičnog ispita, a to je ispitna anksioznost. Naime, iako su stres, anksioznost²¹ i suočavanje s njima urođena iskustva svojstvena svim ljudima, podražaji iz okoline koji uzorkuju emocije stresa i anksioznosti više ne dolaze (samo) od divljih zvijeri, prirodnih katastrofa i sl. kao što je to bilo u počecima ljudske vrste, već dolaze iz društveno-vrjednujućih situacija u današnjem tehnološkom društvu usmjerenu postignućima (Zeidner, 1998). Budući da su ispiti te razni postupci procjenjivanja jedan od najčešćih načina vrjednovanja nečijeg postignuća, ne čudi da, prema Zeidneru (1998), mnogi učenici mogu ostvariti dobre rezultate na ispitima, ali im je izvedba loša zbog iscrpljujućih razina anksioznosti, što dovodi do ograničenja u njihovu obrazovnom i profesionalnom razvoju.

Stoga se ispitna anksioznost može definirati kao „složeni konstrukt koji uključuje kognitivne, afektivne, fiziološke i ponašajne reakcije na situacije procjene“ (Erceg Jugović i Lauri Korajlija, 2012), a te se reakcije javljaju zajedno sa zabrinutosti o mogućim negativnim posljedicama neuspjeha na ispitu ili u sličnim situacijama vrjednovanja (Zeidner, 1998). Cheng i Zheng (2021) definiraju ju kao situacijsku anksioznost koja se očituje ili aktivira tijekom polaganja ispita, uslijed vremenskog pritiska tijekom ispita, rizika koji dotični ispit nosi ili drugih situacijskih čimbenika (Aliakbari i Gheitasi, 2017). Razlozi zašto ispitanici doživljavaju ispitnu anksioznost različiti su – mogu se odnositi na vrstu ispita (npr. kratki ispiti, ispiti iz matematike i dr.), na same ispitanike (npr. očekivanje neuspjeha zbog prijašnjih iskustava), na kontekst u kojemu se odvija ispitivanje (npr. prilikom procesa zapošljavanja) ili na kombinacije varijabla (Urbina, 2004).

Nadalje, ispitna se anksioznost može javiti kao crta osobnosti ili kao stanje (Alderson i Wall, 1993; Erceg Jugović i Lauri Korajlija, 2012; Aliakbari i Gheitasi, 2017) – anksioznost kao crta osobnosti predstavlja uobičajenu reakciju pojedinca na stres, odnosno predodređenost pojedinca za „određeni doživljaj ili za izvođenje određenog ponašanja“, dok je anksioznost kao

²¹ Anksioznost se definira kao „osjećaj nervoze ili nelagode koji utječe na naše kognitivno funkcioniranje, kontrolu pažnje i mehanizme suočavanja“ (Cheng i Zheng, 2021: 178).

stanje povezano s izvršavanjem zadatka, odnosno ona „odražava stvarno reagiranje pojedinca u specifičnoj situaciji“ (Živčić-Bećirević, 2003: 705). Stoga:

Ispitna anksioznost kao stanje predstavlja prolazni doživljaj anksioznosti koji se javlja kad pojedinac percipira stvarne ili zamišljene podražaje vezane uz situacije ispitivanja kao prijeteće ili opasne te na njih reagira napetošću, aktivacijom autonomnoga živčanog sustava i zabrinutošću. Ispitna anksioznost kao osobina predstavlja relativno stabilnu značajku ličnosti pri čemu se mnoštvo podražaja vezanih uz ispitne situacije percipira kao ozbiljna osobna prijetnja uz sklonost reagiranja intenzivnim anksioznim stanjem. (Erceg Jugović i Lauri Korajlija, 2012: 300)

Usto, Erceg Jugović i Lauri Korajlija (2012) napominju da „što osoba ima izraženiju osobinu ispitne anksioznosti, to se s većom sigurnošću može predvidjeti da će ona u ispitnim situacijama reagirati intenzivnijim stanjem anksioznosti“ (str. 300).

Zeidner (1998) navodi da učenici koji imaju ispitnu anksioznost vide situacije vrjednovanja kao osobnu prijetnju, pa reagiraju smanjenim osjećajima samoučinkovitosti, samoponižavajućim poimanjem, iščekivanjem neuspjeha te izraženijim emocionalnim reakcijama i uzbuđenosti na prvu pomisao neuspjeha, no još uvijek nije jasno zašto ove situacije uzorkuju veću anksioznost kod nekih nego kod drugih. I Juretić (2008) napominje kako je ispitna anksioznost jedan od oblika socijalne anksioznosti, za koju navodi da „se pojavljuje kad ljudi postanu zabrinuti kako ih drugi ljudi opažaju i procjenjuju“ (str. 16), a s kojom dijeli mnoga obilježja, a ponajviše „strah od neuspjeha i brigu oko socijalne evaluacije“ te preokupiranost „mislina koje su katastrofične i samooptužujuće“ (str. 17). Također, dalje ističe sljedeće:

Kada osoba proživljava intenzivnu anksioznost prije i tijekom ispita, ona se u stvari ne boji testa ili čak pada na ispitu. Obično je primarno zabrinuta za interpersonalne implikacije lošeg rezultata na testu i kako će izvedba na testu utjecati na to kako će ju drugi ljudi doživjeti i procijeniti. Za većinu su studenata najvažnija „publika“ profesori, roditelji i vršnjaci. (Juretić, 2008: 17)

Nadalje, kako navode Cassady i Johnson (2002) i Živčić-Bećirević (2003), ispitna se anksioznost sastoji od dviju sastavnica – od emocionalne sastavnice, odnosno emocionalnosti (eng. *emotionality*), i kognitivne sastavnice, odnosno kognitivne ispitne anksioznosti (eng. *cognitive test anxiety*). Prvu opisuju kao subjektivnu svjesnost pojačane autonomne uzbuđenosti pojedinca, a tjelesno se očituje kao promjene na koži, znojenje ruku, suha usta, ubrzan rad srca, kratak dah, vrtoglavica, mučnina, nervoza i osjećaj panike. Za drugu navode da se sastoji od kognitivnih reakcija pojedinca na situacije vrjednovanja ili od unutarnjih

dijaloga u vezi s takvim situacijama prije, tijekom i nakon zadataka vrjednovanja, a očituje se u mislima pojedinca koji uspoređuje svoju izvedbu s drugima, razmatra posljedice vlastitog neuspjeha, ima nisku razinu samopouzdanja u svoju izvedbu, prekomjerno brine o vrjednovanju, zabrinut je zbog tuge koju uzrokuje svojim roditeljima, osjeća se nesprenim za ispit te gubi osjećaj samopoštovanja. Živčić-Bećirević (2003) zaključuje kako „za razliku od kognitivne komponente, čini se da emocionalna komponenta ispitne anksioznosti nema većih štetnih učinaka na akademska postignuća“ (str. 706).

Erceg Jugović i Lauri Korajlija (2012) napominju da su ispitna anksioznost i opće akademsko postignuće negativno povezani. Cassady i Johnson (2002) također navode kako je ispitna anksioznost najčešće povezana s pogoršanjem izvedbe ispitanika te ističu da je ta povezanost značajna kod adolescenata i studenata na visokim učilištima, a visoke razine ispitne anksioznosti u negativnoj su korelaciji s kvocijentom inteligencije, s akademskim postignućem u čitanju, engleskomu kao materinskom jeziku, matematici, prirodnim znanostima, stranim jezicima, psihologiji i mehaničkom znanju, zatim s problemskim zadacima, s memorijom te s ocjenama. Usto, pojedinci s visokom razinom ispitne anksioznosti iskusit će značajna pogoršanja svoje izvedbe u situacijama kad osjećaju nisku razinu samopouzdanja za određeni zadatak, kad smatraju da im ispit predstavlja značajnu prijetnju ili kad su svjesni da nisu dovoljno pripremljeni za ispit (ibid.). I Erceg Jugović i Lauri Korajlija (2012) ističu da „blag intenzitet ispitne anksioznosti može služiti kao motivator. Međutim, ako je visokog intenziteta, može imati sasvim suprotan učinak otežavajući kognitivne procese potrebne za uspješno obavljanje zadataka, posebice onih zahtjevnijih“ (str. 300).

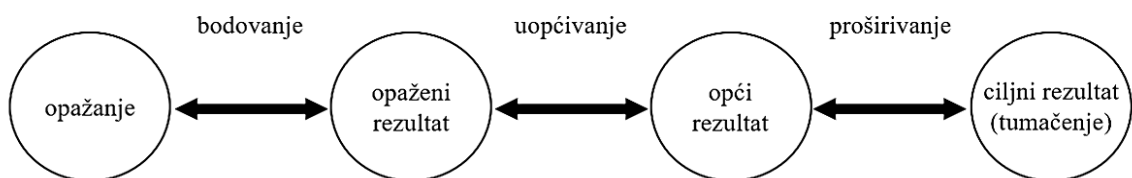
Ipak, unatoč tomu da „brojna istraživanja potvrđuju negativnu povezanost anksioznosti i akademskog uspjeha“ i „da je anksioznost kao stanje negativan prediktor uspjeha na ispitu“ (Živčić-Bećirević, 2003: 705), ispitna anksioznost može imati i drugačiji učinak na ispitanike: prvo, može biti nepovezana s njihovim uspjehom (Živčić-Bećirević, 2003); drugo, prema Zeidneru (1998), umjesto da ograničava njihovu izvedbu, može ju poticati – stoga ne mora nužno biti otežavajuća (eng. *debilitating anxiety*) – ona koja ometa izvedbu na ispitima, već može biti olakšavajuća (eng. *facilitating anxiety*) – ona koja pomaže pri ostvarenju boljih rezultata na ispitima. Te su dvije vrste anksioznosti obrnuto povezane i gotovo da su zrcalne slike u svojem odnosu na izvedbu ispitanika – studenti s visokom razinom iscrpljujuće anksioznosti ostvaruju loše rezultate na studiju, a oni s visokom razinom olakšavajuće anksioznosti ostvaruju relativno dobre rezultate (ibid.). Koja će se od ovih dviju vrsta anksioznosti javiti kod ispitanika ovisi o obilježjima osobnosti (npr.

ekstrovertiranost/introvertiranost, potreba za postignućem, strah od neuspjeha i dr.), kao i o posljedicama određene izvedbe i shvaćanja tih posljedica (Alderson i Wall, 1993).

2.5. POTKRJEPA PROCJENJIVANJA

Uzimajući u obzir prijetnje valjanosti procjenjivanja, prethodno obrađene u potpoglavlju 2.4.5., važno je istaknuti da se valjanost svakako treba potvrditi argumentacijom koja pojašnjava temelje za tumačenje i primjenu rezultata procjenjivanja (Chapelle i Douglas, 2006), a tu je argumentaciju osmislio Bachman (2004; 2005) i nazvao ju **potkrjepom procjenjivanja** (eng. *assessment use argument*). Bachman (2004) je uvidio da postoje mnoga istraživanja valjanosti, ali ne i istraživanja posljedica primjene ispita, pa je stoga ustanovio da postoji potreba za poveznicom između valjanosti i primjene ispita. Također je istaknuo (2005; 2015) da su namijenjene primjene ispita i njihove moguće posljedice svojstvene dionicima, odnosno korisnicima svakog pojedinog procjenjivanja (npr. autorima ispita, ispitanicima, nastavnicima, školama, roditeljima, vladama, poslodavcima, javnosti i dr.) kojima se moraju moći opravdati tumačenja i primjene procjenjivanja jer se tako postiže odgovornost prema njima. Upravo se taj konceptualni okvir koji vodi kroz postupak opravdanja procjenjivanja naziva potkrjepom procjenjivanja. Ovdje valja dodati da se strategija utvrđivanja valjanosti mora razvijati na način da se uvjerenja i želje dotičnih korisnika ispita uzmu u obzir i analiziraju (Cronbach, 1988) te da autori ispita trebaju oblikovati i razraditi posebnu potkrjepu procjenjivanja za svaku namijenjenu primjenu ispita (Bachman i Palmer, 2010).

Bachman (2005) je potkrjepu procjenjivanja osmislio na temelju potkrjepe tumačenja (eng. *interpretive argument*), koju je predstavio Kane (1992), a prema kojoj je rezultat na ispitu premisa, dok su tumačenja tih rezultata zaključci. Kane, Crooks i Cohen (1999) dalje su razvili ovu teoriju i opisali potkrjepu tumačenja kao slijed zaključaka od promatranja izvedbe na ispitu do tumačenja sposobnosti ili predviđanja buduće izvedbe, što je prikazano na slici 1 dolje. Potkrjepa tumačenja, dakle, navodi razloge uključene u primjenu rezultata za donošenje zaključaka i odluka te se primjenjuje svaki put kad se ti rezultati upotrebljavaju (Kane, 2011).

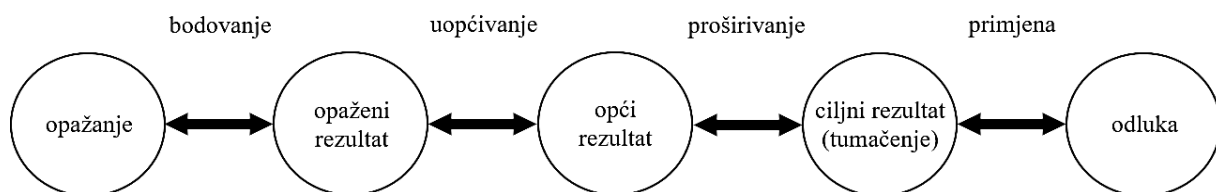


Slika 1. Veze u potkrjepi tumačenja

(Bachman, 2005, prilagođeno prema Kaneu, Crooksu i Cohenu, 1999)

Kako objašnjava Bachman (2004; 2005), prva veza u potkrjepi tumačenja jest bodovanje (eng. *scoring*), odnosno zaključak na temelju opažanja izvedbe na ispitu, koje se temelji na pretpostavci o primjerenosti i dosljednosti bodovanja i uvjeta u kojima je postignut taj rezultat; dokaz u prilog ovom zaključku može biti analiza mjerila i postupaka bodovanja te obilježja ispitnih zadataka. Druga je veza uopćivanje (eng. *generalization*), odnosno zaključak na temelju pretpostavke da se opaženi rezultat na ispitu može ostvariti i na drugim zadacima sličnima onima u ovom ispitu. Treća je veza proširivanje (eng. *extrapolation*), odnosno zaključak na temelju dviju pretpostavaka – prva pretpostavlja da je sposobnost koja se od ispitanika traži na ispitu jednaka onoj koja se traži i na zadacima neobuhvaćenima ispitom, a druga pretpostavlja da zadatci u ispitu predstavljaju zadatke u određenom području ciljne primjene jezika neobuhvaćenom ispitom.

Bachman (2005) je potkrjepu procjenjivanja doradio na temelju Kaneove (2001) potkrjepe tumačenja u koju je dodao još jednu vezu – između ciljnog rezultata i odluka koje se na temelju njega donose, a tu je vezu nazvao primjenom (eng. *utilization*), što je prikazano na slici 2 dolje.



Slika 2. Veze u potkrjepi procjenjivanja (Bachman, 2005)

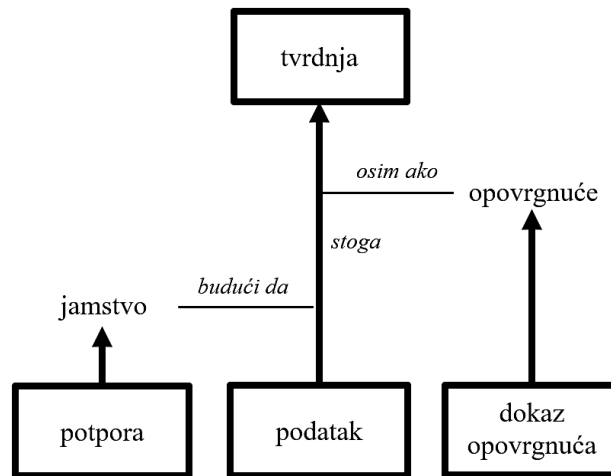
Kane (2001) je dotadašnji dio, koji se odnosi na veze između rezultata i opisnih izjava o ispitanicima, nazvao opisnim dijelom (eng. *descriptive part*), a novi je dio, koji se odnosi na vezu između tih opisnih izjava i odluka koje se donose, nazvao propisnim dijelom (eng. *prescriptive part*). Kasnije je samo prvi dio i dalje nazivao potkrjepom tumačenja – to je dio

koji navodi predložena tumačenja i primjene, a zatim vrjednuje usklađenost ove potkrjepe i vjerodostojnost njezinih zaključaka i pretpostavaka; drugi je dio nazvao potkrjepom valjanosti (eng. *validity argument*) – to je dio koji vrjednuje suvislost potkrjepe tumačenja i prihvatljivost njezinih zaključaka, a uglavnom uključuje iscrpnu analizu i brojna empirijska istraživanja (Kane, 2004, 2011).

Shepard (2016) Kaneovu potkrjepu tumačenja i potkrjepu valjanosti sažima na sljedeći način – za potkrjepu tumačenja kaže da izlaže logički model i suštinske tvrdnje, dok potkrjepa valjanosti podrazumijeva prikupljanje i analizu dokaza kako bi se poduprla vjerodostojnost potkrjepe tumačenja. I sam je Kane (2010) rekao da je svrha potkrjepe tumačenja učiniti tvrdnje i njihove prateće pretpostavke eksplicitnima i stoga dostupnima za provjeru, a da je svrha potkrjepe valjanosti omogućiti ukupno vrjednovanje dokaza za i protiv predloženih tumačenja odnosno primjena, tj. za i protiv potkrjepe tumačenja. Dalje navodi i primjere potkrjepe tumačenja – statistički modeli, psihometrijski modeli, formalne i neformalne teorije o učenju i izvedbi itd. te primjere potkrjepe valjanosti – empirijski dokazi (npr. istraživanja pouzdanosti, odnosa s drugim varijablama i dr.) te prosudbe (npr. kako su se osmislila i primijenila mjerila bodovanja i dr.). Naposljetku, treba imati na umu da „snažni zaključci uglavnom zahtijevaju više dokaza nego slabi zaključci, pa se moraju podastrijeti dokazi potrebni za potporu tvrdnja koje se iznose“ (Kane, 2011: 4).

Bachman (2005) je svoju potkrjepu procjenjivanja dalje razvio na temelju Toulminove (2003) strukture argumenta, koja se sastoji od nekoliko dijelova. Prvi je tvrdnja (eng. *claim*), odnosno izjava o namijenjenim tumačenjima, mjerama koje će se poduzeti, odlukama koje će se donijeti i namijenjenim posljedicama primjene procjenjivanja. Do te se tvrdnje dolazi na temelju određenih podataka (eng. *data*), npr. ostvarenih rezultata na ispitu. Svaka tvrdnja mora biti potkrijepljena jamstvima (eng. *warrant*), odnosno izjavama kojima se opravdava zaključak tvrdnje na temelju podataka. Ta jamstva moraju imati svoju potporu (eng. *backing*), odnosno dokaze koji se prikupljaju radi potkrijepljenja tvrdnje i jamstva. Do tvrdnje se dolazi i tako da se analiziraju moguća opovrgnuća (eng. *rebuttal*), odnosno alternativna objašnjenja ili protutvrdnje namijenjenom tumačenju. Tim opovrgnućima daju se dokazi (eng. *rebuttal data*), odnosno podatci koji određenu protutvrdnju mogu poduprijeti, oslabiti ili odbaciti. Cronbach (1988) ovdje daje dvije usporedbe: navodi da se za utvrđivanje valjanosti treba pripremiti kao za debatu – proučiti temu sa svih kutova pa shvatiti argumente „za“ i „protiv“ kako bi se moglo braniti obje strane, odnosno kao za pravno zastupanje – kad se stranci mora reći koje su jake, a koje slabe točke u njezinu predmetu. Na slici 3 dolje prikazana je struktura potkrjepe procjenjivanja (Bachman, 2005), uz napomenu da može sadržavati više od

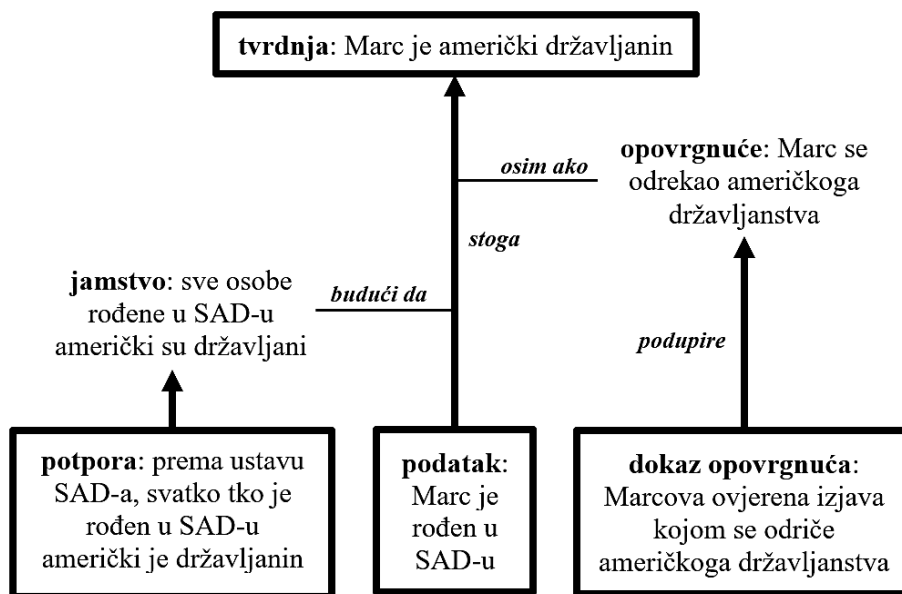
samo jednog podatka te jamstva s pripadajućom potporom i opovrgnuća s odgovarajućim dokazom.



Slika 3. Toulminova struktura argumenta

(Bachman, 2005, prilagođeno prema Mislevyju, Steinberg i Almondu, 2003)

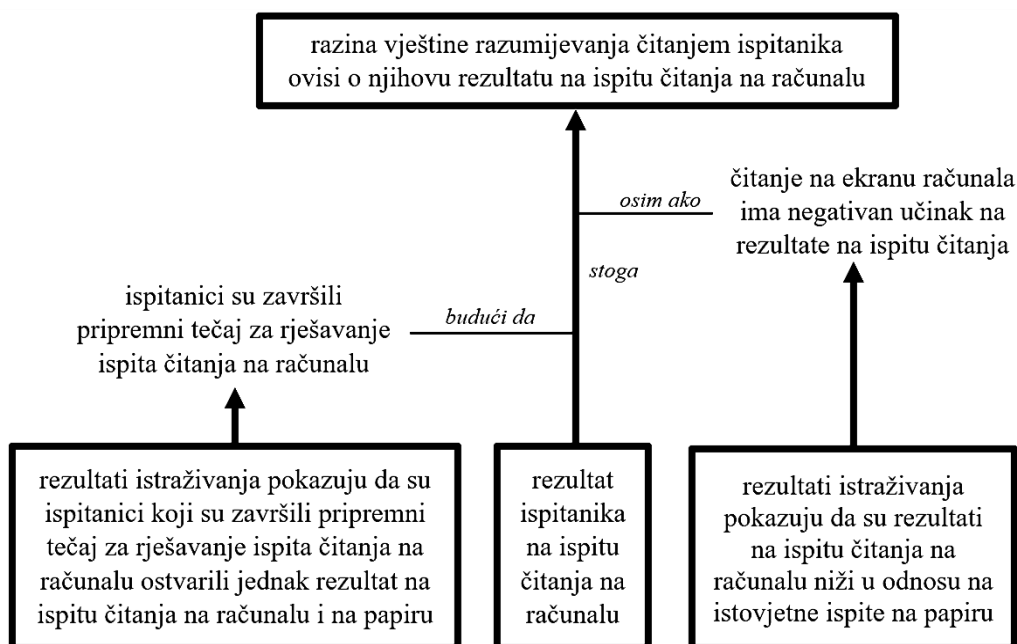
Bachman (2005) navodi i primjer kojime ilustrira Toulminovu strukturu argumenta, a koji je prikazan na slici 4 dolje.



Slika 4. Ilustrativni primjer Toulminove strukture argumenta (Bachman, 2005)

Chapelle i Douglas (2006) napominju da potpore jamstvima i dokazi opovrgnuća predstavljaju empirijske dokaze, pa ako se jamstva ili opovrgnuća samo navode bez potpore

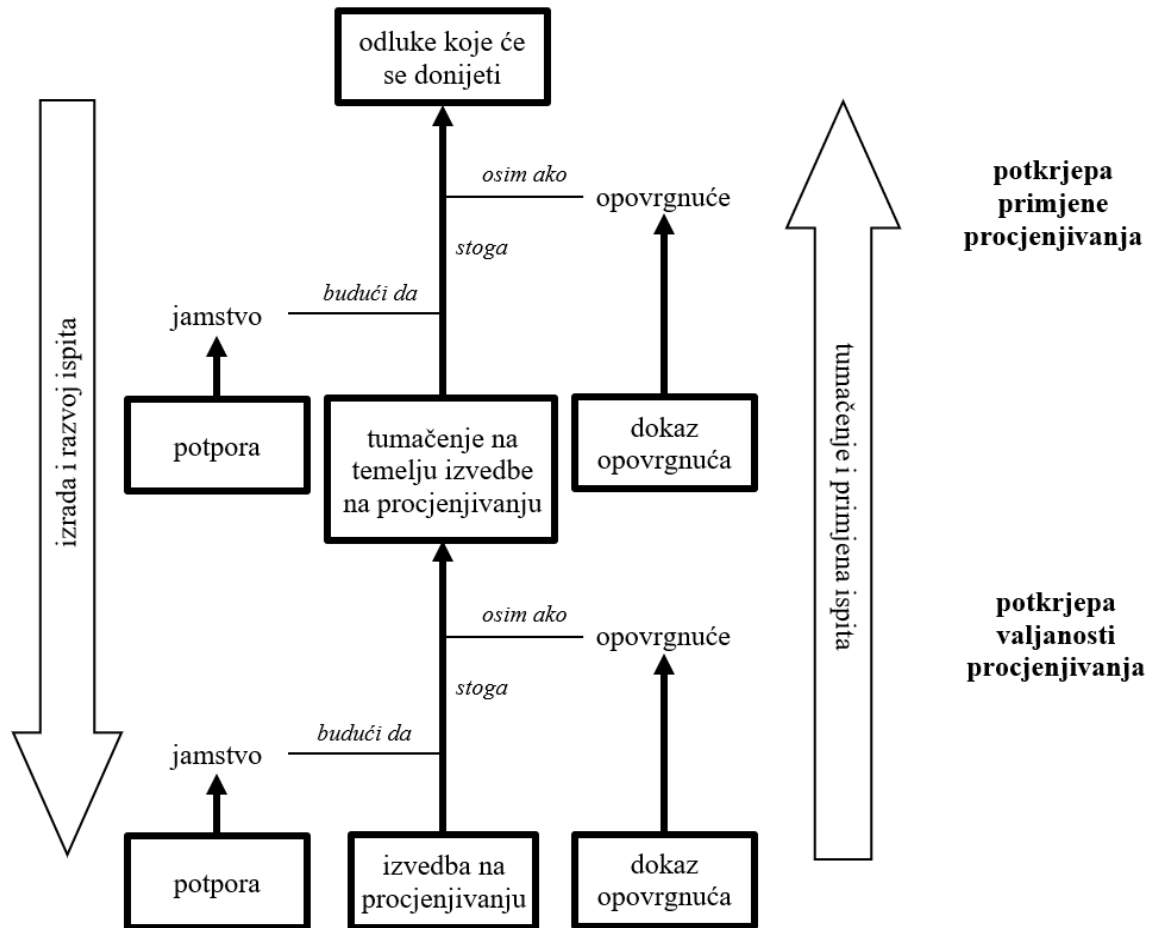
odnosno dokaza koji proizlaze iz istraživanja ili promatranja, ona neće niti potkrijepiti niti oslabiti namijenjena tumačenja. Navode i primjere kojima to ilustriraju te njima podsjećaju da dokazi opovrgnuća predstavljaju empirijske dokaze ranije spomenutih prijetnja valjanosti tumačenja rezultata na ispitu; naglašavaju da što je potpora opovrgnućima jača, to su zaključci o značenju rezultata na ispitu slabiji. Hipoteza mogućih suparnika (eng. *plausible rival hypothesis*) na koju se poziva Cronbach (1988) također pojašnjava ovakvu potkrjepu – prema njoj, nije se dovoljno samo pripremiti za slučajeve gdje naša hipoteza ne odgovara, već i naći alternativno objašnjenje za svoje rezultate te osmisliti istraživanje u kojemu će ta alternativa objašnjenja dovesti do nepodudarnih predviđanja. Na slici 5 dolje prikazan je primjer moguće potkrjepe procjenjivanja koristeći se nekim primjerima Chapelle i Douglasa (2006).



Slika 5. Primjer potkrjepe procjenjivanja (prema Chapelle i Douglasu, 2006)

Polazeći od prethodno spomenute Kaneove podjele potkrjepe na dva dijela (2001; 2004), Bachman (2005) je svoju strukturu potkrjepe procjenjivanja također podijelio na dva dijela – potkrjepu valjanosti procjenjivanja (eng. *assessment validity argument*) i potkrjepu primjene procjenjivanja (eng. *assessment utilization argument*). Potkrjepa valjanosti procjenjivanja povezuje rezultat na ispitu s tumačenjem tih rezultata, a potkrjepa primjene procjenjivanja povezuje tumačenje rezultata na ispitu s odlukama koje se donose na temelju tih tumačenja. Potkrjepe valjanosti i primjene procjenjivanja spojene su u cjelovitu potkrjepu

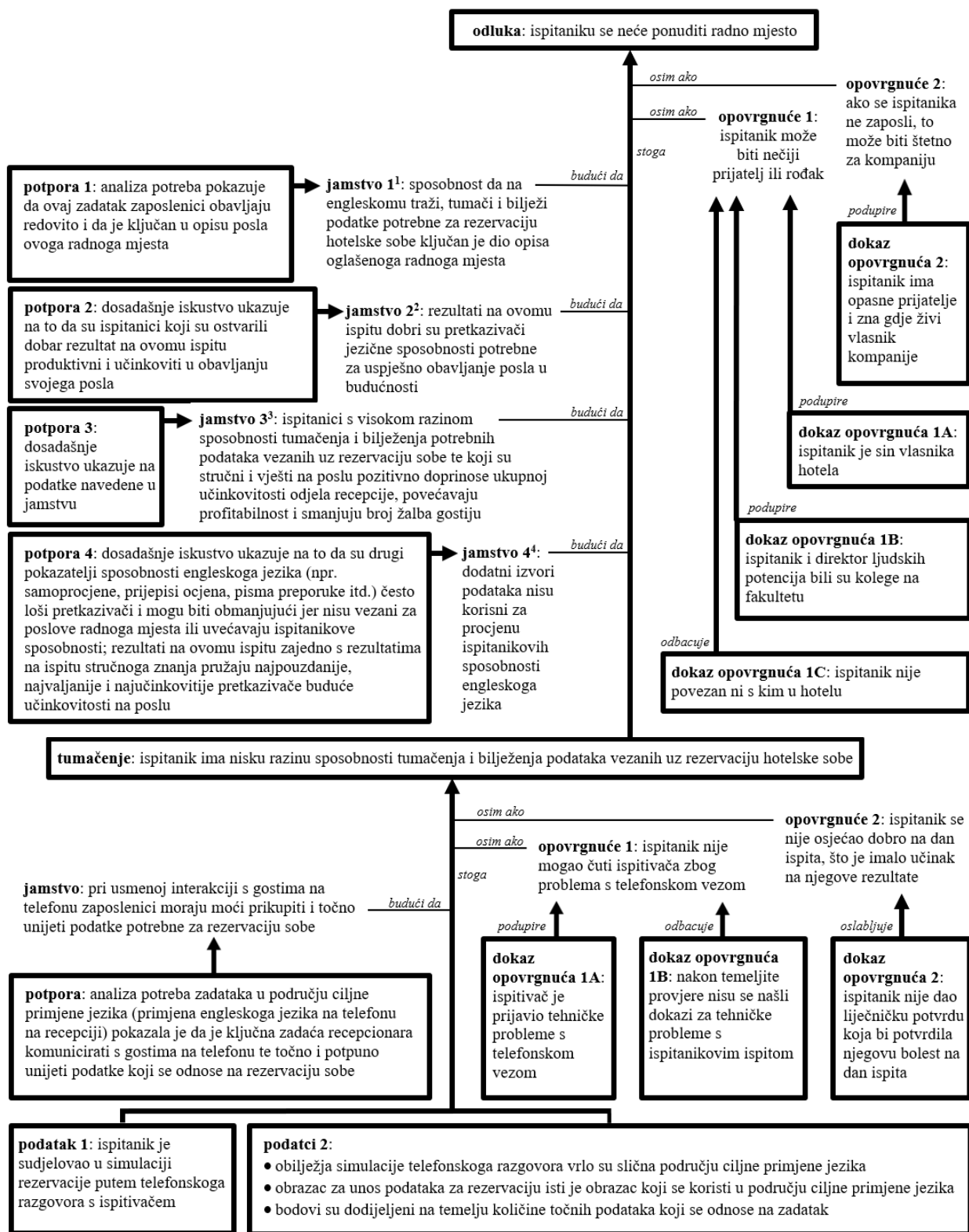
procjenjivanja, koja pruža eksplicitne poveznice od rezultata na ispitu preko njihovih tumačenja do odluka o primjeni ispita. Na slici 6 dolje prikazana je Bachmanova (ibid.) struktura obaju dijelova potkrjepe procjenjivanja.



Slika 6. Struktura potkrjepe procjenjivanja (Bachman, 2005)

Zbog ovog dvostranog pogleda na potkrjepu procjenjivanja, postupak izrade ispita trebao bi voditi autore ispita pri pronalasku načina na koje će njihova izrada dovesti do pozitivnih posljedica procjenjivanja, ali bi i proučavanje pozitivnog učinka ispita trebalo pomoći pri izradi čvrsta konstrukta koji će se mjeriti (Chapelle, Jamieson i Hegelheimer, 2003).

Potkrjepa procjenjivanja ilustrirana je primjerom na slici 7 dolje, koji govori o situaciji u kojoj je međunarodna kompanija tražila nove recepcionare koji mogu u telefonskom razgovoru na engleskom jeziku uzeti podatke od gostiju za rezervaciju sobe; zato je provela simulacije tih razgovora, uz pisani test stručnog znanja i vještina (Bachman, 2005). Bachman na ovom primjeru prikazuje četiri vrste jamstva, označenih eksponentima 1-4, koja potkrjepljuju primjenu procjenjivanja i koja su objašnjena ispod slike.



Slika 7. Primjer potkrjepe procjenjivanja (prema Bachmanu, 2005)

¹ **značajnost** (eng. *relevance*) – tumačenje na temelju rezultata na ispitu značajno je za odluku koja se donosi što se tiče mjere do koje je sposobnost koja se procjenjuje neophodna za

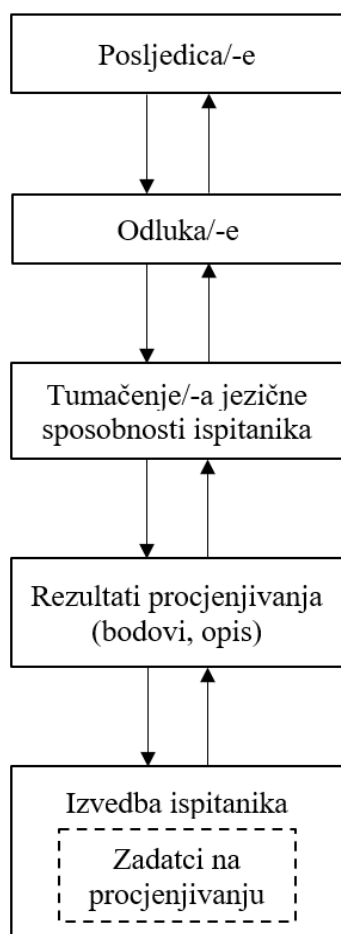
zadatak koji se obavlja u području ciljne primjene jezika i što se tiče stupnja podudarnosti između obilježja ispitnog zadatka i zadatka u ciljnoj primjeni jezika;

² **upotrebljivost** (eng. *utility*) – tumačenje na temelju rezultata na ispitu može se upotrijebiti za odluku koja se donosi, a ta se odluka odnosi na mjeru do koje to tumačenje pruža podatke koji povećavaju vjerojatnost donošenja primjerenih odluka ili izbjegavanja pogrešaka pri odlučivanju;

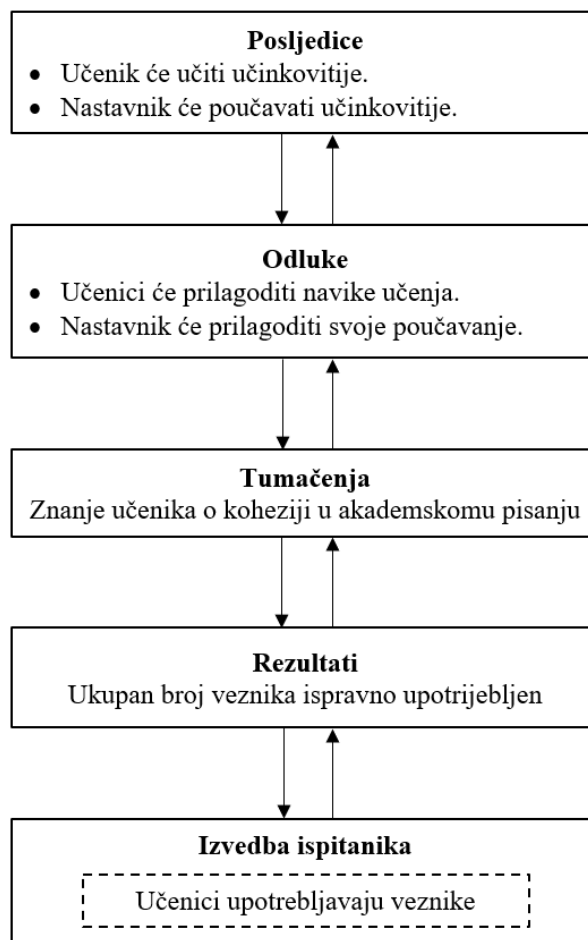
³ **namijenjene posljedice** (eng. *intended consequences*) – posljedice primjene procjenjivanja i donošenje namijenjenih odluka bit će korisne ispitanicima, programu, kompaniji, instituciji, sustavu ili društvu u cjelini;

⁴ **dostatnost** (eng. *sufficiency*) – procjenjivanje pruža dostatne podatke za donošenje odluke ili su uz ispitne rezultate potrebne dodatne informacije.

Bachman (2015) naglašava da su se dotadašnji pristupi analizi valjanosti oslanjali na kvantitativnu tradiciju i nisu uzimali u obzir procjenjivanja koja daju opisne rezultate koji nisu brožani, već verbalni i vizualni, te da se podatci za njih mogu prikupiti kvantitativnom, naturalističkom i dokumentarnom metodologijom. Stoga je nastavio razvijati svoju potkrjepu procjenjivanja (Bachman i Palmer, 2010; Bachman, 2015), koja se i dalje temeljila na Toulminovoj strukturi argumenta (v. sliku 3 gore), vodeći se načelom da je najvažnija primjena jezičnog procjenjivanja prikupljanje podataka kako bi se donijele odluke koje imaju posljedice za dionike tog postupka. Dalje naglašava da nam namijenjene primjene jezičnog procjenjivanja pomažu doći do odluka koje dovode do korisnih posljedica za dionike. Usto, Cronbach (1988) naglašava da pri postupku utvrđivanja valjanosti postoji obveza da se prikladnost posljedica za određene pojedince i institucije analizira te da ih se zaštiti od štetnih posljedica. Slika 8 dolje prikazuje poveznice od izvedbe ispitanika do namijenjenih primjena (tj. odluka odnosno posljedica), a slika 9 daje primjer na zadatku pisane proizvodnje u engleskomu kao drugom jeziku.



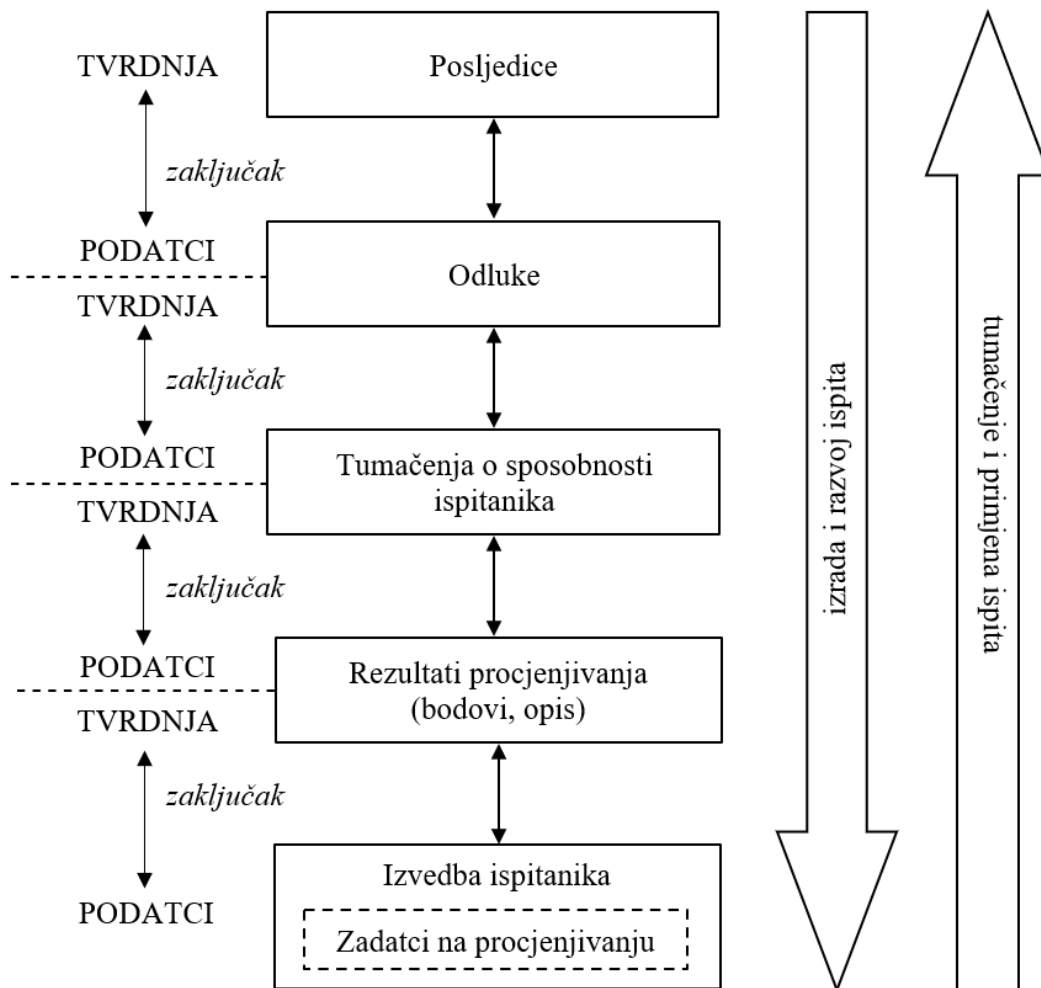
Slika 8. Poveznice između izvedbe ispitanika i posljedica (Bachman i Palmer, 2010)



Slika 9. Poveznice između izvedbe ispitanika i posljedica: primjer zadatka pisanja u engleskomu kao drugom jeziku (Bachman i Palmer, 2010)

Slika 10 dolje prikazuje poveznice između podataka i tvrdnja, odnosno zaključke koji povezuju izvedbu ispitanika s namijenjenim ili stvarnim primjenama procjenjivanja – tumačenjima, odlukama i posljedicama. Tvrdnja do koje se dolazi zaključivanjem na temelju podataka ujedno predstavlja i podatke na temelju kojih se zaključivanjem dolazi do nove tvrdnje. Prema Bachmanu i Palmeru (2010), svaka se tvrdnja sastoji od dvaju dijelova – od ishoda postupka procjenjivanja (eng. *outcome*) i od jednog ili više svojstava tog ishoda (eng. *quality*). Stoga, kao što je prikazano na slici 10 dolje, ako se krene od izvedbe ispitanika na procjenjivanju, on predstavlja podatke, a tvrdnja do koje se dolazi zaključivanjem na temelju tih podataka sastoji se od ishoda, odnosno rezultata procjenjivanja (u obliku bodova, ocjene, opisa, izvještaja i dr.) i od svojstva tog ishoda. U drugom koraku taj rezultat na procjenjivanju predstavlja podatak, a tvrdnja do koje se dolazi zaključivanjem na temelju tih podataka sastoji se od ishoda, odnosno tumačenja i od njegovih svojstava. U trećem koraku ta tumačenja predstavljaju podatke na

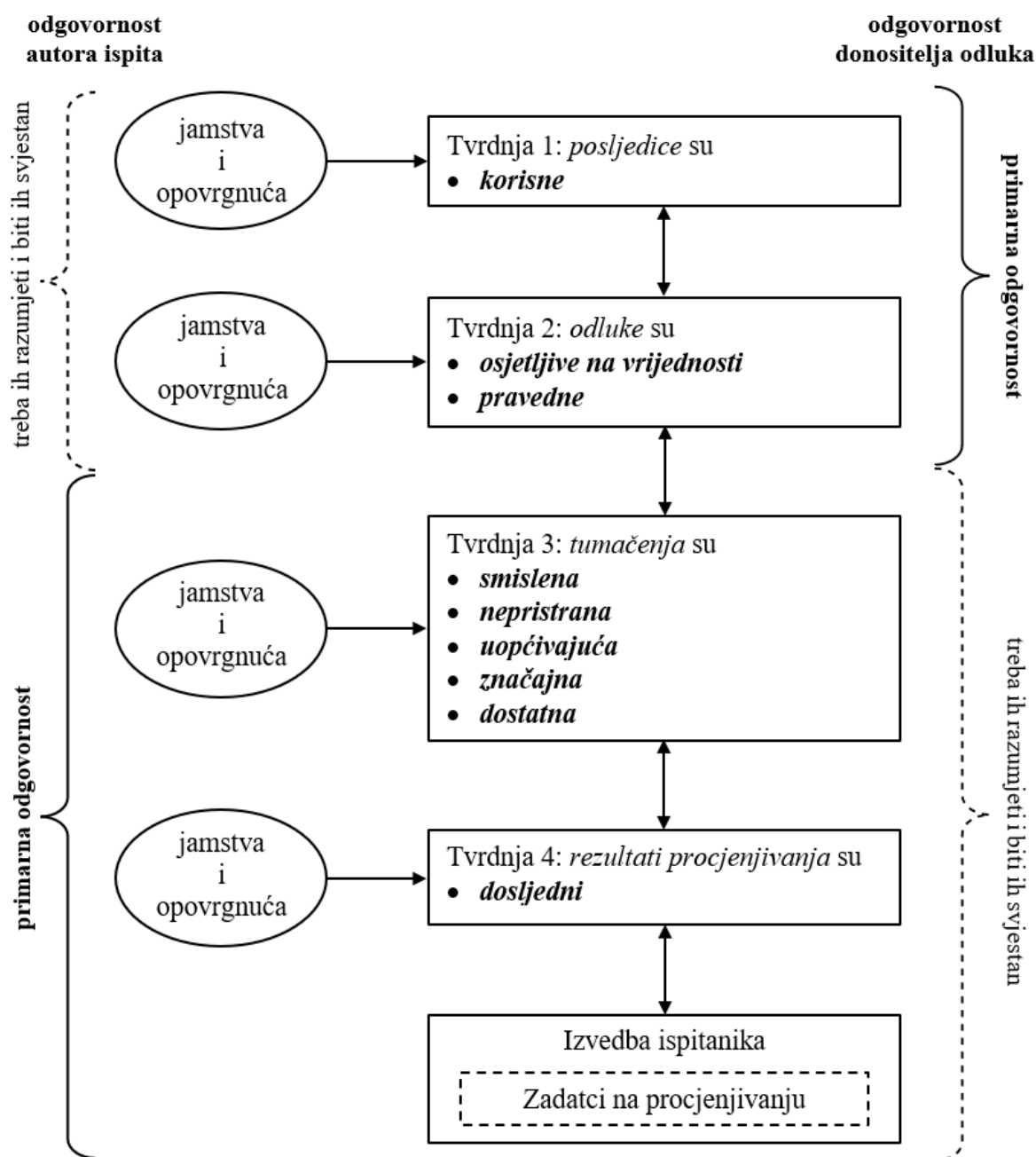
temelju kojih se zaključivanjem dolazi do tvrdnje, koja se sastoji od ishoda, odnosno odluka koje se donose i od njihovih svojstava. U četvrtom koraku te odluke predstavljaju podatke na temelju kojih se zaključivanjem dolazi do tvrdnje, koja se sastoji od ishoda, odnosno posljedica primjene procjenjivanja i donošenja tih odluka te od njegovih svojstava. Naravno, ti se koraci poduzimaju prethodno opisanim redoslijedom ako se radi o tumačenju i primjeni ispita, no njihov je redoslijed obrnut ako se radi o izradi i razvoju ispita.



Slika 10. Poveznice zaključivanja od izvedbe ispitanika do posljedica (prema Bachmanu i Palmeru, 2010; prema Bachmanu, 2015)

Kao što smo prethodno spomenuli, prema Bachmanu (2010) svaka se tvrdnja sastoji od ishoda i njegovih svojstava i ta se svojstva moraju potkrijepiti jamstvima s pripadajućom potporom i opovrgnućima s odgovarajućim dokazima, što čini potkrjepu procjenjivanja. Bachman također navodi da je artikuliranje tvrdnja, jamstava i opovrgnuća jedna skupina aktivnosti u postupku potkrjepe procjenjivanja, dok je druga skupina aktivnosti prikupljanje

potpora za jamstva, odnosno dokaza kojima se opovrgnuća podupiru, oslabljuju ili odbacuju (npr. dokumenti, teorija, prethodna istraživanja, dosadašnje iskustvo, procedure provedbe ispita, mjerila bodovanja, empirijska istraživanja i dr.). Slika 11 dolje (Bachman i Palmer, 2010) prikazuje četiri tvrdnje u potkrjepi procjenjivanja s navedenim ishodom (označenim kurzivom) te njegovim svojstvom, odnosno svojstvima (označenima masnim slovima u kurzivu), zajedno s odgovornostima dvoje dionika procjenjivanja – autora ispita i donositelja odluka.



Slika 11. Tvrdnje i jamstva u potkrjepi procjenjivanja te odgovornosti autora ispita i donositelja odluka (Bachman i Palmer, 2010)

Objašnjenja ovih tvrdnja prema Bachmanu i Palmeru (2010) odnosno Bachmanu (2015) jesu sljedeća:

Tvrdnja 1 – *Posljedice* primjene procjenjivanja i odluka koje se donose **korisne** su dionicima.

Tvrdnja 2 – *Odluke* koje se donose na temelju tumačenja jesu:

- **osjetljive na vrijednosti** zajednice i na bitne zakonske propise;
- **pravedne** za sve dionike na koje te odluke utječu.

Tvrdnja 3 – *Tumačenja* o sposobnosti koja se procjenjuje:

- **smisljena** su u odnosu na određeni nastavni sadržaj, na analizu sposobnosti potrebnih u području ciljne primjene jezika, na opću teoriju jezične sposobnosti ili na bio koju kombinaciju prethodno navedenoga;
- **nepristrana** su u odnosu na skupine ispitanika;
- **uopćivajuća** su na područje ciljne primjene jezika u kojemu će se donijeti odluka;
- **značajna** su za odluku koja se donosi;
- **dostatna** su za donošenje odluke.

Tvrdnja 4 – *Rezultati procjenjivanja* (bodovi, izvješća i dr.) **dosljedni** su u različitim zadacima procjenjivanja, u različitim aspektima postupka procjenjivanja (npr. način, prigoda, ocjenjivači) i u različitim skupinama ispitanika.

Za kraj ovog poglavlja valja još samo napomenuti da iako se pri izradi ispita u obzir uzimaju sva njegova svojstva i iako se utvrdi da je on valjan, u obzir se trebaju uzeti i prijetnje valjanosti procjenjivanja, koje su prethodno predstavljene u potpoglavlju 2.4.5. – Prijetnje valjanosti procjenjivanja.

3. EMPIRIJSKA POLAZIŠTA

Sljedeća predstavljena istraživanja koja su bitna za usporedbu s našime podijeljena su u pet skupina – prva skupina istraživanja bavila se analizom valjanosti e-ispita primjenom metode potkrjepe procjenjivanja, druga je skupina uspoređivala valjanost jezičnih ispita na papiru, na računalu i e-ispita, treća je skupina proučavala povezanost stavova ispitanika o ispitima, e-ispitima i računalima te jezične izvedbe, kao i stavove drugih dionika ispita o e-ispitivanju i e-procjenjivanju, dok je četvrta skupina analizirala moguć učinak ispitne anksioznosti ispitanika na njihovu jezičnu izvedbu i stavove o ispitima.

3.1. PRIMJENA METODE POTKRJEPE PROCJENJIVANJA

Prvu skupinu proučenih istraživanja započinjemo člankom koji daje sustavan pregled onih istraživanja koja su primijenila metodu potkrjepe valjanosti, nakon čega predstavljamo šest nama najznačajnijih istraživanja koja su se poglavito bavila analizom valjanosti e-ispita primjenom Bachmanove metode potkrjepe procjenjivanja, a koja većinom nisu uključena u taj sustavni pregled (osim Llosa, 2008, Oh, 2020 i Park, 2021, koji su u nj bili uključeni kao neobjavljeni doktorski radovi, dok se u našem istraživanju navode kao kasnije objavljeni znanstveni članci). Jedno je od tih šest istraživanja primijenilo teoriju koju su predstavili Bachman i Palmer u svojem radu iz 1996. godine prema kojima je ispit koristan ako uključuje pouzdanost, valjanost konstrukta, prirodnost, poticajnost, utjecajnost i praktičnost, dok se ostalih pet vodilo primjenom potkrjepe procjenjivanja koju je Bachman razrađivao od 2005. godine nadalje.

U prethodno spomenutom sustavnom pregledu istraživanja koja su primijenila metodu potkrjepe valjanosti, Dursun i Li (2021) vodili su se trima istraživačkim pitanjima – koji su trendovi u primjeni potkrjepe valjanosti jezičnog procjenjivanja primijećeni u razdoblju od 2000. do 2018. godine, koji su se zaključci istraživali u tim istraživanjima te koje su metodologije primijenjene za istraživanje svakog od tih zaključaka. Njihovi su izvori uključivali znanstvene časopise *Language Testing*, *Language Assessment Quarterly*, *Assessing Writing*, *Language Testing in Asia*, *Papers in Language Testing and Assessment*, *Applied Linguistics*, *System*, *The Modern Language Journal*, *Language Learning* i *TESOL Quarterly*, zatim bazu podataka *ProQuest Dissertation and Thesis Global* i naposljetku bazu *Research Report Series* koju vodi *Educational Testing Service*, što je nakon pregleda 826 članaka prema ključnim riječima poput valjanosti, utvrđivanja valjanosti i potkrjepe dalo njih 70 koji su se

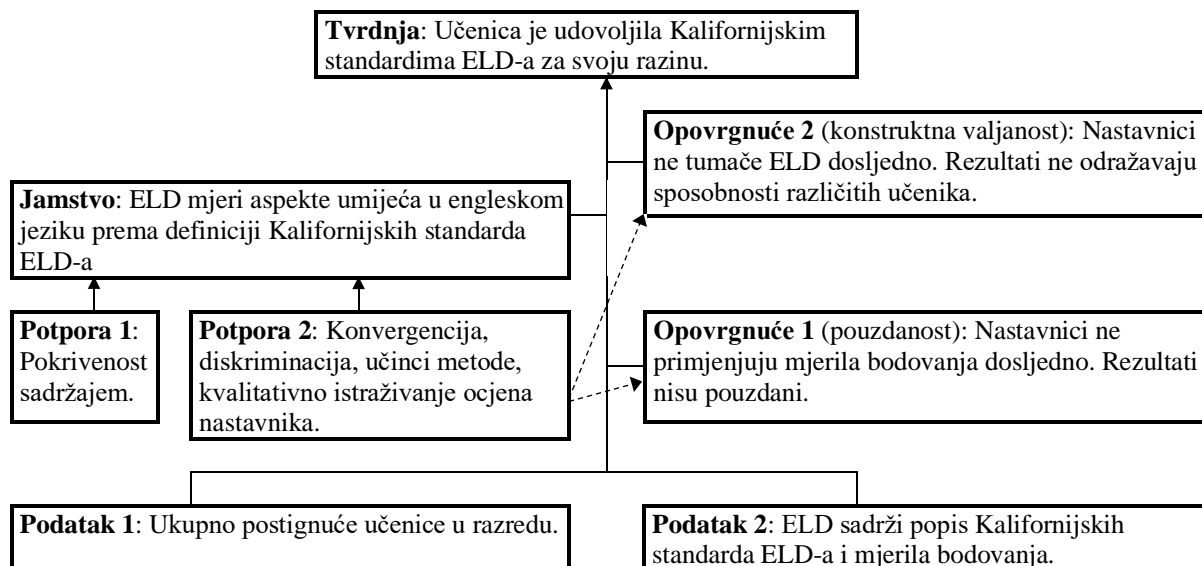
dalje analizirali. Njihovi su rezultati, među ostalim, pokazali da postoji lagan uzlazni trend u ovim istraživanjima od 2005. do 2018. godine, od jednog godišnje (u 2005., 2006., 2007., 2009. i 2012. godini) pa do četrnaest istraživanja godišnje (u 2017. godini). Zatim, 80 % tih istraživanja povezano je s institucijama u Sjevernoj Americi, dok su u Europi njih bila samo dva. Nadalje, najviše su se istraživali standardizirani jezični ispiti poput TOEFL-a (njih 27), a najmanje se istraživalo procjenjivanje u učionici (samo pet istraživanja), dok su instrumenti ostalih istraživanja bili novi ispiti ili ispitni zadatci, lokalni prijemni ispiti te drugi ispiti koje su izradile određene institucije. Što se tiče analize zaključaka u potkrjepama valjanosti ovih istraživanja, neka su uključivala samo jedan zaključak, a neka čak do njih šest. Ti su zaključci ovdje poredani prema učestalosti do kojih su došli istraživači, od najčešće do najrjeđe:

- objašnjavanje, što znači da dokazi pokazuju da bodovi na ispitu odražavaju konstrukt koji se procjenjuje dotičnim ispitom (usp. Barkaoui, 2014; Jun, 2014);
- proširivanje, što znači da su bodovi na ispitu značajni za izvedbu na ne-ispitnim zadacima (usp. Youn, 2013; Gaillard, 2014; Kelly-Riley i Elliot, 2014; Kumazawa i dr., 2016; Riazi, 2016; Harsch i dr., 2017; Santos, 2017; Llosa i Malone, 2018; Oh, 2018);
- uopćivanje, što znači da predstavljeni dokazi pokazuju da bodovi dosljedno odražavaju izvedbu ispitanika (usp. Sawaki i Sinharay, 2013; Chung, 2014);
- bodovanje, što znači da dokazi ukazuju na to da bodovi točno sažimaju značajnu jezičnu izvedbu (usp. Trace i dr., 2017; Li, 2018);
- primjena odnosno posljedice, što znači da se bodovi na ispitu primjenjuju za donošenje odluka i da ove odluke imaju korisne učinke ili posljedice (usp. Santos, 2017);
- definicija područja ispitivanja, što znači da je analiza ciljnog područja jezika povezana s izvedbom na ispitu (usp. Cheng i Sun, 2015; Santos, 2017; Deygers i dr., 2018; Esfandiari i dr., 2018; Checa-García i Guiberson, 2019).

Naposljetku, što se tiče metoda koje su se primijenile radi istraživanja svakog od ovih šest zaključaka, one su bile razne i nisu se primijenile uvijek iste za svaki od njih, a uključivale su analize konstrukta pisanja, silaba i udžbenika (usp. Jun, 2014), analize korpusa ciljnih jezičnih obilježja (usp. Chung, 2014), promatranja u učionici i analizu kurikula (usp. Gaillard, 2014; Jun, 2014), analize potreba i sustavni proces izrade ispitnih zadataka (usp. Park, 2021), analize ispitnih čestica (usp. Kumazawa i dr., 2016), analize razgovora (usp. Youn, 2013), korelacijske analize (usp. Kelly-Riley i Elliot, 2014; Llosa i Malone, 2018), faktorske analize (usp. Sawaki i Sinharay, 2013), Cronbachove vrijednosti α (usp. Chung, 2014), razgovore i upitnike (usp. Santos, 2017) i dr.

Nakon što smo predstavili sustavni pregled istraživanja potkrjepe valjanosti, predstavljamo istraživanje koje su proveli Chapelle, Jamieson i Hegelheimer (2003), koji su, među ostalim, primijenili potkrjepu valjanosti kako bi prikazali pozitivne i negativne značajke primjene ispita engleskoga kao stranog jezika pod nazivom Test Your English. Referirajući se na mjerila koja su izložili Bachman i Palmer (1996), razmotrili su važne aspekte valjanosti dotičnog ispita uključujući i njegove posljedice, odnosno učinak. Razvili su postupak vrjednovanja ispita s dvojakom potkrjepom – pozitivnim i negativnim značajkama ispita – i analizirali ih prema šest svojstava ispita – pouzdanosti, valjanosti konstrukta, prirodnosti, poticajnosti, utjecajnosti i izvedivosti. Također su istražili kakav učinak na izradu ispita imaju posljedice ispita budući da su one predstavljale aspekt potkrjepe valjanosti o kojemu su korisnici ispita (tj. naručitelji izrade ispita i sami ispitanici) najviše znali i koji ih je najviše zanimao. Zaključili su time da su uspjeli utvrditi neke čimbenike koji utječu na način na koji profesionalno znanje motivira odluke koje se tiču izrade i utvrđivanja valjanosti dotičnog ispita te su stoga uvidjeli kako teorija utvrđivanja valjanosti ispita utječe na početne korake izrade ispita.

Sljedećih pet istraživanja primijenilo je Bachmanovu metodu potkrjepe procjenjivanja. Prvo koje izdvajamo jest ono Llose (2008), koje je kao cilj imalo ispitati kvalitetu prosudaba nastavnika u kontekstu standardnog procjenjivanja u razredu koje se primjenjuju za formativno i sumativno odlučivanje visokog rizika o učenicima, pod nazivom English Language Development Classroom Assessment (dalje u tekstu: ELD), koristeći se Bachmanovom (2005) metodom potkrjepe procjenjivanja (slika 12 dolje).



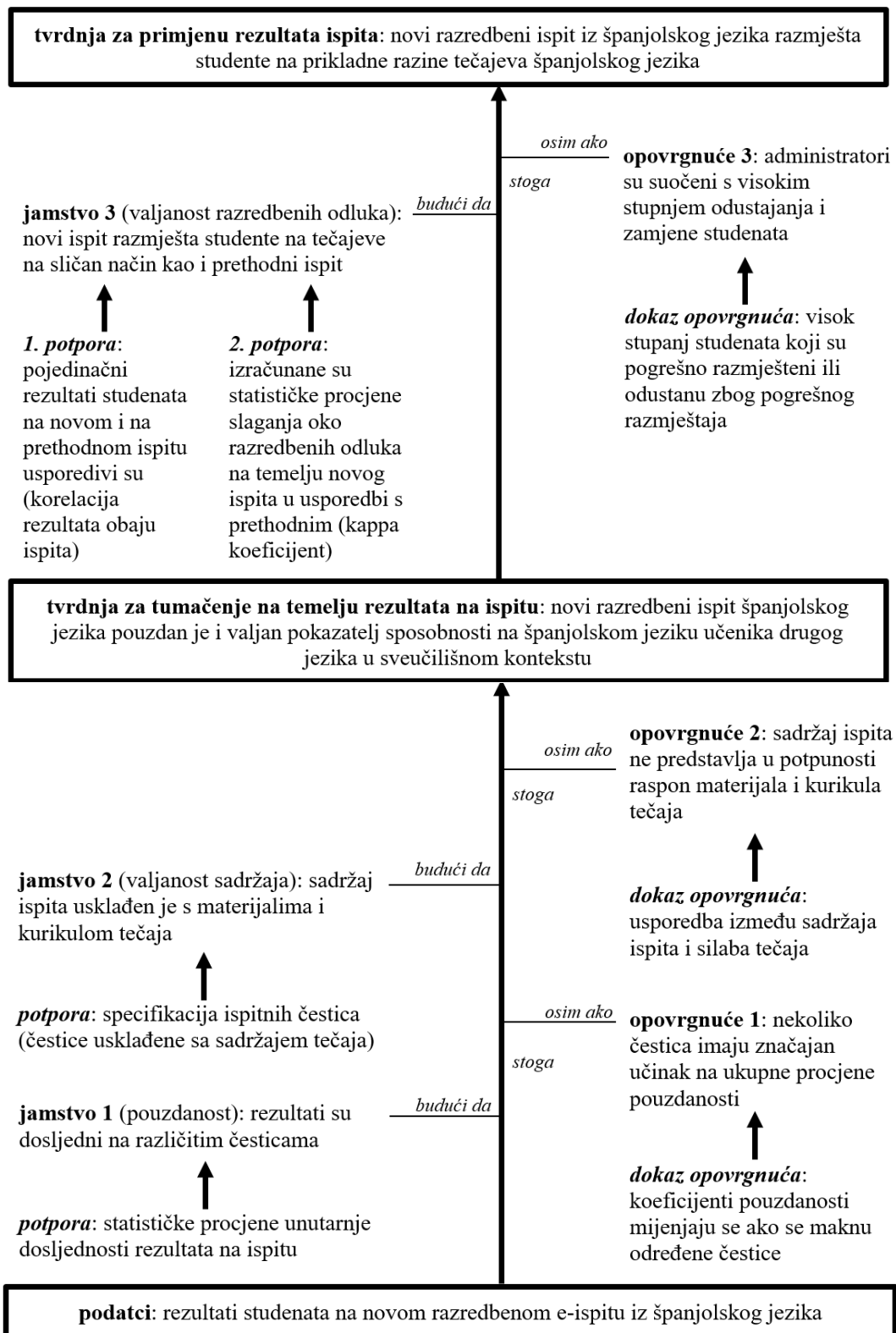
Slika 12. Potkrjepa valjanosti ELD-a (prema Llosi, 2008)

Llosa u ovoj potkrjepi valjanosti ELD-a tvrdi da je ispitanica udovoljila standardima na određenoj razini, a do te tvrdnje dolazi zaključivanjem na temelju jezične izvedbe ispitanice i obilježja postupka procjenjivanja. Kao jamstvo svoje tvrdnje navodi da ELD uistinu mjeri tri aspekta umijeća u engleskom jeziku – slušanje/govorenje, čitanje i pisanje – kao što je to definirano Kalifornijskim standardima ELD-a, a predstavlja i moguća opovrgnuća koja bi mogla oslabiti valjanost namijenjenih zaključivanja – nedosljednu primjenu mjerila bodovanja i nedosljedno tumačenje ELD-a. Potpora njezinu jamstvu dolazi iz postupka izrade ispita (potpora 1) te iz prethodno provedenih kvalitativnih i kvantitativnih analiza (potpora 2), koje su ujedno i dokazi opovrgnuća. U zaključku svojeg istraživanja navodi da dokazi prikupljeni kvantitativnim metodama podupiru sumativnu primjenu bodova, ali ne i primjenu u formativne svrhe. Dalje ističe da rezultati isključivo kvalitativne analize ne bi pokazali da su nastavnici u prosjeku dobri procjenjivači ukupne jezične sposobnosti učenika te da, unatoč varijacijama njihovih tumačenja pojedinačnih standarda, njihova ukupna prosudba može biti vjerodostojna u sumativne svrhe. Također naglašava da je kvalitativno istraživanje pokazalo da ELD ima formativnu ulogu jer je ukazalo na područja u kojima nastavnici mogu napredovati u vezi s procjenjivanjem učenikova ovladavanja pojedinačnim standardima.

Drugo istraživanje koje smatramo značajnim, a koje je također primijenilo metodu potkrjepe procjenjivanja prema Bachmanu, jest istraživanje Wang i sur. (2012), među čijim je autorima i sam Bachman. Istraživači su analizirali Pearsonov ispit akademskog engleskog jezika pod nazivom The Pearson Test of English Academic, odnosno jednu od više njegovih primjena, a ta je primjena za donošenje odluka o prijemu na sveučilište s obzirom na studentovu

pripremljenost za tercijarno obrazovanje u okolini u kojoj se komunicira na engleskom jeziku. Izradili su potkrjepu procjenjivanja koja je uključivala tvrdnje, jamstva s pripadajućom potporom te opovrgnuća s odgovarajućim dokazima, a koju su sačinjavali određeni Pearsonovi dokumenti. U svoju su potkrjepu procjenjivanja uključili četiri tvrdnje – o rezultatima na ispitu, o tumačenju tih rezultata, o odlukama koje se temelje na rezultatima te o posljedicama primjene ispita – i svaku su tvrdnju potkrijepili jamstvima s pripadajućim potporama te opovrgnućima s pripadajućim dokazima. Prva je tvrdnja bila da su rezultati ispita dosljedni. Druga je bila da je ispit smislen (eng. *meaningful*), nepristran (eng. *impartial*), uopćivajuć (eng. *generalizable*) i značajan (eng. *relevant*). Treća je bila da su odluke o prijemu, koje se temelje na rezultatima ispita, osjetljive na lokalne obrazovne i društvene vrijednosti te da su pravedne za sve ispitanike koji jesu ili nisu primljeni na studijske programe za koje su se prijavili. Četvrta je bila da su posljedice primjene ispita korisne ispitanicima i ostalim dionicima koji primjenjuju ovaj ispit, uključujući institucije i organizacije. Naposljetku su zaključili da su za prve dvije tvrdnje prikupili mnogo više dokaza nego za druge dvije, za što navode dva moguća razloga – jedan je to što se prve dvije tvrdnje jedine obrađuju u tradicionalnim pregledima ispita, a drugi je što je potpora za tvrdnje o rezultatima ispita i tumačenju rezultata uglavnom odgovornost autora ispita, dok je potpora za tvrdnje o odlukama i posljedicama ispita uglavnom odgovornost lokalnih donositelja odluka (Bachman i Palmer, 2010). Naposljetku, Wang i sur. (2012) naglašavaju da je odgovornost autora ispita surađivati s korisnicima ispita, odnosno s donositeljima odluka, kako bi se što je jasnije moguće odredilo koja će točno biti primjena dotičnog ispita.

Treće je istraživanje ono koje su proveli Long i sur. (2018), a istraživali su valjanost novih sveučilišnih razredbenih e-ispita iz španjolskog jezika kojima je sveučilište namjeravalo zamijeniti dotadašnje ispite na papiru. Istraživanje je provedeno na uzorku od 2201 novih studenata na prvoj godini koji su riješili e-ispit, a od njih je 1622 riješilo i istovjetnu inačicu na papiru. Istraživačka pitanja koja su ih vodila bila su koliko je razredbeni e-ispit iz španjolskog jezika pouzdan pokazatelj poznavanja španjolskog jezika novih studenata, zatim koliko je taj ispit valjan pokazatelj poznavanja španjolskog jezika u kontekstu kurikula njihova istraživanja te naposljetku koliko se njihov okvir za utvrđivanje valjanosti može uopćiti na druge kontekste stranih jezika na sveučilištu. Njihova je potkrjepa procjenjivanja prikazana na slici 13 niže, a njihovi su rezultati pokazali da je e-ispit pouzdan te da je valjan u smislu sadržaja i odluka o razmještanju na temelju rezultata ispita. Također zaključuju da se temeljna načela njihove metode utvrđivanja valjanosti ispita mogu primijeniti na druge strane jezike u sveučilišnim kontekstima.



Slika 13. Potkrjepa procjenjivanja novim razredbenim e-ispitom iz španj. j. (prema Long i sur., 2018)

U četvrtom istraživanju koje ovdje predstavljamo, Park (2021) analizira, među ostalim, izradu i primjenu ispitnih zadataka za procjenjivanje u području vojnog zrakoplovstva. U istraživanju je sudjelovao 81 muški kontrolor leta vojnog zrakoplovstva u dobi od 19 do 25

godina u jednoj azijskoj državi. Prilikom analize podataka primjenom metode potkrjepe procjenjivanja vodili su se istraživačkim pitanjima koja su dovela do sljedećih zaključaka:

- tvrdnju da se ključne vještine, znanja i procesi potrebni za komunikaciju na engleskom jeziku u kontroli letenja vojnog zrakoplovstva mogu identificirati podupiru značajni dokumenti (točnije, priručnici za osposobljavanje te analiza potreba znanja, vještina i procesa vojnih kontrolora leta);
- tvrdnju da zadatci na procjenjivanju predstavljaju područje engleskoga kao jezika struke, odnosno kontrole letenja, podupire analiza potreba putem upitnika koje su ispunjavali kontrolori leta;
- tvrdnju da se zadatci na procjenjivanju koji zahtijevaju važna znanja, vještine i procese vezane za komunikaciju na engleskom jeziku u području zrakoplovstva mogu simulirati u programu Second Life djelomično podupiru mišljenja dvoje stručnjaka iz njihovih povratnih obavijesti o zadatcima u virtualnom okruženju jer iako su zadatci na procjenjivanju uspješno simulirali prirodne zadatke, dali su i prijedloge za njihovo poboljšanje.

Peto istraživanje ono je Juna (2021), koji je istraživao valjanost ispita kojime se mjeri sposobnost ispitanika da istraže izvore na internetu te da objedine informacije iz tih izvora i njima podupru svoje argumente u pisanom eseju. Od njegovih sedamnaest tvrdnja u potkrjepi valjanosti ovog načina procjenjivanja, izdvajamo njih trinaest koje su bitne za usporedbu s našim istraživanjem. Njegovu prvu tvrdnju, da se ključna znanja, vještine, sposobnosti i procesi u engleskom jeziku potrebni za akademsko pisanje uz korištenje izvorima informacija na sveučilišnim kolegijima mogu identificirati, podupire analiza silaba, udžbenika i usuglašenost stručnjaka tijekom razgovora s nastavnicima. Drugu tvrdnju, da mogući ispitni zadatci predstavljaju područje koje se ispituje, podupire analiza vježba na prvoj godini studija te razgovori s nastavnicima. Treću tvrdnju, da se ispitni zadatci, koji zahtijevaju važne vještine i koji predstavljaju područje koje se ispituje, mogu simulirati i sustavno izraditi, podupire sustavna izrada i oblikovanje ispitnih zadataka i razgovori s nastavnicima. Četvrtu tvrdnju, da su uvjeti provedbe ispita primjereni za pružanje dokaza o ciljanim jezičnim sposobnostima, podupire analiza zadovoljstva ispitanika provedbom ispita. Petu tvrdnju, da su ljestvice za bodovanje eseja prikladne za pružanje dokaza o sposobnosti akademskog pisanja uz korištenje izvorima informacija i da su se primijenile kao što je bilo namijenjeno, djelomično podupiru sustavna izrada ljestvica i razgovori s nastavnicima jer nisu bili prihvaćeni svi njihovi komentari. Šestu tvrdnju, da se nastavnike može obučiti da ne budu pristrani prema određenoj skupini studenata, podupire obuka nastavnika i standardizacija bodovanja. Sedmu tvrdnju, da

je ispitna specifikacija određena tako da se mogu izraditi podudarni zadatci, podupire sustavna izrada ispitne specifikacije i razgovori s nastavnicima. Osmu tvrdnju, da su različita procjenjivanja istog procjenjivača dosljedna, podupire dobar stupanj podudarnosti procjenjivanja jednoga procjenjivača. Devetu tvrdnju, da su procjenjivanja različitih procjenjivača dosljedna, podupire prihvatljiva podudarnost procjenjivača. Desetu tvrdnju, da studenti imaju jednaku mogućnost naučiti ili usvojiti vještinu pisanja uz korištenje izvorima na internetu na kolegijima engleskoga kao drugog jezika, podupire analiza odgovora studenata na pitanja o tomu smatraju li da su imali jednake mogućnosti pripremiti se za ispit. Jedanaestu tvrdnju, da rezultati na ispitu studentima i nastavnicima pružaju korisne i značajne informacije o sposobnostima pisanja studenata te da studenti i nastavnici mogu jasno protumačiti značenje ispitnih rezultata, djelomično podupiru opisnici rezultata, koje se studentima daju zajedno s njihovim rezultatima, a te opisnike mnogi studenti i nastavnici smatraju korisnima, jasnima i laganima za tumačenje, no neki su dali prijedloge koji zahtijevaju daljnju preradu tih opisnika. Dvanaestu tvrdnju, da ispit ima pozitivan učinak na način učenja i poučavanja akademskog pisanja na engleskomu kao drugom jeziku, odnosno da primjena ispita promiče pozitivne povratne učinke, podupiru razgovori s nastavnicima i te razgovori i upitnici na kojima sudjeluju studenti nakon polaganja ispita, i koji pokazuju da nastavnici predviđaju pozitivne povratne učinke primjene ispita na poučavanje, dok su mnogi studenti uključeni u svrsishodne aktivnosti vezane za pripremu za ispit. Trinaesta tvrdnja, da se rezultati pravovremeno objavljuju studentima, podupire dovoljna količina vremena predviđena za procjenjivanje i objavu rezultata studentima.

3.2. JEZIČNI ISPITI NA PAPIRU, NA RAČUNALU I E-ISPITI

Druga skupina proučenih istraživanja bavila se usporedbom valjanosti jezičnih ispita na papiru i ispita na računalu, odnosno e-ispita, a u njoj ćemo predstaviti šest nama najznačajnijih. Prvo je istraživanje Al-Amrija (2007), koji se vodio sljedećim istraživačkim pitanjima: ima li način provedbe ispita učinak na njegovu pouzdanost i valjanost; koliki učinak prethodno poznavanje računala ima na rezultate ispitanika na ispitima na računalima; koliki učinak prethodni stavovi o računalima imaju na rezultat ispitanika na ispitima na računalima; hoće li prethodna sklonost određenom načinu provedbe ispita imati učinak na oba načina ispitivanja te, ako hoće, hoće li ispitanici postići bolje rezultate na ispitu čijem su načinu ispitivanja skloni; hoće li izlaganje ispitivanju na računalu imati učinak na njihovu kasniju sklonost određenom načinu provedbe ispita; hoće li iskustvo s ispitima na računalu imati

pozitivan učinak na stavove prema značajkama ispita na računalu. U istraživanju je sudjelovalo 167 studenata medicine na prvoj godini u Saudijskoj Arabiji, njih 57 % muških i 43 % ženskih. Instrumenti istraživanja bili su sljedeći: TOEFL-ov ispit čitanja, dva zadatka o poznavanju računala, dva upitnika koja su mjerila poznavanje računala, stavove o računalima, sklonost određenom načinu provedbe ispita i stavove prema značajkama ispita na računalu, tri ispita čitanja, polustrukturirani intervjui i protokoli razmišljanja naglas. Rezultati su pokazali da nije bilo značajna učinka načina provedbe ispita na ukupnu pouzdanost i valjanost ispita, da poznavanje računala nije imalo značajan učinak na rezultate na ispitima na računalima te da stavovi o računalima i sklonost određenom načinu provedbe ispita nisu imali značajan učinak na rezultate na ispitima na računalima.

Zatim slijedi istraživanje Hewson i sur. (2007), koji su istraživali moguću ulogu načina provedbe procjenjivanja na jezičnu izvedbu ispitanika na ispitu na papiru i na e-ispitu, zatim ima li povezanosti između njihovih stavova o računalima, koji su uključivali računalnu anksioznost (eng. *computer anxiety*) i povezanost s računalima (eng. *computer engagement*), i izvedbe na e-ispitu te jesu li spol i dob povezani sa stavovima ispitanika o računalima. U istraživanju su sudjelovale tri skupine sveučilišnih studenata psihologije iz UK-a, njih 45 u prvoj (27 ženskih, tri muška i jedan bez odgovora) u dobi od 18 i 55 godina, 49 u drugoj (37 ženskih i osam muških) u dobi od 19 do 55 godina te 21 u trećoj (19 ženskih i 2 muška) u dobi od 21 do 23 godine. Kao instrumenti primijenili su se ispit na papiru i e-ispit s pitanjima višestrukog izbora te upitnik o stavovima o računalima. Rezultati su pokazali da način provedbe procjenjivanja nije imao učinak na njihovu jezičnu izvedbu, a istraživači također nisu došli do značajnih dokaza toga da stavovi o računalima imaju učinak na jezičnu izvedbu na e-procjenjivanju. Ovi su rezultati pružili potporu valjanosti e-procjenjivanja, uz naravno određena ograničenja mogućnosti njihova uopćivanja. Usto, nije ustanovljena povezanost između spola i stavova o računalima, ni između dobi i stavova o računalima.

Treće je istraživanje ono u kojemu su Mohammadi i Barzgaran (2012) pokušali uvidjeti ima li kod iranskih učenika engleskoga kao stranog jezika značajne razlike u rezultatima na ispitu pisanja koje postižu na ispitima na računalu i na ispitima na papiru te koje postižu učenici visoke i niske razine poznavanja računala. U istraživanju je sudjelovalo 80 učenika privatnog jezičnog instituta u Azerbajdžanu u dobi od 15 do 30 godina čiji je materinski jezik većinom bio turski. Instrumenti su bili Cambridgev ispit PET (i to dio slušanja, čitanja i pisanja) te standardizirani upitnik o poznavanju računala. Njihovi su rezultati pokazali da nije bilo statistički značajne razlike između rezultata učenika na ispitu pisanja na računalu i

onih na papiru te da nije bilo statistički značajne razlike između rezultata na ispitu pisanja učenika s visokom razinom poznavanja računala i onih s niskom.

Nakon toga predstavljamo istraživanje Khoshsime i sur. (2017), koji su uspoređivali ispite općeg engleskog jezika na računalu i ispite na papiru te istraživali učinak načina provedbe ispita na rezultate ispita, a također su analizirali stavove ispitanika prema računalima i ispitima na računalu. U istraživanju je sudjelovalo 60 učenika engleskog jezika u privatnoj školi stranih jezika u Iranu u dobi između 15 i 22 godine, a kao instrumenti primijenili su se ispit od 50 pitanja višestrukog izbora na papiru i njegova inačica na računalu te ljestvica stavova o računalima. Njihovi su rezultati pokazali da je bilo značajne razlike između rezultata na ispitu na papiru i na računalu te da su ispitanici bolje rezultate ostvarili na ispitima na papiru. Nadalje, nije uočena značajna korelacija između stavova o računalima i rezultata na ispitu na računalu. Istraživači naposljetku zaključuju da iako su ispitanici postigli bolje rezultate na ispitima na papiru, oblikovali su pozitivne stavove o primjeni računala u ispitivanju nakon što su polagali ispit na računalu.

Peto ovdje izdvojeno istraživanje proveli su Öz i Özturan (2018), koji su se pitali koliko su rezultati ispita na papiru usporedivi s onima na računalu te može li računalna inačica ispita na papiru koji je zadovoljio uvjete pouzdanosti i točnosti mjerenja biti jednako pouzdana i točna. Sudjelovalo je 100 studenata budućih nastavnika engleskog jezika u Turskoj u dobi od 19 do 23 godine. Primijenjeni instrumenti bili su tradicionalni ispit postignuća na papiru, njegova inačica na računalu te upitnik o podacima sudionika i o njihovim stavovima o ispitima na računalu. Rezultati istraživanja pokazali su da su obje inačice ispita imale visoke koeficijente pouzdanosti i visoku unutaraju dosljednost te da ni jedan način provedbe ispita nije imao učinak na rezultate ispitanika. Stoga su istraživači zaključili da je inačica ispita na papiru usporediva s inačicom na računalu.

Naposljetku predstavljamo istraživanje Hewson i Charltona (2019), koji su istraživali valjanost e-procjenjivanja, ali ne uspoređujući ispite na papiru i e-ispite, već analizirajući koju ulogu igraju stavovi o računalima i sklonost određenom načinu procjenjivanja. Nastojali su odgovoriti na tri pitanja – ima li povezanosti između stavova ispitanika o računalima i njihove jezične izvedbe na ispitu na papiru i e-ispitu, je li sklonost ispitanika određenom načinu procjenjivanja povezana s njihovom jezičnom izvedbom na e-ispitu te ima li povezanosti između demografskih varijabla ispitanika (dobi i spola), njihovih stavova o računalima i rezultata njihove jezične izvedbe. U istraživanju je sudjelovalo 89 studenata na magisteriju u UK-u u dobi od 24 do 84 godine, od kojih 63 ženskih i 26 muških. Kao instrumenti primijenili su se interaktivan i automatski bodovan e-ispit gradiva kolegija istraživačkih metoda, upitnik

o stavovima o računalima te upitnik o stavovima o e-procjenjivanju. Rezultati istraživanja pokazali su da ispitanici s manje pozitivnim stavovima o računalima nisu imali lošiju jezičnu izvedbu na e-ispitu te da sklonost ispitanika određenom načinu procjenjivanja nije imala učinak na rezultate njihove jezične izvedbe na e-ispitu. Također, dob i spol ispitanika nisu ukazali na značajnu povezanost s njihovim stavovima o računalima niti s rezultatima njihove jezične izvedbe na ispitu na papiru ni e-ispitu.

3.3. STAVOVI I JEZIČNA IZVEDBA NA ISPITU

Treću skupinu čini pregled istraživanja koja su analizirala stavove dionika procesa jezičnog procjenjivanja. Prvo prikazujemo ona istraživanja u kojima se ispitivao stav ispitanika o jezičnim ispitima, e-ispitima, e-procjenjivanju i računalima te moguć učinak njihovih stavova na izvedbu na jezičnim ispitima. Nakon toga izdvajamo neka istraživanja koja su se bavila analizom stavova drugih dionika, kao što su ispitivači i administratori.

3.3.1. Stavovi ispitanika o ispitima

U trećoj skupini istraživanja izdvajamo njih pet – prvo se bavilo stavovima ispitanika o ispitu na papiru, a druga četiri stavovima o e-ispitu. Prvo ovdje predstavljeno istraživanje proveli su Fan i Ji (2014), koji su istraživali stavove ispitanika prema ispitu Fundan English Test (FET) – ispit umijeća engleskog jezika na sveučilištu. Željeli su utvrditi imaju li stavovi ispitanika o tom ispitu značajan učinak na njihovu izvedbu na njemu. Istraživačka pitanja koja su ih vodila bila su kakav je općenit uzorak stavova ispitanika o FET-u, ima li razlika u stavovima u odnosu na spol i akademsko obrazovanje te koji je odnos između stavova ispitanika o FET-u i njihove izvedbe na ispitu. U istraživanju je sudjelovalo 157 studenata – 85 ženskih i 72 muška – na prvoj i drugoj godini umjetničkog dodiplomskog programa u Kini. Dva su se instrumenta primijenila u ovom istraživanju – strukturirani upitnik i polustrukturirani intervju. Upitnik je imao pet dijelova – stavovi o osmišljavanju ispita, stavovi o podacima o ispitu, stavovi o provedbi ispita, stavovi o ispitu govorenja na računalu te stavovi o povratnom učinku ispita. Rezultati su pokazali da su je općenit uzorak stavova ispitanika o FET-u bio pozitivan, da njihov spol i akademsko obrazovanje nisu imali značajan učinak na njihove stavove te da su samo dva od pet faktora značajno objasnila izvedbu ispitanika na FET-u, a to su bili stavovi o povratnom učinku ispita i o podacima o FET-u.

Što se pak tiče istraživanja stavova ispitanika o e-ispitima, prvo od četiri istraživanja koja ovdje predstavljamo ono je koje je proveo Dizon (2016), koji je u svoje istraživanje uključio 80 japanskih studenata engleskoga kao stranog jezika od prve do četvrte godine sveučilišnih studija u Himejiju u Japanu. Njihove je stavove o e-ispitima istraživao kroz odnos između doživljene korisnosti e-ispita (eng. *perceived usefulness* – PU), doživljene lakoće primjene e-ispita (eng. *perceived ease of use* – PEOU) i bihevioralne namjere primjene e-ispita (eng. *behavioural intention* – BI). Instrumenti njegova istraživanja bili su e-kolokvij i završni e-ispit te upitnik o stavovima o PU-u, o PEOU-u i o BI-u primjene e-ispita. Njegovi rezultati potvrđuju pozitivnu i značajnu povezanost doživljene korisnosti e-ispita i njegove bihevioralne primjene, doživljene lakoće primjene e-ispita i njegove bihevioralne primjene, te doživljene lakoće primjene e-ispita i doživljene korisnosti e-ispita. Usto, utvrdio je i pozitivan stav ispitanika o korisnosti, lakoći primjene i bihevioralne namjere primjene e-ispita.

Drugo istraživanje proveli su Hartshorn i McMurry (2020) sa 153 studenta i 41 nastavnikom na intenzivnom tečaju engleskoga kao drugog jezika u jezičnom centru sveučilišta u Utahu. Među ostalim, ispitivali su učinak pandemije COVID-19 na razine stresa i e-nastavu, a kvantitativni su rezultati pokazali da su studenti naveli da je pandemija COVID-19 imala donekle negativan učinak na njihovu razinu stresa, donekle negativan učinak na učenje engleskog jezika te uglavnom negativan učinak na prelazak na e-nastavu. Nadalje, prije pandemije studenti su nastavu engleskog jezika smatrali vrlo važnom do izuzetno važnom, dok su se nakon pandemije prioritete pomaknuli od stava o nastavi engleskoga kao izuzetno važne do vrlo važne. Zanimljiv je i rezultat koji pokazuje da se 15 % stresa koji su studenti doživjeli pripisuje potrebi učenja engleskog jezika putem interneta, dok varijable poput dobi, spola, bračnog statusa, razine ovladanosti engleskim jezikom, stavova o učinku pandemije na nastavu, odnosno na e-nastavu, engleskog jezika te stavova o važnosti učenja engleskoga prije i poslije pandemije nisu imale statistički značajan učinak na njihovu razinu stresa. Naposljetku, što se tiče učinka pandemije na usavršavanje engleskog jezika studenata, studenti su poboljšali svoju izvedbu u vještinama govorenja i pisanja iako je njihov napredak bio znatno manji kod govorenja nego kod pisanja. Što se tiče kvalitativnih rezultata, oni pokazuju da su, među ostalim, mnogi studenti istaknuli probleme s prilagodbom na primjenu tehnologije u nastavi, neki su napomenuli da im je primjena engleskog jezika postala vrlo teška, posebice zbog manjka prilika za slušanje i govorenje izvan učionice, neki su smatrali da nastavnici nisu bili spremni poučavati putem interneta, neki su istaknuli nemogućnost pohađanja e-nastave zbog drugih ukućana, neki su bili zabrinuti za svoje mentalno zdravlje zbog ograničene ili

nepostojeće interakcije s drugim studentima, nastavnicima i prijateljima, a manji je broj studenata izrazio zabrinutost i za vlastito fizičko zdravlje zbog mogućnosti razbolijevanja.

Treće ovdje predstavljeno istraživanje provelo se na privatnom južnokorejskom sveučilištu s 979 studenata dvanaest različitih fakulteta, a analizirali su se njihovo zadovoljstvo e-kolegijem engleskog jezika, engleskim jezikom na tom kolegiju te e-ispitima koje su na njemu polagali (Chung i Choi, 2021). Rezultati kvantitativne analize do kojih su istraživači došli primjenom upitnika na kojima su sudionici na Likertovoj skali birali vrijednosti od 1 do 5, označavajući u kolikoj se mjeri slažu s ponuđenim tvrdnjama, pokazali su da je većina studenata (74,1 %) bila zadovoljna i vrlo zadovoljna iskustvima na e-kolegiju, da je većina studenata (72,8 %) bila zadovoljna i vrlo zadovoljna količinom engleskog jezika kojime se koristilo na e-kolegiju te da je većina studenata (64,1 %) bila zadovoljna ili vrlo zadovoljna e-ispitima koji su se provodili na tom kolegiju.

Četvrto istraživanje analiziralo je stav sudionika o primjeni e-inačice ispita TOEIC (Test of English for International Communication), a proveli su ga Hoang i sur. (2021). U istraživanju je sudjelovalo 55 studenata koji nisu studirali engleski jezik i koji su svi završili pripremni tečaj za polaganje ispita TOEIC na privatnom sveučilištu u Ho Chi Minhu, u Vijetnamu, a od instrumenata se primijenio upitnik o njihovim stavovima putem Googleovih obrazaca te telefonski intervju o e-ispitivanju i e-procjenjivanju. Rezultati su, među ostalim, pokazali da je većina sudionika smatrala da su im e-ispiti smanjili razinu anksioznosti i brže su dobili rezultate nego što je to slučaj s rezultatima ispita na papiru. Nadalje, većina ih se složila s tvrdnjom da je e-ispit prikladan način provedbe TOEIC-a. Nasuprot tomu, više od polovice sudionika složilo se s tvrdnjom da je znatno teže rukovati e-ispitom nego ispitom na papiru, primjerice zbog količine vremena kojega su morali provesti gledajući u ekran, što im je onemogućilo ostvarivanje boljih rezultata. Nadalje, više od tri četvrtine sudionika smatraju da su mogućnost varanja i moguća nestabilna internetska veza značajni nedostaci e-ispita.

3.3.2. Stavovi ispitanika o računalima

U prethodnim smo potpoglavljima već naveli neka istraživanja koja su se, među ostalim, vodila i ovim istraživačkim pitanjem – Al-Amri (2007), Mohammadi i Barzgaran (2012) te Khoshsima i sur. (2017) proučavali su učinak stavova o računalima na jezičnu izvedbu na ispitima na računalu, dok su Hewson i sur. (2007) te Hewson i Charlton (2019) proučavali učinak stavova o računalima na jezičnu izvedbu na e-ispitima. Ta su istraživanja pokazala da stavovi ispitanika o računalima nemaju značajan učinak na jezičnu izvedbu na

ispitima, a Khoshima i sur. (2017) usto su zaključili da su ispitanici stvorili pozitivne stavove o korištenju računalima prilikom polaganja ispita nakon što su polagali e-ispit.

Još je jedno istraživanje dalo zanimljive rezultate, ono koje je provela Gorsuch (2004). U njemu je nastojala utvrditi hoće li način i uvjeti provedbe ispita te obilježja ispitanika imati učinak na jezičnu izvedbu ispitanika. U istraživanju je sudjelovalo šestoro međunarodnih studenata diplomskog studija sveučilišta u Teksasu u dobi od 24 do 30 godina, a primijenjeni instrumenti bili su ispit slušanja, protokol intervjua, upitnik nakon polaganja ispita te zapisnik opažanja ispitanika prije, tijekom i nakon polaganja ispita. Zaključila je, uz određene komentare o potrebi za daljnjim istraživanjima, da stavovi o računalima i ispitima na računalu ipak mogu imati učinak na rezultate na ispitu na računalu.

3.3.3. Stavovi drugih dionika ispita

Osim ispitanika, određena su istraživanja analizirala i stavove drugih dionika²² ispita, od kojih ih je u ovom potpoglavlju predstavljeno nekoliko bitnih za naše istraživanje.

Prvo prikazujemo istraživanje Winke (2011), koja je ispitivala stavove 267 nastavnika, ravnatelja i administratora u Michiganu (od kojih je 159 bilo nastavnika stranoga jezika, većinom engleskoga kao drugoga jezika) o učinkovitosti ispita The English Language Proficiency Assessment (ELPA) kako bi uvidjela pružaju li njihovi stavovi značajan doprinos širem konceptu valjanosti tog ispita. Upitnik primijenjen u istraživanju uključivao je, uz pitanja o demografskim podacima, četrdeset tvrdnja o ispitu ELPA za koje su ispitanici označivali u kolikoj se mjeri slažu s njima – od 1 do 10 – te je pokraj svake od njih bio ostavljen prostor za komentare, a na kraju je upitnika bilo pet otvorenih pitanja o ispitu. Rezultati kvantitativne analize, među ostalim, pokazali su da se u prosjeku ispitanici najviše nisu slagali s izjavama vezanim za učinkovitu primjenu ispita, zatim s onima vezanima za pojedinačne dijelove ispita – čitanje i pisanje te slušanje. Jedina skupina tvrdnja s kojima se većina ispitanika složila bile su one vezane za učinak ispita na kurikulum i učenje. Nadalje, rezultati kvalitativne analize pokazali su da su ispitanici imali negativne komentare uglavnom o prikladnosti ispita, navodeći da je pretežak i predugačak, dok su komentari o učinku ispita bili i negativni i pozitivni – negativni što se tiče provedbe ispita te njegova učinka na poučavanje i na psihološko stanje učenika, a pozitivni što se tiče učinka na „vidljivost“ učenika engleskog jezika budući da se

²² Pojam dionika u istraživačkom procesu „odnosi se na sve osobe uključene u pripremu i provođenje istraživanja“, a on obuhvaća uloge „istraživača, sudionika, nadležnih osoba institucija u kojima se provodi istraživanje (eng. *gatekeepers*), etičkih povjerenstava, recenzenata i urednika“ (Cergol, 2021:26).

zahvaljujući ELPA-i povećava mogućnost financiranja i obrazovanja učenika u Michiganu. Winke zaključuje time da su stavovi obrazovnih djelatnika, odnosno osoba koje mogu uočiti neplanirane posljedice ispita na poučavanje i učenje, neophodni za utvrđivanje valjanosti ispita te da mogu doprinijeti razumijevanju načina na koji se rezultati ispita trebaju tumačiti i primijeniti. Naposljetku ističe da bi se pri postupku utvrđivanja valjanosti ELPA-e i drugih standardiziranih ispita (umijeća engleskog jezika) trebalo anonimno ispitati obrazovne djelatnike koji ih provode te da bi bilo najbolje da vanjski procjenjivač izradi i provede takvo istraživanje valjanosti, nakon čega bi sažeo i predstavio rezultate javnosti te predložio načine na koje se dotični ispit može poboljšati.

Drugo istraživanje koje ovdje predstavljamo ono je koje su proveli Chung i Choi (2021), u kojemu su analizirali, među ostalim, stavove trinaestero nastavnika engleskoga kao stranog jezika na privatnom južnokorejskom sveučilištu (istraživanje je već navedeno u potpoglavlju 3.3.1. – Stavovi ispitanika o ispitima). Rezultati kvantitativne analize dobiveni na upitniku na kojemu su nastavnici na Likertovoj skali odabirali vrijednosti od 1 do 5 kao bi označili u kolikoj se mjeri slažu s tvrdnjama koje su se odnosile na njihove stavove o e-nastavi i e-ispitima pokazali umjereno visoku vrijednost ($M = 3,9$) ukupnog zadovoljstva e-nastavom, umjereno pozitivnu razinu zadovoljstva dobivenom obukom za e-nastavu ($M = 3,6$) uključenosti studenata u e-nastavu ($M = 3,5$), nisku vrijednost stavova o postignuću njihovih studenata na e-nastavi u usporedbi s tradicionalnom ($M = 2,8$), visoku vrijednost stavova o vremenu potrebnom za pripremu za e-nastavu u odnosu na tradicionalnu ($M = 4,3$) te naposljetku nisku vrijednost zadovoljstva e-ispitima ($M = 2,8$). Rezultati kvalitativne analize dobiveni intervjuima sa sedmero nastavnika pokazali su da su svi smatrali da je planiranje e-nastave i izrada e-ispita od njih zahtijevala značajan dodatan rad i odgovornosti, dok su istovremeno dobivali malo pomoći i obuke, te da je e-procjenjivanje zahtijevalo stalno praćenje studenata i davanje povratnih obavijesti o njihovoj izvedbi kako bi im se olakšalo e-učenje iako su nove metode e-procjenjivanja bile učinkovite i nesumnjivo doprinijele unaprjeđenju i primjeni jezika kod njihovih studenata.

Treće izdvojeno istraživanje proveli su Ghanbari i Nowroozi (2021) na dvadeset iranskih nastavnika engleskog jezika, primijenivši polustrukturirani intervju, a pitali su se s kojim su se preprekama susretali nastavnici tijekom pandemije COVID-19 te koje su strategije suočavanja s tim izazovima primjenjivali. U skladu s njihovim rezultatima jesu i oni iz istraživanja koja su proveli sljedeći istraživači, a koja nisu zasebno prikazana, nego je samo navedeno gdje im se rezultati podudaraju: Hartshorn i McMurry (2020), već navedeno u potpoglavlju 3.3.1. – Stavovi ispitanika o ispitima; Abduh (2021) sa sveučilišnim nastavnicima

engleskoga kao stranog jezika u Saudijskoj Arabiji; Yulianto i Mujtahin (2021) s nastavnicima engleskog jezika u Indoneziji; Zhang i dr. (2021) s nastavnicima engleskoga kao stranog jezika na javnom sveučilištu u Kini; Lučev i dr. (2022) s nastavnicima stranih jezika na visokoškolskim ustanovama u Hrvatskoj. Što se tiče prepreka koje se mogu izdvojiti, one se mogu svrstati u četiri skupine koje su opisane niže.

- **Tehnološke.** Isprva su ove prepreke bile vezane za njihove stavove – primjerice, smatrali su da sustav nije prilagođen e-nastavi u doba pandemije, da tehnologija nije pouzdan alat za procjenjivanje i dr. (Lučev i dr., 2022), a kasnije su bile tehničke prirode – primjerice, problemi s komprimiranjem datoteka i pohranjivanjem u sustav, bojazan od nestanka struje ili internetske veze ili spora internetska veza, nefunkcioniranje opreme studenata, nemogućnost studenata da imaju pristup prikladnim uređajima za e-nastavu i e-ispite i dr. (usp. Abduh, 2021; Yulianto i Mujtahin, 2021; Zhang i dr., 2021; Lučev i dr., 2022). Također su se brinuli oko sigurnosti e-ispita, odnosno varanja studenata, te načina na koji izraditi ispit i izbjeći mogućnost plagiranja (usp. Abduh, 2021; Yulianto i Mujtahin, 2021; Lučev i dr., 2022).
- **Pedagoške.** Ove prepreke uključuju loš mehanizam davanja povratnih obavijesti, zabrinutost u vezi s povratnim učinkom e-procjenjivanja na usvajanje sadržaja kolegija (Lučev i dr., 2022), pojavu novih varijabla u e-procjenjivanju koje utječu na izvedbu studenata na e-ispitima (drugim riječima, učinak na njihovu izvedbu nije imalo samo njihovo znanje), težak zadatak osmišljavanja i ispravljanja e-ispita te nedostatak interaktivna konteksta i vršnjačkog procjenjivanja.
- **Afektivne.** Prepreke koje se svrstavaju u ovu skupinu jesu samoizolacija, bojazan od budućnosti, usklađivanje poslovnog i privatnog života te pripremanje e-ispita (usp. Hartshorn i McMurry, 2020; Abduh, 2021), istovremeno vodeći računa o sadržaju, vremenu i situacijama u kojima se nalaze studenti, što je sve imalo učinak na mentalno zdravlje nastavnika (Hartshorn i McMurry, 2020).
- **Institucijske.** Među ovim preprekama ističu se nepostojanje pravovremenih pravila i smjernica o e-nastavi i e-procjenjivanju na razini sveučilišta (Lučev i dr., 2022), davanje prevelike težine završnim ispitima (drugim riječima, nedovoljna važnost davala se formativnom procjenjivanju), nedostatak usavršavanja za e-nastavu i e-procjenjivanje, izostanak financijske potpore za opremu za e-nastavu (ibid.), neujednačenost između strogih pravila koja su nastavnici morali slijediti tijekom e-nastave i izostanak pravila za nadziranje studenata na e-ispitima, što je imalo učinak na valjanost njihove izvedbe (ibid.).

Strategije suočavanja s prethodno navedenim izazovima također se mogu svrstati u četiri istovjetne skupine, koje su navedene niže.

- **Tehnološke.** Sveučilište je objavilo videe o raznim aspektima e-procjenjivanja kako bi pomoglo nastavnicima, informatička je služba pomagala nastavnicima i studentima kad bi iskrsnuli određeni problemi (Lučev i dr., 2022), nastavnici su isprobali razna tehnološka rješenja za e-procjenjivanje studenata i primijenili ona koja su im bila najprikladnija (usp. Hartshorn i McMurry, 2020; Yulianto i Mujtahin, 2021; Zhang i dr., 2021; Lučev i dr., 2022).
- **Pedagoške.** Strategije u ovoj skupini uključuju snimanje svoje nastave kako bi se uočile moguće greške zbog kojih studenti nisu ispunjavali tražene zadatke, primanje povratnih obavijesti od studenata putem interneta o svojem poučavanju, ograničavanje vremena za rješavanje e-ispita i uskraćivanje mogućnosti studentima da naknadno isprave svoje odgovore, osmišljavanje drugačijih tipova zadataka, primjerice zadataka esejskog tipa, radi smanjivanja mogućnosti varanja (usp. Yulianto i Mujtahin, 2021; Zhang i dr., 2021; Lučev i dr., 2022).
- **Afektivne.** Poruke i planovi koje je sveučilište slalo nastavnicima s vremenom su postajale jasnije, iscrpnije i dosljednije, što je smanjilo razinu stresa kod nastavnika. Druge strategije jesu besplatno e-savjetovanje i e-radionice koje je sveučilište organiziralo za nastavnike i studente kako bi lakše prebrodili probleme nastale zbog ove situacije te tjedni sastanci nastavnika putem interneta gdje su se ponovo mogli čuti sa svojim kolegama.
- **Institucijske.** U ovoj su skupini strategija zahtjevi sveučilišta da nastavnici sustavno pregledavaju objavljene obrazovne smjernice kako bi im se unaprjeđivala kvaliteta, povećanje udjela formativnog procjenjivanja i fleksibilnost dana nastavnicima pri odabiru platforme i oblika tog načina procjenjivanja (Yulianto i Mujtahin, 2021), mogućnost polaganja ispita uz otvorene knjige (Zhang i dr., 2021) te mogućnost ponovnog polaganja ispita zbog tehničkih problema.

Naposljetku predstavljamo istraživanje koje su proveli Nguyen i sur. (2022), a koje se bavilo analizom iskustava administratora i ispitivača u vezi s e-ispitivanjem i e-procjenjivanjem. Istraživače je zanimalo što smatraju da su prednosti, a što mane ovakva načina ispitivanja i procjenjivanja. Istraživanje je provedeno sa 109 sudionika na sveučilištu u Hanoiju, u Vijetnamu, a uključivalo je sveučilišne i srednjoškolske nastavnike engleskog i drugih jezika te administratore ispita. Rezultati prikupljeni putem e-upitnika izrađenog u Googleovim obrascima pokazali su da većina sudionika smatra da e-ispitivanje i e-

procjenjivanje imaju određene prednosti u svim važnijim aspektima, uključujući izvedbu ispitanika te proces provedbe i ocjenjivanja ispita, ali su istovremeno priznali da ovaj način ispitivanja i procjenjivanja nosi sa sobom i određene rizike što se tiče izvedbe ispitanika, ovisnosti o tehnologiji, sigurnosti ispita i procesa ocjenjivanja.

3.4. ISPITNA ANKSIOZNOST I JEZIČNA IZVEDBA NA ISPITU

U petoj skupini istraživanja predstavljamo njih četiri koja su se bavila analizom mogućeg učinka ispitne anksioznosti ispitanika na njihovu jezičnu izvedbu i na njihove stavove o ispitima.

Prvo izdvojeno istraživanje ono je Cassadyja i Johnsona (2002) koji su istraživali pouzdanost i valjanost novog instrumenta – upitnika kognitivne ispitne anksioznosti pod nazivom The Cognitive Test Anxiety Scale – te odnos između kognitivne ispitne anksioznosti i spola, odgađanja (eng. *procrastination*), emocionalnosti i izvedbe ispitanika. U istraživanju je sudjelovalo 168 dodiplomskih studenata psihologije u SAD-u, a instrumenti su bili upitnik kognitivne ispitne anksioznosti, upitnik o odgađanju vezanom za ispite pod nazivom The Test Procrastination Questionnaire, tri ispita iz edukacijske psihologije tijekom njihova studija te njihove ocjene iz ispita SAT (The Scholastic Aptitude Test) – ocjena iz matematičkog dijela, jezičnog dijela i ukupna ocjena. Rezultati istraživanja pokazali su postojanu snažnu korelaciju između izvedbe ispitanika i rezultata na upitniku kognitivne ispitne anksioznosti te slabu ili nepostojanu korelaciju između njihove izvedbe i odugovlačenja. Nadalje, korelacijski koeficijenti između njihovih rezultata na upitniku ispitne anksioznosti i izvedbe nisu ovisili o vremenu ispita, dok je njihovo odugovlačenje koreliralo s izvedbom samo na trećem ispitu, koji se provodio na posljednjem sveučilišnom ispitnom roku. Zatim, ispitanici s visokom razinom ispitne anksioznosti ostvarili su značajno lošije rezultate na jezičnom dijelu ispita SAT od onih s prosječnom i niskom razinom te lošije rezultate na matematičkom dijelu ispita SAT od onih s niskom razinom. Usto, na prvom i trećem od tri ispita iz psihologije ispitanici s visokom razinom ispitne anksioznosti ostvarili su značajno lošije rezultate od onih s prosječnom i niskom razinom, dok su na drugomu ispitanici s visokom i prosječnom razinom ostvarili lošije rezultate od onih s niskom razinom. Što se tiče emocionalnosti, ispitanici s prosječnom razinom emocionalnosti ostvarili su značajno bolje rezultate na prvom i drugom ispitu iz psihologije od onih s visokom razinom, dok na prvom ispitu nije bilo učinka emocionalnosti na njihovu izvedbu. Također, ispitanici s nižom razinom anksioznosti na trima su ispitima iz psihologije dobili bolje ocjene (najviše A i B), a oni s visokom razinom lošije

ocjene (najviše C i D). Na kraju, ispitanice su imale višu razinu emocionalnosti i kognitivne ispitne anksioznosti od ispitanika, ali nije bilo razlike u izvedbi na trima ispitima iz psihologije između ispitanica i ispitanika.

U drugom su istraživanju Chapell i sur. (2005) istraživali odnos između ispitne anksioznosti i prosjeka ocjena 5551 dodiplomskih (63,6 % ženskih i 36,4 % muških) i diplomskih (74,7 % ženskih i 25,3 % muških) studenata triju američkih sveučilišta. Njihovi su rezultati podijeljeni u dva dijela – za dodiplomske i za diplomske studente, kao što je opisano niže.

- Dodiplomske studentice imale su višu razinu ispitne anksioznosti i viši prosjek ocjena od dodiplomskih studenata. Također je uočen značajan negativan odnos između ispitne anksioznosti i prosjeka ocjena – uočene su značajne razlike u prosjeku ocjena kod studentica i studenata s niskom, umjerenom i visokom ispitnom anksioznosti. Naime, studentice s niskom razinom ispitne anksioznosti imale su viši prosjek ocjena od onih s umjerenom i visokom razinom, a one s umjerenom razinom imale su viši prosjek od onih s visokom, dok su studenti s niskom razinom ispitne anksioznosti imali viši prosjek od onih s umjerenom i visokom razinom, ali nije bilo značajne razlike između onih s umjerenom i visokom razinom. Naposljetku, 5 % studentica s najnižom razinom ispitne anksioznosti imale su značajno viši prosjek ocjena od 5 % studentica s najvišom razinom, a jednaki su rezultati bili i kod studenata u tim dvjema skupinama.
- Diplomske studentice imale su značajno višu razinu ispitne anksioznosti i viši prosjek ocjena od diplomskih studenata. Također je uočen značajan negativan odnos između ispitne anksioznosti i prosjeka ocjena – uočene su značajne razlike u prosjeku ocjena kod studentica s niskom, umjerenom i visokom ispitnom anksioznosti. Naime, studentice s niskom razinom ispitne anksioznosti imale su viši prosjek ocjena od onih s umjerenom i visokom razinom, a one s umjerenom razinom imale su viši prosjek od onih s visokom, dok kod studenata nije bilo značajnih razlika u prosjeku ocjena kod onih s niskom, umjerenom i visokom razinom. Naposljetku, 5 % studentica s najnižom razinom ispitne anksioznosti imale su značajno viši prosjek ocjena od 5 % studentica s najvišom razinom, dok kod studenata u tim dvjema skupinama nije uočena značajna razlika.

Treće je istraživanje socijalne i ispitne anksioznosti te percepcije samoefikasnosti kao prediktora ishoda ispitne situacije, koje je provela Juretić (2008) na 78 studenata prvih, drugih i trećih godina studija Filozofskog, Pravnog i Ekonomskog fakulteta Sveučilišta u Rijeci. Rezultati pokazuju „da je socijalno anksioznija osoba također anksioznija u ispitnoj situaciji, ima nižu samoefikasnost, a prilikom polaganja ispita više je zaokupljena negativnim

automatskim mislima“, a usto i da „student s višom anksioznošću u ispitnoj situaciji ima manju samoefikasnost, prilikom polaganja ispita više je zaokupljen negativnim automatskim mislima i mislima koje pokazuju nedostatak motivacije i interesa“ (str. 24). Dalje zaključuje da „porastom ispitne anksioznosti kao crte ličnosti opada vjerojatnost prolaska ispita“ (str. 25).

Četvrto je istraživanje ono koje su proveli Aliakbari i Gheitasi (2017), koji su se pitali koliko su srednjoškolski učenici anksiozni prije ispita iz engleskog jezika, koji je njihov stav o učenju engleskog jezika, ima li povezanosti između njihove anksioznosti vezane za ispit iz engleskog jezika i njihovih stavova prema učenju engleskog jezika te ima li u tomu razlike između muških i ženskih učenika i između učenika različitih smjerova. U istraživanju je sudjelovalo 400 učenika srednje škole u Iranu u dobi od 15 do 19 godina, od kojih 200 muških i 200 ženskih. Instrumenti koji su se primijenili bili su strukturirani upitnik o ispitnoj anksioznosti Westside i upitnik o stavovima o učenju jezika. Rezultati su pokazali da je 60 % ispitanika imalo iznadprosječnu razinu ispitne anksioznosti, no da nije bilo značajne razlike između muških i ženskih ispitanika, niti između učenika različitih smjerova, a smjer također nije imao učinak na stavove ispitanika. Također je uočena pozitivna, ali slaba povezanost između razine ispitne anksioznosti ispitanika i njihovih stavova o učenju engleskog jezika te između dobi ispitanika i razine njihove ispitne anksioznosti, no nije uočena značajna povezanost između njihove dobi i stavova prema učenju engleskog jezika.

4. VALJANOST E-PROCJENJIVANJA U ENGLISKOM U KAO JEZIKU STRUKE

U prethodnim poglavljima teorijskog dijela ovog rada predstavili su se pojmovi prema područjima koja se isprepliću u našem istraživanju, a koja su navedena i u samom njegovu naslovu – Analiza e-procjenjivanja u engleskomu kao jeziku struke. Prvo se opisao engleski kao jezik struke te se prikazao njegov razvoj i trenutna situacija; nakon toga bavilo se problematikom pojma procjenjivanja, navele su se njegove vrste, a zatim se prikazao razvoj jezičnog procjenjivanja u užem smislu i primjena tehnologije u ovom području; sljedeće se objasnio pojam e-ispita, njihove prednosti i nedostaci te primjena, izrada i svojstva; naposljetku su se obradili pojmovi valjanosti i utvrđivanja valjanosti te se prikazala valjanost procjenjivanja i čimbenici koji joj predstavljaju moguće prijetnje. Ti su se pojmovi uvrstili i u glosar (prilog 7), koji nastoji usustaviti prijevod dotičnih pojmova s engleskog jezika i njihovu primjenu, ali i kojime se u hrvatski jezik uvode oni pojmovi koji u njemu dosad nisu postojali u području procjenjivanja.

Na valjanost se više ne gleda kao na skup njezinih različitih vrsta, već kao na jedinstven koncept koji objedinjuje njezine razne aspekte (usp. Anastasi, 1976; Messick, 1980, 1988; Angoff, 1988; Cronbach, 1988; Douglas, 2001; Anderson i Banerjee, 2002; Urbina, 2004; Davies i Elder, 2005; Fulcher i Davidson, 2007; Milas, 2009). Valjanost je također svojstvena namijenjenoj primjeni određenog ispita, pa stoga nikad nije potpuna, već se na nju gleda u smislu stupnja na kojemu prikupljeni podatci podupiru tumačenje rezultata za dotičnu primjenu ispita (usp. Anastasi, 1976; Angoff, 1988; Messick, 1989; Anderson i Banerjee, 2001; Urbina, 2004; Bachman, 2004; Davies i Elder, 2005; Weir, 2005; McNamara i Roever, 2006; Fulcher i Davidson, 2007; Milas, 2009; Fitzpatrick i Clenton, 2010; Green, 2014; Shepard, 2016; Chapelle, 2021; Green i Fulcher, 2021). Utvrđivanje valjanosti podrazumijeva prikupljanje dokaza kojima se opravdavaju tumačenja rezultata koje ispitanici ostvaruju prilikom svoje izvedbe na ispitima (usp. Messick, 1989; McNamara, 2000; Urbina, 2004; Davies i Elder, 2005; Weir, 2005; Fulcher i Davidson, 2007; Llosa, 2008; Green, 2014), a sam se postupak utvrđivanja valjanosti treba voditi namijenjenom primjenom ispita (usp. Read i Chapelle, 2001; Chapelle, Jamieson i Hegelheimer, 2003; Shepard, 2016). U teorijskom dijelu ovog rada uslijedio je prikaz upravo jednog takvog konceptualnog okvira koji se u vidu metode primjenjuje za utvrđivanje valjanosti ispita, a naziva se potkrjepom procjenjivanja (usp. Bachman, 2005, 2015; Bachman i Palmer, 2010). Ovom metodom Bachman (2015) zauzima stav suprotan dotadašnjem prevladavajućem kvantitativnom pristupu analize valjanosti,

ističući da rezultati procjenjivanja nisu nužno samo brojčani, već i opisni, odnosno verbalni i vizualni. Naposljetku, valja istaknuti stajalište Hughesa i Hughesa (2020) koje je bitno za naše istraživanje, a koje glasi da iako za ispite niskog rizika koji će se primjenjivati više puta unutar određene institucije tijekom određenog vremenskog razdoblja utvrđivanje valjanosti nije ključno, ipak je poželjno utvrditi njihovu valjanost barem u manjem opsegu.

Na kraju teorijskog dijela rada dan je pregled nama najbitnijih istraživanja u području analize valjanosti procjenjivanja, i to onih istraživanja koja su primijenila metodu potkrjepe procjenjivanja, onih koja su uspoređivala ispite na papiru s e-ispitima i ispitima na računalima, te onih koja su istraživala moguć učinak stavova i ispitne anksioznosti na jezičnu izvedbu na ispitima. Kao što pokazuje pregled dosadašnjih istraživanja valjanosti e-procjenjivanja primjenom metode potkrjepe procjenjivanja, ovih istraživanja ima samo određen broj, a u hrvatskom kontekstu iznimno su rijetka – u ovom se stoljeću samo nekoliko desetaka njih u nekom obliku bavilo metodom potkrjepe valjanosti, i to velikom većinom u Sjevernoj Americi (usp. Dursun i Li, 2021). Metodu potkrjepe procjenjivanja u svojim je istraživanjima valjanosti ispita primijenio određen broj istraživača (usp. Chappelle, Jamieson i Hegelheimer, 2003; Llosa, 2008; Wang i sur., 2012; Sawaki i Sinharay, 2013; Youn, 2013; Barkaoui, 2014; Chung, 2014; Gaillard, 2014; Jun, 2014; Kelly-Riley i Elliot, 2014; Cheng i Sun, 2015; Kumazawa i dr., 2016; Riazi, 2016; Harsch i dr., 2017; Santos, 2017; Trace i dr., 2017; Deygers i dr., 2018; Esfandiari i dr., 2018; Li, 2018; Llosa i Malone, 2018; Long i sur., 2018; Oh, 2018; Checa-García i Guiberson, 2019; Park, 2021; Jun, 2021), dok u Hrvatskoj ne postoje analize valjanosti ispita primjenom metode potkrjepe procjenjivanja. Nadalje, određeni je broj istraživanja koja su uspoređivala valjanost jezičnih ispita na papiru s ispitima na računalu, odnosno s e-ispitima (usp. Al-Amri, 2007; Hewson i sur., 2007; Mohammadi i Barzgaran, 2012; Khoshshima i sur., 2017; Öz i Özturan, 2018; Hewson i Charlton, 2019) te onih koja su analizirala moguć učinak stavova ispitanika na njihovu izvedbu na jezičnim ispitima (usp. Gorsuch, 2004; Fan i Ji, 2014; Dizon, 2016; Hartshorn i McMurry, 2020; Chung i Choi, 2021; Hoang i sur., 2021), kao i moguć učinak razine ispitne anksioznosti na njihovu izvedbu (usp. Cassady i Johnson, 2002; Chapell i sur., 2005; Juretić, 2008; Aliakbari i Gheitasi, 2017). Iako istraživanja u potonjim dvjema skupinama nisu primijenila metodu potkrjepe procjenjivanja u svojim analizama, bitna su za naše istraživanje jer se njihovi rezultati mogu usporediti s našima prilikom davanja potpore jamstvima i pružanja dokaza opovrgnućima tvrdnja do kojih dolazimo u našoj potkrjepi valjanosti e-procjenjivanja u engleskomu kao jeziku struke (v. potpoglavlje 4.3.4.). Također valja istaknuti i određena istraživanja koja su analizirala stavove drugih dionika procesa procjenjivanja, kao što su ispitivači, nastavnici, ravnatelji i administratori, jer nam je

to, među ostalim, bitno i za raspravu o mogućim glotodidaktičkim primjenama rezultata našeg istraživanja (usp. Winke, 2011; Abduh, 2021; Chung i Choi, 2021; Ghanbari i Nowroozi, 2021; Yulianto i Mujtahin, 2021; Zhang i dr., 2021; Lučev i dr., 2022; Nguyen i sur., 2022).

U sljedećim se poglavljima predstavlja istraživanje valjanosti e-procjenjivanja u engleskomu kao jeziku struke i njegovi rezultati o kojima se na kraju raspravlja u svjetlu prijašnjih i mogućih budućih istraživanja.

4.1. CILJEVI ISTRAŽIVANJA, ISTRAŽIVAČKA PITANJA I HIPOTEZE

Glavni je cilj našeg istraživanja analizirati valjanost e-procjenjivanja u engleskomu kao jeziku struke, a njegovi su pojedinačni ciljevi sljedeći:

1. usporediti rezultate na e-ispitu i ispitu na papiru;
2. ispitati stavove sudionika o e-ispitima, ispitima općenito i računalima;
3. ispitati razinu ispitne anksioznosti sudionika.

Iz prethodno navedenih ciljeva istraživanja izvode se sljedeća istraživačka pitanja (IP):

- IP1 Ispituje li se e-ispitima sposobnost razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u engleskomu kao jeziku struke?
- IP2 Jesu li rezultati na e-ispitu povezani sa stavovima sudionika o e-ispitima, ispitima općenito i računalima?
- IP3 Jesu li rezultati na e-ispitu povezani s razinom ispitne anksioznosti sudionika?

Uzimajući u obzir teorijska i empirijska polazišta ovog istraživanja, našim su se istraživanjem ispitala sljedeće hipoteze (H):

- H1A: E-ispitom ispituje se sposobnost razumijevanja čitanjem u engleskomu kao jeziku struke.
- H1B: E-ispitom ispituje se sposobnost razumijevanja i primjene vokabulara u engleskomu kao jeziku struke.
- H1C: E-ispitom ispituje se sposobnost pisane proizvodnje u engleskomu kao jeziku struke.
- H2A: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o e-ispitima.
- H2B: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o ispitima općenito.
- H2C: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o računalima.
- H3: Rezultati na e-ispitu nisu povezani s razinom ispitne anksioznosti sudionika.

4.2. METODOLOGIJA ISTRAŽIVANJA

4.2.1. Sudionici

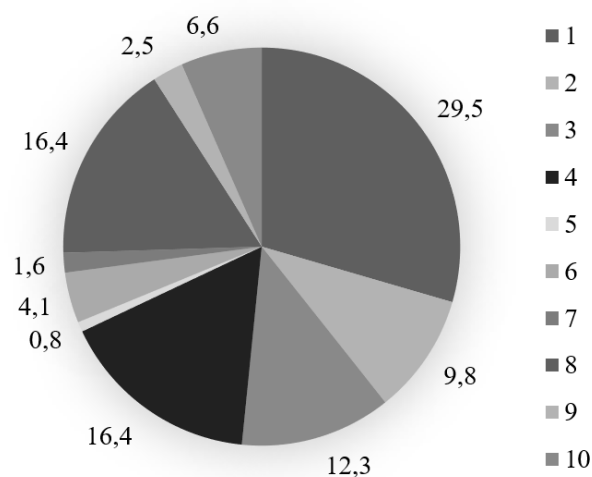
Sudionici²³ u ovom istraživanju bili su studenti prve godine redovnog diplomskog studija ekonomije koji su učili engleski kao jezik struke, točnije poslovni engleski jezik, na razini B1-B2 prema ZEROJ-u. Drugim riječima, uzorak je bio prigodni budući da su ga činili sudionici „u okviru nekoga vida institucionalizirane nastave jezika“ i dobrovoljački²⁴, budući da su ga činili „dobrovoljci koji su se javili na istraživačev oglas“ (Kraš i Miličević, 2015: 33), odnosno u našem slučaju na poziv istraživača za sudjelovanjem u istraživanju.

Prikupljanje podataka od sudionika odvijalo se u dvama koracima, u međusobnom razmaku od tri i pol mjeseca i u razdoblju strogih epidemioloških mjera, pa je od početnog ukupnog broja od 210 sudionika koji su sudjelovali u prvom koraku njih samo 122 sudjelovalo i u drugomu, od kojih su bile 82 žene i 40 muškaraca, odnosno 67,2 % ženskih i 32,8 % muških sudionika. Ovaj ukupan broj sudionika u skladu je s općim okvirnim pravilom „da svaki uzorak u eksperimentalnom istraživanju treba imati najmanje 30 ispitanika budući da je s manjim brojem teško dobiti raspodjelu podataka kakva je neophodna za velik broj statističkih testova“ (Kraš i Miličević, 2015: 34), i u skladu je preporukama za područje obrazovanja „gdje se preporučuje 100 ispitanika za deskriptivna istraživanja, 50 za korelacijska, a 15-30 po skupini za eksperimentalna“ (ibid.).

Nadalje, kao što se vidi na slici 14, sudionici su bili studenti različitih nastavnika, od kojih je nastavnik 1 ujedno autor ovog istraživanja i jezičnog ispita koji se primijenio kao jedan od instrumenata istraživanja. Iz iste je slike također razvidno kako su većina sudionika bili studenti drugih nastavnika.

²³ Koristimo se pojmom *sudionik*, a ne depersonaliziranim pojmovima *subjekt*, *ispitanik* ili *informant*, jer se on odnosi na osobe koje istraživaču pružaju značajne podatke, ali u istraživanju imaju aktivnu ulogu budući da ga oplemenjuju svojim točkama gledišta i na taj način istraživaču pružaju nove uvide u problematiku istraživanja (Cergol, 2021).

²⁴ Dobrovoljački su uzorci nereprezentativni „zbog toga što ne postoji jednaka vjerojatnost javljanja na oglas za svakoga člana populacije [...] Ipak, takvi su uzorci iz praktičnih razloga vrlo česti, između ostaloga i u istraživanjima usvajanja drugoga jezika“ (Kraš i Miličević, 2015: 33).



Slika 14. Zastupljenost (%) sudionika studenata određenog nastavnika (1-10) u uzorku

Uz same sudionike, ovo je istraživanje uključivalo još dionika. Tako je u njemu također sudjelovalo dvoje sustručnjaka, nastavnika engleskoga kao jezika struke, odnosno poslovnog engleskog jezika, koji su procjenjivali zadatak pisane proizvodnje u jezičnom ispitu.

4.2.2. Instrumenti

Instrumenti koji su se primijenili u istraživanju i koji su iscrpnije predstavljeni niže jesu jezični ispit, odnosno ispit Poslovnog engleskog jezika 1 (i to njegova e-inačica te istovjetna inačica na papiru) te strukturirani upitnici za studente – Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika, Upitnik o stavovima o ispitima općenito, Upitnik o stavovima o računalima i Upitnik o ispitnoj anksioznosti Westside.

4.2.2.1. Jezični ispit

Prvi instrument primijenjen u istraživanju bio je nestandardizirani jezični e-ispit Poslovnog engleskog jezika 1 (prilog 1.), koji je sastavio istraživač u skladu sa silabom kolegija Poslovni engleski jezik 1 i svim nastavnim materijalima (udžbenicima, radnim priručnikom, bilješkama s nastave i ostalim materijalima koje nastavnici objavljuju u Googleovoj učionici). Ispit se izradio u skladu s ispitima kakve studenti na ovom kolegiju i inače polažu, s jednakim tipovima zadataka, jednake dužine i jednakih mjerila bodovanja.

Pri izradi ispita vodilo se sljedećim načelima:

- ispit obuhvaća one sposobnosti koje ima namjeru ispitivati;

- sposobnosti koje ispit ispituje predstavljene su i obrađene na nastavi;
- sposobnosti obuhvaćene ovim ispitom ispituju se zadacima istog tipa koji se javljaju u ispitima koje studenti polažu na kraju kolegija Poslovni engleski 1;
- sudionici imaju pristup silabu kolegija i svim nastavnim materijalima;
- provedba ispita ne traje duže od sat vremena.

Izradile su se dvije inačice ispita – e-ispit u Googleovim obrascima i tiskani ispit na papiru pri čemu su vrsta, sadržaj i oblikovanje zadataka ispita na papiru u najvećoj mogućoj mjeri istovjetni onima u e-ispitu.

Ispit se sastojao od triju dijelova: razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje. Prva dva dijela ukupno su nosila dvadeset bodova, a treći dio isto dvadeset, što ukupno daje četrdeset mogućih bodova koje su sudionici mogli ostvariti na cijelom jezičnom ispit. Slijedi iscrpniji prikaz dijelova ispita.

1. **Razumijevanje čitanjem.** Ovaj se dio ispita sastojao od jednog zadatka zatvorenog tipa, odnosno zadatka razumijevanja čitanjem u kojemu je sudionicima dan tekst od 499 riječi na jednu od tema obrađenih na nastavi – u ovom slučaju, temu oglašavanja. Navedeni je broj riječi odabran jer se tekstovi slične dužine pojavljuju u priručnicima i materijalima koji se primjenjuju u svrhu razvoja sposobnosti razumijevanja čitanjem u jeziku struke na razini B1-B2. Nakon teksta pet je ispitnih čestica ispitivalo sposobnost razumijevanja čitanjem sudionika zadatkom višestrukog izbora – za svako od pet pitanja bila su ponuđena četiri moguća odgovora, od kojih je samo jedan bio točan. Ukupan broj bodova koje su sudionici mogli ostvariti u ovom dijelu ispita bio je pet. Na e-ispitu ovaj je zadatak automatski procjenjivao Googleov obrazac, a na ispit na papiru procjenjivao ga je istraživač ručno prema jednakim mjerilima bodovanja kao što je to bio slučaj u Googleovim obrascima za e-ispit.
2. **Razumijevanje i primjena vokabulara.** Ovaj dio ispita imao je dva zadatka zatvorenog tipa – prvi je zadatak uključivao sedam ispitnih čestica, a drugi osam, i na svakoj od njih sudionici su mogli ostvariti po jedan bod, odnosno ukupno petnaest mogućih bodova u ovom dijelu ispita. Prvi je zadatak također bio zadatak višestrukog izbora – svaka je od sedam ispitnih čestica imala po jednu rečenicu u kojoj je nedostajala jedna riječ, a za svaku od njih bila su ponuđena četiri moguća odgovora, od kojih je samo jedan bio točan. I ovaj je zadatak, kao i prethodni, na e-ispitu automatski procjenjivao Googleov obrazac, a na ispit na papiru istraživač ručno, slijedeći mjerila bodovanja istovjetna onima na e-ispitu u Googleovim obrascima. Nadalje, u drugom zadatku sudionicima je dan tekst od 116 riječi iz kojega se izostavilo osam riječi, a oni su od dvanaest ponuđenih riječi morali

odabrati njih osam i umetnuti ih na odgovarajuće mjesto u tekstu, s time da su svaku riječ mogli upotrijebiti samo jednom. Na e-ispitu ovaj je zadatak također automatski procjenjivao Googleov obrazac jer su sudionici riječi birali iz padajućeg izbornika, dok je na ispitu na papiru njihovu izvedbu ručno procjenjivao istraživač, ali se vodilo računa da se ne oduzimaju bodovi zbog pravopisnih pogrešaka u slučaju pogrešno napisane riječi, sve dok se moglo vidjeti koju je riječ pojedini sudionik napisao na odgovarajuće mjesto u tekstu, budući da se pravopis nije procjenjivao, a na e-ispitu sudionici nisu ni upisivali riječi, već ih birali među ponuđenima. Količina teksta u ovim dvama zadacima – sedam rečenica u prvomu i jedan tekst od 116 riječi u drugomu – također je odgovarala zadacima za uvježbavanje razumijevanja i primjene vokabulara koji se pojavljuju u priručnicima i materijalima koji se primjenjuju u poučavanju jezika struke na razini B1-B2.

3. **Pisana proizvodnja.** Ovaj je dio ispita sadržavao jedan zadatak otvorenog tipa. U njemu su sudionici morali napisati pismo žalbe prema predstavljenoj situaciji i slijedeći zadane upute koje su bile podijeljene u tri niže predstavljene kategorije – oblik, sadržaj i jezik. Uz svaku kategoriju bio je naveden ukupan broj bodova koji su sudionici mogli ostvariti (od njih ukupno dvadeset) i to prema smjernicama u natuknicama za svaku kategoriju kako bi lakše provjerili jesu li slijedili svaku smjernicu.
 - a. **Oblik.** Pismo žalbe moralo je slijediti poslovni stil korespondencije i biti primjereno oblikovano u odlomke te sadržavati sljedeće sastavnice: adresu pošiljatelja i primatelja, datum, predmet pisma, pozdrav na početku i na kraju pisma, završnu rečenicu i potpis. U ovoj kategoriji sudionici su mogli ostvariti najviše devet bodova.
 - b. **Sadržaj.** Pismo žalbe moralo je uključiti odgovarajuć sadržaj – razlog pisanja, objašnjenje svoje žalbe te zahtjev za primateljevo neposredno djelovanje. Opći podatci mogli su se pronaći u kratku opisu situacije u uputama zadataka, ali se od sudionika tražilo da ih prošire i dodaju svoj sadržaj. U ovoj kategoriji sudionici su mogli ostvariti najviše šest bodova.
 - c. **Jezik.** Jezična kvaliteta pisane proizvodnje sudionika procjenjivala se prema dvjema smjernicama – rasponu i primjerenosti vokabulara te pravopisnoj i gramatičkoj točnosti, a u svakoj od njih mogli su ostvariti najviše 2,5 boda, odnosno ukupno pet bodova u ovoj kategoriji. Prvo, vokabular se smatrao širokim ako nisu primjenjivali izraze koji su već bili u uputama zadatka, već su dodali svoje odgovarajuće izraze. Naime, nisu im se oduzimali bodovi ako su ponovili izraze iz zadatka, ali su sva 2,5 boda mogli dobiti samo ako su primjenjivali izraze kojih nije bilo u zadatku. Drugo, vokabular se smatrao primjerenim ako je bio poslovnog registra i bez skraćenica.

Pravopis se smatrao točnim ako su se dosljedno slijedila sva pravopisna pravila engleskog jezika, a gramatika ako su se dosljedno slijedila sva pravila gramatike engleskog jezika.

Usto, broj riječi koje su sudionici trebali napisati u ovom zadatku pisane proizvodnje nije bio određen, ali im je napomenuto da slijede upute i uključe sve dijelove koje dobro napisano poslovno pismo mora sadržavati. Nadalje, ovaj je zadatak procjenjivalo dvoje stručnjaka, nastavnika poslovnog engleskog jezika, koji su slijedili ista mjerila bodovanja koja je izradio istraživač (a koja su i sami sudionici imali navedena u uputama na jezičnom ispitu) i koja su objašnjenja u trima prethodnim odlomcima – od ukupno mogućih 20 bodova, dodjeljivali su se u trima kategorijama – oblik (mogućih 9 bodova), sadržaj (mogućih 6 bodova) i jezik (mogućih 5 bodova). Procjenjivači su ih prije procjenjivanja prvo zajedno analizirali i prokomentirali kako bi se usuglasili oko toga razumiju li ih na jednak način. Zatim je svaki od njih procjenjivao zadatak pisane proizvodnje svih 122 sudionika na e-ispitu, a nakon toga isti taj zadatak na ispitu na papiru. Ako su se kod određenog sudionika bodovi procjenjivača u pojedinoj kategoriji razlikovali za jedan bod ili više, procjenjivači su ponovo procijenili taj zadatak i sporazumno došli do konačnog broja bodova za dotičnu kategoriju. Podudarnost procjenjivača iscrpnije je analizirana niže u potpoglavlju 4.3.1.3. Konačni rezultat izvedbe svakog sudionika na ovom dijelu jezičnog ispita predstavlja prosjek bodova dvoje procjenjivača.

4.2.2.2. Upitnik o stavovima o e-ispitima PEJ-a

Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika (prilog 2.) preuzet je od Fan i Ji (2014). Nakon što je troje sustručnjaka prevelo sve 24 čestice izvornoga upitnika (prilog 2.1.), izostavilo se njih pet (čestice 15-19 u izvornom upitniku) jer se odnose na ispit govorenja putem računala bez ispitivača, koji nije bio dio našeg istraživanja, a dodala se jedna (čestica 9 u našoj inačici upitnika), koja se odnosi na zadatke čitanja i poznavanja vokabulara. Ovaj se prevedeni i prilagođeni upitnik dao trima studentima kako bi se vidjelo hoće li njegove čestice shvatiti na jednak način na koji upućuje izvornik. Svaka od ukupno 20 čestica Upitnika o stavovima o e-ispitima PEJ-a sadrži jednu tvrdnju za koju su sudionici odabrali koliko se slažu s njom, označujući na Likertovoj skali jednu vrijednost od 1 (*Nikako se ne slažem s tvrdnjom.*) do 6 (*Sasvim se slažem s tvrdnjom.*), pri čemu je za svaku tvrdnju veći broj označivao pozitivniji stav o e-ispitima (prilog 2.2.). Budući da je navedeni upitnik koji se primijenio u ovom istraživanju u određenoj mjeri izmijenjen u odnosu na izvornu inačicu,

provjerili smo faktorsku strukturu upitnika eksploratornom faktorskom analizom, metodom glavnih komponenata. Uzimajući u obzir karakteristične korijenove veće od 1, *scree plot* i interpretabilnost rezultata, ekstrahiran je jedan faktor (prilozi 2.3. i 2.4.). Ovim faktorom objašnjeno je 36,65 % varijance ukupnih rezultata. Pouzdanost tipa unutarnje konzistencije iznosila je $\alpha = ,903$. Iako su određene čestice (točnije 9, 11, 18, 19 i 20) bile u znatnijoj mjeri saturirane i drugim faktorima, uvidom u podatke o pouzdanosti skale, ako bi se neka od ovih čestica isključila iz daljnje obrade, primjećujemo da bi se izbacivanjem bilo koje od njih pouzdanost snizila, što je snažan argument za zadržavanje ovih čestica (tablica 2 u potpoglavlju 4.3.1.2.2. – Odgovori na Upitniku o stavovima o e-ispitima PEJ-a). Ukupni se rezultat stoga računa kao zbroj odgovora sudionika na svih 20 čestica podijeljen s brojem čestica.

4.2.2.3. *Upitnik o stavovima o ispitima općenito*

Upitnik o stavovima o ispitima općenito (prilog 3.) jest cjelovit prijevod Ljestvice o stavovima o ispitima (prilog 3.1.), koji je izradio Dodeen (2008 u Muñoz, 2017). Pri prijevodu ovog upitnika na hrvatski jezik savjetovalo se troje stručnjaka te se prevedeni upitnik, kao i prethodni, dao trima studentima radi provjere razumijevanja čestica. Upitnik o stavovima o ispitima općenito sastoji se od 17 čestica. Svaka čestica sadrži jednu tvrdnju za koju su sudionici odabrali koliko se slažu s njom, označujući na Likertovoj skali jednu vrijednost od 1 (*Nikako se ne slažem s tvrdnjom.*) do 5 (*Sasvim se slažem s tvrdnjom.*), pri čemu je veći broj označivao pozitivniji stav o ispitima, a rekodirale su se čestice broj 9, 11 i 13 (prilog 3.2.). Eksploratorna faktorska analiza provedena metodom glavnih komponenata pokazala je da rezultate na česticama najbolje objašnjava jedan faktor (prilozi 3.3. i 3.4.), što je sukladno ranijim istraživanjima (*ibid.*). Dobiveni faktor objašnjavao je 26,14 % varijance. Stoga se ukupni rezultat računa kao zbroj odgovora sudionika na svih 17 čestica podijeljen s brojem čestica. Pouzdanost unutarnje konzistencije iznosila je $\alpha = ,804$, a podatci o pouzdanosti skale prikazani su u tablici 3 u potpoglavlju 4.3.1.2.3 – Odgovori na Upitniku u stavovima o ispitima općenito.

4.2.2.4. *Upitnik o stavovima o računalima*

Upitnik o stavovima o računalima (prilog 4.) sadrži 53 čestice koje ispituju sudionike o njihovim stavovima o računalima, a svaka čestica sadrži tvrdnju za koju su sudionici odabrali koliko se slažu s njom, označujući na Likertovoj skali jednu vrijednost od 1 (*Nikako se ne*

slažem s tvrdnjom.) do 5 (*Sasvim se slažem s tvrdnjom.*), uz dva dodatna pitanja o posjedovanju i učestalosti korištenja računalom (prilog 4.2.). Upitnik, koji je dio Istraživanja stavova o računalima (prilog 4.1.), osmislili i primijenili u svojim istraživanjima Hewson i sur. (2007) te Hewson i Charlton (2019). U našem se istraživanju navedeni upitnik preuzeo u cijelosti i suradnjom troje sustručnjaka preveo na hrvatski, osim izostavljanja izraza „na televiziji“ u čestici 32 jer ona kao medij više nije najzastupljenija. Kako bi se provjerilo razumijevanje čestica prevedenog upitnika, čestice je pročitao i o njihovu razumijevanju istraživača izvijestilo troje studenata. Faktorskom analizom autori su dobili tri faktora: anksioznost, uključenost i ovisnost. U svojem smo istraživanju provjerili faktorsku strukturu upitnika eksploratornom faktorskom analizom, metodom glavnih komponentata. Uzimajući u obzir karakteristične korijenove veće od 1, *scree plot* i interpretabilnost rezultata, ekstrahirana su dva faktora (prilozi 4.3. i 4.4.), koji su opisani niže.

- Prvi faktor, pod nazivom anksioznost vezana za računala, objašnjavao je 15,602 % varijance i uključivao je 24 čestice (čestice 1, 2, 5, 7, 8, 11, 15, 18, 19, 21, 22, 23, 25, 26, 28, 30, 33, 38, 41, 44, 48, 49, 50 i 53 iz Upitnika 3 u prilogu 4.). Rekodirale su se čestice 7, 22, 25, 26, 27, 30, 33, 38, 41 i 49. Veći rezultat označivao je veću anksioznost vezanu za računala. Pouzdanost tipa unutarnje konzistencije iznosila je $\alpha = ,882$.
- Drugi faktor, pod nazivom ovisnost o računalima, objašnjavao je 12,954 % varijance i uključivao je 19 čestica (čestice 9, 10, 12, 13, 17, 24, 29, 34, 35, 36, 37, 39, 40, 42, 43, 46, 47, 51 i 52 iz Upitnika 3 u prilogu 4.). Rekodirale su se čestice 35, 36, 37 i 47. Veći rezultat označivao je veću ovisnost o računalima. Pouzdanost tipa unutarnje konzistencije iznosila je $\alpha = ,805$.

Deset čestica izbačeno je iz daljnje obrade (čestice 3, 4, 6, 14, 16, 20, 27, 31, 32 i 45) zbog podjednake saturacije obama faktorima, zbog čega nisu bile interpretabilne. U prilog ovoj odluci govorio je i porast pouzdanosti ako se navedene čestice izbace (tablice 4 i 5 u potpoglavlju 4.3.1.2.4. – Odgovori na Upitniku o stavovima o računalima).

4.2.2.5. Upitnik o ispitnoj anksioznosti

Zadnji strukturirani upitnik u ovom istraživanju jest Upitnik o ispitnoj anksioznosti, koji se sastoji od deset čestica (prilog 5.) i koji predstavlja cjelovit prijevod izvornika pod nazivom Ljestvica ispitne anksioznosti Westside (prilog 5.1.), koju je izradio Driscoll (2007 u Aliakbari i Gheitasi, 2017). Postupak prijevoda i pregleda bio je istovjetan postupku primijenjenom kod prethodno opisana tri upitnika. Svaka čestica sadrži jednu tvrdnju za koju

su sudionici odabrali koliko se slažu s njom, označujući na Likertovoj skali jednu vrijednost od 1 (*Nikako se ne slažem s tvrdnjom.*) do 5 (*Sasvim se slažem s tvrdnjom.*), pri čemu je za svaku tvrdnju veći broj označivao veću razinu ispitne anksioznosti (prilog 5.2.). Eksploratorna faktorska analiza provedena metodom glavnih komponenata pokazala je da rezultate na česticama najbolje objašnjava jedan faktor (prilozi 5.3. i 5.4.), što je sukladno ranijim istraživanjima (ibid.). Dobiveni faktor objašnjavao je 46,76 % varijance. Stoga se ukupni rezultat računa kao zbroj odgovora sudionika na svih deset čestica i zatim se taj zbroj podijeli s brojem čestica (tj. dijeli se s 10). Rezultati na skali tumače se na sljedeći način (ibid.):

- 1,0—1,9 = niska ispitna anksioznost
- 2,0—2,5 = uobičajena/prosječna ispitna anksioznost
- 2,5—2,9 = visoka prosječna ispitna anksioznost
- 3,0—3,4 = umjereno visoka ispitna anksioznost
- 3,5—3,9 = visoka ispitna anksioznost
- 4,0—5,0 = izrazito visoka ispitna anksioznost.

Pouzdanost unutarne konzistencije iznosila je $\alpha = ,861$, a podatci o pouzdanosti skale prikazani su u tablici 6 u potpoglavlju 4.3.1.2.5. – Odgovori na Upitniku o razni ispitne anksioznosti.

4.2.3. Prikupljanje podataka

Prikupljanje podataka počelo je provedbom jezičnog e-ispita Poslovnog engleskog jezika u 2021. godini. Ovaj e-ispit proveo se u ispitnim uvjetima prema naputcima fakulteta u kojima se polažu e-ispiti, što znači da ga studenti pišu putem interneta, na računalu u Googleovu obrascu, a snimaju se pametnim telefonom iz dovoljne udaljenosti da profesor koji ih nadgleda u svakom trenutku vidi njihovo lice i ruke te tipkovnicu i ekran računala. Sudionicima su se uz e-ispit dala i dva upitnika, također u Googleovim obrascima – Upitnik o stavovima o računalima (prilog 4.) i Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika (prilog 2.). U ovom prvom dijelu prikupljanja podataka, kad su se podatci prikupljali putem interneta, sudjelovalo je 210 sudionika.

U drugom dijelu prikupljanja podataka, koje se provodilo tri i pol mjeseca nakon prvog dijela, podatci su se prikupljali uživo u otežanim okolnostima slijedeći epidemiološke mjere. Tada se broj sudionika smanjio s njih 210 koliko ih je sudjelovalo u prvom dijelu na njih 122 koji su sudjelovali i u drugom dijelu. Oni su riješili isti ispit, ali ovaj put na papiru.

Uz taj ispit na papiru dala su im se i preostala dva upitnika, također na papiru – Upitnik o stavovima o ispitima općenito (prilog 3.) i Upitnik o razini ispitne anksioznosti (prilog 5.).

Što se tiče obavješćivanja sudionika o sadržaju ispita, njihovo sudjelovanje u istraživanju i polaganje ovog ispita predstavilo im se kao ponavljanje gradiva za redovan ispit ovog kolegija na kraju semestra. Također im se napomenulo da se za ispit ne moraju posebno pripremati budući da je njegova svrha prikupiti podatke za istraživanje, a ne koristiti se njihovim rezultatima za vrjednovanje njihove jezične izvedbe i dodjeljivanja ocjene. Obavješćivanje sudionika iscrpnije je opisano u potpoglavlju 4.2.5. – Etička načela.

4.2.4. Analiza podataka

Analiza prikupljenih podataka u istraživanju bila je kvantitativna i kvalitativna²⁵ kako bi se utvrdila valjanost e-procjenjivanja. Kvantitativna se analiza provela postupcima statističke analize, a kvalitativna prema Bachmanovoj metodi potkrjepe procjenjivanja (Bachman, 2005, 2015; Bachman i Palmer, 2010).

4.2.4.1. Postupci statističke analize

Podatci do kojih se došlo u ovom istraživanju analizirali su se primjenom postupaka deskriptivne i inferencijalne statistike²⁶ – korelacijskom analizom i analizom varijance, što je predstavljeno u poglavlju 4.3. niže.

Deskriptivnom statistikom dobiven je sažet opis prikupljenih podataka, odnosno razina pojedinih varijabla, a rezultati su prikazani i grafički. Ovim su se statističkim analizama usporedili ukupni rezultati koje su sudionici ostvarili na e-ispitu s onima na ispitu na papiru, te njihovi rezultati u pojedinim dijelovima ispita – razumijevanju čitanjem, razumijevanju i

²⁵ Kraš i Miličević (2015) kvalitativnu i kvantitativnu znanstvenu paradigmu, odnosno skup načela i postupaka u znanosti, definiraju na sljedeće načine: za kvalitativna istraživanja navode da su to istraživanja koja „odlikuje opažanje u prirodnim uvjetima, određena doza subjektivnosti, orijentiranost ka procesu istraživanja i nemogućnost generalizacije“, dok za kvantitativna istraživanja navode da su to istraživanja koja odlikuje „kontrolirano mjerenje pojava, objektivnost, orijentiranost ka ishodu i mogućnost generalizacije“ (str. 10). Kvalitativne metode koje se primjenjuju u postupcima utvrđivanja valjanosti jezičnih ispita jesu introspekcija i retrospekcija putem verbalnih protokola, promatranja, upitnici, intervjui, analize teksta, razgovora i/ili diskursa primijenjenih na instrumentima i prilikom izvedbe u sklopu procjenjivanja te razni analitički modeli (Banerjee i Luoma, 1997: 277).

²⁶ Deskriptivne i inferencijalne analize Kraš i Miličević (2015) opisuju na sljedeći način: „prve opisuju rezultate dobivene na nekom uzorku (tu spada npr. prosjek), dok druge preko niza različitih testova procjenjuju s kolikom se pouzdanošću rezultati dobiveni u uzorku mogu uopćiti i izvan njega“ (str. 10). Također ističu da inferencijalna analiza „istraživaču ne pruža konačan dokaz, već je probabilističke prirode, odnosno zasniva se na procjeni vjerojatnosti određenoga ishoda“ (str. 13).

primjeni vokabulara te pisanoj proizvodnji. Također su se analizirala psihometrijska obilježja čestica četiriju primijenjenih upitnika – Upitnika o stavovima o e-ispitima PEJ-a, Upitnika o stavovima o ispitima općenito, Upitnika o stavovima o računalima i Upitnika o razini ispitne anksioznosti.

Inferencijalnom statistikom odgovoreno je na postavljena istraživačka pitanja. S ciljem utvrđivanja međusobnog učinka dvaju faktora (način ispitivanja, tj. e-ispit i ispit na papiru, te sadržaj ispita, tj. njegova prethodno spomenuta tri dijela), odnosno kako bi se utvrdilo ovisi li utjecaj jednog faktora o razinama drugog faktora, provedena je dvofaktorska analiza varijance. Ovo je bio primaran interes prvog istraživačkog pitanja, koje je nastojalo odgovoriti ispituje li se e-ispitima sposobnost razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u EJS-u. Nadalje, kako bismo utvrdili postoji li sukladnost u variranju varijabla, odnosno njihov međusoban odnos, a što je ispitivano drugim i trećim istraživačkim pitanjem, provedena je korelacijska analiza. Budući da su ta pitanja nastojala odgovoriti jesu li rezultati na e-ispitu povezani sa stavovima sudionika o e-ispitima, ispitima općenito i računalima te s razinom njihove ispitne anksioznosti, ova je analiza ispitala međuovisnost ukupnih i pojedinačnih rezultata sudionika na objema inačicama ispita te njihovih odgovora na primijenjenim upitnicima.

4.2.4.2. Metoda potkrjepe procjenjivanja

Metodu potkrjepe procjenjivanja osmislio je Bachman (Bachman, 2005, 2015; Bachman i Palmer, 2010) na temelju Kaneove potkrjepe tumačenja (Kane, 1992, 2001, 2004; Kane, Crooks i Cohen, 1999) i Toulminove strukture argumenata (Toulmin, 2003). Cilj mu je bio povezati valjanost ispita s njihovim namijenjenim primjenama, koje su svojstvene dionicima svakog pojedinog procjenjivanja – od autora ispita, preko ispitanika i procjenjivača, do nastavnika, škola, poslodavaca i šire. Rezultati koje ispitanici ostvaruju na temelju svoje izvedbe na ispitima moraju se moći protumačiti dotičnim dionicima, a namijenjene primjene ispita moraju se opravdati, što se postiže upravo metodom potkrjepe procjenjivanja. U njoj se na temelju podataka oblikuju tvrdnje koje se potkrjepljuju jamstvima s pripadajućom potporom, a moguća se opovrgnuća tih tvrdnja odbacuju, oslabljuju ili podupiru odgovarajućim dokazima (usp. Cronbach, 1988; Chapelle i Douglas, 2006; Shepard, 2016). Od jedne do druge tvrdnje dolazi se zaključivanjem na temelju dostupnih jamstava i opovrgnuća, a prilikom utvrđivanja valjanosti određenog ispita prvo se na temelju rezultata koji ispitanici ostvare na ispitu zaključivanjem dolazi do tvrdnje o sposobnosti dotičnih ispitanika, a zatim se mora

istražiti razumnost te tvrdnje (Davis i Elder, 2010). Ovom metodom Bachman (2015) zauzima stav suprotan dotadašnjem prevladavajućem kvantitativnom pristupu analize valjanosti, ističući da rezultati procjenjivanja nisu nužno samo brožčani, već i opisni, odnosno verbalni i vizualni. Stoga i naše istraživanje kombinira kvantitativnu i kvalitativnu obradu i analizu podataka.

Potkrjepa procjenjivanja sastoji se od dvaju dijelova – potkrjepe valjanosti procjenjivanja i potkrjepe primjene procjenjivanja (Bachman, 2005, 2015; Chapelle, Jamieson i Hegelheimer, 2003; Bachman i Palmer, 2010), od kojih prva povezuje izvedbu na ispitu s tumačenjem ostvarenih rezultata ispitanika, a druga povezuje to tumačenje s odlukama koje se donose na temelju njih i posljedica do kojih te odluke dovode. Budući da je tema našeg istraživanja analiza valjanosti e-procjenjivanja u engleskomu kao jeziku struke, ono se usredotočilo na prvi dio potkrjepe procjenjivanja, odnosno na potkrjepu valjanosti procjenjivanja. Svaka se tvrdnja, koja se mora poduprijeti odgovarajućim jamstvima i mogućim opovrgnućima, u potkrjepi procjenjivanja sastoji od ishoda i njegovih svojstava, a u dijelu potkrjepe valjanosti procjenjivanja te dvije tvrdnje glase da su rezultati na e-ispitu dosljedni te da su tumačenja tih rezultata smislena, nepristrana, uopćivajuća, značajna i dostatna (kao što je prikazano na slici 11 u potpoglavlju 2.5. – Potkrjepa procjenjivanja).

Jezična izvedba na e-ispitu dovela je do tumačenja rezultata uz određena jamstva i opovrgnuća, što predstavlja dio potkrjepe procjenjivanja koji se naziva potkrjepom valjanosti procjenjivanja. Ta tumačenja dalje dovode do odluka, uz određena jamstva i opovrgnuća, što predstavlja dio potkrjepe procjenjivanja koji se naziva potkrjepom primjene procjenjivanja, no ona nije dio našeg istraživanja, već je to samo potkrjepa valjanosti procjenjivanja. U cijelom postupku potkrjepe procjenjivanja, pa tako i u dijelu potkrjepe valjanosti procjenjivanja, na temelju dobivenih podataka (tj. jezične izvedbe, tumačenja te izvedbe i odluka donesenih na temelju tih tumačenja), oblikovale su se tvrdnje koje predstavljaju kvalitativne opise valjanosti e-procjenjivanja. Neki od dokaza tih tvrdnja, odnosno njihovih jamstava i mogućih opovrgnuća, proizašli su iz kvalitativnih podataka (tj. analiza ishoda i silaba kolegija, nastavnih materijala, Googleove učionice, procedura provedbe i procjenjivanja ispita, ispitne specifikacije, mjerila bodovanja, sadržaja ispitnih čestica te stručnosti procjenjivača), a neki iz kvantitativnih podataka (tj. rezultata na ispitu i rezultata upitnika).

4.2.5. Etička načela

Načela etike istraživanja s ljudima uzela su se u obzir tijekom provedbe našeg istraživanja, obrade podataka i prikazivanja rezultata, i to istovremeno **makroetička načela**²⁷ i **mikroetička načela**²⁸ budući da „ta dva pogleda nisu proturječni pristupi etičkoj dilemi. Naprotiv, potpuno su nadopunjujući. Dobra etička odluka u bilo kojem određenom slučaju mora uključivati dijalektiku između obaju gledišta“ (Truog i sur., 2015: 12). Stoga su se u ovom istraživanju primijenila sljedeća načela:

- opća etička načela, i to četiri od deset načela Nürnberškog kodeksa, koja su ovdje navedena u kraćem obliku (Shuster, 1997): načelo 1 – dobrovoljni pristanak sudionika posve je neophodan, načelo 2 – istraživanje treba dati rezultate koji će biti na dobrobit društva, načelo 8 – istraživanje trebaju provoditi znanstveno osposobljene osobe, te načelo 9 – tijekom istraživanja sudionik je slobodan odustati od istraživanja u bilo kojem trenutku;
- istraživač je istovremeno bio odgovoran prepoznati nepredvidive situacije tijekom prikupljanja podataka, tj. provedbe jezičnih ispita, i mogao je prekinuti istraživanje u određenom trenutku da je vidio da su sudionici bili anksiozni, da su se osjećali nelagodno, da su podbacili u svojem doprinosu istraživanju i sl. (Cergol, 2021).

U skladu s prethodno navedenim načelima, za predmetno je istraživanje istraživač pripremio dvije obavijesti – **molbu nadležnim osobama**, u kojoj je u pisanu obliku zatražio suglasnost za provođenje istraživanja od čelnika ustanove, pročelnika katedre i kolega nastavnika, te **obaviješteni pristanak sudionika**, u kojemu se sudionike unaprijed pisano obavijestilo o istraživanju te se zatražila njihova suglasnost o sudjelovanju u istraživanju (usp. Cergol, 2021; Kraš i Miličević, 2015). U ovoj su se obavijesti sudionicima naveli razlozi i ciljevi istraživanja, način provedbe istraživanja, činjenica da mogu u svakom trenutku odustati od sudjelovanja u istraživanju bez navođenja razloga te im se jamčila anonimnost – prikupljeni će podatci biti dostupni isključivo istraživaču, koji će ih unositi i obrađivati pod šifrom, što je u skladu s Uredbom (EU) 2016/679 Europskog parlamenta i Vijeća od 27. travnja 2016. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka te o stavljanju izvan snage Direktive 95/46/EZ (OUZP, 2016), o čemu govore i Cergol (2021)

²⁷ Makroetička načela definiraju se kao „skup općih etičkih smjernica uvrštenih u etičke kodekse i smjernice institucijskih etičkih povjerenstava“ (Cergol, 2021: 18).

²⁸ Mikroetička načela definiraju se kao odluke „koje istraživači svakodnevno donose susrećući se s neočekivanim situacijama u praksi te nastojeći donijeti etički ispravne odluke, koje nisu predviđene markoetičkim načelima“ (Cergol, 2021: 19).

te Kraš i Miličević (2015). Sudionici su naposljetku, prema svojoj prosudbi, potpisali obavješteni pristanak za sudjelovanje u istraživanju.

Kako bi se osiguralo da svi ispitanici polažu ispite u jednakim uvjetima, u ovom su se istraživanju slijedili određeni **ispitni rituali**, kako ih opisuje Fulcher (2010: 5-6), čija je svrha „jednakost mogućnosti“ i „meritokracija“. Prilikom polaganja ispita na papiru sudionici su sjedili sami na međusobnoj udaljenosti od najmanje jednog metra jedan iza drugoga i najmanje metar i pol jedan kraj drugoga, nije im bilo dozvoljeno unositi zabranjene predmete (ništa osim penkale), nisu smjeli izlaziti iz ispitne prostorije za vrijeme trajanja ispita u kojoj je trebala biti potpuna tišina za vrijeme ispita, a nadziratelj ih je pomno nadzirao za vrijeme cijelog trajanja ispita. Za vrijeme polaganja e-ispita nastojalo se primijeniti gotovo istovjetna pravila, samo što nije bilo onih vezanih za ispitnu prostoriju, već onih uz prostoriju u kojoj su pisali ispit na svojem računalu – u njoj nije smjelo biti nikoga osim njih te su se morali snimati kamerom iz dovoljne udaljenosti da se vidi njihovo lice, ruke, tipkovnica i ekran. Također, kako bi se zajamčila nepristranost procjenjivača prilikom procjenjivanja zadatka pisane proizvodnje i na ispitu na papiru i na e-ispitu, procjenjivačima nisu bila poznata imena sudionika.

U ovom su se istraživanju također slijedila **načela istraživačke čestitosti**, odnosno preuzimanje istraživačeve potpune odgovornosti za vjerodostojnost njegova istraživanja, upoznatost s pravilima i politikama koje se odnose na istraživanja te njihovo poštivanje, primjena prikladnih istraživačkih metoda, utemeljenost zaključaka na kritičkoj analizi rezultata, vođenje jasne i točne evidencije istraživanja kako bi drugi mogli taj rad provjeriti i ponoviti (Resnik i Shamoo, 2011). Usporedno s time, izbjegavali su se praksa znanstvenog nepoštenja i upitni istraživački postupci, što, među ostalim, uključuje izmišljanje podataka ili rezultata, krivotvorenje rezultata i plagijat pri izradi istraživačkog nacrtu i provođenju istraživanja i dr. (Steneck, 2006). Nadalje, nastojalo se zajamčiti **autonomiju sudionika** jer je istraživač istovremeno bio nastavnik nekim sudionicima pa je stoga bio u poziciji moći (Cergol, 2021). Zato se posebna pažnja posvetila činjenici da su sudionici samostalno donosili odluke o sudjelovanju i da nisu snosili nikakve posljedice sudjelovanjem ili nesudjelovanjem u istraživanju.

S druge strane, sudionike se stimuliralo da sudjeluju u istraživanju određenim **nagradama**, koje se dijele prema intrinzičnoj i ekstrinzičnoj koristi – prva podrazumijeva „osjećaj zadovoljstva koje neki sudionici dožive zbog toga što sudjeluju u istraživanju [...] (n)eki sudionici neke vrste istraživanja doživljavaju terapijskim [...] (n)eki sudionici razumiju i poštuju dobrobit koju će rezultati istraživanja donijeti zajednici“ (Cergol, 2021: 89), dok druga može biti intelektualna, poput unaprjeđivanja znanja stranog jezika ako se

istraživanje bavi tom tematikom, ili materijalna, poput novčane nagrade, odrađivanja obveze na određenom kolegiju, slatkiša za djecu i sl. (ibid.). Kao što je već prethodno navedeno u potpoglavlju 4.2.3. – Prikupljanje podatka, nagrada sudionicima ovog istraživanja bila je u obliku intelektualne koristi jer su sudjelovanjem u istraživanju ponavljali gradivo koje je kasnije bilo uključeno u završni ispit kolegija Poslovni engleski jezik 1, a ovisno o predmetnom nastavniku dobili su dodatne bodove za ispit i/ili za domaću zadaću. Također im se napomenulo da se nije potrebno pripremati za ovo istraživanje ni na kakav način jer se njihova izvedba neće ocjenjivati niti će o njoj ovisiti njihova ocjena na kraju semestra. No zbog tog je razloga istovremeno i upitna njihova motivacija, pa se u poglavlju 4.4. – Diskusija rezultata istraživanja i ona razmatra kao jedan od mogućih učinaka na njihovu izvedbu budući da nisu bili motivirani ocjenom. Međutim, dobili su i materijalne nagrade jer su u nagradnom izvlačenju na Googleovu Meetu imali priliku osvojiti nekoliko nagrada – prva je bila poklon-bon u trgovačkom centru, dvije druge nagrade bili su poklon-bonovi za kino, a tri treće nagrade bile su poklon-bonovi na internetskoj stranici za kupnju s popustom.

4.3. REZULTATI ISTRAŽIVANJA

4.3.1. Preliminarne provjere i statističke analize podataka

Kao preduvjet provedbe statističkih analiza kojima će se provjeriti postavljene hipoteze, provedene su određene preliminarne analize i provjere podataka, što uključuje analizu podataka koji nedostaju u bazi, deskriptivnu statistiku, detekciju i tretman univarijatno ili multivarijatno odstupajućih podataka te analizu osnovnih deskriptivnih pokazatelja i matrice bivarijatnih korelacija kompozitnih varijabla.

4.3.1.1. Analiza i tretman podataka koji nedostaju te univarijatna i multivarijatna odstupanja rezultata

Analizom podataka koji nedostaju (eng. *missing value analysis*), u bazi s rezultatima 122 sudionika, zaključeno je da problem izostanka pojedinih odgovora nije izražen, odnosno da podataka koji nedostaju nema budući da su sudionici davali odgovore na način koji je jamčio ispunjavanje svih upitnika (putem Googleova obrasca i na papiru).

Univarijatno odstupajući podatci (eng. *outliers*) slučajevi su s neočekivanom vrijednosti na jednoj varijabli, dok su multivarijatna odstupanja slučajevi s neočekivanim ili

neuobičajenim kombinacijama odgovora na više varijabla (Tabachnick i Fidell, 2013). Analizom univarijatnih odstupanja u primijenjenim varijablama primjećujemo da nema ekstremno odstupajućih rezultata za koje je izgledno da su uistinu odstupajući rezultati, a ne samo visok ili nizak rezultat za koji je očekivano da se u setu podataka pojavi. Zamjećujemo slučaj sudionika 9 i 39, koji na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa su njihovi podatci isključeni iz daljnje obrade budući da bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Kako bismo provjerili postojanje odstupajućih kombinacija varijabla (multivarijatnih aberacija), izračunane su Mahalanobisove udaljenosti. Mahalanobisova udaljenost ukazuje na udaljenost vrijednosti pojedinog sudionika od centroida ostalih slučajeva, kad je centroid točka sjecišta središnjih vrijednosti svih varijabla u zamišljenom multivarijatnom prostoru (Tabachnick i Fidell, 2013). Uz mjerilo $p < ,001$ te $df = 11$ (broj varijabla za koje se provodi provjera) zaključeno je da ne postoje sudionici s multivarijatnim odstupanjima iako su sudionici 33 i 67 bili granični slučajevi. Stoga analize nastavljamo s ukupnim brojem sudionika ($N = 120$).

4.3.1.2. Deskriptivni rezultati

4.3.1.2.1. Izvedba sudionika na jezičnom ispitu

Kao što je prethodno opisano u potpoglavlju 4.2.3. – Prikupljanje podataka, sudionici su jezičnom ispitu pristupili dva puta – prvi put njegovoj inačici na internetu, odnosno e-ispitu, a drugi put njegovoj inačici na papiru. Njihova se jezična izvedba procjenjivala na obama ispitima, i to na trima dijelovima u svakom od njih – razumijevanju čitanjem, razumijevanju i primjeni vokabulara te pisanoj proizvodnji. Ukupni rezultat njihove izvedbe na jezičnom ispitu zbroj je njihove izvedbe na svakom od tih triju dijelova – ukupno su mogli ostvariti najviše 40 bodova (na prvomu pet, na drugomu petnaest, a na trećemu dvadeset bodova).

Budući da dijelovi razumijevanja čitanjem te razumijevanja i primjene vokabulara sadrže zadatke zatvorenog tipa, na njima je izvedbu sudionika na e-ispitu automatski procjenjivao Googleov obrazac, a ručno istraživač na ispitu na papiru. S druge strane, budući da dio pisane proizvodnje sadrži zadatak otvorenog tipa, na njemu je izvedbu sudionika procjenjivalo dvoje procjenjivača prema mjerilima bodovanja opisanima u potpoglavlju 4.2.2.1. – Jezični ispit.

Rezultati izvedbe sudionika na jezičnom ispitu iscrpnije su prikazani u prilogu 1.1., a niže u tablici 1 prikazani su prema trima kategorijama i ukupno. Iz ove se tablice vidi da je srednja vrijednost ukupne izvedbe sudionika veća za e-ispit nego za ispit na papiru (razlika $M = 0,38$), kao i srednje vrijednosti njihove izvedbe na dijelu razumijevanja čitanjem (razlika $M = 0,08$) i na dijelu pisane proizvodnje (razlika $M = 0,97$). S druge strane, srednja je vrijednost njihove izvedbe na ispitu na papiru veća na dijelu razumijevanja i primjene vokabulara (razlika $M = 0,67$).

Tablica 1

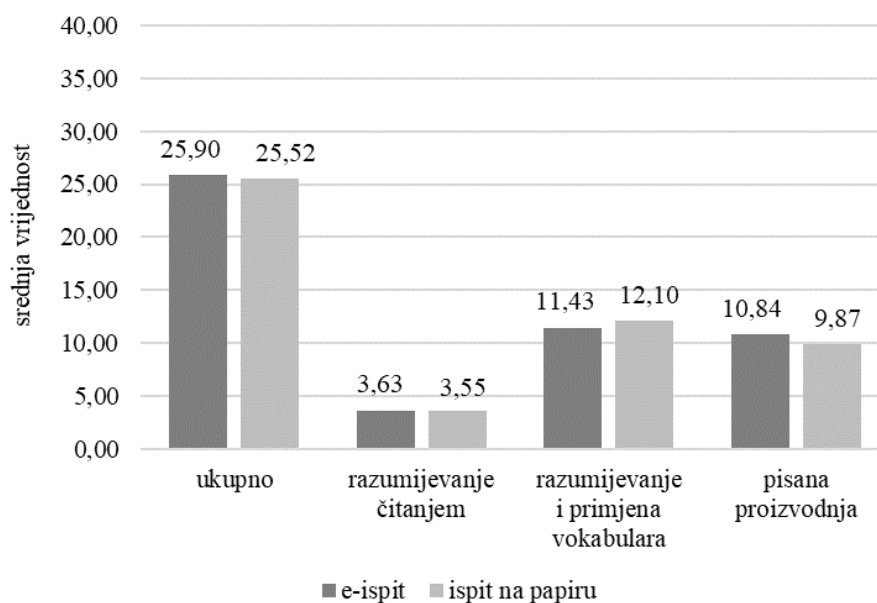
Deskriptivna statistika izvedbe sudionika na jezičnom ispitu ($N = 120$)

		najmanja vrijednost	najveća vrijednost	<i>M</i>	<i>SD</i>	asimetričnost distribucije ^a	spljoštenost distribucije ^b
e-ispit	ukupno	10,50	34,00	25,8958	5,06518	-0,620	0,133
	razumijevanje čitanjem	1,00	5,00	3,6250	1,07736	-0,432	-0,472
	razumijevanje i primjena vokabulara	3,00	15,00	11,4333	2,73395	-0,993	0,775
	pisana proizvodnja	1,25	17,00	10,8375	3,30547	-0,436	-0,093
ispit na papiru	ukupno	5,00	34,50	25,5208	5,44294	-1,105	1,791
	razumijevanje čitanjem	1,00	5,00	3,5500	1,06787	-0,280	-0,859
	razumijevanje i primjena vokabulara	2,00	17,00	12,1000	2,31691	-1,076	2,434
	pisana proizvodnja	,00	16,50	9,8708	3,47318	-0,974	1,345

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Radi preglednosti njihovi su rezultati grafički prikazani na slici 15 iz koje je razvidno kako su srednje vrijednosti ukupnih rezultata sudionika na e-ispitu te na njegovim dijelovima razumijevanje čitanjem i pisana proizvodnja veće od njihovih rezultata na ispitu na papiru, dok su veće na ispitu na papiru nego na e-ispitu samo u dijelu razumijevanja i primjene vokabulara.



Slika 15. Usporedba jezične izvedbe sudionika na e-ispitu i ispitu na papiru ($N = 120$)

4.3.1.2.2. Odgovori na Upitniku o stavovima o e-ispitima Poslovnog engleskog jezika

Na Upitniku o stavovima o e-ispitima Poslovnog engleskog jezika (prilog 2.) sudionici su označivali koliko se slažu s ponuđenih dvadeset tvrdnja, od vrijednosti 1 do 6, a veća je vrijednost značila njihov pozitivniji stav o e-ispitima. Odgovori sudionika iscrpno su prikazani u prilogu 2.2., a u tablici 2 prikazana su psihometrijska obilježja čestica Upitnika. Iz te je tablice vidljivo da su za sve čestice srednje vrijednosti u gornjoj polovici mogućih vrijednosti, i to od $M = 3,56$ do $M = 5,69$, pri čemu je najveća srednja vrijednost odgovora sudionika na čestici 11 (*Upoznat(a) sam s internetskom stranicom, odnosno Googleovom učionicom u kojoj su svi podatci vezani za e-ispite PEJ-a.*), a najmanja na čestici 18 (*Mislim da me e-ispiti PEJ-a potiču da više slušam emisije na engleskom jeziku na radiju, TV-u, internetu i dr.*).

Tablica 2

Psihometrijska obilježja čestica Upitnika o stavovima o e-ispitima PEJ-a ($N = 120$)

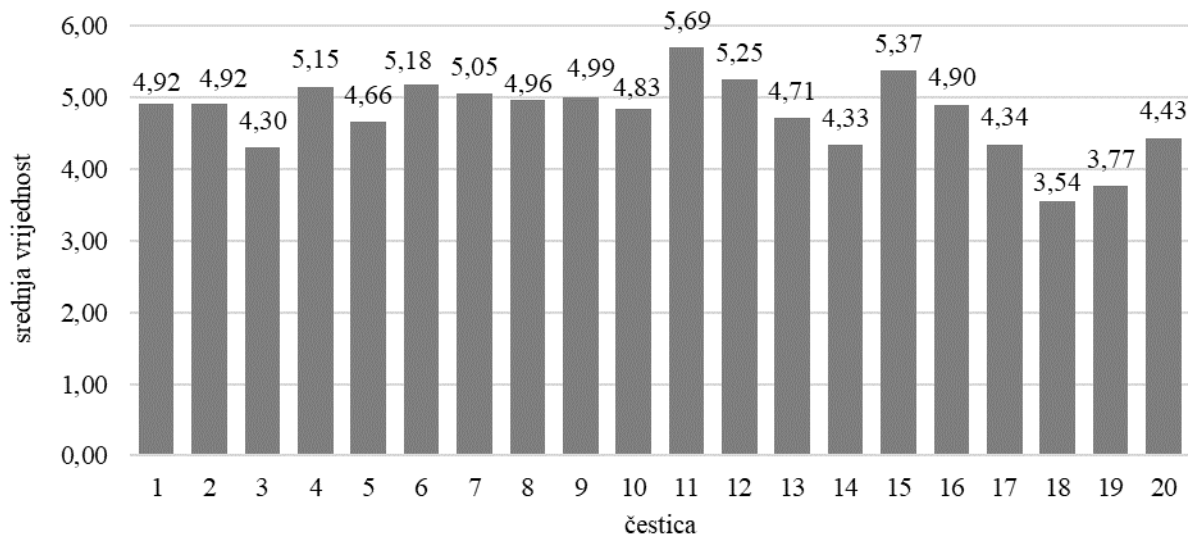
čestica	najmanja vrijednost	najveća vrijednost	M	SD	asimetričnost distribucije ^a	spljoštenost distribucije ^b	korelacija čestice i ukupnog rezultata	pouzdanost bez čestice
1.	2	6	4,93	1,038	-0,718	-0,247	,616	,897
2.	1	6	4,93	1,083	-1,321	2,096	,582	,898
3.	1	6	4,32	1,328	-0,559	-0,350	,631	,896
4.	2	6	5,15	1,090	-1,215	0,579	,572	,898
5.	1	6	4,64	1,308	-0,613	-0,583	,559	,898
6.	1	6	5,17	1,221	-1,638	2,306	,470	,900
7.	3	6	5,07	0,989	-0,789	-0,452	,580	,898
8.	2	6	4,98	1,159	-0,855	-0,293	,502	,900
9.	2	6	5,02	1,092	-0,860	-0,280	,448	,901
10.	1	6	4,86	1,279	-0,981	0,418	,481	,900
11.	1	6	5,69	0,848	-3,731	15,897	,380	,902
12.	2	6	5,25	0,998	-1,193	0,487	,513	,900
13.	1	6	4,74	1,254	-1,007	0,444	,663	,895
14.	1	6	4,36	1,419	-0,497	-0,604	,579	,898
15.	1	6	5,37	1,004	-2,010	4,465	,458	,901
16.	1	6	4,91	1,277	-1,181	0,900	,588	,897
17.	1	6	4,38	1,561	-0,646	-0,776	,633	,896
18.	1	6	3,56	1,709	-0,034	-1,204	,471	,902
19.	1	6	3,78	1,716	-0,276	-1,210	,539	,900
20.	1	6	4,47	1,517	-0,758	-0,477	,582	,898

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Srednje vrijednosti odgovora sudionika prikazane su također grafički na slici 16. Iz te je slike vidljivo da su se sudionici, nakon tvrdnje 11, najviše složili s tvrdnjom 15 (*Smatram da nastavnici dobro nadgledaju e-ispite PEJ-a.*), a potom s tvrdnjom 12 (*Nastavnici mi daju dovoljno informacija o e-ispitima PEJ-a.*); nasuprot tomu, nakon tvrdnje 18 najmanje su se složili s tvrdnjom 19 (*Mislim da me e-ispiti PEJ-a potiču da više čitam tekstove na engleskom*

jeziku.), a zatim s tvrdnjom 3 (Vjerujem da bodovi na e-ispitima PEJ-a uistinu odražavaju moju razinu poznavanja engleskog jezika.).



Slika 16. Srednje vrijednosti odgovora sudionika na Upitniku o stavovima o e-ispitima PEJ-a (N=120)

Zanimljivo je primijetiti da se tvrdnje s kojima su se sudionici najmanje složili prvenstveno odnose na povratni učinak ispita (čestice 17, 18 i 19), zatim na stav o tomu u kojoj mjeri e-ispit PEJ-a odražava njihovu razinu poznavanja engleskoga jezika (čestica 3) te na uvjete u kojima polažu taj ispit (čestica 14). S druge strane, tvrdnje s kojima su se najviše složili odnose se prvenstveno na transparentnost informacija o ispitu PEJ-a (čestice 6, 11 i 12), zatim na provedbu ispita (čestica 15) te na stav o tomu u kojoj mjeri ovaj ispit ispituje one jezične sposobnosti i jezične vještine koje im trebaju za studij i posao (čestica 4).

4.3.1.2.3. Odgovori na Upitniku o stavovima o ispitima općenito

Na Upitniku o stavovima o ispitima općenito (prilog 3.) sudionici su označivali koliko se slažu s ponuđenih sedamnaest tvrdnja, u rasponu vrijednosti od 1 do 5, s time da što je vrijednost veća, to je njihov stav o ispitima općenito pozitivniji, osim kod tvrdnja 9, 11 i 13, gdje je njihov stav pozitivniji ako je vrijednost manja (prilog 3.2.). Nakon što su se vrijednosti tih triju čestica rekodirale, pozitivniji stav o ispitima općenito bio je izražen većom vrijednosti, u skladu s ostalim česticama Upitnika (tablica 3). Kao što se može vidjeti iz te tablice, srednje su vrijednosti odgovora sudionika u rasponu od $M = 1,95$ do $M = 4,78$, pri čemu se najveća

odnosi na česticu 10 (*Znam kako se primjereno ponašati na ispitu.*), a najmanja na česticu 16 (*Više volim kolegije koji imaju mnogo ispita.*).

Tablica 3

Psihometrijska obilježja čestica Upitnika o stavovima o ispitima općenito ($N = 120$)

čestica	najmanja vrijednost	najveća vrijednost	<i>M</i>	<i>SD</i>	asimetričnost distribucije ^a	spljoštenost distribucije ^b	korelacija čestice i ukupnog rezultata	pouzdanost bez čestice
1.	1	5	3,67	0,989	-0,420	-0,596	,449	,791
2.	1	5	3,43	0,959	-0,679	0,118	,461	,790
3.	1	5	3,52	1,085	-0,667	-0,158	,503	,787
4.	1	5	3,39	1,087	-0,395	-0,419	,491	,787
5.	1	5	2,97	1,119	0,123	-0,738	,431	,792
6.	1	5	3,22	1,238	-0,260	-0,931	,322	,801
7.	1	5	3,48	1,020	-0,390	-0,362	,453	,790
8.	1	5	3,77	1,002	-0,942	0,774	,379	,795
9.*	1	5	2,77	1,193	0,101	-0,902	,297	,802
10.	2	5	4,78	0,510	-2,629	8,258	-,060	,812
11.*	1	5	3,27	0,959	-0,154	-0,032	,421	,793
12.	1	5	3,02	1,247	-0,021	-1,039	,479	,788
13.*	1	5	3,55	0,969	-0,228	-0,673	,100	,812
14.	1	5	2,04	0,991	0,601	-0,468	,465	,790
15.	1	5	3,46	0,869	-0,730	0,326	,543	,786
16.	1	4	1,95	0,969	0,665	-0,625	,339	,798
17.	1	5	3,01	0,930	-0,208	-0,297	,540	,786

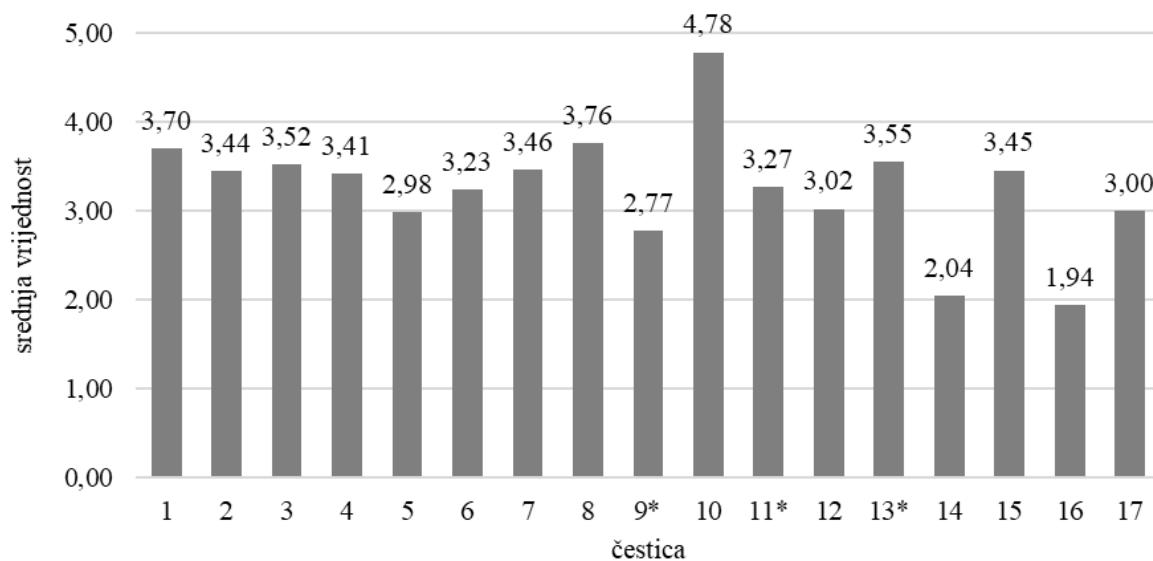
* Rekodirane čestice.

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Srednje vrijednosti odgovora sudionika na ovom upitniku prikazani su i grafički (v. sliku 17). Iz grafičkog je prikaza vidljivo da su se nakon tvrdnje 10 sudionici najviše složili s tvrdnjom 8 (*Tijekom ispita učim mnoge korisne stvari (npr. organizaciju, upravljanje vremenom, kako raditi brže itd.)*), a zatim s tvrdnjom 1 (*Ispiti me motiviraju da više učim.*); s

druge strane, nakon tvrdnje 16 najmanje su se složili s tvrdnjom 14 (*Uživam polagati ispite.*), a potom s tvrdnjom 9 (*Za mene je polaganje ispita bolno iskustvo.* – rekodirano).



* Rekodirane čestice.

Slika 17. Srednje vrijednosti odgovora sudionika na Upitniku o stavovima o ispitima općenito
($N = 120$)

Ono što je zanimljivo primijetiti iz ovih rezultata jest vrlo visoka srednja vrijednost za tvrdnju 10, koja se jedina odnosi na ponašanje ispitanika tijekom polaganja ispita, i koja se od druge najveće srednje vrijednosti (čestica 8) razlikuje za čak $M = 1,02$, druga se od treće (čestica 1) razlikuje za samo $M = 0,06$, a treća od četvrte (čestica 3) za samo $M = 0,18$. Usto, te se tri potonje čestice sve odnose na učenje i organiziranje vremena za učenje prije i tijekom ispita. Nasuprot tomu, tvrdnje s kojima su se sudionici najmanje složili odnose se na kolegije s mnogo ispita (čestica 16), na polaganje ispita (čestice 9 i 14) te naposljetku na uvid u širu sliku o onomu što se uči zahvaljujući ispitu (čestica 5).

4.3.1.2.4. Odgovori na Upitniku o stavovima o računalima

Na Upitniku o stavovima o računalima (prilog 4.) sudionici su označivali koliko se slažu s ponuđene 53 tvrdnje, a svoj stupanj slaganja mogli su izraziti vrijednostima od 1 do 5, pri čemu je veća značila negativniji stav o računalima, osim kod tvrdnja 7, 22, 25, 26, 27, 30, 33, 35, 36, 37, 38, 41, 47 i 49, gdje je negativniji stav bio izražen manjom vrijednosti (prilog 4.2.). Nakon što su se vrijednosti tih četrnaest čestica rekodirale, negativniji stav o računalima

bio je izražen većom vrijednosti, u skladu s ostalim česticama Upitnika. Njihov je stav podijeljen na dva dijela i prikazan u tablicama 4 i 5 niže, sukladno faktorskoj analizi (opisanoj u potpoglavlju 4.2.2.4. – Upitnik o stavovima o računalima):

- anksioznost vezana za računala (24 čestice: 1, 2, 5, 7, 8, 11, 15, 18, 19, 21, 22, 23, 25, 26, 28, 30, 33, 38, 41, 44, 48, 49, 50 i 53), pri čemu je veći rezultat označivao veću anksioznost;
- ovisnost o računalima (19 čestica: 9, 10, 12, 13, 17, 24, 29, 34, 35, 36, 37, 39, 40, 42, 43, 46, 47, 51 i 52), pri čemu je veći rezultat označivao veću ovisnost.

U tim se tablicama može primijetiti da nisu predstavljeni rezultati za deset čestica (čestice 3, 4, 6, 14, 16, 20, 27, 31, 32 i 45), a to je zato što se one nisu obrađivale jer su bile podjednako saturirane obama faktorima.

Kao što se može vidjeti iz tablice 4, srednje su vrijednosti odgovora sudionika na faktoru anksioznosti koja je vezana za računala u rasponu od $M = 1,38$ do $M = 3,18$, pri čemu se najveća odnosi na česticu 26 (*Ne brine me mogućnost neuspjeha kad radim neki zadatak na računalu.* – rekodirano), a najmanja na česticu 25 (*Važno mi je da budem vješt/-a na računalu.* – rekodirano). Što se tiče srednjih vrijednosti odgovora sudionika na faktoru ovisnosti o računalima (tablica 5), one su u rasponu od $M = 1,53$ do $M = 3,35$, pri čemu se najveća odnosi na česticu 35 (*Računala mi nisu važna u životu.* – rekodirano), a najmanja na česticu 13 (*Osjećam da nemam kontrolu kad se koristim računalom.*).

Tablica 4

Psihometrijska obilježja čestica Upitnika o stavovima o računalima –
faktor „anksioznost vezana uz računala“ ($N = 120$)

čestica	najmanja vrijednost	najveća vrijednost	<i>M</i>	<i>SD</i>	asimetričnost distribucije ^a	spljoštenost distribucije ^b	korelacija čestice i ukupnog rezultata	pouzdanost bez čestice
1.	1	5	2,05	1,099	0,788	-0,257	,518	,876
2.	1	5	1,63	1,037	1,616	1,788	,577	,875
5.	1	5	1,86	0,910	1,034	0,744	,538	,876
7.*	1	5	1,73	0,838	1,063	1,049	,452	,878
8.	1	4	1,48	0,830	1,579	1,401	,550	,876
11.	1	5	2,08	1,094	0,516	-0,961	,402	,879
15.	1	5	2,65	1,050	0,258	-0,197	,488	,877
18.	1	4	1,73	0,932	0,939	-0,339	,519	,876
19.	1	5	2,50	1,223	0,420	-0,733	,532	,876
21.	1	5	2,86	1,416	0,183	-1,216	,320	,883
22.*	1	5	1,93	1,153	1,117	0,327	,452	,878
23.	1	4	1,98	0,884	0,346	-0,998	,370	,880
25.*	1	4	1,38	0,662	1,711	2,248	,286	,882
26.*	1	5	3,18	1,243	-0,249	-0,770	,264	,884
28.	1	5	1,92	0,913	1,043	1,133	,452	,878
30.*	1	5	2,88	1,385	0,078	-1,189	,409	,880
33.*	1	5	2,44	1,346	0,473	-1,021	,552	,875
38.*	1	5	2,53	0,987	0,009	-0,339	,447	,878
41.*	1	5	2,70	1,241	0,188	-0,882	,340	,882
44.	1	5	1,95	1,158	1,154	0,493	,469	,877
48.	1	5	2,00	1,085	0,844	-0,114	,598	,874
49.*	1	5	1,78	0,842	0,860	0,475	,650	,874
50.	1	5	1,87	1,100	1,002	-0,116	,542	,875
53.	1	5	1,74	0,874	1,145	1,074	,523	,877

* Rekodirane čestice.

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Tablica 5

Psihometrijska obilježja čestica Upitnika o stavovima o računalima –
faktor „ovisnost o računalima“ (N = 120)

čestica	najmanja vrijednost	najveća vrijednost	M	SD	asimetričnost distribucije ^a	spljoštenost distribucije ^b	korelacija čestice i ukupnog rezultata	pouzdanost bez čestice
9.	1	5	2,24	1,335	0,775	-0,616	,305	,802
10.	1	5	1,97	1,115	0,955	0,081	,431	,794
12.	1	5	1,74	1,126	1,603	1,748	,556	,787
13.	1	5	1,53	0,961	2,007	3,575	,548	,789
17.	1	5	1,97	1,236	0,987	-0,285	,336	,799
24.	1	5	1,72	0,936	1,472	2,246	,508	,791
29.	1	5	1,65	0,984	1,453	1,390	,519	,790
34.	1	5	2,35	1,157	0,274	-0,899	,385	,796
35.*	1	5	3,35	1,294	-0,207	-1,051	,526	,787
36.*	1	5	3,00	1,341	0,000	-1,115	,529	,787
37.*	1	5	2,42	1,345	0,465	-1,056	,365	,798
39.	1	5	1,78	0,909	1,196	1,413	,462	,794
40.	1	5	2,78	1,325	0,226	-1,006	,425	,794
42.	1	5	1,67	0,929	1,292	0,955	,347	,799
43.	1	5	2,32	1,472	0,705	-0,940	,522	,787
46.	1	5	2,03	1,163	0,994	0,161	,492	,790
47.*	1	5	2,00	1,366	1,067	-0,289	-,140	,830
51.	1	5	1,78	1,070	1,448	1,577	,539	,788
52.	1	5	2,50	1,243	0,387	-0,739	-,144	,827

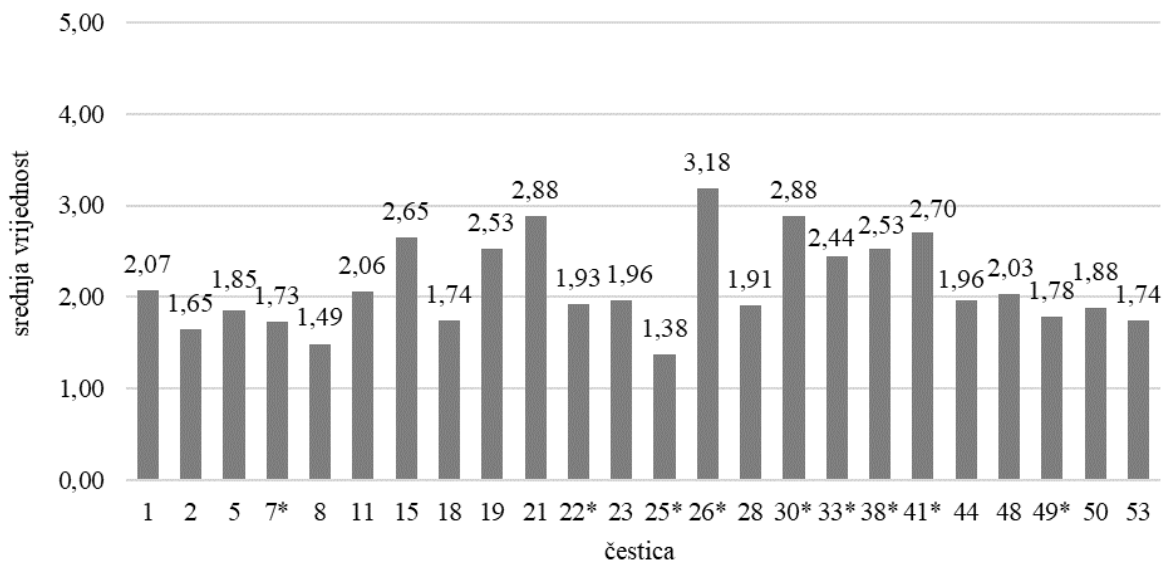
* Rekodirane čestice.

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Grafički prikazi srednjih vrijednosti odgovora sudionika na Upitniku o stavovima o računalima (v. slike 18 i 19) donose jasniji uvid u stavove sudionika o računalima. Na faktoru anksioznosti vezane za računala (slika 18) nakon tvrdnje 26 sudionici su se najviše složili s tvrdnjama 21 (*Radije se koristim olovkom i papirom nego programom poput Worda kad pripremam završnu inačicu nekog pisanog rada.*) i 30 (*Nikad nisam pokušao/-la izbjeći*

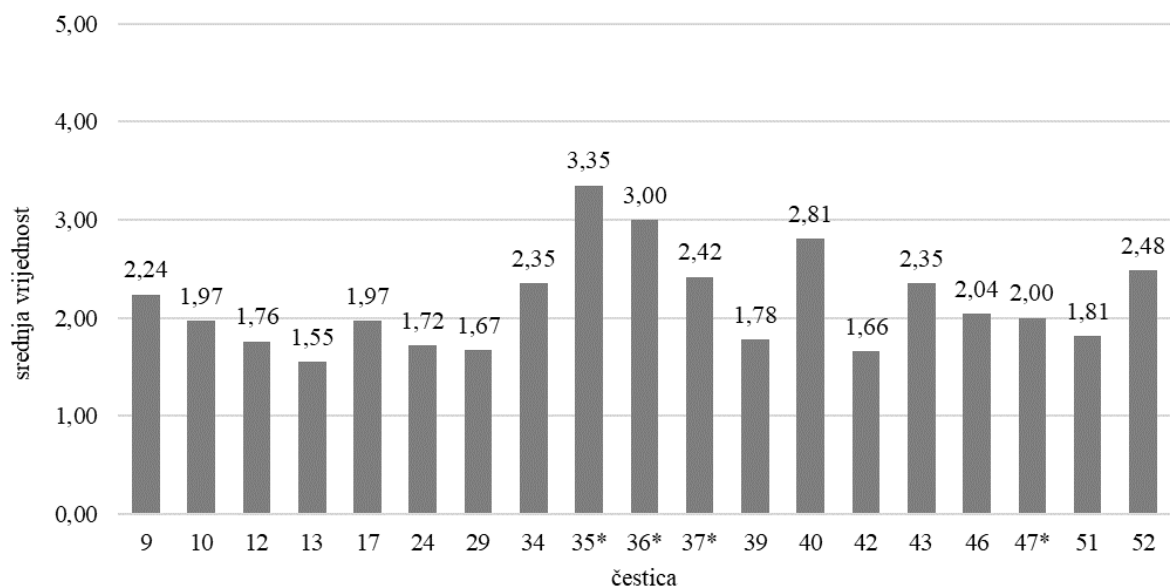
koristiti se računalom.); suprotno tomu, nakon tvrdnje 25 najmanje su se složili s tvrdnjom 8 (*Računala me straše.*), a potom s tvrdnjom 2 (*Poduzimam sve što mogu da se izbjegnem koristiti računalom.*). Uvidom u ove rezultate valja istaknuti kako, iako je većina sudionika samouvjereni prilikom rješavanja zadatka na računalu i važno im je da su u tomu vješti te se ne izbjegavaju koristiti njime, ipak se radije koriste papirom i olovkom dok pripremaju završne inačice pisanih radova.



* Rekodirane čestice.

Slika 18. Srednje vrijednosti odgovora sudionika na Upitniku o stavovima o računalima – faktor „anksioznost vezana za računala“ ($N = 120$)

Na faktoru ovisnosti o računalima (slika 19) nakon tvrdnje 35 najviše su se složili s tvrdnjom 36 (*Malen dio slobodna vremena provodim na računalu.* – rekodirano), zatim s tvrdnjom 40 (*Aktivnosti na računalu ponekad ometaju moje obveze (studiranje, posao i dr.).*); s druge strane, nakon tvrdnje 13 najmanje su se složili s tvrdnjom 42 (*Kad se ne koristim računalom, često se osjećam nemirno.*), a potom s tvrdnjom 29 (*Kad vidim računalo, osjećam kako me privlači.*). Uvidom u prethodno navedene čestice s kojima se složila većina sudionika, primjećujemo da su im računala važna u životu i da na njima provode puno slobodnog vremena, što ponekad ometa njihove obveze, no kad se pogledaju navedene čestice s kojima su se najmanje složili, vidimo da sudionike istovremeno računala ne privlače kad ih vide te da tijekom rada na njima osjećaju kontrolu, a ne nemir.



* Rekodirane čestice.

Slika 19. Srednje vrijednosti odgovora sudionika na upitniku o stavovima o računalima – faktor „ovisnost o računalima“ (N = 120)

4.3.1.2.5. Odgovori na Upitniku o razini ispitne anksioznosti

Na Upitniku o razini ispitne anksioznosti (prilog 5.) sudionici su označivali koliko se slažu s ponuđenih deset tvrdnja, odabравši pritom vrijednosti od 1 do 5, pri čemu veća vrijednost označuje veću razinu ispitne anksioznosti. Njihovi su odgovori iscrpnije prikazani u prilogu 5.2., dok tablica 6 niže prikazuje psihometrijska obilježja čestica Upitnika. Razvidno je da su srednje vrijednosti svih čestica u gornjoj polovici mogućih vrijednosti, i to od $M = 2,88$ do $M = 4,20$, pri čemu je najveća srednja vrijednost odgovora sudionika na čestici 2 (*Kad učim za ispit, brinem se da se neću sjetiti gradiva na samom ispitu.*), a najmanja na čestici 10 (*Borim se s pisanim zadacima ili ih izbjegavam jer smatram da štogod napravio/napravila, to neće biti dovoljno dobro. Želim da to bude savršeno.*).

Tablica 6

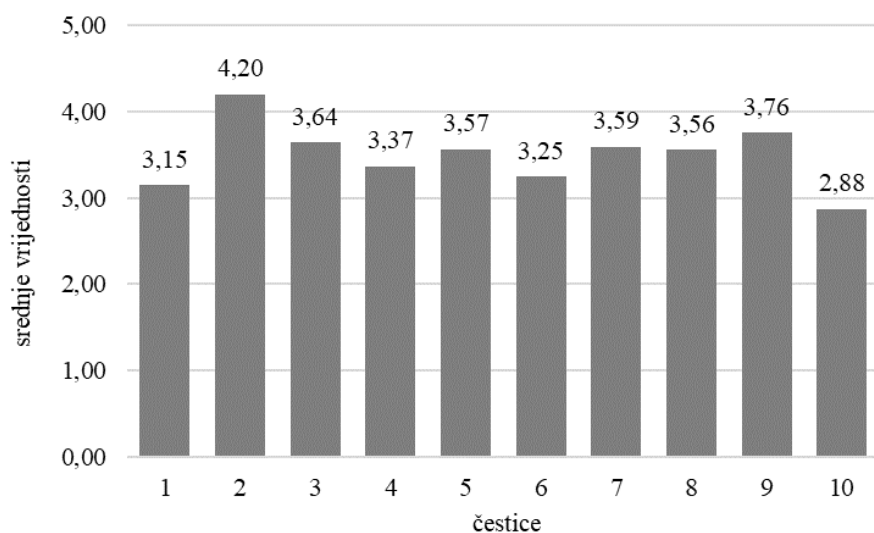
Psihometrijska obilježja čestica Upitnika o razini ispitne anksioznosti ($N = 120$)

čestica	najmanja vrijednost	najveća vrijednost	M	SD	asimetričnost distribucije ^a	spljoštenost distribucije ^b	korelacija čestice i ukupnog rezultata	pouzdanost bez čestice
1.	1	5	3,16	1,223	-0,112	-0,939	,436	,860
2.	1	5	4,20	,940	-1,399	2,096	,640	,844
3.	1	5	3,62	1,138	-0,680	-0,260	,696	,838
4.	1	5	3,35	1,120	-0,255	-0,634	,770	,831
5.	1	5	3,57	1,098	-0,502	-0,466	,575	,848
6.	1	5	3,24	1,264	-0,112	-1,123	,766	,830
7.	1	5	3,59	1,192	-0,494	-0,760	,611	,845
8.	1	5	3,58	1,228	-0,495	-0,927	,319	,870
9.	1	5	3,77	1,150	-0,978	0,242	,364	,865
10.	1	5	2,88	1,142	0,284	-0,648	,600	,846

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Iz grafičkog prikaza srednjih vrijednosti odgovora sudionika na ovom upitniku (slika 20) vidljivo je da su se nakon tvrdnje 2 sudionici najviše složili s tvrdnjom 9 (*Nakon ispita brinem se o tomu jesam li dao/dala sve od sebe.*), a zatim s tvrdnjom 3 (*Za vrijeme važnih ispita smatram da mi ide loše i da ću pasti.*); nasuprot tomu, nakon tvrdnje 10 najmanje su se složili s tvrdnjom 1 (*Što sam bliže ispitu, teže mi je usredotočiti se na gradivo.*), a potom s tvrdnjom 6 (*Toliko se brinem prije važna ispita da sam toliko izmoren(a) da ne mogu dati sve od sebe na ispitu*). Rezultati pokazuju da je većina sudionika anksiozna i prilikom učenja za ispit, i tijekom polaganja ispita, i nakon ispita, no da ne gube usredotočenost niti se više brinu kako se ispit približava, a tijekom polaganja ne izbjegavaju zadatke pisanja težeći savršenstvu.



Slika 20. Srednje vrijednosti odgovora sudionika na Upitniku o razini ispitne anksioznosti ($N = 120$)

4.3.1.3. *Formiranje kompozitnih varijabla, deskriptivni rezultati i normalitet distribucija kompozitnih varijabla*

U našem istraživanju oblikovano je nekoliko kompozitnih varijabla. Stav sudionika prema e-ispitima oblikovan je kao prosječna vrijednost na česticama upitnika kojima se ispitivao ovaj objekt stava. Na isti su način izračunani rezultati za varijable ukupnih rezultata stavova o ispitima općenito i računalima te ukupni rezultat na ispitnoj anksioznosti.

Prije provedbe analiza potrebnih za odgovor na postavljena istraživačka pitanja analizirani su deskriptivni pokazatelji kompozitnih varijabla primijenjenih u ovom istraživanju (tablica 7). Osim navedenih kompozitnih varijabla primijenjene su i ostale varijable izravnih rezultata (nisu kompozitne), a čija su metrijska obilježja navedena u tablici 8.

Kako su na Upitniku o stavovima o e-ispitima Poslovnog engleskog jezika (PEJ) mogući odgovori sudionika sezali u rasponu od 1 do 6, a na česticama preostalih kompozitnih mjera od 1 do 5, primjećujemo da su prosječne vrijednosti kompozitne varijable stavova prema e-ispitima PEJ-a iznadprosječne ($M = 4,78$; $SD = ,752$), što upućuje na općenito pozitivan stav sudionika prema ovakvu načinu ispitivanja. Prosječna vrijednost stavova prema ispitima općenito bliža je središnjoj vrijednosti skale ($M = 3,25$; $SD = ,502$), što upućuje na srednje pozitivan stav sudionika prema ispitima općenito. Stavovi prema računalima u prosjeku su ispod središnje vrijednosti skale, što upućuje na sniženu anksioznost sudionika vezanu za korištenje računalima ($M = 2,12$; $SD = ,556$), kao i nisku ovisnost o računalima ($M = 2,15$;

$SD = ,531$). Nadalje, sudionici izražavaju umjerene vrijednosti ispitne anksioznosti ($M = 3,5$; $SD = ,767$).

Uvidom u podatke o asimetričnosti i spljoštenosti distribucija kompozitnih varijabla, ali i histograme distribucija kompozitnih varijabla (prilog 6.) uočavamo da su distribucije simetrične, što je i u skladu s ranije analiziranim prosječnim vrijednostima. Analizom podataka o asimetričnosti i spljoštenosti distribucija iz tablice 7, gdje je vidljivo da nema vrijednosti koje prelaze $\pm 1,5$, zaključujemo da distribucije ne odstupaju od normaliteta. Formalni statistički testovi normaliteta distribucije nisu primaran izbor zbog velike strogosti koja dovodi do odbacivanja nul-hipoteze i u slučajevima vrlo malenih odstupanja distribucije od normalne (Tabachnick i Fidell, 2013), već su u slučaju većih uzoraka važniji veličina indeksa asimetričnosti i oblik distribucije nego razina statističke značajnosti. Kako navodi Waternaux (1976, prema Tabachnick i Fidell, 2013), pozitivna ili negativna odstupanja u spljoštenosti distribucije dovode do podcjenjivanja varijance, no taj učinak pozitivne spljoštenosti nestaje s više o 100 sudionika, a učinak negativne spljoštenosti nestaje u uzorcima većim od 200. Gravetter i Wallnau (2014) te George i Mallery (2010) ističu kao prihvatljiv raspon vrijednosti asimetričnosti i spljoštenosti distribucije ± 2 kako bismo distribuciju mogli smatrati normalnom. Stoga možemo reći da su rezultati kompozitnih varijabla primijenjenih u ovom istraživanju normalno distribuirani. Pouzdanosti mjera stavova prema ispitima općenito i prema računalima pokazuju zadovoljavajuću pouzdanost iskazanu Cronbachovim alfa-koeficijentom.

Tablica 7

Deskriptivna statistika kompozitnih varijabla primijenjenih u istraživanju ($N = 120$)

	najmanja vrijednost	najveća vrijednost	<i>M</i>	<i>SD</i>	asimetričnost distribucije ^a	spljoštenost distribucije _b	Cronbach <i>α</i>
stavovi o e-ispitu - ukupno	2,75	6,00	4,7788	0,75213	-0,527	-0,409	,903
stavovi o ispitima općenito - ukupno	2,00	4,59	3,2525	0,50206	-0,010	-0,250	,804
stavovi o računalima - anksioznost	0,96	3,88	2,1123	0,53370	0,434	0,066	,855
stavovi o računalima - ovisnost	1,05	4,47	2,1461	0,57903	0,815	1,169	,822
razina ispitne anksioznosti	1,00	5,00	3,4950	0,76859	-0,446	0,057	,861

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Uvidom u tablicu 8 gdje su sadržane deskriptivne vrijednosti preostalih (nekompozitnih) varijabla primijenjenih u istraživanju, uočavamo da mjerilima Geoga i Malleryja (2010) ne udovoljavaju dvije varijable (budući da im vrijednosti spljoštenosti distribucije premašuju ± 2), a to su „ispit na papiru - pisana proizvodnja - procjenjivač A - jezik“ i „ispit na papiru - razumijevanje i primjena vokabulara“. Stoga možemo reći da se spomenute dvije varijable ne distribuiraju normalno. Međutim, znatne razlike u distribucijama rezultata usporednih varijabla procjenjivača A i procjenjivača B navode nas na zaključak o mogućim znatnijim razlikama u procjenama ovo dvoje procjenjivača. Stoga uspoređujemo rezultate dvoje procjenjivača na korespondentnim mjerama (usporedni histogrami u prilogu 1.2.). Uočavamo određene sličnosti, ali i određene razlike. Prvo, zabrinjavajuća je distribucija varijabla „e-ispit - pisana proizvodnja - oblik“ koja nalikuje bimodalnoj, što može remetiti zaključke daljnjih statističkih analiza. Varijabla „e-ispit - pisana proizvodnja - sadržaj“ kod procjenjivača A pokazuje manje kontinuiran varijabilitet, odnosno nije jednolična distribucija kao kod procjenjivača B. Isto je zamijećeno i kod „ispit na papiru - pisana proizvodnja - sadržaj“, pa možemo zaključiti da je riječ o stabilnoj razlici u procjenjivanju sadržaja istog tipa kod procjenjivača A i B.

Tablica 8

Deskriptivna statistika varijabla primijenjenih u istraživanju ($N = 120$)

	najmanja vrijednost	najveća vrijednost	<i>M</i>	<i>SD</i>	asimetričnost distribucije ^a	spljoštenost distribucije ^b
e-ispit - pisana proizvodnja - procjenjivač A - oblik	0,00	8,50	4,229	2,3594	-0,102	-1,107
e-ispit - pisana proizvodnja - procjenjivač A - sadržaj	0,00	6,00	3,700	1,0874	-0,592	0,893
e-ispit - pisana proizvodnja - procjenjivač A - jezik	0,00	5,00	2,879	,9830	-0,364	0,531
e-ispit - pisana proizvodnja - procjenjivač A - ukupno	1,00	17,00	10,808	3,3388	-0,411	-0,227
e-ispit - pisana proizvodnja - procjenjivač B - oblik	0,00	8,50	4,129	2,3023	-0,096	-1,240
e-ispit - pisana proizvodnja - procjenjivač B - sadržaj	0,00	6,00	3,908	1,2451	-0,312	0,362
e-ispit - pisana proizvodnja - procjenjivač B - jezik	0,00	5,00	2,829	1,0720	-0,503	0,169
e-ispit - pisana proizvodnja - procjenjivač B - ukupno	1,00	18,00	10,867	3,4132	-0,347	-0,063
ispit na papiru - pis. proizv. - procjenjivač A - oblik	0,00	8,00	3,550	1,9655	0,009	-0,514
ispit na papiru - pis. proizv. - procjenjivač A - sadržaj	0,00	6,00	3,404	1,3320	-0,561	0,418
ispit na papiru - pis. proizv. - procjenjivač A - jezik	0,00	5,00	3,075	0,9715	-1,466	3,132
ispit na papiru - pis. proizv. - procjenjivač A - ukupno	0,00	16,50	10,029	3,4454	-0,985	1,696
ispit na papiru - pis. proizv. - procjenjivač B - oblik	0,00	7,50	3,408	1,8618	-0,029	-0,609
ispit na papiru - pis. proizv. - procjenjivač B - sadržaj	0,00	6,00	3,483	1,4141	-0,519	0,168
ispit na papiru - pis. proizv. - procjenjivač B - jezik	0,00	5,00	2,821	1,1054	-0,703	0,458
ispit na papiru - pis. proizv. - procjenjivač B - ukupno	0,00	16,50	9,713	3,6065	-0,845	0,792
e-ispit – pisana proizvodnja - ukupno	1,25	17,00	10,8375	3,30547	-0,436	-0,093
e-ispit - razumijevanje čitanjem	1,00	5,00	3,6250	1,07736	-0,432	-0,472
e-ispit - razumijevanje i primjena vokabulara	3,00	15,00	11,4333	2,73395	-0,993	0,775
e-ispit - ukupno	10,50	34,00	25,8958	5,06518	-0,620	0,133
ispit na papiru - pisana proizvodnja - ukupno	0,00	16,50	9,8708	3,47318	-0,974	1,345
ispit na papiru - razumijevanje čitanjem	1,00	5,00	3,5500	1,06787	-0,280	-0,859
ispit na papiru - razumijevanje i primjena vokabulara	2,00	17,00	12,1000	2,31691	-1,076	2,434
ispit na papiru - ukupno	5,00	34,50	25,5208	5,44294	-1,105	1,791

^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,221.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 0,438.

Analizirane su korelacije ocjena procjenjivača A i B na korespondentnim mjerama (tablica 9) te su provedeni t-testovi za ponovljena mjerenja kako bismo utvrdili postojanje razlika u procjenama u pojedinim kategorijama. Iako su korelacije većinom vrlo visoke, kakve i trebaju biti, odstupaju varijable „e-ispit - pisana proizvodnja - jezik“ i „ispit na papiru - pisana proizvodnja - jezik“, gdje su korelacije između dvaju procjenjivača znatno niže. Kako je riječ o istom tipu sadržaja, možemo zaključiti da u procjeni jezika dolazi do veće subjektivnosti procjena. Provedenim t-testovima za ponovljena mjerenja (zavisne uzorke) utvrđeno je statistički značajno razlikovanje dvaju procjenjivača na četirima varijablama. Na varijabli „e-ispit - pisana proizvodnja - sadržaj“ ($t = -2,795$; $df = 120$; $p < ,01$) procjene procjenjivača B su bile statistički značajno više ($M = 3,876$; $SD = 1,290$) nego procjene procjenjivača A ($M = 3,669$; $SD = 1,34$). Na varijabli „ispit na papiru - pisana proizvodnja - oblik“ ($t = 2,609$; $df = 120$; $p < ,01$) procjene procjenjivača A su bile statistički značajno više ($M = 3,525$; $SD = 1,977$) nego procjene procjenjivača B ($M = 3,384$; $SD = 1,873$). Na varijabli „ispit na papiru - pisana proizvodnja - jezik“ ($t = 3,742$; $df = 120$; $p < ,01$) procjene procjenjivača B su bile statistički značajno niže ($M = 2,822$; $SD = 1,101$) nego procjene procjenjivača A ($M = 3,079$; $SD = 0,968$). Procjenjivači su se razlikovali i na ukupnoj mjeri pisanog ispita „ispit na papiru - pisana proizvodnja - ukupno“ ($t = 2,899$; $df = 120$; $p < ,01$), pri čemu je procjenjivač A imao više procjene ($M = 10,004$; $SD = 3,422$) u odnosu na procjenjivača B ($M = 9,682$; $SD = 3,607$).

Tablica 9

Korelacije korespondentnih procjena procjenjivača A i B ($N = 120$)

		korelacija	p
Par 1	e-ispit - pisana proizvodnja - procjenjivač A - oblik i e-ispit - pisana proizvodnja - procjenjivač B - oblik	,968	,000
Par 2	e-ispit - pisana proizvodnja - procjenjivač A - sadržaj i e-ispit - pisana proizvodnja - procjenjivač B - sadržaj	,782	,000
Par 3	e-ispit - pisana proizvodnja - procjenjivač A - jezik i e-ispit - pisana proizvodnja - procjenjivač B - jezik	,768	,000
Par 4	e-ispit - pisana proizvodnja - procjenjivač A - ukupno i e-ispit - pisana proizvodnja - procjenjivač B - ukupno	,924	,000
Par 5	ispit na papiru - pisana proizvodnja - procjenjivač A - oblik i ispit na papiru - pisana proizvodnja - procjenjivač B - oblik	,954	,000
Par 6	ispit na papiru - pisana proizvodnja - procjenjivač A - sadržaj i ispit na papiru - pisana proizvodnja - procjenjivač B - sadržaj	,873	,000
Par 7	ispit na papiru - pisana proizvodnja - procjenjivač A - jezik i ispit na papiru - pisana proizvodnja - procjenjivač B - jezik	,742	,000
Par 8	ispit na papiru - pisana proizvodnja - procjenjivač A - ukupno i ispit na papiru - pisana proizvodnja - procjenjivač B - ukupno	,941	,000

Budući da nismo utvrdili podudarnost procjena dvaju procjenjivača, primijenio se unutarklasni koeficijent korelacije (eng. *Intraclass Correlation Coefficient*) – ICC (tablica 10). Prema Koo i Lee (2016), vrijednosti ICC-a ispod 0,5 označuju nisku pouzdanost, između 0,5 i 0,75 umjerenu, između 0,75 i 0,9 dobru, a iznad 0,9 izvrsnu. Ova analiza pokazuje rezultate većinom u skladu s rezultatima provedenih t-testova, te je podudarnost procjena dvaju procjenjivača izvrsna ili dobra na svim varijablama osim na varijabli „ispit na papiru - jezik“, gdje je ona umjerena. Usto, tri varijable s najnižom podudarnosti procjenjivača („e-ispit - sadržaj“, „e-ispit - jezik“ i „ispit na papiru - jezik“) ujedno su varijable na kojima su utvrđene nešto niže korelacije među procjenama procjenjivača A i B, ali ipak zadovoljavajuće – drugim riječima, može se reći da su procjene u cijelosti gledano pouzdane.

Tablica 10

Unutarklasni koeficijent korelacije za ispitivanje podudarnosti procjena procjenjivača A i B

	ICC	Interval pouzdanosti 95%		F	df1	df2	p
		donja granica	gornja granica				
e-ispit - pisana proizvodnja - oblik	0,968	,954	,978	62,612	120	120	,000
e-ispit - pisana proizvodnja - sadržaj	0,783	,698	,846	8,629	120	120	,000
e-ispit - pisana proizvodnja - jezik	0,780	,699	,841	8,052	120	120	,000
e-ispit - pisana proizvodnja - ukupno	0,930	,901	,950	27,210	120	120	,000
ispit na papiru - pisana proizvodnja - oblik	0,952	,930	,967	42,364	120	120	,000
ispit na papiru - pisana proizvodnja - sadržaj	0,878	,829	,913	15,380	120	120	,000
ispit na papiru - pisana proizvodnja - jezik	0,733	,618	,814	7,080	120	120	,000
ispit na papiru - pisana proizvodnja - ukupno	0,940	,913	,959	34,375	120	120	,000

4.3.2. Bivarijatne povezanosti varijabla primijenjenih u istraživanju

Uvid u bivarijatne povezanosti varijabla primijenjenih u istraživanju preduvjet je provedbe analiza kojima se testiraju hipoteze. U tablici 11 mogu se vidjeti bivarijatne povezanosti kompozitnih varijabla primijenjenih u našem istraživanju. Budući da su varijable normalno distribuirane, a one koje odstupaju od normaliteta ne odstupaju značajno, primijenjen je Pearsonov koeficijent korelacije. Vidimo da su obrasci povezanosti očekivani (npr. koreliraju međusobno različiti aspekti poznavanja engleskog jezika, ispitivani različitim

metodama), a zapažene povezanosti nisu previsoke. Stoga možemo reći da su zadovoljeni preduvjeti za provođenje daljnjih analiza. U tablici 11 niže rabe se skraćeni nazivi varijabla (pisanje = zadatak pisane proizvodnje; papir = ispit na papiru).

Tablica 11

Bivarijatne povezanosti varijabla primijenjenih u istraživanju ($N = 120$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. e-ispit - pisanje - ukupno	1													
2. e-ispit - razumij. čitanjem	,176	1												
3. e-ispit - raz. i primj. vokabulara	,175	,284**	1											
4. e-ispit - ukupno	,785**	,481**	,714**	1										
5. papir - pisanje - ukupno	,457**	,181*	,028	,352**	1									
6. papir - razumij. čitanjem	,250**	,298**	,303**	,390**	,227*	1								
7. papir - raz. i primj. vokabulara	,370**	,274**	,386**	,508**	,477**	,344**	1							
8. papir - ukupno	,498**	,291**	,242**	,517**	,885**	,487**	,797**	1						
9. e-ispit - pisanje - oblik	,865**	,080	,058	,613**	,413**	,201*	,328**	,443**	1					
10. e-ispit - pisanje - sadržaj	,564**	,211*	,200*	,521**	,250**	,197*	,170	,270**	,150	1				
11. e-ispit - pisanje - jezik	,715**	,172	,235**	,630**	,293**	,152	,287**	,339**	,398**	,440**	1			
12. papir - pisanje - oblik	-,037	-,128	-,075	-,092	-,141	-,082	-,082	-,141	,006	-,042	,091	1		
13. papir - pisanje - sadržaj	,348**	,255**	,135	,355**	,808**	,272**	,361**	,723**	,261**	,325**	,198*	,443**	1	
14. papir - pisanje - jezik	,354**	,192*	,162	,359**	,807**	,225*	,462**	,756**	,231*	,233*	,398**	,505**	,661**	1

** $p < ,01$, * $p < ,05$

Kako bismo provjerili oblik povezanosti, za kompozitne varijable koje imaju ne-nulte povezanosti napravljeni su dijagrami raspršenja²⁹, koji upućuju na linearne povezanosti. Kako bismo provjerili postojanje multikolinearnosti, primijenjen je VIF-pokazatelj (eng. *variance inflation factor*). Kako su svi VIF-pokazatelji imali vrijednost ispod 2,5, a graničnom vrijednosti indikativnom za multikolineranost smatra se 3 (Field, 2013), zaključujemo kako ne postoji problem multikolinearnosti. U prilog tomu govori i nepostojanje visokih koeficijenata korelacije u tablici 12. Problem singulariteta također nije prisutan budući da kompoziti nisu formirani tako da uključuju druge kompozite ili njihove elemente. U tablici 12 niže rabe se skraćeni nazivi varijabla (pisanje = zadatak pisane proizvodnje; papir = ispit na papiru).

Tablica 12

Bivarijatne povezanosti kompozitnih varijabla primijenjenih u istraživanju ($N = 120$)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. e-ispit - pisanje - ukupno	1												
2. e-ispit - razumij. čitanjem	,176	1											
3. e-ispit - raz. i primj. vokabulara	,175	,284**	1										
4. e-ispit - ukupno	,785**	,481**	,714**	1									
5. papir - pisanje - ukupno	,457**	,181*	,028	,352**	1								
6. papir - razumij. čitanjem	,250**	,298**	,303**	,390**	,227*	1							
7. papir - raz. i primj. vokabulara	,370**	,274**	,386**	,508**	,477**	,344**	1						
8. papir - ukupno	,498**	,291**	,242**	,517**	,885**	,487**	,797**	1					
9. stavovi o e-ispitima PEJ-a - ukupno	,045	-,001	,009	,034	,006	-,196*	-,024	-,045	1				
10. stavovi o ispitima općenito - ukupno	,173	,150	,008	,149	,150	,111	,075	,150	,269**	1			
11. stavovi o računalima - anksioznost	,104	,044	,072	,116	-,010	-,100	-,025	,037	-,154	-,104	1		
12. stavovi o računalima - ovisnost	-,057	-,121	-,080	-,107	-,143	-,092	-,086	-,146	,003	-,067	,092	1	
13. razina ispitne anksioznosti - ukupno	-,066	-,205*	-,122	-,153	-,096	-,200*	-,109	-,147	-,133	-,306**	,392**	,164	1

** $p < ,01$, * $p < ,05$

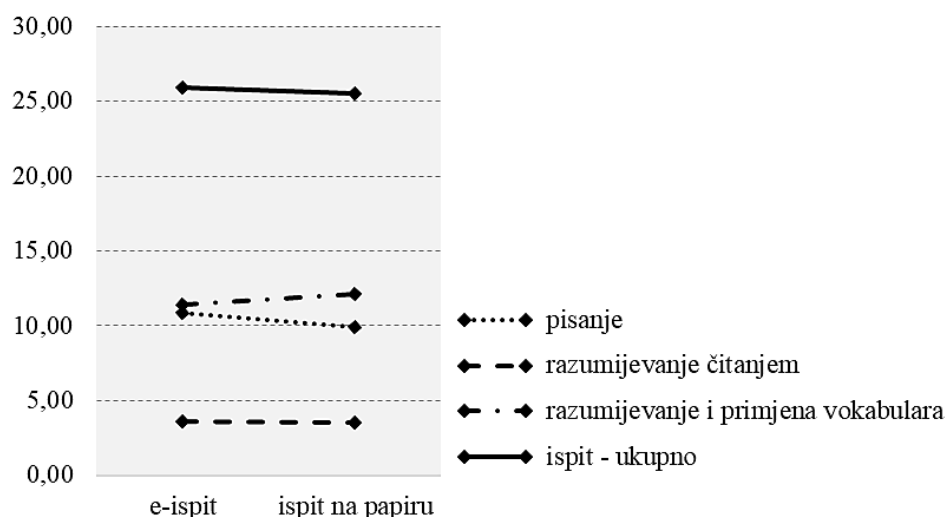
²⁹ Dijagrami raspršenja ne prikazuju se u radu jer ne stoje u fokusu analiza, a njihov bi cjelovit prikaz zauzimao previše prostora u radu.

4.3.3. Testiranje hipoteza

Prvo istraživačko pitanje odnosilo se na to ispituje li se e-ispitima sposobnost razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u engleskomu kao jeziku struke (EJS). Pretpostavili smo da se e-ispitom ispituje sposobnost razumijevanja čitanjem u EJS-u (H1A), sposobnost razumijevanja i primjene vokabulara u EJS-u (H1B) te sposobnost pisane proizvodnje u EJS-u (H1C). Niske do umjerene povezanosti između odgovarajućih dijelova e-ispita i ispita na papiru (istaknuto masnim brojevima u tablici 12), upućuju na zaključak da se ne ispituju u potpunosti iste stvari, već dio varijance rezultata ostaje neobjašnjen. Navedeni rezultati potaknuli su provedbu dodatne provjere razlika.

Kao dodatna provjera razlika u rezultatima između e-ispita i ispita na papiru provedena je dvosmjerna analiza varijance za ponovljena mjerenja. Utvrđeno je da glavni učinak načina ispitivanja (e-ispit ili ispit na papiru) nije statistički značajan ($F(1, 216,627) = 7,819, p = ,001, \eta^2_p = ,005$). Drugi glavni učinak bio je dio ispita, odnosno njegov sadržaj (pisanje, razumijevanje čitanjem, razumijevanje i primjena vokabulara, ispit ukupno). Ovaj se faktor pokazao statistički značajnim ($F(1,805, 216,627) = 2246,748, p = ,001, \eta^2_p = ,949$), kao i međudjelovanje spomenutih dvaju faktora ($F(2,106, 252,707) = 6,554, p = ,001, \eta^2_p = ,052$). Grafički prikaz glavnih učinaka i međudjelovanja može se vidjeti na slici 21 niže.

Rezultati provedene analize varijance idu u prilog hipotezama H1A, H1B i H1C da e-ispit i ispit na papiru ispituju iste stvari budući da nije utvrđeno postojanje glavnog učinka načina ispitivanja. Nepostojanje glavnog učinka načina ispitivanja upućuje na zaključak da za ishod nije značajno ispituje li se sadržaj e-ispitom ili ispitom na papiru. Statistički značajan faktor sadržaja ispita govori nam da sudionici nisu bili podjednako uspješni u raznim ispitivanim područjima poznavanja jezika (razumijevanje čitanjem, razumijevanje i primjena vokabulara, pisana proizvodnja te ispit ukupno). Statistički značajno međudjelovanje dvaju glavnih učinaka govori nam da odnos uspješnosti u pojedinim dijelovima ispita nije jednak u ispitu na papiru i e-ispitu, a što je vidljivo i iz slike 21.



Slika 21. Grafički prikaz glavnih učinaka sadržaja ispita i načina ispitivanja te njihovo međudjelovanje (N=120)

Sukladno grafičkom prikazu na slici 21, provedeni t-testovi za zavisne uzorke dovode do zaključka da postoje razlike u uspješnosti istih sudionika na mjerama „pisanje - ukupno“ ($t = 2,996$; $df = 119$; $p < 0,01$) i „razumijevanje i primjena vokabulara“ ($t = -2,589$; $df = 119$; $p < 0,05$). Na mjeri „pisanje - ukupno“ ispitanici postižu bolji rezultat na e-ispitu ($M = 10,837$; $SD = 3,305$) u odnosu na ispit na papiru ($M = 9,871$; $SD = 3,473$). Na mjeri „razumijevanje i primjena vokabulara“ sudionici postižu slabiji rezultat na e-ispitu ($M = 11,433$; $SD = 2,734$) u odnosu na ispit na papiru ($M = 12,1$; $SD = 2,317$).

Drugo istraživačko pitanje ispitivalo je jesu li rezultati na e-ispitu povezani sa stavovima sudionika o e-ispitima Poslovnog engleskog jezika, o ispitima općenito i o računalima, a treće istraživačko pitanje jesu li rezultati na e-ispitima povezani s ispitnom anksioznosti. Stoga su ispitane bivarijatne povezanosti koje možemo iščitati iz ranije navedene tablice 12. Uočavamo da stavovi sudionika o e-ispitima Poslovnog engleskog jezika, o ispitima općenito, o računalima, kao ni ispitna anksioznost, ne koreliraju statistički značajno niti s jednim dijelom e-ispita osim s dijelom razumijevanja čitanjem. Utvrđena je povezanost nezatna i negativna³⁰ ($r = -,205$; $p < ,05$; u tablici 12 istaknuto kurzivom), pri čemu osobe niže ispitne anksioznosti postižu bolje rezultate na ovom dijelu ispita. Stoga možemo reći da su hipoteze H2A („Rezultati na e-ispitu nisu povezani sa stavovima sudionika o e-ispitima.“),

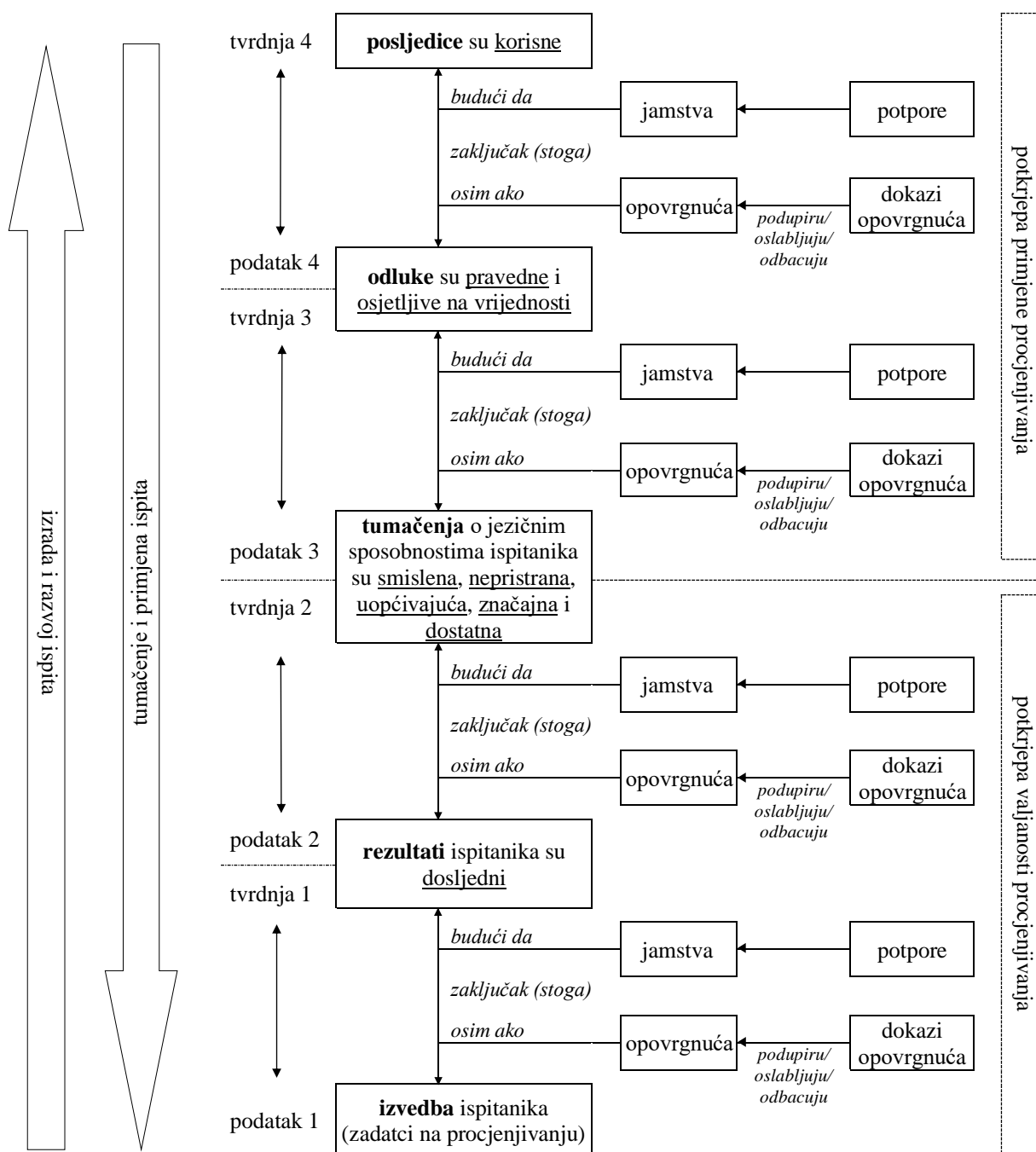
³⁰ Hrvatska enciklopedija (2021) za opis korelacije rabi sljedeće izraze: -1 = potpuna i negativna; od -1 do $-0,75$ = vrlo dobra do izvrsna i negativna; od $-0,75$ do $-0,50$ = umjerena i negativna; od $-0,50$ do $-0,25$ = slaba i negativna; od $-0,25$ do 0 = nezatna i negativna; od 0 = nema korelacije; od 0 do $0,25$ = nezatna i pozitivna; od $0,25$ do $0,50$ = slaba i pozitivna; od $0,50$ do $0,75$ = umjerena i pozitivna; od $0,75$ do 1 = vrlo dobra do izvrsna i pozitivna; $+1$ = potpuna i pozitivna.

H2B („Rezultati na e-ispitu nisu povezani sa stavovima sudionika o ispitima općenito.“) i H2C („Rezultati na e-ispitu nisu povezani sa stavovima sudionika o računalima.“) potvrđene, a da je hipoteza H3 („Rezultati na e-ispitu nisu povezani s razinom ispitne anksioznosti sudionika). uglavnom potvrđena.

4.3.4. Potkrjepa valjanosti e-procjenjivanja u engleskomu kao jeziku struke

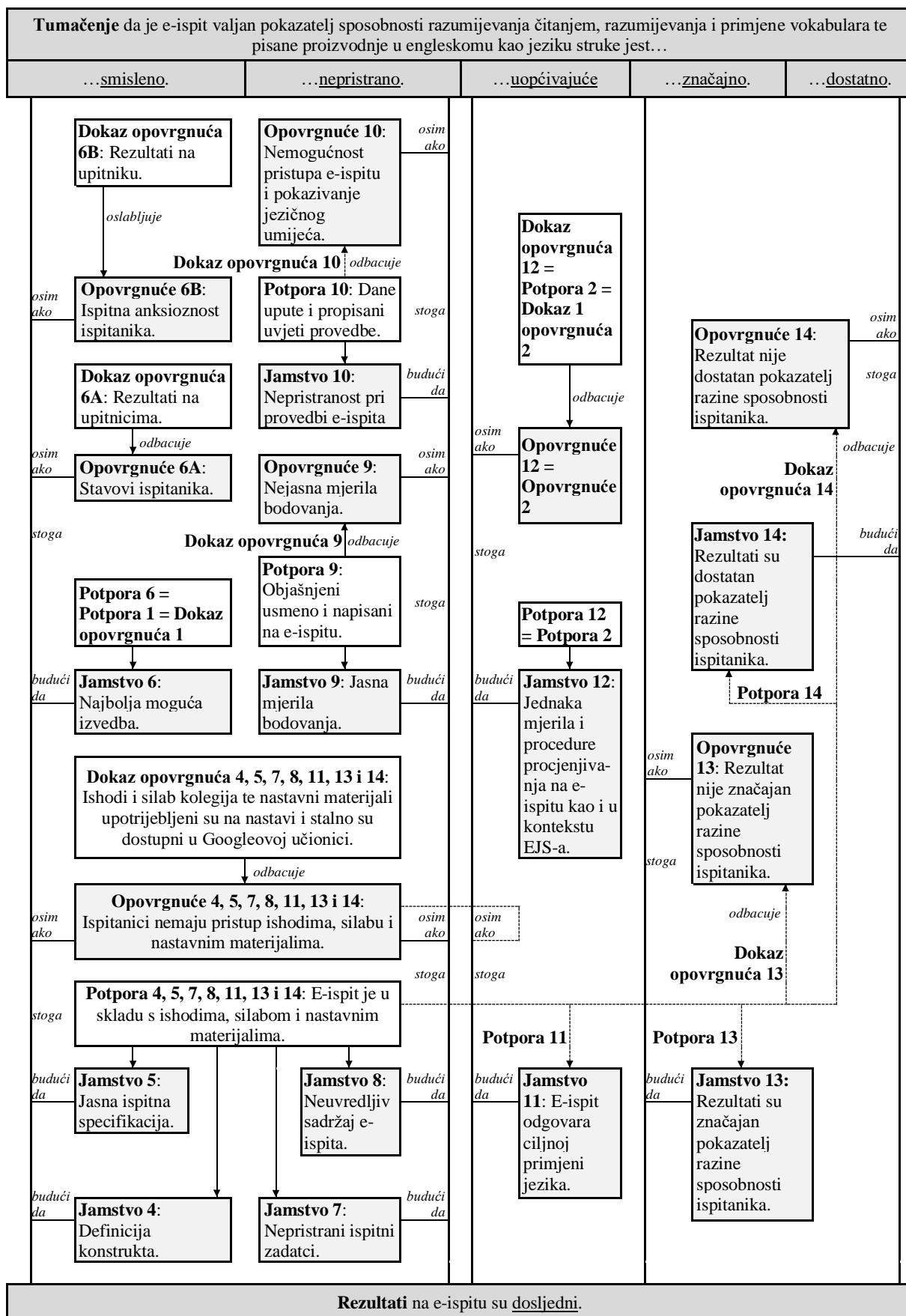
Analiza rezultata valjanosti e-procjenjivanja provela se prema Bachmanovoj metodi potkrjepe procjenjivanja (Bachman, 2005, 2015; Bachman i Palmer, 2010). Kao što je prikazano na slici 22 niže, jezična izvedba sudionika na e-ispitu Poslovnog engleskog jezika (PEJ-a) predstavlja prve podatke na temelju kojih se zaključivanjem dolazi do prve tvrdnje, odnosno rezultata na tom e-ispitu. Ti rezultati tada postaju podatci na temelju kojih se zaključivanjem dolazi do druge tvrdnje, odnosno tumačenja tih rezultata o jezičnim sposobnostima sudionika. Ovaj prvi dio potkrjepe procjenjivanja jest potkrjepa valjanosti procjenjivanja. Drugi se dio potkrjepe procjenjivanja naziva potkrjepom primjene procjenjivanja. Ona počinje prethodno spomenutim tumačenjima koja sada predstavljaju podatke na temelju kojih se zaključivanjem dolazi do treće tvrdnje, odnosno namijenjenih odluka koje se mogu donijeti. Naposljetku, te odluke predstavljaju podatke na temelju kojih se zaključivanjem dolazi do četvrte tvrdnje, odnosno mogućih posljedica donesenih odluka. Svaka se od tih četiriju tvrdnja cjelokupne potkrjepe procjenjivanja sastoji od ishoda (koji su označeni masnim slovima na slici 22) i njihovih svojstava (koja su podcrtana na slici 22). Ta su svojstva potkrijepljena jamstvima s pripadajućom potporom i opovrgnućima s odgovarajućim dokazima koji ih odbacuju, oslabljuju ili podupiru.

Budući da je tema ovog istraživanja analiza valjanosti e-procjenjivanja u engleskomu kao jeziku struke (EJS-u), u našoj se analizi primijenio dio Bachmanove potkrjepe procjenjivanja koji povezuje izvedbu sudionika na e-ispitu PEJ-a s tumačenjem tih rezultata, odnosno potkrjepa valjanosti procjenjivanja (donji dio slike 22, ispod isprekidane crte).

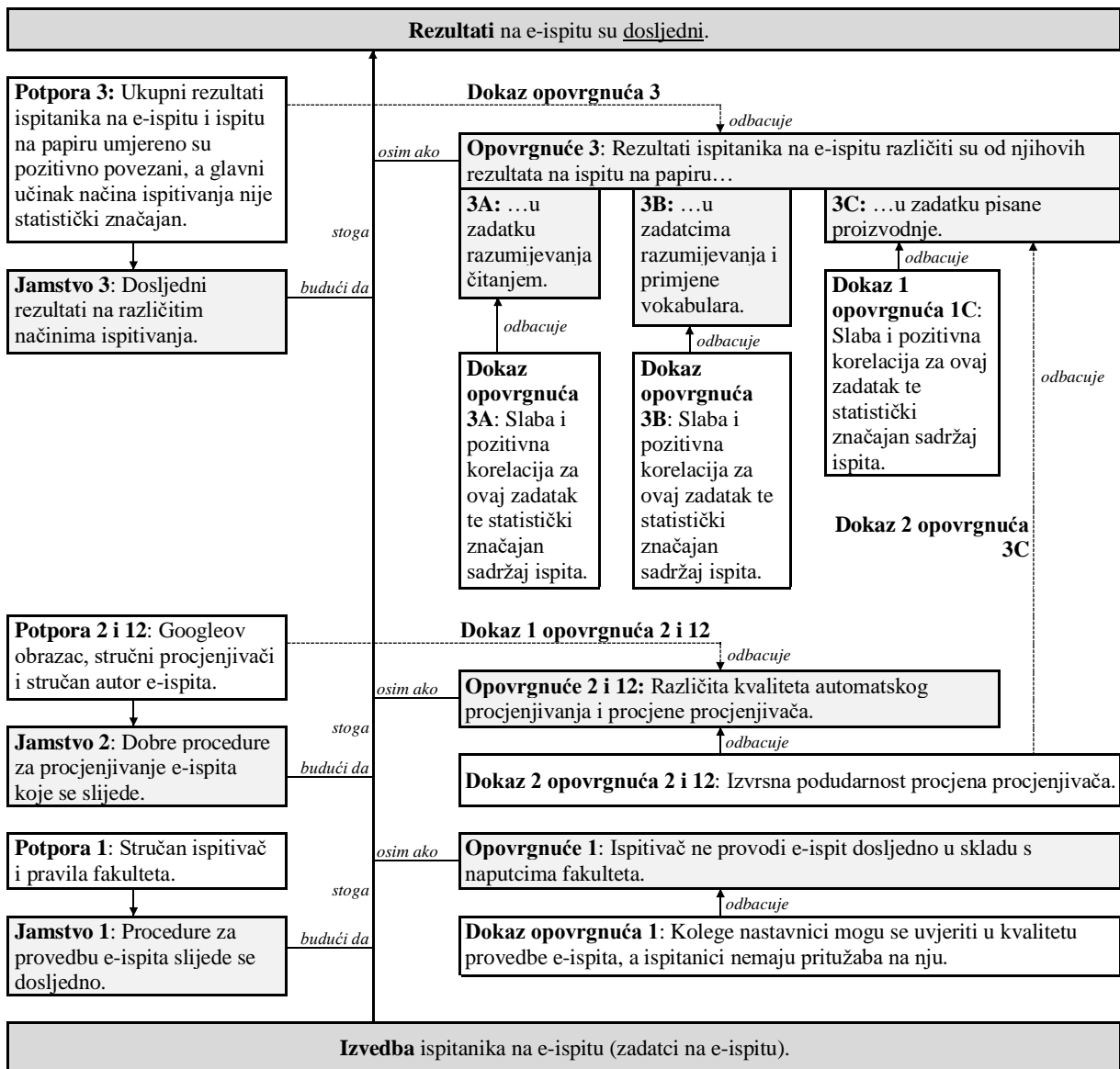


Slika 22. Objedinjeni prikaz potkrjepa procjenjivanja (prema Bachmanu, 2005, 2015; prema Bachmanu i Palmeru, 2010)

Na slici 23 dolje, koja je podijeljena na dva dijela i od kojih je svaki na jednoj stranici, prikazana je naša potkrjepa valjanosti procjenjivanja u EJS-u. Ispod slike ta je potkrjepa iscrpnije objašnjena, počevši od izvedbe sudionika na e-ispitu PEJ-a, koja se nalazi na dnu slike, dakle na drugom dijelu dijagrama. Na slici 23 i u kasnijem objašnjenju primjenjujemo izraz ispitanik kao općenit pojam koji označuje i osobu koja pristupa ispitu i sudionike našeg istraživanja.



Slika 23. Potkrjepa valjanosti e-procjenjivanja u engleskomu kao jeziku struke (1. dio)



Slika 23. Potkrjepa valjanosti e-procjenjivanja u engleskomu kao jeziku struke (2. dio)

4.3.4.1. Prva tvrdnja – rezultati koje su ispitanici ostvarili na e-ispitu dosljedni su

Prvi podatak u našoj potkrjepi valjanosti procjenjivanja u EJS-u jest izvedba ispitanika na zadacima na e-ispitu PEJ-a. Na temelju tog podatka oblikujemo prvu tvrdnju, koja glasi da su rezultati koje su ispitanici ostvarili na e-ispitu dosljedni, a to zaključujemo na temelju jamstava 1, 2 i 3 s pripadajućim potporama, uzimajući u obzir moguća opovrgnuća s odgovarajućim dokazima. Ta su jamstva, zajedno sa svojim opovrgnućima, slična onima iz istraživanja Wang i sur. (2012), no u našem se istraživanju odnose na e-ispit PEJ-a, dok se kod njih odnose na Pearsonov ispit akademskog engleskog jezika. Također, neka od jamstava i opovrgnuća koja su oni oblikovali kao tvrdnje o rezultatima ispita zaključujući na temelju

izvedbe ispitanika u svojem istraživanju nisu oblikovana u našem jer nisu primjenjiva ili nisu u skladu s našim ciljevima istraživanja.

- **Jamstvo 1** prve tvrdnje glasi da se procedure za provedbu e-ispita slijede dosljedno u svim prilikama i za sve skupine ispitanika, što znači da je svaka skupina ispitanika e-ispit polagala u jednakim uvjetima te da je ispitivač koji ih je nadgledao bio stručan i da je na jednak način slijedio pravila fakulteta o provedbi e-ispita za sve skupine ispitanika, što predstavlja **potporu** ovom jamstvu. Ova je potpora ujedno u skladu s načelom 8 Nürnberškog kodeksa, koje navodi da istraživanje trebaju provoditi znanstveno osposobljene osobe (Shuster, 1997), i s načelima istraživačke čestitosti (Steneck, 2006; Resnik i Shamoo, 2011) – naime, ispitanici e-ispit u Googleovim obrascima rješavaju na vlastitom računalu i snimaju se pametnim telefonom iz dovoljne udaljenosti da ispitivač u svakom trenutku može vidjeti njihovo lice i ruke te tipkovnicu i ekran računala. Ova su pravila u skladu s ispitnim ritualima koji se primjenjuju radi pružanja jednakih mogućnosti svim ispitanicima (Fulcher, 2010). Moguće **opovrgnuće** ovog jamstva jest da se ne slijede pravila fakulteta o provedbi e-ispita ili da se ona ne primjenjuju jednako kod svih skupina ispitanika. To se opovrgnuće odbacuje **dokazom** da se e-ispit provodi na internetu u Googleovoj učionici s objavljenom poveznicom, a provedbi mogu pristupiti i drugi kolege nastavnici kako bi se uvjerali u kvalitetu provedbe e-ispita; usto, nije zabilježena nikakva pritužba ni jednog ispitanika na bilo koji aspekt njegove provedbe.
- **Jamstvo 2** glasi da su procedure za procjenjivanje e-ispita dobro opisane i da se slijede, a **potpora** je ovom jamstvu tvrdnja da Googleov obrazac automatski boduje jedan dio e-ispita (tj. zadatak razumijevanja čitanjem te zadatke razumijevanja i primjene vokabulara) i da ispitivač, odnosno autor ispita, može dodatno provjeriti odgovore ispitanika te prije objave rezultata popraviti moguće propuste u pogledu predviđanja svih točnih odgovora koje obrazac prihvaća prilikom automatskog bodovanja. Nadalje, potpora je ovom jamstvu i tvrdnja da autor ispita, koji je stručnjak u području PEJ-a, osmišljava opisnike procjenjivanja trećeg dijela e-ispita (tj. zadatka pisane proizvodnje), koje su u skladu sa silabom kolegija Poslovni engleski jezik 1, a dvoje procjenjivača, također stručnjaka u ovom području, o njima raspravlja te se oko njih usuglašava. Ova je potpora, kao i prethodno opisana potpora 1, u skladu s načelom 8 Nürnberškog kodeksa. Tvrdnja koja bi mogla **opovrgnuti** ovo jamstvo jest da kvaliteta procjene dvoje procjenjivača nije jednaka kvaliteti automatskog procjenjivanja ostalih dvaju dijelova ispita, no **dokaz 1** koji odbacuje ovo opovrgnuće upravo je tvrdnja da su procjenjivači stručnjaci PEJ-a, što je već navedeno u potpori ovom jamstvu. Kao **dokaz 2** ovog opovrgnuća primjenjuju se

statističke analize budući da se one, prema Chapelle i Douglasu (2006), primjenjuju kako bi se opovrgnuća u potkrjepi valjanosti poduprla, oslabjela ili odbacila. Opovrgnuće 2 ovim se dokazom odbacuje zato što statistička analiza, kao što je prethodno predstavljeno u tablici 10, ukazuje na izvrsnu podudarnost pri ukupnom procjenjivanju zadatka ($ICC = 0,93$), no i kad se usporede tri zasebne kategorije procjenjivanja, njihove su podudarnosti izvrsne za kategoriju oblik ($ICC = 0,97$), a dobre za kategorije sadržaj i jezik ($ICC = 0,78$).

- **Jamstvo 3** glasi da su rezultati ispitanika dosljedni na različitim načinima ispitivanja, odnosno na e-ispitu i na njegovoj istovjetnoj inačici na papiru, čemu su **potpora** provedene daljnje statističke analize, a što se podudara s istraživanjima koja su proveli Al-Amri (2007), Hewson i sur. (2007), Mohammadi i Barzgaran (2012) te Öz i Özturan (2018). Rezultati naše statističke analize koja se primijenila kao potpora ovom jamstvu prethodno su prikazani u tablici 12 iz koje se vidi da Pearsonov koeficijent korelacije ukazuje na umjerenu i pozitivnu povezanost ukupnog rezultata ispitanika na e-ispitu i na ispitu na papiru ($r = ,517$, $p < ,01$), a dvosmjernom analize varijance za ponovljena mjerenja utvrđeno je da glavni učinak načina ispitivanja (e-ispit ili ispit na papiru) nije statistički značajan ($F(1, 216,627) = 7,819$, $p = ,001$, $\eta^2p = ,005$). Ovi su rezultati istovremeno i **dokaz** kojime se odbacuje moguće opovrgnuće ovog tumačenja, odnosno **opovrgnuće 3**, a koje glasi da isti ispitanici na istom ispitu, samo u njegovoj inačici na papiru, ostvaruju različite rezultate. Prethodno spomenuto opovrgnuće 3 odnosi se na ukupni rezultat e-ispita, dok se opovrgnuća 3A, 3B i 3C odnose svako na jedan od triju dijelova e-ispita, kao što je opisano u nastavku.
 - **Opovrgnuće 3A** odnosi se na zadatak razumijevanja čitanjem, a glasi da su rezultati ispitanika na e-ispitu različiti od njihovih rezultata na ispitu na papiru u tom zadatku. Kao **dokaz** ovog opovrgnuća također se primjenjuju rezultati statističkih analiza. Iz prethodno predstavljene tablice 12 vidljivo je da Pearsonov koeficijent korelacije ukazuje na slabu i pozitivnu povezanost rezultata ispitanika u zadatku razumijevanja čitanjem na e-ispitu i na ispitu na papiru ($r = ,298$, $p < ,01$). Nadalje, dvosmjernom analize varijance za ponovljena mjerenja utvrđeno je da je sadržaj ispita (odnosno razumijevanje čitanjem, razumijevanje i primjena vokabulara te pisana proizvodnja) statistički značajan ($F(1,805, 216,627) = 2246,748$, $p = ,001$, $\eta^2p = ,949$) te da je međudjelovanje načina ispitivanja i sadržaja ispita također statistički značajno ($F(2,106, 252,707) = 6,554$, $p = ,001$, $\eta^2p = ,052$). Stoga ovaj dokaz odbacuje opovrgnuće 3A.

- **Opovrgnuće 3B** odnosi se na zadatke razumijevanja i primjene vokabulara te glasi da su rezultati ispitanika na e-ispitu različiti od njihovih rezultata na ispitu na papiru u tim zadacima. I za ovo se opovrgnuće kao **dokaz** primjenjuju rezultati statističkih analiza. U istoj tablici 12 vidi se da Pearsonov koeficijent korelacije ukazuje na slabu i pozitivnu povezanost rezultata ispitanika u zadacima razumijevanja i primjene vokabulara na e-ispitu i na ispitu na papiru ($r = ,386$, $p < ,01$). Nadalje, kao kod dokaza opovrgnuća 3B, dvosmjernom analize varijance za ponovljena mjerenja utvrđeno je da je sadržaj ispita statistički značajan te da je međudjelovanje načina ispitivanja i sadržaja ispita također statistički značajno. Stoga ovaj dokaz odbacuje opovrgnuće 3B.
- **Opovrgnuće 3C** odnosi se na zadatak pisane proizvodnje i glasi da su rezultati ispitanika u zadatku pisane proizvodnje na e-ispitu različiti od njihovih rezultata na ispitu na papiru. Dva su dokaza ovog opovrgnuća. Kao **dokaz 1** također su se primijenili rezultati statističkih analiza provedenih u ovom istraživanju, a tablica 12 prikazuje Pearsonov koeficijent korelacije koji ukazuje na slabu i pozitivnu povezanost rezultata ispitanika u zadatku pisane proizvodnje na e-ispitu i na ispitu na papiru ($r = ,457$, $p < ,01$). Nadalje, istovjetno dokazima opovrgnućima 3A i 3B, dvosmjernom analize varijance za ponovljena mjerenja utvrđeno je da je sadržaj ispita statistički značajan te da je međudjelovanje načina ispitivanja i sadržaja ispita također statistički značajno. Stoga ovaj dokaz odbacuje opovrgnuće 3C. Kao **dokaz 2** razmatra se protutvrdnja da su različiti rezultati ispitanika u zadatku pisane proizvodnje posljedica nepodudarnosti procjena dvoje procjenjivača pri procjenjivanju zadatka pisane proizvodnje, kao što je to u svojem istraživanju tvrdila Llosa (2008). Međutim, prethodno opisani dokaz 2 opovrgnuća 2 odbacuje opovrgnuće 3C.

4.3.4.2. Druga tvrdnja – tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke smisleno je, nepristrano, uopćivajuće, značajno i dostatno

Tvrdnja da su rezultati koje su ispitanici ostvarili na e-ispitu dosljedni sada predstavlja drugi podatak u našoj potkrjepi valjanosti procjenjivanja u EJS-u. Na temelju tog podatka oblikuje se druga tvrdnja, odnosno smisleno, nepristrano, uopćivajuće, značajno i dostatno tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i

primjene vokabulara te pisanja u engleskomu kao jeziku struke. Do te druge tvrdnje dolazi se zaključivanjem na temelju jamstava 4-14 s pripadajućim potporama i na temelju opovrgnuća tih jamstava s odgovarajućim dokazima. I u ovom dijelu potkrjepe procjenjivanja neka su jamstva slična onima koja su oblikovali Wang i sur. (2012), s prethodno spomenutom razlikom u ispitu koji se primijenio kao instrument istraživanja. Ova se tvrdnja dijeli na pet dijelova, prema pet svojstava njezina ishoda, kao što je prikazano niže.

1. Tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke je smisljeno.

- **Jamstvo 4** glasi da se sposobnosti definirane u konstrukt temelje na izvornoj primjeni PEJ-a te da se ovaj konstrukt jasno razlikuje od drugih koji su mu slični, npr. općeg engleskog ili engleskog jezika drugih struka, što je u skladu s definicijom konstrukta (usp. McNamara, 2000; Fulcher i Davidson, 2007; Fulcher, 2010; Chapelle, 2021; Green i Fulcher, 2021). Nadalje, **jamstvo 5** glasi da ispitna specifikacija jasno opisuje obilježja ispitnih zadataka koje ispitanici rješavaju u e-ispitu, što je pak u skladu s definicijom ispitne specifikacije (usp. McNamara, 2000; Udier i Jelaska, 2008; Fulcher, 2010; Green, 2014; Hughes i Hughes, 2020). **Potpورا** ovim dvama jamstvima jest tvrdnja da su čestice e-ispita usklađene sa sadržajem (silabom) kolegija Poslovni engleski jezik 1, što se podudara s jamstvima i potporama u potkrjepama procjenjivanja Llose (2008) te Long i sur. (2018), koje su svoja jamstva tumačenja rezultata ispita poduprle sukladnosti ispita s kurikulumom, odnosno državnim standardima. Također, sve se čestice izrađuju koristeći se materijalima za nastavu – udžbenikom, radnim priručnikom, prezentacijama s predavanja te materijalima za ponavljanje i pripremu za ispit. Usto, ishodi kolegija predstavljaju se ispitanicima na nastavi na prvom predavanju, a e-ispit izrađuje se u skladu s njima. Moguće **opovrgnuće** jamstava 4 i 5 jest da ispitanici nemaju pristup silabu kolegija, njegovim ishodima i/ili nastavnim materijalima, no **dokaz** koji ga odbacuje jest da se svim materijalima koristi na nastavi te da su dostupni ispitanicima u svakom trenutku u Googleovoj učionici. Naime, gradivo se obrađuje prema silabu, a i prije e-ispita ispitanicima se daju upute o tomu da će sadržavati sve što je navedeno u silabu te da se pripreme za ispit koristeći se svim materijalima za nastavu.
- **Jamstvo 6** glasi da procedure za provedbu e-ispita omogućuju ispitanicima njihovu najbolju moguću izvedbu prilikom koje mogu pokazati svoje sposobnosti razumijevanja

čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u PEJ-u. **Potpore** ovom jamstvu jednaka je tvrdnja koja podupire jamstvo 1 (odnosno potpore 1) i dokazu koji odbacuje opovrgnuće te tvrdnje (odnosno dokazu opovrgnuća 1) – provedbu e-ispita nadgleda stručan ispitivač koji slijedi pravila fakulteta o provedbi e-ispita, a provedbi mogu pristupiti drugi nastavnici radi uvida u njezinu kvalitetu te na nju nema pritužaba ni od jednog ispitanika ni tijekom ni nakon nje.

- **Opovrgnuće 6A**, odnosno prva tvrdnja koja bi mogla opovrgnuti jamstvo 6, jest da su rezultati na e-ispitu povezani sa stavovima ispitanika o e-ispitima PEJ-a, o ispitima općenito i o računalima. U našem istraživanju, kao **dokaz** ovog opovrgnuća statističkim su se analizama ispitala povezanosti rezultata ispitanika na e-ispitu i njihovih odgovora na trima upitnicima – Upitniku o stavovima o e-ispitima Poslovnog engleskog jezika (prilog 2.), Upitniku o stavovima o ispitima općenito (prilog 3.) i Upitniku o stavovima o računalima (prilog 4.). Iako je stav ispitanika prema e-ispitima PEJ-a općenito pozitivan (iznadprosječna vrijednost 4,78 od moguće najveće vrijednosti 6,00), iako je njihov stav o ispitima općenito srednje pozitivan (srednja vrijednost od 3,25 od moguće najveće vrijednosti 5,00) te iako su im anksioznost vezana za korištenje računalima i ovisnost o računalima niske (niske vrijednosti od 2,11 za anksioznost i 2,14 za ovisnost od moguće najveće vrijednosti 5,00), kao što je prethodno prikazano u tablici 7 u potpoglavlju 4.3.1.3., nakon što se ispituju bivarijatne povezanosti (tablica 12), uočava se da stavovi ispitanika o e-ispitima PEJ-a, o ispitima općenito te o računalima nisu statistički značajno povezani ni s jednim dijelom e-ispita. Stoga ovaj dokaz odbacuje opovrgnuće 6A, što je u skladu s rezultatima istraživanja koje su proveli Al-Amri (2007), Hewson i sur. (2007), Mohammadi i Barzgaran (2012), Khoshsimi i sur. (2017) te Hewson i Charlton (2019), a djelomično u skladu s rezultatima do kojih su došli Fan i Ji (2014).
- **Opovrgnuće 6B**, odnosno druga tvrdnja koja bi mogla opovrgnuti jamstvo 6, jest da su rezultati na e-ispitu povezani s razinom ispitne anksioznosti ispitanika. Kao **dokaz** ovog opovrgnuća primjenjuje se statistička analiza ispitivanja povezanosti rezultata ispitanika na e-ispitu i njihovih odgovora na Upitniku o razini ispitne anksioznosti (prilog 5.). Iako srednja vrijednost njihove ispitne anksioznosti iznosi 3,5 od najviše 5,00 (tablica 7), što znači da imaju umjereno visoku ispitnu anksioznost prema visokoj ispitnoj anksioznosti (Driscoll, 2007 u Aliakbari i Gheitasi, 2017), nakon ispitivanja bivarijatnih povezanosti (tablica 12) uočava se da razina njihove ispitne anksioznosti nije statistički značajno povezana ni s jednim dijelom e-ispita osim s dijelom

razumijevanja čitanjem, gdje je ona neznatna i negativna ($r = -,205; p < ,05$). To znači da ispitanici s nižom razinom ispitne anksioznosti postižu bolje rezultate u zadatku razumijevanja čitanjem. Stoga ovaj dokaz oslabljuje opovrgnuće 6B, što je djelomično u skladu s rezultatima istraživanja koje su proveli Cassady i Johnson (2002), Chapell i sur. (2005), Juretić (2008) te Erceg Jugović i Lauri Korajlija (2012).

2. Tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke je nepristrano.

- **Jamstvo 7** glasi da e-ispit ne sadržava ispitne zadatke čiji sadržaj ili traženi odgovori idu u korist samo određenim ispitanicima. E-ispit također ne uključuje sadržaj koji bi nekim ispitanicima mogao biti uvredljiv – tematski, vjerski, kulturno, jezično i dr., što predstavlja **jamstvo 8**. Ova dva jamstva podupire **potpora** jednaka već prethodno navedenoj potpori jamstvima 4 i 5, a ta je da su svi nastavni materijali u skladu sa silabom i ishodima kolegija te da se e-ispit izrađuje u skladu s njima. I **opovrgnuće** ovih jamstava jednako je prethodno opisanom opovrgnuću jamstava 4 i 5, odnosno tvrdnji da ispitanici nemaju pristup silabu kolegija, njegovim ishodima i/ili nastavnim materijalima, što se odbacuje također već predstavljenim **dokazom** opovrgnuća 4 i 5 – da se njima koristi na nastavi i da su im u svakom trenutku dostupni u Googleovoj učionici.
- **Jamstvo 9** glasi da su procedure i mjerila bodovanja zadataka u e-ispitu jasno opisani i razumljivi svim ispitanicima, čija je **potpora** tvrdnja da su upute o rješavanju svakog zadatka te bodovi koji se na njima mogu ostvariti napisani na samom ispitu, i to na engleskom jeziku razine B1-B2, koja se očekuje od studenata kolegija Poslovni engleski jezik 1, te da im se upute daju usmeno prije početka e-ispita na engleskom jeziku iste razine. Ovo se jamstvo može **opovrgnuti** tvrdnjom da su mjerila bodovanja nejasna, no potpora ovom jamstvu ujedno je i **dokaz** kojime se ovo opovrgnuće odbacuje.
- **Jamstvo 10** glasi da se prema ispitanicima postupa jednako prilikom svih aspekata provedbe e-ispita – svi imaju jednak pristup informacijama o proceduri provedbe e-ispita, svi imaju jednak pristup e-ispitu što se tiče uvjeta, troškova, lokacije i opreme vezanih za polaganje e-ispita, te svi mogu na jednak način pokazati svoje jezično umijeće u PEJ-u. **Potpورا** ovom jamstvu jesu tvrdnje da se upute o provedbi e-ispita daju na nastavi, ali i da ih svi ispitanici imaju dostupne u nastavnim materijalima u Googleovoj učionici te da e-ispit polažu putem interneta pod jednakim uvjetima koje propisuje fakultet na kojem

studiraju, a koji ne diskriminiraju ni jednog ispitanika, pa ni one kojima je potrebna prilagodba ispitnih materijala i/ili tehnologije. Moguće **opovrgnuće** ovog jamstva jest tvrdnja da neki ispitanici možda ne mogu pristupiti e-ispitu jer nemaju potrebnu opremu za nj (tj. računalo i kameru) ili osigurane uvjete za neometano polaganje e-ispita (tj. zasebnu prostoriju bez prisutnih trećih osoba te pouzdanu internetsku vezu) zbog čega ne mogu pokazati svoje jezično umijeće. Međutim, ovo se opovrgnuće odbacuje **dokazom** koji je istovremeno i potpora ovom jamstvu, a to je tvrdnja da su im se sve upute o provedbi pravovremeno daju na nastavi i u nastavnim materijalima stalno dostupnim na Googleovoj učionici te da uvjete provedbe ispita propisuje njihov fakultet.

3. Tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke je uopćivajuće.

- **Jamstvo 11** glasi da obilježja e-ispita odgovaraju području ciljne primjene jezika, odnosno području PEJ-a, a **potpora** tom jamstvu jednaka je prethodno objašnjenjnoj potpori 4, 5, 7 i 8, a ta je da se ispit izrađuje prema ishodima i silabu kolegija Poslovni engleski jezik 1 te nastavnim materijalima, a i da se svi oni predstavljaju ispitanicima na nastavi i da su im stalno dostupni u Googleovoj učionici. Ovo se jamstvo opovrgava već predstavljanim **opovrgnućem** jamstava 4, 5, 7 i 8, koje glasi da ispitanici nemaju pristup silabu i ishodima kolegija i/ili nastavnim materijalima, a to se opovrgnuće odbacuje također prethodno objašnjenim **dokazom** dotičnog opovrgnuća.
- **Jamstvo 12** glasi da mjerila i procedure procjenjivanja odgovora na e-ispitu odgovaraju onima koji se primjenjuju pri procjenjivanju jezične izvedbe u kontekstu PEJ-a, a **potpora** mu je jednaka kao i već opisana potpora 2, odnosno da e-ispit djelomično automatski procjenjuje Googleov obrazac, a djelomično dvoje procjenjivača koji su stručnjaci u području PEJ-a. **Opovrgnuće** jamstva 12 također je jednako opovrgnuću 2, a to je da se kvaliteta automatskog procjenjivanja i procjene dvoje procjenjivača razlikuju, no potpora 2, koja je istovremeno dokaz opovrgnuća 2, ujedno je **dokaz** kojime se opovrgnuće 12 odbacuje.

4. Tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke je značajno.

- **Jamstvo 13** glasi da tumačenja konstrukta e-ispita daju informacije koje su značajan pokazatelj razine sposobnosti ispitanika kako bi im se mogle dodijeliti ocjene na temelju rezultata koje su ostvarili prilikom svoje jezične izvedbe. Kao **potpora** ovom jamstvu uzima se tvrdnja da se prema ishodima i silabu kolegija Poslovni engleski jezik 1 izrađuje e-ispit i da su postotci prema kojima se ispitanicima dodjeljuje određena ocjena propisani i objašnjeni na nastavi te da su oni navedeni u nastavnim materijalima koje imaju uvijek dostupne u Googleovoj učinioci, a ta je potpora istovjetna potpori 4, 5, 7, 8 i 11. Moguće **opovrgnuće** ovog jamstva jest tvrdnja da rezultat ostvaren na e-ispitu nije značajan za ocjenjivanje ispitanika jer ocjena ne odražava razinu njihove jezične sposobnosti u PEJ-u, no to se odbacuje **dokazom** koji je ujedno i potpora ovom jamstvu.

5. Tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u engleskomu kao jeziku struke je dostatno.

- **Jamstvo 14** glasi da su tumačenja konstrukta e-ispita dostatna za utvrđivanje razine sposobnosti ispitanika i kako bi na temelju rezultata ostvarenih prilikom svoje jezične izvedbe mogli dobiti ocjene. **Potpora** ovom jamstvu ista je kao i potpora 4, 5, 7, 8, 11 i 13, što znači da se e-ispit izrađuje prema ishodima i silabu kolegija Poslovni engleski jezik 1 i da se ispitanicima dodjeljuje određena ocjena prema postotcima koji su propisani i objašnjeni na nastavi, a koji se nalaze i u uvijek dostupnim nastavnim materijalima u Googleovoj učinioci. Moguće **opovrgnuće** ovog jamstva jest tvrdnja da rezultat ostvaren na e-ispitu nije dostatan za ocjenjivanje ispitanika jer ocjena ne odražava razinu njihove jezične sposobnosti u PEJ-u, no to se odbacuje **dokazom** koji je ujedno i potpora ovom jamstvu.

4.4. DISKUSIJA REZULTATA ISTRAŽIVANJA

Glavni je cilj našeg istraživanja analizirati valjanost e-procjenjivanja u engleskomu kao jeziku struke (EJS), a njegovi su pojedinačni ciljevi usporediti rezultate na e-ispitu i ispitu

na papiru, ispitati stavove sudionika o e-ispitima, ispitima općenito i računalima, te ispitati razinu ispitne anksioznosti sudionika. Svaki od tih triju ciljeva postavlja po jedno istraživačko pitanje, od kojih svako izvodi svoje hipoteze, koje se raspravljaju u narednim potpoglavljima.

Procjenjivanje čija se valjanost analizira u ovom istraživanju jest procjenjivanje u širem smislu, što znači da se njime rasuđuje i prikupljaju dokazi kako bi se donijeli zaključci o ispitanicima (McNamara i Roever, 2006) na temelju kvalitativnih i kvantitativnih pokazatelja njihovih jezičnih sposobnosti (usp. Brown i Abeywickrama, 2010), ali i zaključci o kvaliteti poučavanja, pri čemu se procjenjuje jesu li ishodi učenja u skladu s obrazovnim programima, silabima i sl. (usp. Ghanbari i Nowroozi, 2021). Budući da se naše istraživanje bavi analizom valjanosti procjenjivanja, usredotočilo se na prvi dio prethodne definicije procjenjivanja, odnosno na dio koji povezuje izvedbu ispitanika s tumačenjem rezultata koje su ostvarili na ispitima prilikom svoje izvedbe. To je u skladu s prvim dijelom metode potkrjepe procjenjivanja, jednom od mogućih metoda utvrđivanja valjanosti koja je primijenjena i u našem istraživanju, i koja se dijeli na dva dijela – prvi je dio potkrjepa valjanosti procjenjivanja, a drugi je dio potkrjepa primjene procjenjivanja (Bachman, 2005, 2015; Bachman i Palmer, 2010).

4.4.1. Usporedba rezultata na e-ispitu i ispitu na papiru

Sukladno prvom pojedinačnom cilju našeg istraživanja postavilo se prvo istraživačko pitanje – ispituje li se e-ispitima sposobnost razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u EJS-u. Naime, valjanost e-procjenjivanja u EJS-u podrazumijeva da se e-ispitom ispituje konstrukt EJS-a, odnosno strukture koje osmišljavaju autori e-ispita poslovnog engleskog jezika (PEJ) koje odgovaraju njegovoj stvarnoj primjeni (Chapelle, 2021), koristeći se znanjem i vještinama u PEJ-u koje su tim konstruktom određene i koje on ispituje (McNamara, 2000). Ispitanicima su potrebne određene jezične sposobnosti za primjenu jezika kao instrumenta izvan učionice, pa zato autori silaba, autori ispita i nastavnici moraju znati koje su te sposobnosti kako bi osigurali da sadržaj nastave, ali i ispita koji se primjenjuju u procjenjivanju, to pokrivaju (usp. Basturkmen, 2010). I u svojem su preglednom radu Dursun i Li (2021) predstavili zaključke 70 članaka koji su primijenili potkrjepu procjenjivanja, od kojih su najčešći bili oni da bodovi ostvareni na ispitu odražavaju konstrukt koji se procjenjuje dotičnim ispitom, da su bodovi na ispitu značajni za izvedbu na ne-ispitnim zadacima te da bodovi dosljedno odražavaju izvedbu ispitanika. Stoga bi se rezultati koje su sudionici našeg istraživanja ostvarili na e-ispitu PEJ-a trebali tumačiti kao

pokazatelj njihovih sposobnosti u PEJ-u, i to njih tri – sposobnosti razumijevanja čitanjem u PEJ-u, sposobnosti razumijevanja i primjene vokabulara u PEJ-u te sposobnosti pisane proizvodnje u PEJ-u (usp. Bachman, 2004).

Metoda potkrjepe valjanosti procjenjivanja primijenjena je u našoj analizi kako bi se pronašao odgovor na prvo istraživačko pitanje, i prema toj se metodi na temelju izvedbe sudionika na e-ispitu PEJ-a može zaključiti da su njihovi rezultati dosljedni neovisno o načinu ispitivanja, odnosno neovisno o tomu polažu li e-ispit ili ispit na papiru, što je u skladu s jamstvom 3 na slici 23 u potpoglavlju 4.3.4. Međutim, u obzir valja uzeti i jamstva 1-2, 4-5 i 7-14 iako se ona ne odnose na način ispitivanja nego na druge aspekte ispita koji imaju moguć učinak na izvedbu sudionika. Analizom tih jamstava s pripadajućim potporama, opovrgnućima i dokazima tih opovrgnuća može se zaključiti da su rezultati sudionika dosljedni i zato što se procedure za provedbu e-ispita slijede dosljedno i zato što su procedure za procjenjivanje e-ispita dobre i također se slijede dosljedno. Nadalje, može se zaključiti da je e-ispit valjan pokazatelj sudionikovih sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u PEJ-u zato što su konstrukt i specifikacija e-ispita u skladu sa stalno dostupnim ishodima, silabom i nastavnim materijalima koji odgovaraju ciljnoj primjeni jezika, zato što su sadržaj ispita i ispitni zadatci nepristrani i neuvredljivi, zato što su mjerila bodovanja jasna, zato što je provedba ispita nepristrana, zato što procjenjivanje na ispitu PEJ-a odgovara procjenjivanju u kontekstu EJS-a te zato što su rezultati značajan i dostatan pokazatelj razine ispitanikovih sposobnosti u PEJ-u.

Nadalje, kad se uspoređuje valjanost e-ispita s valjanosti ispita na papiru, one su u teoriji istovrijedne, no treba uzeti u obzir moguće probleme utvrđivanja valjanosti tih dvaju načina ispitivanja (usp. Green, 1988; Isbell i Kremmel, 2020; Oh, 2020). Od mogućih problema utvrđivanja valjanosti jedan može biti taj da su sudionici našeg istraživanja dva puta rješavali potpuno jednak ispit, koji je sadržavao istovjetne ispitne čestice, što je također moglo imati učinak na njihovu izvedbu. Međutim, prema preporuci psihologa, ostavilo se dovoljno vremena između rješavanja tih dvaju ispita – tri i pol mjeseca, što je otklonilo ovaj mogući negativan učinak. Još jedan mogući problem utvrđivanja valjanosti može biti činjenica da su rješavali isti ispit dva puta, a da im nije donosio ocjenu, što je moglo imati učinak na njihovu motivaciju i, posljedično, kvalitetu izvedbe. Ipak, njihovo se zanimanje za rješavanje ovih ispita potaknulo nagradama u obliku intelektualne koristi, odnosno ponavljanjem gradiva za završni ispit te dodatnim bodovima za ispit i/ili domaću zadaću, kao i nagradama u obliku materijalne koristi, odnosno poklon-bonovima.

Stoga, nakon što se u potkrjepi valjanosti procjenjivanja dalo jamstvo 3 s pripadajućom potporom i nakon što su se pružili dokazi za moguća opovrgnuća tog jamstava (tj. opovrgnuća 3A, 3B i 3C), a i nakon što su se analizirala jamstva i opovrgnuća 1-2, 4-5 i 7-14, može se doći do niže navedenih zaključaka o hipotezama H1A, H1B i H1C, koje su se izvele iz prvog pojedinačnog cilja.

- **H1A: E-ispitom ispituje se sposobnost razumijevanja čitanjem u engleskomu kao jeziku struke.** Ova je hipoteza potvrđena budući da je tvrdnja opovrgnuća 3A („Rezultati ispitanika na e-ispitu različiti su od njihovih rezultata na ispitu na papiru u zadatku razumijevanja čitanjem.“) odbačena dokazom opovrgnuća 3A, koji dokazuje da je u zadatku razumijevanja čitanjem korelacija između e-ispita i ispita na papiru slaba i pozitivna, a sadržaj e-ispita statistički je značajan.
- **H1B: E-ispitom ispituje se sposobnost razumijevanja i primjena vokabulara u engleskomu kao jeziku struke.** Ova je hipoteza potvrđena budući da je tvrdnja opovrgnuća 3B („Rezultati ispitanika na e-ispitu različiti su od njihovih rezultata na ispitu na papiru u zadacima razumijevanja i primjene vokabulara.“) odbačena dokazom opovrgnuća 3B, koji dokazuje da je u zadacima razumijevanja i primjene vokabulara korelacija između e-ispita i ispita na papiru slaba i pozitivna, a sadržaj e-ispita statistički je značajan.
- **H1C: E-ispitom ispituje se sposobnost pisane proizvodnje u engleskomu kao jeziku struke.** Ova je hipoteza potvrđena budući da je tvrdnja opovrgnuća 3C („Rezultati ispitanika na e-ispitu različiti su od njihovih rezultata na ispitu na papiru u zadatku pisane proizvodnje.“) odbačena obama dokazima opovrgnuća 3C – dokazom 1 opovrgnuća 3C, koji ukazuje na slabu i pozitivnu korelaciju između e-ispita i ispita na papiru u zadatku pisane proizvodnje te na statistički značajan sadržaj ispita, te dokazom 2 opovrgnuća 3C, koji ukazuje na izvrsnu podudarnost procjena dvoje procjenjivača.

Način provedbe ispita jedan je od mogućih čimbenika koji utječu na valjanost ispita, a našim su se rezultatima potvrdile gornje hipoteze jer smo zaključili da način provedbe nema učinak na izvedbu sudionika i na rezultate njihove izvedbe, što je u skladu s rezultatima do kojih su u svojim istraživanjima došli Al-Amri (2007) primijenivši kao instrument TOEFL-ov ispit čitanja, Hewson i sur. (2007) primijenivši sveučilišni ispit iz psihologije, zatim

Mohammadi i Barzgaran (2012) primijenivši Cambridgev ispit slušanja, čitanja i pisanja PET, i naposljetku Öz i Özturan (2018) primijenivši tradicionalni ispit postignuća u engleskom jeziku, koji su svi također uspoređivali rezultate ispitanika na e-ispitu i njegovoj inačici na papiru. Stoga možemo reći da su rezultati ovog istraživanja, koje je provedeno u hrvatskom kontekstu, značajni jer su usporedivi s ostalim istraživanjima koja su se provodila u različitim državama i koja su primjenjivala različite instrumente. Ipak, rezultati su nam se razlikovali od onih Khoshshime i sur. (2017), koji su primijenili ispit općeg engleskog jezika i uvidjeli da su ispitanici bolje rezultate ostvarili na ispitima na papiru nego na e-ispitima, no do ove je razlike u rezultatima možda došlo zbog razlike konteksta u kojemu su provodili svoje ispite – naime, broj sudionika bio je upola manji od našega i svi su bili iz samo jedne privatne škole stranih jezika.

Uzimajući u obzir činjenicu da valjanost ispita nikad nije potpuna, već je na određenom stupnju, te da smo kao jedan od dokaza prilikom utvrđivanja valjanosti e-ispita PEJ-a primijenili usporedbu s rezultatima na njegovoj inačici na papiru, možemo tvrditi da naši rezultati povisuju stupanj valjanosti e-ispita PEJ-a (usp. Anastasi, 1976; Angoff, 1988; Messick, 1989; Anderson i Banerjee, 2001; Urbina, 2004; Bachman, 2004; Davies i Elder, 2005; Weir, 2005; McNamara i Roever, 2006; Fulcher i Davidson, 2007; Milas, 2009; Green, 2014; Fitzpatrick i Clenton, 2010; Shepard, 2016; Chapelle, 2021; Green i Fulcher, 2021). Ovaj naš doprinos istraživanju valjanosti e-procjenjivanja u engleskomu kao jeziku struke od značajne je koristi jer su dionici procesa e-procjenjivanja PEJ-a, od autora kurikula, silaba i ispita, preko predmetnih nastavnika, do samih ispitanika odnosno studenata, dobili potkrjepu za primjenu ovog načina procjenjivanja. Također, rezultati našeg istraživanja mogu se primijeniti i pri procjenjivanju drugih kolegija na visokoškolskim ustanovama, ali i u srednjoškolskom i osnovnoškolskom kontekstu, odnosno kad god je dionicima bitno utvrditi valjanost instrumenta kojime se koriste.

4.4.2. Ispitivanje stavova sudionika o e-ispitima, o ispitima općenito i o računalima

Sukladno drugom pojedinačnom cilju našeg istraživanja postavilo se drugo istraživačko pitanje – jesu li rezultati na e-ispitu povezani sa stavovima sudionika o e-ispitima, ispitima općenito i računalima. Naime, stavovi ispitanika predstavljaju moguću prijetnju valjanosti e-procjenjivanja, pogotovo ako oni ovise o načinu provedbe ispita putem interneta i/ili na računalu (usp. Green, 1998, Chapelle i Douglas, 2006; Isbell i Kremmel, 2020; Li i sur., 2021). Budući da stavovi mogu utjecati na reakciju, djelovanje i/ili ponašanje ispitanika (usp.

Allport, 1971; Eiser, 1986; Petz, 2005) te da mogu utjecati na učenje i postignuće u stranom jeziku (Gardner, 1985), a povezani su i s motivacijom za učenjem jezika (Aliakbari i Gheitasi, 2017), naše je istraživanje nastojalo ispitati jesu li stavovi sudionika povezani s njihovim rezultatima na e-ispitu, i to tri skupine stavova – o e-ispitima PEJ-a, o ispitima općenito i o računalima. Nadalje, stavovi ispitanika utječu na njihovu motivaciju za učenjem i učinkovitost učenja (Hutchinson i Waters, 1987), a ispitanici su motiviraniji kad ih se poučava EJS-u nego općem engleskom zbog značajnih tema na koje su usredotočeniji (usp. Dudley-Evans i St John, 1998).

Metodom potkrjepe valjanosti procjenjivanja primijenjenoj u našoj analizi nastojao se pronaći odgovor i na drugo istraživačko pitanje. Na temelju tvrdnje da su rezultati ispitanika na e-ispitu PEJ-a dosljedni može se zaključiti da je tumačenje da je e-ispit valjan pokazatelj njihovih sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u PEJ-u smisleno budući da je omogućena najbolja izvedba ispitanika na e-ispitu, neovisno o njihovim stavovima (v. jamstvo 6 na slici 23 u potpoglavlju 4.3.4.). Nakon što se u potkrjepi valjanosti procjenjivanja dalo jamstvo 6 s pripadajućom potporom i nakon što je pružen dokaz za moguće opovrgnuće tog jamstva, odnosno opovrgnuće 6A, o hipotezama H2A, H2B i H2C, koje su se izvele iz drugog pojedinačnog cilja, mogu se dati niže navedeni zaključci.

- **H2A: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o e-ispitima.** Ova je hipoteza potvrđena budući da je dio tvrdnje opovrgnuća 6A („Rezultati na e-ispitu povezani su sa stavovima ispitanika o e-ispitima PEJ-a [...]“) odbačen prvim od triju dijelova dokaza opovrgnuća 6A, odnosno korelacijskom analizom rezultata sudionika na e-ispitu i njihovih odgovora na Upitniku o stavovima o e-ispitima PEJ-a. Naime, ona ukazuje na to da stavovi sudionika o e-ispitima PEJ-a nisu statistički značajno povezani s njihovim rezultatima ni na jednom dijelu e-ispita.
- **H2B: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o ispitima općenito.** Ova je hipoteza potvrđena budući da je dio tvrdnje opovrgnuća 6A („Rezultati na e-ispitu povezani su sa stavovima ispitanika [...] o ispitima općenito [...]“) odbačen drugim dijelom dokaza opovrgnuća 6A, odnosno korelacijskom analizom rezultata sudionika na e-ispitu i njihovih odgovora na Upitniku o stavovima o ispitima općenito. Rezultati ove analize ukazuju na to da stavovi sudionika o ispitima općenito nisu statistički značajno povezani s njihovim rezultatima ni na jednom dijelu e-ispita.

Naši su rezultati na temelju kojih donosimo zaključke za hipoteze H2A i H2B djelomično u skladu s rezultatima Fan i Ji (2014), čemu u prilog može ići i činjenica da se u tom istraživanju nije ispitala povezanost stavova ispitanika o e-ispitima s njihovim rezultatima, već se analizirala povezanost njihovih stavova o određenom ispitu engleskog jezika, premda na papiru. Naime, navedeno istraživanje ukazuje na to da dva od pet faktora značajno objašnjavaju izvedbu ispitanika na dotičnom ispitu – stavovi ispitanika o povratnom učinku ispita i o podacima o ispitu. S druge strane, do zaključaka suprotnih našima u svojim su istraživanjima došli Gorsuch (2004), koja navodi da stavovi o ispitima na računalu mogu imati učinak na rezultate, te Hartshorn i McMurry (2020), čiji rezultati ukazuju na to da su njihovi ispitanici uslijed pandemije COVID-19 poboljšali svoju izvedbu na ispitu engleskog jezika u vještinama govorenja i pisanja unatoč svojim stavovima o važnosti nastave engleskoga, o učinku pandemije na njihovu razinu stresa, učenje engleskoga, prelazak na e-nastavu, potrebu učenja engleskoga putem interneta i dr. Do razlika u rezultatima našeg i potonjih dvaju istraživanja možda je došlo zato što je prvo analiziralo samo ispit slušanja i to na uzorku od samo šestero sudionika, dok se drugo prvenstveno usredotočilo na stavove sudionika uslijed pandemije COVID-19. Što se tiče usporedbe naših rezultata s rezultatima ostalih istraživanja koje smo izdvojili u ovom radu, ne možemo ih u potpunosti usporediti s našima jer smo ispitali povezanost stavova ispitanika s njihovim rezultatima na e-ispitu – nismo samo analizirali njihove stavove o njemu. Stoga možemo samo istaknuti da su rezultati ostalih istraživanja ukazali na pozitivne stavove ispitanika o korisnosti i primjeni e-ispita (Dizon, 2016), na zadovoljstvo ispitanika e-nastavom i provedbom e-ispita (Chung i Choi, 2021) te s jedne strane na nižu razinu anksioznosti i brže dobivanje rezultata, a s druge strane na teže rukovanje i veću mogućnost varanja kod e-ispita u odnosu na ispite na papiru (Hoang i sur., 2021).

- **H2C: Rezultati na e-ispitu nisu povezani sa stavovima sudionika o računalima.** Ova je hipoteza potvrđena budući da je dio tvrdnje opovrgnuća 6A („Rezultati na e-ispitu povezani su sa stavovima ispitanika [...] o računalima.“) odbačen trećim dijelom dokaza opovrgnuća 6A, odnosno korelacijskom analizom rezultata sudionika na e-ispitu i njihovih odgovora na Upitniku o stavovima o računalima. Ova analiza ukazuje na to da stavovi sudionika o računalima nisu statistički značajno povezani s njihovim rezultatima ni na jednom dijelu e-ispita. Naši su rezultati u skladu s rezultatima istraživanja koji pokazuju da ni poznavanje računala ni stavovi o računalima nemaju značajan učinak na rezultate na ispitima na računalima (Al-Amri, 2007; Khoshsima i sur., 2017) odnosno e-ispitima

(Hewson i sur., 2007; Hewson i Charlton, 2019) te da razina poznavanja računala nema učinak na rezultate na e-ispitu pisanja (Mohammadi i Barzgaran, 2012). Podudarnost naših rezultata s prethodno navedenima ukazuje na značajnost ovog istraživanja jer ga stavlja u globalni kontekst, kojemu daje svoj doprinos provedenim analizama. Nasuprot tomu, rezultati nam se razlikuju od zaključka koji iznosi Gorsuch (2004), prema kojoj stavovi o računalima mogu imati učinak na rezultate na ispitu na računalu, no tomu su mogući razlozi možda već prethodno spomenuti malen uzorak i ograničen instrument.

4.4.3. Ispitivanje razine ispitne anksioznosti sudionika

Sukladno trećem pojedinačnom cilju našeg istraživanja postavilo se treće istraživačko pitanje – jesu li rezultati na e-ispitu povezani s razinom ispitne anksioznosti sudionika. Naime, kao što prethodno raspravljeni stavovi ispitanika predstavljaju moguću prijetnju valjanosti e-procjenjivanja, tako i ispitna anksioznost može imati učinak na izvedbu ispitanika, i to posebice ako se radi o ispitu koji se provodi putem interneta i/ili na računalu (usp. Chapelle i Douglas, 2006). Anksioznost koja se javlja tijekom ili zbog polaganja ispita (Cheng i Zheng, 2021), zbog toga što je ona crta osobnosti ispitanika ili je stanje koje je povezano sa situacijom polaganja ispita (Alderson i Wall, 1993; Živčić-Bećirević, 2003; Erceg Jugović i Lauri Korajlija, 2012), može dovesti do loše odnosno pogoršane izvedbe na ispitu (Zeidner, 1998; Cassady i Johnson, 2002) i može biti negativno povezana s općim akademskim postignućem (Erceg Jugović i Lauri Korajlija, 2012). Međutim, ako je blaga intenziteta, može motivirati ispitanike (ibid.) pa čak i potaknuti njihovu izvedbu, zbog čega će ostvariti relativno dobre rezultate (Zeidner, 1998). Zbog ovih je razloga naše istraživanje ispitivalo je li razina ispitne anksioznosti sudionika povezana s njihovim rezultatima na e-ispitu.

Metoda potkrjepe valjanosti procjenjivanja primijenjena je i u sklopu trećeg istraživačkog pitanja. Tvrdnja da su rezultati ispitanika na e-ispitu PEJ-a dosljedni dovodi do zaključka koji je jednak zaključku kod prethodnog istraživačkog pitanja, odnosno da je tumačenje da je e-ispit valjan pokazatelj njihovih sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u PEJ-u smisleno budući da je omogućena najbolja izvedba ispitanika na e-ispitu te da na nju nema učinak ni razina njihove ispitne anksioznosti (v. jamstvo 6 na slici 23 u potpoglavlju 4.3.4.). Nakon što se u potkrjepi valjanosti procjenjivanja dalo prethodno navedeno jamstvo 6 s pripadajućom potporom, no ovaj put s pruženim dokazom za opovrgnuće 6B tog jamstva, o hipotezi H3, koja se izvela iz trećeg pojedinačnog cilja, može se dati niže naveden zaključak.

- **H3: Rezultati na e-ispitu nisu povezani s razinom ispitne anksioznosti sudionika.** Ova je hipoteza djelomično potvrđena budući da je tvrdnja opovrgnuta 6B („Rezultati na e-ispitu povezani su s razinom ispitne anksioznosti ispitanika.“) oslabljena dokazom opovrgnuta 6B, odnosno korelacijskom analizom rezultata sudionika na e-ispitu i njihovih odgovora na Upitniku o ispitnoj anksioznosti. Naime, rezultati ukazuju na to da razina ispitne anksioznosti sudionika nije statistički značajno povezana ni s jednim dijelom ispita osim s dijelom razumijevanja čitanjem, gdje je ona neznatna i negativna, što znači da sudionici s nižom razinom ispitne anksioznosti ostvaruju bolje rezultate na zadatku razumijevanja čitanjem. Ovi se rezultati djelomično podudaraju sa zaključkom Erceg Jugović i Lauri Korajlije (2012), koje su pregledom više istraživanja došle do zaključka da su ispitna anksioznost i opće akademsko postignuće negativno povezani, zatim sa zaključkom Cassidyja i Johnsona (2002), koji ističu da je kognitivna sastavnica ispitne anksioznosti ispitanika najdosljednije povezana s pogoršanjem njihove izvedbe, kao i s onima do kojih su došli Chapell i sur. (2005), prema kojima ispitna anksioznost ima učinak na prosjek ocjena, te naposljetku s onima Juretić (2008), koja ističe povezanost socijalne i ispitne anksioznosti, razine ispitne anksioznosti i samoučinkovitosti te ispitne anksioznosti kao crte ličnosti i prolaska ispita. Stoga možemo zaključiti da, budući da je naše istraživanje ispitalo učinak ispitne anksioznosti na izvedbu ispitanika na e-ispitu, potrebna su daljnja istraživanja kako bi se analizirao njezin učinak upravo na ovaj način procjenjivanja.

4.4.4. Zaključak diskusije

Valjanost je stvar prosuđivanja, a ne neko matematičko svojstvo (McNamara i Roeber, 2006) i ne odnosi se na određeni ispit, nego na tumačenje koje se temelji na njegovoj primjeni (Milas, 2009). To tumačenje treba biti kvalitetno (Bachman, 2004) i potkrijepljeno dokazima (Cronbach, 1988), a postupak utvrđivanja valjanosti neprekidan je i stupnjevit (Fitzpatrick i Clenton, 2010), pa nijedan ispit nije valjan u potpunosti (Davies i Elder, 2005). U našem je istraživanju kao postupak utvrđivanja valjanosti ispita PEJ-a primijenjena metoda potkrjepe procjenjivanja, i to njezin dio potkrjepe valjanosti procjenjivanja, kojom smo nastojali utvrditi valjanost tog ispita zaključujući na temelju dviju skupina podataka – prve je podatke predstavljala izvedba ispitanika na e-ispitu PEJ-a, a druge su podatke predstavljali dosljedni rezultati koje su ispitanici ostvarili na tom ispitu. U sklopu metode potkrjepe procjenjivanja oblikovale su se dvije tvrdnje – prethodno navedena druga skupina podataka u

potkrjepi valjanosti procjenjivanja istovremeno je predstavljala prvu tvrdnju, koja je glasila da su rezultati ispitanika dosljedni, dok je druga tvrdnja glasila da je tumačenje da je e-ispit valjan pokazatelj ispitanikovih sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u EJS-u smisleno, nepristrano, uopćivajuće, značajno i dostatno. Te su se dvije tvrdnje potkrijepile pomoću četrnaest jamstava s pripadajućim potporama, uzimajući u obzir četrnaest opovrgnuća tih jamstava s odgovarajućim dokazima, jer je za donošenje snažnih zaključaka potrebno više dokaza nego u slučaju slabih zaključaka (Kane, 2011). Stoga, budući da su gotovo sva opovrgnuća odbačena, a samo je jedno od njih oslabljeno danim dokazom, možemo tvrditi da su naši zaključci o značenju rezultata na e-ispitu snažni jer, kako napominju Chapelle i Douglas (2006), što je potpora opovrgnućima jača, to su zaključci o značenju rezultata na ispitu slabiji. Sukladno tomu, na temelju tvrdnje da su rezultati e-ispita PEJ-a dosljedni, može se donijeti relativno snažan zaključak da je e-ispit PEJ-a valjan pokazatelj ispitanikovih sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisane proizvodnje u EJS-u. Naposljetku valja istaknuti da naši rezultati i metoda potkrjepe procjenjivanja kojom se do njih došlo mogu imati široku primjenu u polju glotodidaktike, što predstavljamo o sljedećem poglavlju.

4.5. GLOTODIDAKTIČKE IMPLIKACIJE ISTRAŽIVANJA

Polje primjene e-procjenjivanja sve se više širi u nastavi kod svih dionika procesa jezičnog učenja, poučavanja i ispitivanja, i to, među ostalim, zbog želja autora da ispiti budu lakši za izradu i distribuciju, zbog želja nastavnika da budu praktičniji za provedbu i ocjenjivanje te zbog želja ispitanika odnosno učenika da budu prilagodljiviji za rješavanje na raznovrsnim uređajima. Te želje mogu biti posljedice raznih uzroka – od želja i potreba određenih ustanova i pojedinaca za digitalizacijom nastavnih procesa, preko uvođenja raznovrsnijih metoda učenja, poučavanja i procjenjivanja, od kojih je jedno i automatsko procjenjivanje jezične proizvodnje, do uzroka na koje dionici ovih procesa ne mogu utjecati, a zbog kojih e-procjenjivanje postaje jedini dostupan način jer tradicionalno procjenjivanje na papiru nije moguće provesti, kao što su pandemije, ratovi i dr. Kao jedni od dionika procesa e-procjenjivanja, nastavnici se susreću s brojnim tehnološkim, pedagoškim, afektivnim i institucijskim preprekama te razvijaju raznovrsne strategije kojima se suočavaju s tim izazovima (usp. Hartshorn i McMurry, 2020; Chung i Choi, 2021; Abduh, 2021; Ghanbari i Nowroozi, 2021; Yulianto i Mujtahin, 2021; Zhang i dr., 2021; Lučev i dr., 2022; Nguyen i sur., 2022).

Uzimajući u obzir prethodno navedeno, važno je da e-ispiti kojima se nastavnici koriste ili koje sami izrađuju budu valjani, odnosno da rezultati koje ispitanici na njima postižu budu valjani pokazatelji sposobnosti koje taj ispit ispituje. Autori ispita mogu primijeniti potkrjepu valjanosti procjenjivanja tako da tumačenje rezultata ispita o ispitanikovim sposobnostima na temelju njihove izvedbe podupru jamstvima i da odbace moguća opovrgnuća tih jamstava primjerenim dokazima. Ona se može primijeniti na sve kolegije i predmete, na razne vrste ispita – od onih niskog rizika, kojima se prati napredak učenika i studenata te im se daju određene povratne obavijesti o njihovoj izvedbi, do onih srednjeg rizika, kojima se procjenjuje njihovo postignuće u određenom području. Na ovaj će se način poboljšati kvaliteta e-procjenjivanja na visokoškolskim ustanovama na ispitima srednjeg rizika poput završnih semestralnih ispita, ali i u srednjim školama u pojedinom nastavnom predmetu na ispitima srednjeg rizika prilikom procjenjivanja njihova postignuća ili na ispitima niskog rizika radi praćenja učenika u usvajanju nastavnog gradiva. Ispiti čija se valjanost utvrdi metodom potkrjepe valjanosti procjenjivanja mogu imati formativnu ulogu i za nastavnike budući da im mogu ukazati na područja koja mogu poboljšati prilikom procjenjivanja učenikova ovladavanja određenim standardima (usp. Llosa, 2008).

Naposljetku, važno je istaknuti da je potkrjepa procjenjivanja koju smo primijenili u svojem istraživanju jedna od mogućih metoda utvrđivanja valjanosti (usp. Messick, 1980, 1989; Angoff, 1988; McNamara, 2000; Urbina, 2004; Davies i Elder, 2005, 2010; Weir, 2005; Llosa, 2008; Shepard, 2016). Međutim, temeljna načela ove metode utvrđivanja valjanosti ispita mogu se primijeniti na druge sveučilišne kolegije, ali i druge nastavne predmete na drugim ustanovama, kao što su to svojim istraživanjem dokazali Long i sur. (2018), koji ističu da se njihova temeljna načela metode utvrđivanja valjanosti novog e-ispita španjolskog jezika mogu primijeniti na druge strane jezike u sveučilišnim kontekstima.

5. ZAKLJUČAK

Primjena e-procjenjivanja i njegova valjanost još su uvijek nedovoljno istražena područja gdje se svakodnevno dolazi do novih saznanja kojima ih istraživači obogaćuju. Naše istraživanje pruža znanstvenu utemeljenost primjene e-procjenjivanja i obogaćuje znanstvenu literaturu vezanu za valjanost e-procjenjivanja engleskoga kao jezika struke (EJS) iako se rezultati do kojih smo došli mogu primijeniti na druge jezike, ali i druge kolegije i nastavne predmete. Ono što je od ključne važnosti jest da koji god bio razlog uvođenja e-procjenjivanja, ono mora biti valjano, neovisno o tomu gleda li se na valjanost kao na jedinstveni koncept koji objedinjuje više njezinih aspekata, što je i naše stajalište, ili o tomu smatra li se svaki od njih njezinom zasebnom vrstom (Anastasi, 1976; Messick, 1980, 1988; Angoff, 1988; Cronbach, 1988; Urbina, 2004; Davies i Elder, 2005; Weir, 2005; Milas, 2009). U svakom se slučaju ona mora moći dokazati, imajući na umu da nikad nije potpuna, već da je na određenu stupnju (Anastasi, 1975, 1976; Angoff, 1988; Messick, 1989; Urbina, 2004; Bachman, 2004; Weir, 2005; Milas, 2009; Fitzpatrick i Clenton, 2010; Shepard, 2016).

Međutim, nije dovoljno samo tvrditi da je određeno procjenjivanje valjano, nego se te tvrdnje moraju potkrijepiti primjerenim jamstvima koja moraju imati odgovarajuću potporu, a istovremeno se moraju sagledati i sva moguća opovrgnuća dotične tvrdnje koja se moraju odbaciti zahvaljujući odgovarajućim empirijskim dokazima (Messick, 1980; Davies, 1984; Angoff, 1988; Cronbach, 1988; Messick, 1989; Kane, 1992, 2010; Bachman i Palmer, 1996; Kane, Crooks i Cohen, 1999; McNamara, 2000; Chapelle, Jamieson i Hegelheimer, 2003; Toulmin, 2003; Bachman, 2004, 2005, 2015; Urbina, 2004; Davies i Elder, 2005, 2010; Weir, 2005; Chapelle i Douglas, 2006; McNamara i Roever, 2006; Milas, 2009; Bachman i Palmer, 2010; Shepard, 2016). Ipak, budući da valjanost nikad nije potpuna, neki dokazi neće u potpunosti odbaciti moguća opovrgnuća, već će ih samo oslabiti ili čak poduprijeti. Stoga se mora težiti tomu da potpore i dokazi valjanosti e-procjenjivanja budu što snažniji i da u što većoj mjeri potkrjepljuju dotično procjenjivanje. Taj konceptualni okvir potkrjepe procjenjivanja osmislio je Bachman (2005, 2015) odnosno Bachman i Palmer (2010), i podijeli ga na dva dijela – potkrjepu valjanosti procjenjivanja i potkrjepu primjene procjenjivanja. Mi smo se koristili potkrjepom valjanosti procjenjivanja u svojem istraživanju te smo ujedno uveli nove termine i obogatili pojmovlje u hrvatskom jeziku u području procjenjivanja, što je prikazano u glosaru na kraju ovog rada (prilog 7). I drugi su istraživači primijenili potkrjepu procjenjivanja te poduprli svoja jamstva i pružili odgovarajuće dokaze mogućih opovrgnuća za zaključke do kojih su došli na temelju rezultata sudionika na ispitima primijenjenima u

istraživanjima (Chapelle, Jamieson i Hegelheimer, 2003; Llosa, 2008; Wang i sur., 2012; Long i sur., 2018; Jun, 2021; Park, 2021).

U našem istraživanju, primijenivši potkrjepu valjanosti e-procjenjivanja u EJS-u, pružila se potpora jamstvima tvrdnje da su rezultati koje su sudionici ovog istraživanja ostvarili prilikom svoje jezične izvedbe na e-ispitu Poslovnog engleskog jezika (PEJ) dosljedni jer se procedure za provedbu e-ispita slijede dosljedno, jer su procedure za procjenjivanje e-ispita dobre i slijede se dosljedno te jer su njihovi rezultati dosljedni neovisno o načinu ispitivanja, odnosno neovisno o tomu polažu li e-ispit ili ispit na papiru. Sva moguća opovrgnuća ovih jamstava odbačena su dokazima koji su potvrdili da se te procedure slijede dosljedno, a zahvaljujući empirijskim dokazima do kojih se došlo primjenjujući rezultate korelacijskih analiza odbacilo se opovrgnuće tvrdnje da su rezultati sudionika na e-ispitu različiti od njihovih rezultata na ispitu na papiru u zadatku razumijevanja čitanjem, u zadacima razumijevanja i primjene vokabulara te u zadatku pisane proizvodnje, kao i opovrgnuće tvrdnje da su rezultati na e-ispitu povezani sa stavovima sudionika o e-ispitima, o ispitima općenito i o računalima.

Nadalje, pružila se potpora jamstvima tvrdnje da je tumačenje da je e-ispit valjan pokazatelj sposobnosti razumijevanja čitanjem, razumijevanja i primjene vokabulara te pisanja u EJS-u smisleno, nepristrano, uopćivajuće, značajno i dostatno jer je konstrukt e-ispita jasno definiran, jer je njegova ispitna specifikacija jasna, jer je omogućena najbolja moguća izvedba ispitanika, jer su ispitni zadatci nepristrani, jer je sadržaj e-ispita neuvredljiv, jer su jasna mjerila bodovanja, jer je provedba e-ispita nepristrana, jer su uvjeti provedbe ispita propisani i dani prilikom provedbe, jer e-ispit odgovara ciljnoj primjeni jezika, jer su jednaka mjerila bodovanja na e-ispitu kao i u kontekstu EJS-a te jer su rezultati značajan i dostatan pokazatelj razine sposobnosti ispitanika. I kod ove tvrdnje, kao i kod prethodne, sva moguća opovrgnuća njezinih jamstava odbačena su odgovarajućim dokazima. Ti su dokazi potvrdili da je sadržaj e-ispita usklađen s ishodima i silabom kolegija te nastavnim materijalima kolegija, da su ishodi, silab i nastavni materijali upotrijebljeni na nastavi i da su stalno dostupni ispitanicima u Googleovoj učionici, da su mjerila bodovanja objašnjena usmeno i napisana na e-ispitu, da su dane upute i propisani uvjeti provedbe e-ispita te da su kvaliteta automatskog procjenjivanja u Googleovu obrascu i kvaliteta procjene procjenjivača jednake. Nadalje, empirijski dokazi do kojih se došlo korelacijskom analizom također su potvrdili da stavovi ispitanika o e-ispitima PEJ-a, o ispitima općenito i o računalima nisu povezani s rezultatima na e-ispitu. Jedini dokaz kojime se opovrgnuće ove tvrdnje nije odbacilo nego oslabilo jesu rezultati korelacijske analize koji pokazuju da su rezultati na e-ispitu povezani s razinom ispitne anksioznosti sudionika budući da ona nije statistički značajno povezana ni s jednim dijelom e-ispita osim s dijelom

razumijevanja čitanjem, gdje je neznatna i negativna. Ovo ukazuje na to da sudionici s nižom razinom ispitne anksioznosti postižu bolje rezultate u zadatku razumijevanja čitanjem.

Zato možemo reći da je naš zaključak djelomično u skladu sa zaključkom Fitzgerald i Clentona (2010), koji navode da je njihova potkrjepa valjanosti snažna jer su im sve hipoteze o valjanosti (pouzdanost rezultata, poboljšanje u jezičnom znanju, rezultati usporedivi s rezultatima na sličnim ispitima) potkrijepljene empirijskim istraživanjima. Nadalje, iako navode da njihovi rezultati podupiru primjenu ispita koji su istraživali u istraživačke svrhe, čime je dokazana njegova pragmatična valjanost, nisu uspjeli dokazati njegovu korisnost nastavnicima i učenicima, za što su potrebna dodatna istraživanja. Međutim, budući da se naše istraživanje bavilo analizom valjanosti, a ne i primjenom e-procjenjivanja u EJS-u, te smo u skladu s time potkrjepljivali valjanost procjenjivanja, možemo tvrditi da su ispiti kojima smo se koristili u ovom istraživanju valjani. Primjena e-procjenjivanja u EJS-u mogla bi se također istražiti Bachmanovom metodom potkrjepe procjenjivanja, i to dijelom koji se odnosi na potkrjepu primjene procjenjivanja, što je jedno od mogućih daljnjih istraživanja u ovom području. Na taj bi se način pokušalo vidjeti hoće li rezultati tog istraživanja biti u skladu s rezultatima Chapelle, Jamieson i Hegelheimera (2003), čije je istraživanje ukazalo na važnost koju utvrđivanje valjanosti ima na početne korake izrade ispita, te s onima Wang i sur. (2012), koji ističu da bi autori ispita trebali surađivati s donositeljima odluka radi jasna određivanja primjene dotičnog ispita.

Što se tiče drugih daljnjih istraživanja analize same valjanosti e-procjenjivanja, prije nego što ih navedemo valja razmotriti **ograničenja** našega istraživanja. **Prvo ograničenje** jest uzorak. Sudjelovanje u istraživanju bilo je ponuđeno svim studentima prve godine svih nastavnika kolegija PEJ 1, pa i samog istraživača. Iako je pojedinačno najviše broj sudionika bilo istraživačevih, oni su ipak predstavljali samo 29,5 % ukupnog uzorka, dok je preostalih 71,5 % bilo neravnomjerno raspoređeno među ostalih devetero nastavnika. Uzorak ovog istraživanja odabran je kako bi se na njemu mogao primijeniti i prvi instrument – jezični ispit PEJ-a, koji je ujedno i **drugo ograničenje** istraživanja. Radi praktičnosti provedbe istraživanja, odlučeno je da će se primijeniti ispit čiji konstrukt odgovara gradivu koje sudionici obrađuju na nastavi kako bi ih se potaknulo da sudjeluju u istraživanju. Ovaj je instrument izradom i odabirom odgovarajućih čestica morao odgovarati ne samo gradivu kolegija, nego i pravilima provedbe ispita na matičnom fakultetu, pa je stoga njegovo rješavanje trajalo sat vremena. Nadalje, instrument je izradio istraživač, koji je jedan od predmetnih nastavnika, budući da na toj ustanovi svaki nastavnik samostalno izrađuje svoje ispite u skladu sa silabom. **Treće ograničenje** istraživanja jest procjenjivanje izvedbe ispitanika na jezičnom ispitu. Procjena

pisane proizvodnje nije bila automatska jer trenutačno dostupna tehnologija nije na zadovoljavajućoj razini. Autor ispita bio je istovremeno i jedan od dvoje procjenjivača zadatka pisane proizvodnje, što je također odlučeno radi praktičnosti provedbe istraživanja, no ni njemu, kao ni drugom procjenjivaču, nije bio poznat identitet pojedinog sudionika prilikom procjenjivanja jer je pri procjenjivanju ispita na papiru taj zadatak bio na posebnoj stranici na kojoj nije navedeno ime sudionika, a pri procjenjivanju e-ispita taj je zadatak bio u zasebnom odjeljku Googleova obrasca. **Četvrto ograničenje** odnosi se na druge instrumente, odnosno na odabir primijenjenih upitnika do kojih se došlo pregledom dostupne literature. U istraživanju su primijenjeni dotični standardizirani upitnici jer se njima već koristilo u drugim istraživanjima i bili su u skladu s ciljevima našega.

Slijedom prethodno navedenoga, moguće **smjernice za daljnja istraživanja** opisane su niže.

- U uzorak uključiti veći broj sudionika – ne samo one na prvoj godini studija, već na svim godinama studija. U tomu bi se slučaju trebao izraditi i novi instrument koji bi obuhvaćao gradivo svih godina studija, odnosno koji bi ispitivao vještine koje su zajedničke silabima svih godina. Također, u tom bi se slučaju mogla analizirati korelacija između e-ispita i ispita na papiru studenata različitih godina, čime bi se dodala nova hipoteza koju bi se trebalo ispitati.
- U uzorak uključiti podjednak broj sudionika svih nastavnika. U tom bi se slučaju mogla analizirati još jedna korelacija – ona između e-ispita i ispita na papiru studenata različitih nastavnika, čime bi se dodala još jedna nova hipoteza za analizu.
- U uzorak uključiti i nastavnike, ispitivače, administratore ispita kako bi se ispitali i njihovi stavovi o ispitu čija se valjanost utvrđuje, što je u skladu sa zaključcima istraživanja Winke (2011), koja navodi stavovi da su stavovi obrazovnih djelatnika uključenih u proces ispitivanja neophodni za utvrđivanje valjanosti dotičnog ispita.
- Ispit kojime se koristilo kao instrumentom u ovom istraživanju može biti veći i sadržavati više čestica u svakom dijelu, a posebice u dijelu razumijevanja čitanjem budući da smo u svojem istraživanju došli do rezultata koji ukazuju na to da je samo ovaj dio ispita statistički značajno povezan s razinom ispitne anksioznosti sudionika, što nije bio slučaj za ostale dijelove ispita. Također, ispit može sadržavati i nove dijelove – razumijevanje slušanjem te govornu proizvodnju i interakciju – čime bi se također povećao broj hipoteza istraživanja.
- Pri procjenjivanju zadatka/zadataka jezične proizvodnje može se uključiti više procjenjivača.

- Procjenjivanje zadatka jezične proizvodnje (tj. pisanja i/ili govorenja) može se provesti postupkom automatskog procjenjivanja jezične proizvodnje, ovisno o stupnju napretka te tehnologije u trenutku provedbe istraživanja, vodeći računa o tomu da ispitivanje ne bude vođeno tehnologijom, nego da se ona primijeni radi njegova unaprjeđenja (Brunfaut, 2023).
- Kao instrumenti mogli bi se primijeniti drugi upitnici kojima se ispituju stavovi sudionika o e-ispitima, o ispitima općenito i o računalima te razina njihove ispitne anksioznosti, vodeći računa o tomu da su ti instrumenti također primijenjeni u prijašnjim istraživanjima.
- Ako je moguće, usporediti silabe drugih kolegija, i to ne samo stranih jezika, te izraditi instrument na hrvatskom jeziku koji bi ispitivao vještine koje su zajedničke silabima dotičnih kolegija.
- Budući da vrlo rijetko valjanost ispita utvrđuje sam autor ispita jer on može imati određeni subjektivan interes za ishod tog postupka te može nevoljko tražiti dokaze protiv namijenjenih tumačenja dotičnog ispita (Davies i Elder, 2005), osigurati da autor ispita nije ista osoba koja utvrđuje i njegovu valjanost. Ova je moguća smjernica za daljnja istraživanja također u skladu sa zaključkom Winke (2011), koja ističe neutralnost vanjske osobe koja utvrđuje valjanost dotičnog ispita.

Naše je istraživanje nastojalo pružiti empirijski dokaz za valjanost primjene e-procjenjivanja engleskoga kao jezika struke te se njime upotpunjuje znanstvena literatura na globalnoj razini, a posebice u hrvatskom kontekstu budući da istraživanja valjanosti primjenom metode potkrjepe procjenjivanja ne postoje, kao ni pojmovlje u području ovog konceptualnog okvira (v. glosar u prilogu 7). Upravo ovakvi konceptualni alati za razvoj potkrjepe valjanosti pomažu istraživačima uvidjeti gdje su potrebna daljnja istraživanja (Chapelle, 2011). U našem se slučaju rezultatima nastojalo poboljšati kvalitetu e-procjenjivanja ispitima srednjeg rizika na visokoškolskim ustanovama, ali u srednjim i osnovnim školama koje primjenjuju e-ispite kojima se procjenjuje izvedba ispitanika.

POPIS LITERATURE

- Abduh, M. Y. M. (2021).** Full-time Online Assessment during COVID-19 Lockdown: EFL Teachers' Perceptions. *Asian EFL Journal Research Articles*, 28(1).
- Al-Amri, S. (2007).** *Computer-based vs. Paper-based Testing: Does the test administration mode matter?* Proceedings of the BAAL Conference 2007.
- Alderson, J. C. (1988).** New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing*, 5, 220-232.
- Alderson, J. C. i Wall, D. (1993).** Does washback exist? *Applied Linguistics* 14(2), 115-129.
- Alderson, J. C. i Banerjee, J. (2001).** Language testing and assessment (Part 1). *Language Teaching*, 34, 213-236.
- Alderson, J. C. i Banerjee, J. (2002).** Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- Aliakbari, M. i Gheitasi, M. (2017).** Investigating students' test anxiety and attitude toward foreign language learning in secondary school in Ilam. *Journal of Pan-Pacific Association of Applied Linguistics*, 20(2), 51-69.
- Allport, G. W. (1971).** Attitudes. U Thomas, K. (ur.), *Attitudes and Behaviour*. Harmondsworth: Penguin.
- Anastasi, A. (1976).** *Psychological Testing*. New York: Macmillan
- Angoff, W. H. (1988).** Validity: An Evolving Concept. U Wainer, H. i Braun, H. (ur.), *Test validity* (str. 19-32). Hillsdale, NJ: Lawrence Erlbaum.
- Bachman, L. F. (1990).** *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004).** *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. (2005).** Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L. F. (2015).** Justifying the Use of Language Assessments: Linking Test Performance with Consequences. *JLTA Journal*, 18, 3-22.
- Bachman, L. F. i Palmer, A. S. (1996).** *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F. i Palmer, A. S. (2010).** *Language assessment in practice* (2. izdanje). Oxford: Oxford University Press.
- Bae, Y. i Lee, Y-S. (2010).** The validation of parallel test forms: 'Mountain' and 'beach' picture series for assessment of language skills. *Language Testing* 28(2), 155-177.
- Bagarić Medve, V. (2011).** Jezično ispitivanje danas. *Strani jezici*, 40(4), 319-342.
- Bagarić Medve, V. (2023).** Jezični ispiti u nastavi stranih jezika. U Bagarić Medve, V. i Pavičić Takač, V. (ur.), *Učenje i poučavanje stranih jezika*. Osijek: Filozofski fakultet, 461-500.
- Bagarić Medve, V. i Škarica, I. (2023).** Vrednovanje u nastavi stranih jezika. U Bagarić Medve, V. i Pavičić Takač, V. (ur.), *Učenje i poučavanje stranih jezika*. Osijek: Filozofski fakultet, 451-459.
- Barkaoui, K. (2014).** Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL iBT writing tasks. *Language Testing*, 31(2), 241-259.
- Banerjee, J. i Luoma, S. (1997).** Qualitative Approaches to Test Validation. U Clapham, C. i Corson, D. (ur.), *Encyclopedia of Language and Education, vol. 7 – Language Testing and Assessment* (str. 275-287). Dordrecht, Boston i London: Kulwer Academic Publishers.

- Basturkmen, H. (2010).** *Developing Courses in English for Specific Purposes*. Basingstoke: Palgrave Macmillan.
- Brinton, D. M. (2013).** Content-Based Instruction in English for Specific Purposes. U Chapelle, C. A. (ur.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.
- Brooks, R. L. (2017).** Language Assessment in the US Government. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 62-76). New York: Springer.
- Brown, H. D. i Abeywickrama, P. (2010).** *Language Assessment: Principles and Classroom Practices*. Pearson Education.
- Brunfaut, T. (2023).** Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, 40(1) 15-23.
- Brunfaut, T. i Clapham, C. (2013).** Attitudes and language learning. U Byram, M. i Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, 59-65. London: Routledge.
- Cassady, J. C. i Johnson, R. E. (2002).** Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, 27, 270-295.
- Cergol, K. (2021).** *Etika istraživanja u primijenjenoj lingvistici*. Zagreb : Srednja Europa.
- Chalhoub-Deville, M. (2001).** Language testing and technology: Past and future. *Language Learning & Technology*, 5(2), 95-98.
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A. i McCann, N. (2005).** Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology*, 97, 268-274.
- Chapelle, C. A. (2011).** Validity argument for language assessment: The framework is simple... *Language Testing*, 29(1), 19-27.
- Chapelle, C. A. (2021).** Validity in Language Assessment. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 11-20). New York i London: Routledge.
- Chapelle, C. A., Jamieson J. i Hegelheimer, V. (2003).** Validation of a web-based ESL test. *Language Testing*, 20; 409-439.
- Chapelle, C. A. i Douglas, D. (2006).** *Assessing Language Through Technology*. Cambridge: Cambridge University Press.
- Chapelle, C. A. i Lee, H. (2021).** Understanding Argument-Based Validity in Language Testing. U Chapelle, C. A. i Voss, E. (ur.), *Validity Argument in Language Testing – Case Studies of Validation Research* (str. 19-44). Cambridge University Press.
- Chapelle, C. A. i Voss, E. (2016).** 20 years of technology and language assessment in Language Learning & Technology. *Language Learning & Technology*, 20(2), 116-128.
- Chapelle, C. A. i Voss, E. (2017).** Utilizing Technology in Language Assessment. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 149-161). New York: Springer.
- Chapelle, C. A. i Voss, E. (ur.) (2021).** *Validity Argument in Language Testing – Case Studies of Validation Research*. Cambridge University Press.
- Checa García, I. i Guiberson, M. (2019).** Test validity in morphosyntactic measures for typical and SLI incipient Spanish English bilinguals. *Language Testing*, 36(1), 77-100.
- Chen, C.-F. E. i Cheng, W.-Y. E. (2008).** Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94-112.

- Cheng, L. i Sun, Y. (2015).** Interpreting the impact of the Ontario Secondary School Literacy Test on second language students within an argument based validation framework. *Language Assessment Quarterly*, 12(1), 50-66.
- Cheng, L. i Zheng, Y. (2021).** Measuring Anxiety. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 177-186). New York i London: Routledge.
- Chung, S.-J. i Choi, L.-J. (2021).** The Development of Sustainable Assessment during the COVID-19 Pandemic: The Case of the English Language Program in South Korea. *Sustainability*, 13, 4499.
- Chung, Y. R. (2014).** *A test of productive English grammatical ability in academic writing: Development and validation*. Neobjavljen doktorski rad, Iowa State University, Ames, IA.
- Cotos, E. (2013).** Potential of Automated Writing Evaluation Feedback. *CALICO Journal*, 28(2), 420-459.
- Council of Europe (2001).** *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cronbach, L. J. (1988).** Five perspectives on validation argument. U Wainer, H. i Braun, H. (ur.), *Test validity* (str. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Davies, A. (1984).** Validating three tests of English language proficiency. *Language Testing*, 1, 50-69.
- Davies, A. (2011).** Kane, validity and soundness. *Language Testing*, 29(1) 37-42.
- Davies, A. i Elder, C. (2005).** Validity and validation in language testing. U Hinkel, E. (ur.), *Handbook of research in second language teaching and learning* (str. 795-814). London: Routledge.
- Department for Education (2015).** *Technical Awards for 14 to 16 Year Olds: 2017 and 2018 performance tables: Technical guidance for awarding organisations*. London: Department for Education.
- Deygers, B., van den Branden, K., i van Gorp, K. (2018).** University entrance language tests: A matter of justice. *Language Testing*, 35(4), 449-476.
- Dizdar, D. (2006).** *Kvantitativne metode*. Zagreb: Kineziološki fakultet Sveučilišta u Zagrebu.
- Dizon, J. (2016).** Measuring Japanese EFL Student Perceptions of Internet-Based Tests with the Technology Acceptance Model. *The Electronic Journal for English as a Second Language*, 20(2).
- Dobson, A. (2013).** Evaluation. U Byram, M. i Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, 233-239. London: Routledge.
- Douglas, D. (2001).** Performance consistency in second language acquisition and language testing research: a conceptual gap. *Second Language Research*, 17(4), 442-456.
- Dudley-Evans, T. i St John M. J. (1998).** *Developments in English for Specific Purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Dunkel, P. A. (1999).** Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning & Technology*, 2(2), 77-93.
- Dursun, A. i Li, Z. (2021).** A Systematic Review of Argument-Based Validation Studies in the Field of Language Testing (2000-2018). U Chapelle, C. A. i Voss, E. (ur.), *Validity Argument in Language Testing – Case Studies of Validation Research* (str. 45-70). Cambridge University Press.
- Educational Testing Service (2023).** *The Criterion® Online Writing Evaluation Service*. Pristupljeno 8. travnja 2023. na: <https://www.ets.org/criterion.html>
- Eiser, J. R. (1986).** *Social psychology – Attitudes, cognition and social behaviour*. Cambridge: Cambridge University Press.

- Erceg Jugović, I. i Lauri Korajlija, A. (2012).** Povezanost ispitne anksioznosti s perfekcionizmom. *Psihologijske teme*, 21(2), 299-316.
- Esfandiari, M. R., Riasati, M. J., Vaezian, H. i Rahimi, F. (2018).** A quantitative analysis of TOEFL iBT using an interpretive model of test validity. *Language Testing in Asia*, 8(1), 7.
- Falchikov, N. (2004).** Involving students in assessment. *Psychology Learning and Teaching*, 3(2), 102-108.
- Fan, J. i Ji, P. (2014).** Test candidates' attitudes and their test performance: The case of the Fudan English Test. *University of Sydney Papers in TESOL*, 9, 1-35.
- Field, A. (2013).** *Discovering statistics using SPSS (4th edition)*. SAGE Publications Ltd.
- Fitzpatrick, T. i Clenton, J. (2010).** The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27(4), 537-554.
- Fox, J. (2017).** Using Portfolios for Assessment/Alternative Assessment. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 135-147). New York: Springer.
- Fulcher, G. (2010).** *Practical Language Testing*. London: Hodder Education.
- Fulcher, G. i Davidson, F. (2007).** *Language testing and assessment: An advanced resource book*. Abingdon: Routledge.
- Gaillard, S. (2014).** *The elicited imitation task as a method for French proficiency assessment in institutional and research settings*. Neobjavljen doktorski rad, University of Illinois at Urbana Champaign, Champaign, IL.
- García Laborda, J. (2007).** On the net: Introducing standardized EFL/ESL exams. *Language Learning & Technology*, 11(2), 3-9.
- Gardner, R. (1985).** *Social Psychology and Second Language Learning: The Role of Attitude and Motivation*. London: Edward Arnold.
- Garrett, N. (2009).** Computer-Assisted Language Learning Trends and Issues Revisited: Integrating Innovation. *The Modern Language Journal*, 93, 719-740.
- George, D. i Mallery, M. (2010).** *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.)*. Pearson.
- Glavaš, G. (2011).** *Automatizirano programsko inženjerstvo temeljeno na algoritmima pretraživanja i obradi prirodnog jezika*. Rad za kvalifikacijski doktorski ispit na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Prisupljeno 10. travnja 2023. na <https://www.researchgate.net/publication/228414281>
- Ghanbari, N. i Nowroozi, S. (2021).** The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, 11(27).
- Godwin-Jones, R. (2008).** Web-writing 2.0: Enabling, documenting, and assessing writing online. *Language Learning & Technology*, 12(2), 7-13.
- Gorsuch, G. (2004).** Test Takers' Experiences with Computer-administered Listening Comprehension Tests: Interviewing for Qualitative Explorations of Test Validity. *CALICO Journal*, 21(2), 339-371.
- Gravetter, F. i Wallnau, L. (2014).** *Essentials of Statistics for the Behavioral Sciences (8th ed.)*. Wadsworth.
- Green, B. F. (1988).** Construct Validity of Computer-Based Tests. U Wainer, H. i Braun, H. (ur.), *Test Validity* (str. 19-32). Hillsdale, NJ: Lawrence Erlbaum.
- Green, A. (2014).** *Exploring Language Assessment and Testing*. London i New York: Routledge.
- Green, A. i Fulcher, G. (2021).** Test Design Cycle. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 69-77). New York i London: Routledge.

- Grotjahn, R. (1986).** Test validation and cognitive psychology: some methodological considerations. *Language Testing*, 3, 159-185.
- Gruba, P. i Corbel, C. (1997).** Computer-Based Testing. U Clapham, C. i Corson, D. (ur.), *Encyclopedia of Language and Education*, vol. 7 – *Language Testing and Assessment* (str. 141-149). Dordrecht, Boston i London: Kulwer Academic Publishers.
- Halonja, A. i Mihaljević, M. (2012).** Nazivi sa sastavnicom e- u hrvatskome jeziku. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 38(1), 55-86.
- Harsch, C. i Malone, M. E. (2021).** Language Proficiency Frameworks and Scales. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 33-44). New York i London: Routledge.
- Harsch, C., Ushioda, E. i Ladroue, C. (2017).** Investigating the predictive validity of TOEFL iBT® test scores and their use in informing policy in a United Kingdom University setting. *ETS Research Report Series*, (1), 1-80.
- Hartshorn, K. J. i McMurry B. L. (2020).** The Effects of the COVID-19 Pandemic on ESL Learners and TESOL Practitioners in the United States. *International Journal of TESOL Studies*, 2(2), 140-156.
- Hewson, C. i Charlton, J. P. (2019).** An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences. *Journal of Computer Assisted Learning*, 35(1), 51-60.
- Hewson, C., Charlton, J. i Brosnan, M. (2007).** Comparing online and offline administration of multiple choice question assessments to psychology undergraduates: Do assessment modality or computer attitudes influence performance? *Psychology Learning and Teaching*, 6(1), 37-46.
- Hoang, G. T. L. i Kunnan, A. J. (2016).** Automated Essay Evaluation for English Language Learners: A Case Study of MY Access. *Language Assessment Quarterly*, 13(4), 359-376.
- Hoang, T. D., Phan, T. L., i Le, H. P. (2021).** Non-English Major Students' Perceptions Towards TOEIC Online Learning and Testing. *AsiaCALL Online Journal*, 12(5), 114-128.
- Hrvatska enciklopedija, mrežno izdanje. (2021).** Leksikografski zavod Miroslav Krleža. Pristupljeno 26. ožujka 2023. na:
<http://www.enciklopedija.hr/Natuknica.aspx?ID=71291>
- Hughes, A. i Hughes, J. (2020).** *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hutchinson, T. i Waters, A. (1987).** *English for Specific Purposes: A learning-centered approach*. Cambridge: Cambridge University Press.
- Isbell, D. R. i Kremmel, B. (2020).** Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 37(4), 600-619.
- Jelaska, Z. i Cvikić, L. (2008).** Procjenjivanje i vrjednovanje u hrvatskome jeziku. *Lahor*, 1(5), 115-125.
- Johns, A. M. (2013).** The History of English for Specific Purposes Research. U Paltridge, B. i Starfield, S. (ur.), *The Handbook of English for Specific Purposes* (str. 5-30). Chichester, UK: Wiley-Blackwell.
- Jun, H. (2021).** Justifying the Interpretation and Use of an ESL Writing Final Examination. U Chapelle, C. A. i Voss, E. (ur.), *Validity Argument in Language Testing – Case Studies of Validation Research* (str. 235-263). Cambridge University Press.
- Jun, H. S. (2014).** *A validity argument for the use of scores from a web search permitted and web source based integrated writing test*. Neobjavljen doktorski rad, Iowa State University, Ames, IA.

- Juretić, J. (2008).** Socijalna i ispitna anksioznost te percepcija samoeфикаsnosti kao prediktori ishoda ispitne situacije. *Psihologijske teme*, 17(1), 15-36.
- Kamenov, Ž. (2017).** *Valjanost u istraživanjima*. Prezentacija na kolegiju Metodologija i prezentacija znanstvenoga rada na Poslijediplomskom doktorskom studiju Glotodidaktike, Filozofski fakultet Sveučilišta u Zagrebu.
- Kane, M. (1992).** An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Kane, M. (2001).** Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319-342.
- Kane, M. (2004).** Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2, 135-170.
- Kane, M. (2010).** Validity and fairness. *Language Testing*, 27(2), 177-182.
- Kane, M. (2011).** Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing*, 29(1) 3-17.
- Kane, M., Crooks, T. i Cohen, A. (1999).** Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5-17.
- Kelly Riley, D. i Elliot, N. (2014).** The WPA Outcomes Statement, validation, and the pursuit of localism. *Assessing Writing*, 21, 89-103.
- Kessler, G. (2013).** Authoring tools for language assessment. U Chapelle, C. A. (ur.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.
- Khoshsima, H. i sur. (2017).** The Impact of Digital Item Presentation on the Intermediate Level Language Learners' Testing Performance and Attitudes towards Onscreen Test in Private Language Learning Context. *Journal of Applied Linguistics and Language Research*, 4(3), 196-211.
- Koo, T. K. i Li, M. Y. (2016).** A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15, 155-163.
- Kramer, J. (2013).** Assessment and testing. U Byram, M. i Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, 52-57. London: Routledge.
- Kraš, T. i Miličević, M. (2015).** *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika*. Rijeka: Filozofski fakultet Sveučilišta u Rijeci.
- Kuiken, F. i Vedder, I. (2021).** Scoring Approaches – Scales/Rubrics. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 125-134). New York i London: Routledge.
- Kumazawa, T., Shizuka, T., Mochizuki, M. i Mizumoto, M. (2016).** Validity argument for the VELC Test® score interpretations and uses. *Language Testing in Asia*, 6(2).
- Kuna, D. i Ferčec, I. (2023).** Strani jezik za posebne namjene. U Bagarić Medve, V. i Pavičić Takač, V. (ur.), *Učenje i poučavanje stranih jezika* (str. 589-604). Osijek: Filozofski fakultet.
- Lasagabaster, D. (2013).** Attitudes and Motivation in Bilingual Education. U Chapelle, C. A. (ur.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.
- Lesiak-Bielawska, E. D. (2015).** Technology in ESP Pedagogy. *English for Specific Purposes World*, 48(16).
- Li, M., Luo, L., Sikdar, S., Nizam, N. I., Gao, S., Shan, H., Kruger, M., Kruger, U., Mohamed, H., Xia, L. i Ge Wang, G. (2021).** Optimized collusion prevention for online exams during social distancing. *npj Science of Learning*, 6(1).
- Li, S. (2018).** Developing a test of L2 Chinese pragmatic comprehension ability. *Language Testing in Asia*, 8(1), 3.
- Lim, H. i Kahng, J. (2012).** Review of the software CRITERION. *Language Learning & Technology*, 16(2), 38-45.

- Liu, S. i Kunnan, A. (2016).** Investigating the Application of Automated Writing Evaluation to Chinese Undergraduate English Majors: A Case Study of WriteToLearn. *CALICO Journal*, 33(1), 71-91.
- Llosa, L. (2008).** Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice*, 27(3), 32-42.
- Llosa, L. i Malone, M. E. (2018).** Comparability of students' writing performance on TOEFL iBT and in required university writing courses. *Language Testing*, 36(2), 235-263.
- Long, A. Y., Shin, S.-Y., Geeslin, K. i Willis, E. V. (2018).** Does the test work? Evaluating a web-based language placement test. *Language Learning & Technology*, 22(1), 137-156.
- Lučev, I., Lukić Jakopčević, L. i Špiljak, V. (2022).** Iskustva predavača stranih jezika struke u visokom školstvu Republike Hrvatske u online nastavi 2020. *Strani jezici*, 51, 23-54
- Matić, A. (2017).** Obilježja pokreta očiju i modeli čitanja utemeljeni na uporabi uređaja za praćenje pokreta očiju. *Hrvatska revija za rehabilitacijska istraživanja*, 53(2), 130-141.
- McNamara, T. (1996).** *Measuring second language performance*. London: Longman.
- McNamara, T. (2000).** *Language Testing*. Oxford: Oxford University Press.
- McNamara, T. i Roever, C. (2006).** *Language Testing: The Social Dimension*. Malden, MA: Blackwell.
- Medved Krajnović, M. (2010).** *Od jednojezičnosti do višejezičnosti – Uvod u istraživanja procesa ovladavanja inim jezikom*. Zagreb: Leykam International.
- Messick, S. (1980).** Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1988).** The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement. U Wainer, H. i Braun, H. (ur.), *Test validity* (str. 33-45). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S. (1989).** Validity. U Linn, R. L. (ur.), *Educational measurement* (str. 13-103). New York: Macmillan.
- Milas, G. (2009).** *Istraživačke metode u psihologiji i drugim društvenim znanostima*. 2. izd. Jastrebarsko: Naklada Slap.
- Milošević, M. (2010).** *Izrada mjernog instrumenta stresa na radnom mjestu bolničkih zdravstvenih djelatnika i procjena njegove uporabne vrijednosti*. Doktorska disertacija.
- Ministarstvo znanosti i obrazovanja (2019).** *Pravilnik o izmjenama i dopuni Pravilnika o načinima, postupcima i elementima vrednovanja učenika u osnovnim i srednjim školama*. Narodne novine, br. 82/2019.
- Ministarstvo znanosti, obrazovanja i športa (2010).** *Pravilnik o načinima, postupcima i elementima vrednovanja učenika u osnovnoj i srednjoj školi*. Narodne novine, br. 112/2010.
- Mohammadi, M. i Barzgaran, M. (2012).** Comparability of Computer-based and Paper-based Versions of Writing Section of PET in Iranian EFL Context. *Journal of Foreign Language Teaching and Translation Studies*, 1(2), 1-20.
- Muñoz, R. E. (2017).** The Effect of Washback on EFL Learners' Attitudes toward Tests. *Studies in English Language Teaching*, 5(3), 516-530.
- Nguyen, T. Q. Y., Tran, T. T. H., Nguyen, T. N. Q., Nguyen, T. P. T., Nguyen, T. C., Bui, T. S., i Nguyen, Q. H. (2022).** Online Language Testing and Assessment in the Pandemic: Opinions from Test Administrators and Examiners. *Proceedings of the AsiaCALL International Conference*, 1, 30-45.

- Oh, S. (2020).** Second language learners' use of writing resources in writing assessment. *Language Assessment Quarterly*, 17(1), 60-84.
- Oller, J. W., Jr. (1980).** Language testing research 1979-1980. U Kaplan, R. B., Jones, R. L. i Tucker, G. R. (ur.), *Annual Review of Applied Linguistics*, 1 (str. 124-150). Rowley, Massachusetts: Newbury House.
- OUZP (2016).** Uredba (EU) 2016/679 Europskog parlamenta i Vijeća od 27. travnja 2016. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka te o stavljanju izvan snage Direktive 95/46/EZ. Pristupljeno 7. travnja 2023. na: <https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:32016R0679>
- Öz, H. i Özturan, T. (2018).** Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, 14(1), 67-85.
- Paltridge, B. i Starfield, S. (ur.) (2013).** *The Handbook of English for Specific Purposes*. Chichester, UK: Wiley-Blackwell.
- Park, M. (2021).** Domain Definition Inference for a Virtual Interactive Aviation English Test (VIAET) for Military Air Traffic Controllers. U Chapelle, C. A. i Voss, E. (ur.), *Validity Argument in Language Testing – Case Studies of Validation Research* (str. 73-95). Cambridge University Press.
- Pearson (2023).** *Pearson Assessments*. Pristupljeno 8. travnja 2023. na: <https://www.pearsonassessments.com/>
- Petz, B. (2005).** *Psihologijski rječnik*. Jastrebarsko: Naklada Slap.
- Poehner, M. E., Davin, K. J. i Lantolf, J. P. (2017).** Dynamic Assessment. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 243-256). New York: Springer.
- Polio, C. (2001).** Review of the software TEST PILOT. *Language Learning & Technology*, 5(2), 34-37.
- Read, J. i Chapelle C. A. (2001).** A framework for second language vocabulary assessment. *Language Testing* 18, 1-32.
- Resnik, D. B. i Shamoo, A. E. (2011).** The Singapore Statement on Research Integrity. *Account Res*, 18(2), 71-75.
- Riazi, A. M. (2016).** Comparing writing performance in TOEFL iBT and academic assignments: An exploration of textual features. *Assessing Writing*, 28, 15-27.
- Robinson, P. (1991).** *ESP Today: A Practitioner's Guide*. Hemel Hempstead: Prentice Hall International.
- Roehr-Brackin, K. (2021).** Measuring Aptitude. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (str. 147-156). New York i London: Routledge.
- Roever, C. (2001).** Web-based language testing. *Language Learning & Technology*, 5(2), 84-94.
- Santos, V. (2017).** *A computer adaptive test of productive and contextualized academic vocabulary breadth in English (CAT PAV): Development and validation*. Neobjavljen doktorski rad, Iowa State University, Ames, IA.
- Sawaki, Y. (2001).** Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, 5(2), 38-59.
- Sawaki, Y. i Sinharay, S. (2013).** Investigating the value of section scores for the TOEFL iBT® TEST. *ETS Research Report Series*, (2), i-113.
- Shepard, L. A. (2016).** Evaluating test validity: reprise and progress. *Assessment in Education: Principles, Policy & Practice*, 23(2). 268-280.
- Shuster, E. (1997).** Fifty years later: the significance of the Nuremberg Code. *The New England Journal of Medicine*, 337(20), 1436-1440.

- Spolsky, B. (1977).** Language testing: Art or science. U Nickel, G. (ur.), *Proceedings of the Fourth International Congress of Applied Linguistics*, 3, 7-28. Stuttgart: Hochschulverlag.
- Spolsky, B. (2017).** History of Language Testing. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 375-384). New York: Springer.
- Steneck, N. H. (2006).** Fostering Integrity in Research: Definitions, Current Knowledge, and Future Directions. *Science and Engineering Ethics*, 12(1), 53-74.
- Šuman, S. (2021).** Pregled metoda obrade prirodnih jezika i strojnog prevođenja. *Zbornik Veleučilišta u Rijeci*, 9(1), 371-384.
- Tabachnick, B. G. i Fidell, L. S. (2013).** *Using multivariate statistics (6. izd.)*. Pearson Education Inc.
- Toulmin, S. E. (2003).** *The uses of argument – Updated edition*. Cambridge, England: Cambridge University Press.
- Trace, J., Janssen, G. i Meier, V. (2017).** Measuring the impact of rater negotiation in writing performance assessment. *Language Testing*, 34(1), 3-22.
- Truchot, C. (2013).** Languages for specific purposes. U Byram, M. i Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, 391-395. London: Routledge.
- Truog R. D., Brown, S. D., Browning, D., Hundert, E. M., Rider, E. A. Bell, S. K. i Meyer, E. C. (2015).** Microethics: The Ethics of Everyday Clinical Practice. *Hastings Center Report* 45(1), 11-17.
- Tulasi, L. i Murthy, N. S. R. (2022).** A Review of Linguistic and Communicative Competence; An ESP Approach. *The Review of Contemporary Scientific and Academic Studies*, 2(3). Pristupljeno 22. siječnja 2023. na: <https://doi.org/10.55454/rcsas.2.3.2022.002>
- Turnbull, D., Chugh, R. i Luck, J. (2021).** Transitioning to E-Learning during the COVID-19 pandemic: How have Higher Education Institutions responded to the challenge? *Education and Information Technologies*, 26, 6401-6419.
- Udier, S. L. i Jelaska, Z. (2008).** Službena provjera poznavanja inojezičnoga hrvatskoga jezika. *Lahor*, 6(2). 237-255.
- Upton, T. A. i Connor, U. (2013).** Language for Specific Purposes: Overview. U Chapelle, C. A. (ur.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.
- Urbina, S. (2004).** *Essentials of psychological testing*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Vantage Labs (2023).** *IntelliMetric®*. Pristupljeno 8. travnja 2023. na: <https://www.intellimetric.com/>
- Vantage Learning (2016).** *MyAccess!®*. Pristupljeno 8. travnja 2023 na: <https://www.vantagelearning.com/products/my-access-school-edition/>
- Vitello, S. i Williamson, J. (2017).** Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England. *London Review of Education*, 15(3), 536-548.
- Wall, D. (1997).** Impact and Washback in Language Testing. U Clapham, C. i Corson, D. (ur.), *Encyclopedia of Language and Education*, vol. 7 – *Language Testing and Assessment* (str. 141-149). Dordrecht, Boston i London: Kulwer Academic Publishers.
- Wang, H., Choi, I., Schmidgall, J. i Bachman, L. F. (2012).** Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing*, 29(4), 603-619.
- Weir, C. J. (2005).** *Language Testing and Validation*. New York: Palgrave Macmillan.
- Winke, P. (2011).** Evaluating the Validity of a High-Stakes ESL Test: Why Teachers' Perceptions Matter. *Tesol Quarterly*, 45(4), 628-660.

- Winke, P. i MacGregor, D. (2001).** Review of the software Hot Potatoes. *Language Learning & Technology*, 5(2), 28-33.
- Woodin, J. i Müller-Hartmann, A. (2013).** Task-based teaching and assessment. U Byram, M. i Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, 687-692. London: Routledge.
- Xi, X. (2010).** Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27(3). 291-300.
- Xi, X. i Sawaki, Y. (2017).** Methods of Test Validation. U Shohamy, E., Or, I. G. i May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (str. 193-209). New York: Springer.
- Youn, S. J. (2013).** *Validating task based assessment of L2 pragmatics in interaction using mixed methods*. Neobjavljen doktorski rad, University of Hawai'i at Manoa.
- Yulianto, D. i Mujtahin, N. M. (2021).** Online Assessment during Covid-19 Pandemic: EFL Teachers' Perspectives and Their Practices. *Journal of English teaching*, 7(2), 229-242.
- Zeidner, M. (1998).** *Test Anxiety – The State of the Art*. New York: Kluwer Academic Publishers.
- Zhang, C., Yan, X., Wang, J. (2021).** EFL Teachers' Online Assessment Practices During the COVID-19 Pandemic: Changes and Mediating Factors. *Asia-Pacific Edu Res*, 30(6), 499-507.
- Živčić-Bećirević, I. (2003).** Uloga automatskih misli i ispitne anksioznosti u uspjehu studenata. *Društvena istraživanja: časopis za opća društvena pitanja*, 5(67), 703-720.

PRILOZI

PRILOG 1. Ispit Poslovnog engleskog jezika

Ovaj ispit i upitnik sastavni su dijelovi istraživanja valjanosti e-procjenjivanja i stavova studenata o ispitima.

Svi prikupljeni podatci koristit će se isključivo za izradu doktorskog rada.

Sudjelovanje u istraživanju u potpunosti je dobrovoljno i u svakom trenutku imate pravo odustati bez navođenja razloga.

Prikupljeni podatci bit će dostupni isključivo istraživaču koji će ih unositi i obrađivati pod šifrom, pa će vaš identitet biti u potpunosti zaštićen i vaša anonimnost zajamčena.

Dajem svoj obaviješteni pristanak na sudjelovanje u ovom istraživanju stavova o e-ispitima i računalima. Pročitao/la sam i razumijem ovaj dio o pristanku (ZAOKRUŽITE).

DA – NE

Vaše prezime: _____

Vaše ime: _____

Test structure

There are two sections in this test:

- Reading Comprehension and Use of Vocabulary (Tasks 1, 2 and 3)
- Writing (Task 4).

Each section is worth 20 points.

The time allowed time for the whole test is 60 minutes.

Reading Comprehension and Use of Vocabulary

TASK 1

Read this passage and answer the questions below using the information from the text.

Advertising and viral marketing

How companies advertise

Advertising informs customers about the existence and benefits of products and services, and attempts to persuade them to buy them. Most companies use advertising agencies to produce their advertising for them. They give the agency a statement of the objectives of the advertising campaign, known as a brief, an overall advertising strategy concerning the message to be communicated to the target customers, and a budget. The agency creates advertisements (often abbreviated to adverts or ads), and develops a media plan, specifying which media – newspapers, magazines, the Internet, radio, television, cinema, posters, mail, etc. – will be used and in which proportions.

Advertising spending and sales

It is always difficult to know how much to spend on advertising. Increased ad spending can increase sales, but many companies just spend a fixed percentage of current sales revenue, or simply spend as much as their competitors (the comparative-parity method). On the other hand, lots of creative and expensive advertising campaigns, including television commercials that lots of people see and remember, and which win prizes awarded by the advertising industry for the best ads, don't lead to increased sales.

Potential drawbacks of advertising

Advertising is widely considered to be essential for launching new consumer products. Combined with sales promotion such as free samples, price reduction and competitions, advertising may generate the initial trial of a new products. But traditional advertising is expensive, it doesn't always reach the target customers, and it isn't always welcome if it does reach them. People might choose to look at posters in the street or on public transport (and virtual ones in computer and video games), or look at the ads in newspapers and magazines, but many other ads interrupt them when they're trying to do something else, like read a web page, listen to the radio, or watch a TV programme or a film.

Word-of-mouth advertising and viral marketing

This is why the best form of advertising has always been word-of-mouth advertising: people telling their friends about good products and services. For example, at the end of the last century, more and more people were saying to their friends 'Have you used Google? It's great.' Today, word-of-mouth has developed into viral marketing: companies succeed in getting people to spread commercial messages, like a virus, via peer-to-peer (P2P) networks on the Internet. The classic example is Hotmail, which added a little advertisement for itself at the bottom of every email sent using a Hotmail address. In the mid-1990s, the number of users increased from 500,000 to 12 million within a year. More and more companies are trying new strategies like setting up blogs or online forums, commenting on other people's blogs and social networking websites, making podcasts, and putting videos on YouTube, and hoping that people will use the 'Share' function to send a link to all their contacts. Viral marketing allows companies to inform and persuade, and create a 'buzz', so that an idea spreads very quickly, at very little cost.

CIRCLE THE CORRECT LETTER.

- 1. What function of advertising can be found in the text?**
 - a) To update customers on any changes to products and services.
 - b) To make sure products and services are sold to well-informed customers.
 - c) To emphasise the benefits of products so that the customers find them appealing.
 - d) To convince the customers to buy their products and services by letting them know about them.

- 2. According to the text, what role of advertising agencies is mentioned?**
 - a) To design ads and develop a strategy for the media.
 - b) To brief the company on the advertising strategy they will use.
 - c) To give a statement of the objectives of an advertising campaign.
 - d) To communicate the advertising campaign to the target customers.

3. Which method of determining advertising spending is NOT given in the text?
- a) Spending an amount equal to their competitors'.
 - b) Spending a set percentage of existing sales revenue.
 - c) Increasing current spending in order to increase sales.
 - d) Using inexpensive but award-winning advertising campaigns.
4. According to the text, which of the following advertising methods is the most effective?
- a) Viral marketing.
 - b) Sales promotion.
 - c) Newspaper and magazine advertisements.
 - d) Posters in the street or on public transport.
5. Which option is mentioned in the text as a way of using the Internet to advertise?
- a) Likes.
 - b) Links.
 - c) Blogs.
 - d) Viruses.

TASK 2

Complete each sentence by selecting the most appropriate word to fill in the gap. Only one answer is correct for each gap. **CIRCLE THE CORRECT LETTER.**

1. When there is a _____, a company has to recruit a new employee to fill it, and most companies have a defined selection process when recruiting new people.
- a) post
 - b) vacancy
 - c) position
 - d) advertisement
2. If an individual employee has a complaint, a company usually has a _____ procedure to deal with it.
- a) lawsuit
 - b) staffing
 - c) grievance
 - d) arbitration
3. Most manufacturers divide their products into product _____ – groups of closely related products, sold to the same customer groups, and marketed through the same outlets.
- a) lines
 - b) items
 - c) mixes
 - d) ranges
4. At the bottom of the Maslow's 'hierarchy of needs' there are _____ needs.
- a) safety
 - b) physiological
 - c) psychological
 - d) self-actualisation

5. Market _____ is dividing a market into distinct groups of buyers who have different requirements or buying habits.
 - a) targeting
 - b) grouping
 - c) separation
 - d) segmentation

6. _____ is an element of the marketing mix that is related only to services, not products.
 - a) Price
 - b) Place
 - c) Process
 - d) Promotion

7. The ‘_____ concept’ assumes that the producer’s task is to find wants and fill them, which means you don’t sell what you make – you make what will be bought.
 - a) selling
 - b) skimming
 - c) marketing
 - d) advertising

TASK 3

Complete the text with EIGHT DIFFERENT expressions of the twelve provided - you must use each expression only ONCE in the text. WRITE THE EXPRESSIONS ON THE LINES.

start up
start-up
venture
creators

invented
founders
launched
innovative

enterprises
competitive
entrepreneurs
entrepreneurial

This is a story about student entrepreneurs, Ana and Josipa, the **1)**_____ of a medical **2)**_____ called IDerma. Not only were they business students from our Faculty, but they were also **3)**_____ entrepreneurs who in 2012 **4)**_____ special software that solves the problems in communication between doctors and patients. They can create a personalised white label medical application in only 48 hours for a **5)**_____ price of \$99. These **6)**_____ young women would like to encourage students to build their own projects and **7)**_____ a business. The founders claim that a start-up is worth the risk because even if a new business fails, young **8)**_____ will gain some valuable experience and benefit from new business contacts.

Writing

TASK 4

Write a letter of complaint following the instructions mentioned further below. Your letter will be evaluated according to the following criteria:

Use the block style of letter writing (9 points):

- paragraphs
- addresses
- date
- greeting
- subject
- closing remark
- signing off
- signature

Use all the necessary functions for a letter of complaint (6 points):

- stating the reason for writing
- explaining the reasons
- asking for immediate action

Pay attention to the quality of your English (5 points):

- vocabulary range and appropriateness
- spelling and grammar accuracy

You work for Asda (Great Wilson St / Leeds LS11 5AD / UK), one of Britain's leading retailers. You ordered a shipment of organic vegetables from the British Growers Association (Nottingham Road / Louth / LN11 0WB / UK) ten days ago, but it has not yet arrived, although it was supposed to be at your warehouse five days ago. Demand the shipment within two days. Make a strong complaint.

Write your letter here:

Prilog 1.1. Rezultati izvedbe sudionika na jezičnom ispitu

SUDIONIK			E-ISPIT											ISPIT NA PAPIRU												
Red. br.	Spol	Nastavnik	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Ukupno (od 20)	Pisanje								Ukupno (od 40)	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Ukupno (od 20)	Pisanje								Ukupno (od 40)
						Procjenjivač A				Procjenjivač B								Procjenjivač A				Procjenjivač B				
						oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno					oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno	
1.	1	1	3,00	13,00	14,00	4,00	5,00	4,50	13,50	4,50	6,00	4,00	14,50	30,00	3,00	12,00	7,50	2,00	3,00	2,00	7,00	1,50	3,00	3,50	8,00	22,50
2.	1	1	4,00	12,00	13,00	5,50	4,00	4,50	14,00	5,00	3,00	4,00	12,00	29,00	4,00	13,00	13,00	4,50	3,00	4,50	12,00	5,00	4,00	5,00	14,00	30,00
3.	1	1	2,00	12,00	16,50	7,00	5,00	4,00	16,00	7,00	6,00	4,00	17,00	30,50	5,00	13,00	12,50	4,50	3,50	4,50	12,50	4,00	4,00	4,50	12,50	30,50
4.	1	1	4,00	8,00	5,75	0,00	2,00	3,50	5,50	0,00	3,00	3,00	6,00	17,75	2,00	5,00	7,75	0,00	4,00	3,50	7,50	1,00	4,50	2,50	8,00	14,75
5.	1	2	3,00	5,00	13,50	6,50	4,00	3,00	13,50	6,00	5,50	2,00	13,50	21,50	5,00	13,00	13,00	4,50	5,00	3,00	12,50	5,00	5,50	3,00	13,50	31,00
6.	2	2	4,00	13,00	15,25	8,50	4,00	4,00	16,50	8,00	3,00	3,00	14,00	32,25	4,00	11,00	10,00	4,50	3,00	2,50	10,00	4,50	3,00	2,50	10,00	25,00
7.	2	2	3,00	13,00	13,50	7,00	4,00	3,50	14,50	6,00	3,00	3,50	12,50	29,50	4,00	14,00	12,75	5,00	4,00	3,50	12,50	5,50	4,00	3,50	13,00	30,75
8.	1	1	4,00	10,00	12,75	5,00	4,00	4,00	13,00	5,00	3,50	4,00	12,50	26,75	3,00	12,00	8,75	3,00	2,00	3,50	8,50	2,50	2,50	4,00	9,00	23,75
9.*	2	4	3,00	8,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	11,00	3,00	10,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	13,00
10.	2	1	3,00	9,00	13,25	5,00	4,00	4,50	13,50	6,00	3,00	4,00	13,00	25,25	2,00	15,00	15,00	7,50	4,00	3,50	15,00	7,50	4,00	3,50	15,00	32,00
11.	2	3	4,00	10,00	8,75	1,00	3,00	4,00	8,00	1,00	4,00	4,50	9,50	22,75	4,00	12,00	9,75	3,50	2,00	3,50	9,00	4,50	2,00	4,00	10,50	25,75
12.	2	3	3,00	7,00	12,50	7,00	4,00	3,00	14,00	7,00	2,50	1,50	11,00	22,50	4,00	12,00	16,50	6,50	6,00	4,00	16,50	6,50	6,00	4,00	16,50	32,50
13.	1	3	3,00	15,00	6,75	2,00	2,00	3,00	7,00	2,00	2,50	2,00	6,50	24,75	4,00	14,00	12,50	5,50	4,00	4,50	14,00	4,50	3,50	3,00	11,00	30,50
14.	1	3	3,00	14,00	15,75	6,50	4,00	5,00	15,50	6,50	5,00	4,50	16,00	32,75	3,00	14,00	6,25	3,00	1,00	2,50	6,50	2,50	1,50	2,00	6,00	23,25
15.	2	3	4,00	14,00	13,50	8,00	2,00	3,50	13,50	7,00	3,00	3,50	13,50	31,50	5,00	15,00	14,00	6,50	3,50	3,50	13,50	6,00	4,00	4,50	14,50	34,00
16.	2	3	5,00	14,00	11,25	5,00	3,00	2,50	10,50	6,00	4,50	1,50	12,00	30,25	5,00	15,00	10,25	3,50	4,00	3,00	10,50	2,00	5,00	3,00	10,00	30,25
17.	2	4	4,00	13,00	13,50	6,00	3,00	3,50	12,50	6,50	4,00	4,00	14,50	30,50	4,00	14,00	12,75	5,00	4,00	3,50	12,50	5,00	4,50	3,50	13,00	30,75
18.	2	5	5,00	12,00	13,00	5,00	4,00	3,50	12,50	5,50	4,50	3,50	13,50	30,00	5,00	15,00	12,25	4,50	5,00	3,00	12,50	4,00	4,50	3,50	12,00	32,25
19.	2	3	4,00	10,00	9,75	1,00	5,00	3,00	9,00	0,50	6,00	4,00	10,50	23,75	3,00	13,00	13,25	7,00	3,00	3,50	13,50	6,50	3,50	3,00	13,00	29,25
20.	1	3	3,00	13,00	12,00	7,50	2,00	2,50	12,00	6,50	3,50	2,00	12,00	28,00	4,00	13,00	9,75	4,50	3,00	2,00	9,50	4,00	2,50	3,50	10,00	26,75
21.	2	3	3,00	12,00	11,25	4,50	2,50	3,00	10,00	5,00	4,00	3,50	12,50	26,25	3,00	12,00	6,50	1,00	2,00	4,00	7,00	1,00	2,50	2,50	6,00	21,50
22.	1	3	5,00	14,00	13,50	6,00	4,00	3,50	13,50	5,00	5,00	3,50	13,50	32,50	4,00	12,00	7,25	2,50	2,50	2,50	7,50	2,50	3,50	1,00	7,00	23,25
23.	2	3	4,00	12,00	7,50	1,50	4,00	2,00	7,50	1,00	4,00	2,50	7,50	23,50	3,00	12,00	9,50	3,50	4,00	3,00	10,50	2,50	3,50	2,50	8,50	24,50
24.	2	6	4,00	15,00	14,25	6,50	3,00	3,00	12,50	7,50	4,00	4,50	16,00	33,25	5,00	15,00	13,25	4,50	5,00	4,00	13,50	3,50	5,50	4,00	13,00	33,25
25.	2	6	4,00	5,00	14,25	8,50	4,00	2,50	15,00	7,50	4,00	2,00	13,50	23,25	3,00	10,00	13,50	7,00	4,00	3,50	14,50	6,00	3,50	3,00	12,50	26,50
26.	2	1	4,00	13,00	12,25	6,00	3,00	3,50	12,50	5,50	3,00	3,50	12,00	29,25	2,00	15,00	12,00	5,50	4,00	3,00	12,50	4,50	3,00	4,00	11,50	29,00
27.	2	6	5,00	13,00	16,00	6,50	5,00	4,50	16,00	6,50	5,00	4,50	16,00	34,00	4,00	15,00	14,75	6,50	4,00	3,50	14,00	6,50	4,50	4,50	15,50	33,75
28.	1	1	3,00	9,00	11,25	5,50	5,00	2,50	13,00	4,50	3,50	1,50	9,50	23,25	4,00	9,00	9,25	5,50	2,00	2,50	10,00	4,50	2,00	2,00	8,50	22,25
29.	2	1	3,00	10,00	15,00	8,00	4,00	3,50	15,50	7,50	4,00	3,00	14,50	28,00	3,00	14,00	14,25	8,00	3,50	3,50	15,00	7,00	4,00	2,50	13,50	31,25
30.	1	6	5,00	9,00	6,50	1,50	2,50	2,50	6,50	1,50	2,50	2,50	6,50	20,50	5,00	13,00	9,25	4,00	4,00	2,00	10,00	3,50	3,50	1,50	8,50	27,25

SUDIONIK			E-ISPLIT												ISPIT NA PAPIRU													
Red. br.	Spol	Nastavnik	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Pisanje										Ukupno (od 40)	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Pisanje										Ukupno (od 40)
					Ukupno (od 20)	Procjenjivač A				Procjenjivač B				Ukupno (od 20)				Procjenjivač A				Procjenjivač B						
						oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno					oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno			
31.	2	1	4,00	13,00	13,50	6,50	4,00	2,50	13,00	7,00	4,00	3,00	14,00	30,50	5,00	15,00	14,50	7,50	3,50	3,50	14,50	7,50	3,50	3,50	14,50	34,50		
32.	2	1	3,00	7,00	9,50	5,50	2,00	2,00	9,50	5,00	2,50	2,00	9,50	19,50	4,00	11,00	11,25	5,00	3,00	2,50	10,50	4,50	4,00	3,50	12,00	26,25		
33.	2	6	1,00	3,00	13,75	7,50	4,00	3,50	15,00	7,00	3,50	2,00	12,50	17,75	2,00	12,00	7,25	4,00	1,50	2,00	7,50	3,50	1,50	2,00	7,00	21,25		
34.	2	3	2,00	7,00	8,25	1,50	4,00	2,50	8,00	1,50	4,00	3,00	8,50	17,25	4,00	11,00	13,00	6,00	3,00	2,50	11,50	6,00	4,50	4,00	14,50	28,00		
35.	1	3	4,00	15,00	10,75	1,00	5,00	4,00	10,00	1,00	6,00	4,50	11,50	29,75	5,00	12,00	10,25	3,00	5,00	3,00	11,00	2,50	4,50	2,50	9,50	27,25		
36.	2	3	5,00	5,00	14,50	7,50	4,00	3,00	14,50	7,00	3,50	4,00	14,50	24,50	3,00	12,00	9,00	2,50	3,50	2,50	8,50	3,00	3,50	3,00	9,50	24,00		
37.	1	3	3,00	8,00	6,75	0,00	3,00	3,00	6,00	1,00	3,50	3,00	7,50	17,75	3,00	8,00	5,50	1,00	2,00	2,00	5,00	1,00	2,50	2,50	6,00	16,50		
38.	1	4	4,00	13,00	12,50	5,50	4,00	2,50	12,00	5,50	4,00	3,50	13,00	29,50	2,00	13,00	11,75	3,50	4,00	3,00	10,50	5,00	4,50	3,50	13,00	26,75		
39.*	2	4	5,00	15,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	20,00	4,00	11,00	6,50	0,50	3,00	3,50	7,00	0,50	2,50	3,00	6,00	21,50		
40.	2	4	4,00	12,00	8,00	2,00	4,00	2,50	8,50	2,00	3,00	2,50	7,50	24,00	2,00	17,00	10,50	5,00	2,50	3,00	10,50	5,50	1,50	3,50	10,50	29,50		
41.	2	4	5,00	15,00	12,25	5,50	4,00	2,50	12,00	5,50	3,00	4,00	12,50	32,25	5,00	13,00	10,50	2,00	4,00	4,50	10,50	2,50	4,50	3,50	10,50	28,50		
42.	2	4	5,00	14,00	12,25	4,00	4,00	3,00	11,00	5,50	5,00	3,00	13,50	31,25	4,00	13,00	10,00	4,50	2,50	3,00	10,00	4,50	3,50	2,00	10,00	27,00		
43.	2	1	3,00	12,00	8,50	2,50	4,00	1,50	8,00	3,00	3,50	2,50	9,00	23,50	2,00	9,00	12,00	5,00	3,50	3,00	11,50	4,50	4,50	3,50	12,50	23,00		
44.	2	4	4,00	13,00	13,75	7,00	3,00	2,50	12,50	8,00	3,50	3,50	15,00	30,75	4,00	12,00	10,50	3,00	3,50	3,50	10,00	3,50	4,00	3,50	11,00	26,50		
45.	1	4	4,00	13,00	10,25	2,50	4,00	4,00	10,50	2,00	4,00	4,00	10,00	27,25	3,00	12,00	12,50	4,00	4,00	4,50	12,50	3,00	5,50	4,00	12,50	27,50		
46.	1	4	5,00	14,00	12,75	4,00	5,00	3,50	12,50	3,50	6,00	3,50	13,00	31,75	5,00	13,00	12,50	3,50	5,00	3,50	12,00	3,50	5,00	4,50	13,00	30,50		
47.	2	4	2,00	13,00	10,50	5,50	4,00	2,00	11,50	6,00	3,00	0,50	9,50	25,50	2,00	12,00	7,50	1,50	2,00	3,50	7,00	1,50	3,00	3,50	8,00	21,50		
48.	1	4	5,00	14,00	13,75	5,50	4,00	3,50	13,00	5,50	4,50	4,50	14,50	32,75	4,00	14,00	9,00	1,50	3,50	3,00	8,00	2,00	5,00	3,00	10,00	27,00		
49.	2	4	1,00	8,00	6,75	3,00	3,00	1,50	7,50	2,50	2,00	1,50	6,00	15,75	3,00	12,00	12,75	6,00	4,00	3,50	13,50	6,00	3,00	3,00	12,00	27,75		
50.	2	4	3,00	13,00	9,50	5,00	3,00	1,50	9,50	5,00	3,00	1,50	9,50	25,50	2,00	13,00	11,50	3,50	5,00	3,00	11,50	4,00	5,00	2,50	11,50	26,50		
51.	2	4	5,00	14,00	9,75	1,50	5,00	3,00	9,50	1,50	6,00	2,50	10,00	28,75	3,00	14,00	12,75	4,00	5,00	3,50	12,50	3,00	6,00	4,00	13,00	29,75		
52.	2	4	4,00	14,00	8,75	3,50	2,00	2,50	8,00	4,00	3,50	2,00	9,50	26,75	5,00	13,00	9,00	3,50	3,00	3,00	9,50	3,50	3,00	2,00	8,50	27,00		
53.	1	4	5,00	12,00	15,75	8,00	4,00	3,00	15,00	8,00	5,50	3,00	16,50	32,75	4,00	15,00	8,50	3,00	3,00	3,00	9,00	3,00	3,00	2,00	8,00	27,50		
54.	1	7	1,00	9,00	9,50	2,50	6,00	2,00	10,50	2,50	5,50	0,50	8,50	19,50	4,00	11,00	8,25	3,00	4,00	2,00	9,00	3,50	3,00	1,00	7,50	23,25		
55.	1	7	5,00	10,00	9,25	1,00	5,00	2,50	8,50	1,50	5,00	3,50	10,00	24,25	4,00	10,00	10,25	2,00	4,00	4,00	10,00	2,50	4,50	3,50	10,50	24,25		
56.	1	1	5,00	11,00	6,50	1,00	3,00	2,50	6,50	1,50	3,00	2,00	6,50	22,50	4,00	12,00	12,50	5,00	4,00	2,50	11,50	6,00	3,50	4,00	13,50	28,50		
57.	2	4	4,00	13,00	12,00	4,50	4,00	2,00	10,50	5,50	5,00	3,00	13,50	29,00	4,00	14,00	10,00	2,00	5,00	3,00	10,00	2,00	4,50	3,50	10,00	28,00		
58.	1	1	4,00	11,00	7,50	1,50	4,00	1,50	7,00	1,00	4,50	2,50	8,00	22,50	4,00	13,00	11,25	2,50	5,00	3,00	10,50	3,00	5,50	3,50	12,00	28,25		
59.	2	1	2,00	12,00	5,50	0,50	2,00	3,00	5,50	0,50	2,50	2,50	5,50	19,50	3,00	11,00	6,50	1,50	2,00	3,50	7,00	1,50	2,00	2,50	6,00	20,50		
60.	2	1	4,00	14,00	5,00	0,50	2,00	2,00	4,50	0,50	2,50	2,50	5,50	23,00	3,00	14,00	8,50	4,00	2,00	3,00	9,00	3,50	2,00	2,50	8,00	25,50		
61.	1	1	4,00	12,00	11,50	5,50	2,50	3,50	11,50	5,00	2,50	4,00	11,50	27,50	5,00	13,00	13,25	5,00	4,00	4,00	13,00	5,00	4,00	4,50	13,50	31,25		
62.	2	8	4,00	14,00	15,00	6,00	5,00	4,00	15,00	6,50	5,00	3,50	15,00	33,00	3,00	13,00	14,00	4,50	6,00	4,50	15,00	3,50	6,00	3,50	13,00	30,00		
63.	1	8	4,00	13,00	12,00	5,50	3,00	2,50	11,00	5,50	4,00	3,50	13,00	29,00	4,00	12,00	5,25	1,00	2,00	3,00	6,00	1,00	2,00	1,50	4,50	21,25		
64.	2	8	4,00	11,00	9,75	3,50	4,00	2,50	10,00	3,00	5,00	1,50	9,50	24,75	4,00	10,00	9,50	3,00	4,00	3,00	10,00	3,00	4,00	2,00	9,00	23,50		

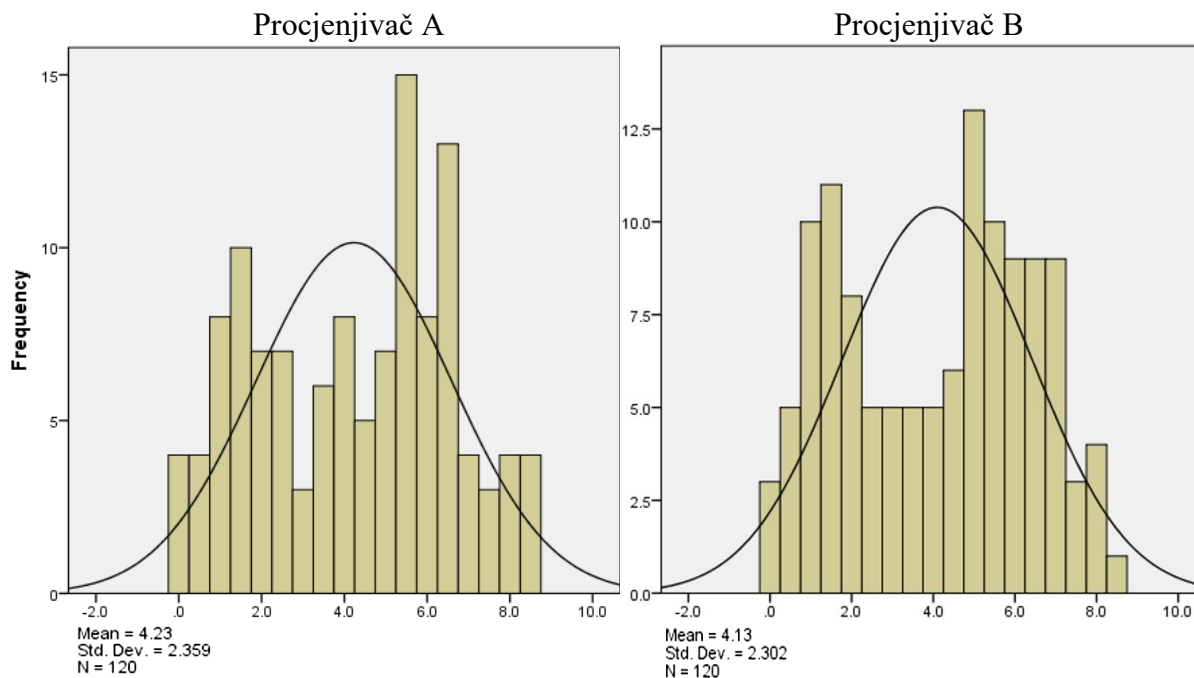
SUDIONIK			E-ISPIT												ISPIT NA PAPIRU											
Red. br.	Spol	Nastavnik	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Ukupno (od 20)	Pisanje								Ukupno (od 40)	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Ukupno (od 20)	Pisanje								Ukupno (od 40)
						Procjenjivač A				Procjenjivač B								Procjenjivač A				Procjenjivač B				
						oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno					oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno	
65.	2	8	5,00	14,00	13,00	5,50	5,00	3,50	14,00	5,00	3,50	3,50	12,00	32,00	5,00	15,00	9,25	0,50	5,00	3,50	9,00	0,50	5,00	4,00	9,50	29,25
66.	2	8	4,00	13,00	12,25	4,00	4,00	4,00	12,00	3,00	5,50	4,00	12,50	29,25	4,00	14,00	10,50	5,00	3,00	3,00	11,00	4,00	3,00	3,00	10,00	28,50
67.	2	9	5,00	4,00	1,50	1,00	0,00	0,00	1,00	2,00	0,00	0,00	2,00	10,50	3,00	7,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	10,00
68.	2	8	5,00	12,00	11,00	3,00	5,50	2,00	10,50	2,50	6,00	3,00	11,50	28,00	4,00	13,00	7,25	1,50	4,00	3,00	8,50	1,50	2,50	2,00	6,00	24,25
69.	1	8	5,00	15,00	13,50	5,50	5,00	3,00	13,50	5,00	6,00	2,50	13,50	33,50	5,00	14,00	11,75	3,50	5,00	3,00	11,50	3,50	6,00	2,50	12,00	30,75
70.	2	1	5,00	15,00	3,75	0,00	3,00	1,50	4,50	0,50	2,00	0,50	3,00	23,75	4,00	10,00	5,50	0,50	2,00	3,50	6,00	0,50	2,50	2,00	5,00	19,50
71.	2	1	4,00	6,00	8,00	0,50	5,00	2,50	8,00	1,50	4,00	2,50	8,00	18,00	4,00	12,00	12,25	4,50	3,00	3,50	11,00	5,00	4,50	4,00	13,50	28,25
72.	2	8	3,00	12,00	9,25	4,50	3,00	2,50	10,00	4,00	2,00	2,50	8,50	24,25	3,00	13,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	16,00
73.	2	8	4,00	13,00	7,50	2,00	3,00	3,00	8,00	1,50	3,00	2,50	7,00	24,50	3,00	11,00	9,50	2,00	5,00	4,00	11,00	1,00	3,50	3,50	8,00	23,50
74.	2	1	5,00	10,00	16,00	6,50	5,00	5,00	16,50	6,50	4,50	4,50	15,50	31,00	4,00	13,00	15,25	4,50	6,00	5,00	15,50	4,50	6,00	4,50	15,00	32,25
75.	2	8	2,00	15,00	10,25	3,50	3,00	2,50	9,00	5,00	4,00	2,50	11,50	27,25	4,00	13,00	9,25	3,00	4,00	3,00	10,00	3,00	3,00	2,50	8,50	26,25
76.	2	8	2,00	12,00	8,25	2,50	3,00	3,00	8,50	1,50	4,00	2,50	8,00	22,25	3,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00
77.	2	8	5,00	15,00	10,50	4,00	5,00	2,00	11,00	3,00	4,50	2,50	10,00	30,50	5,00	15,00	10,25	3,50	4,00	2,50	10,00	3,50	3,50	3,50	10,50	30,25
78.	2	8	5,00	14,00	9,25	2,50	4,00	2,00	8,50	2,50	5,00	2,50	10,00	28,25	4,00	13,00	12,00	4,50	3,00	4,50	12,00	5,00	3,00	4,00	12,00	29,00
79.	2	8	3,00	13,00	13,75	6,50	3,00	4,00	13,50	7,00	3,00	4,00	14,00	29,75	4,00	15,00	15,00	6,50	5,00	4,00	15,50	5,50	4,50	4,50	14,50	34,00
80.	1	8	4,00	15,00	13,25	5,50	4,00	4,50	14,00	6,00	3,00	3,50	12,50	32,25	4,00	14,00	10,50	4,50	3,00	3,50	11,00	5,50	2,50	2,00	10,00	28,50
81.	2	8	4,00	13,00	12,75	5,00	4,00	4,00	13,00	4,50	5,50	2,50	12,50	29,75	3,00	10,00	8,00	3,00	2,00	3,50	8,50	3,00	1,50	3,00	7,50	21,00
82.	2	8	3,00	10,00	3,75	2,00	2,00	0,00	4,00	1,50	2,00	0,00	3,50	16,75	3,00	12,00	9,25	4,00	2,00	3,50	9,50	4,50	2,50	2,00	9,00	24,25
83.	1	8	2,00	11,00	6,25	1,00	2,50	3,00	6,50	1,00	2,00	3,00	6,00	19,25	3,00	14,00	5,25	0,50	2,00	3,00	5,50	0,50	2,00	2,50	5,00	22,25
84.	2	8	3,00	14,00	9,00	4,00	2,50	3,00	9,50	3,50	3,00	2,00	8,50	26,00	5,00	15,00	7,00	1,00	4,00	3,00	8,00	1,00	3,00	2,00	6,00	27,00
85.	2	8	2,00	13,00	9,50	3,00	3,50	3,00	9,50	3,50	4,00	2,00	9,50	24,50	4,00	13,00	9,75	2,50	4,00	3,50	10,00	2,50	4,00	3,00	9,50	26,75
86.	1	1	2,00	10,00	13,25	5,50	4,00	4,50	14,00	4,50	5,00	3,00	12,50	25,25	2,00	10,00	11,25	4,00	4,00	4,00	12,00	4,50	3,50	2,50	10,50	23,25
87.	2	10	4,00	14,00	5,50	0,50	4,00	1,50	6,00	0,50	3,50	1,00	5,00	23,50	4,00	10,00	5,75	0,50	2,50	3,50	6,50	0,50	2,00	2,50	5,00	19,75
88.	2	10	3,00	11,00	6,50	2,00	2,00	1,50	5,50	2,00	3,00	2,50	7,50	20,50	2,00	9,00	7,25	3,00	1,50	3,00	7,50	3,00	2,00	2,00	7,00	18,25
89.	2	9	2,00	9,00	12,50	6,50	4,00	3,50	14,00	5,00	4,00	2,00	11,00	23,50	2,00	12,00	12,75	5,00	4,00	3,50	12,50	5,00	4,00	4,00	13,00	26,75
90.	2	1	3,00	15,00	8,00	0,00	5,00	3,00	8,00	0,00	5,00	3,00	8,00	26,00	4,00	11,00	6,25	0,00	2,50	3,50	6,00	0,00	3,00	3,50	6,50	21,25
91.	1	1	5,00	8,00	8,75	1,50	4,50	2,00	8,00	1,50	6,00	2,00	9,50	21,75	2,00	15,00	7,75	2,00	3,00	3,00	8,00	2,50	2,00	3,00	7,50	24,75
92.	2	10	3,00	11,00	9,25	2,50	4,00	3,00	9,50	2,00	4,00	3,00	9,00	23,25	4,00	11,00	7,50	0,50	5,00	3,00	8,50	0,50	3,50	2,50	6,50	22,50
93.	1	10	3,00	12,00	9,50	2,50	4,00	2,50	9,00	2,50	5,00	2,50	10,00	24,50	5,00	15,00	10,00	2,00	4,00	3,50	9,50	2,50	4,50	3,50	10,50	30,00
94.	2	1	4,00	14,00	15,50	6,50	4,00	4,00	14,50	6,50	5,50	4,50	16,50	33,50	5,00	15,00	11,75	5,00	3,00	3,50	11,50	4,50	4,00	3,50	12,00	31,75
95.	2	10	4,00	13,00	10,75	4,00	4,00	1,50	9,50	4,00	5,00	3,00	12,00	27,75	3,00	13,00	9,75	2,00	5,00	3,50	10,50	3,00	4,00	2,00	9,00	25,75
96.	2	1	4,00	10,00	7,50	2,00	2,50	3,00	7,50	2,00	3,00	2,50	7,50	21,50	2,00	11,00	12,00	5,00	4,00	3,50	12,50	3,50	4,00	4,00	11,50	25,00
97.	1	4	2,00	12,00	7,00	1,50	5,50	1,00	8,00	1,00	4,00	1,00	6,00	21,00	4,00	9,00	6,75	2,50	4,00	1,50	8,00	2,00	3,00	0,50	5,50	19,75
98.	1	9	3,00	8,00	1,25	1,50	0,00	0,00	1,50	1,00	0,00	0,00	1,00	12,25	3,00	8,00	2,75	1,00	2,00	1,00	4,00	0,50	0,50	0,50	1,50	13,75

SUDIONIK			E-ISPIT												ISPIT NA PAPIRU																
Red. br.	Spol	Nastavnik	Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Pisanje												Raz. čit. (od 5,00)	Raz. i primj. vok. (od 15,00)	Pisanje												Ukupno (od 40)
					Ukupno (od 20)	Procjenjivač A				Procjenjivač B				Ukupno (od 20)	Procjenjivač A				Procjenjivač B												
						oblik	sadržaj	jezik	ukupno	oblik	sadržaj	jezik	ukupno		oblik	sadržaj			jezik	ukupno	oblik	sadržaj	jezik	ukupno							
99.	2	2	3,00	8,00	17,00	6,50	5,00	4,50	16,00	7,00	6,00	5,00	18,00	28,00	5,00	8,00	9,75	3,50	3,00	2,50	9,00	3,50	4,50	2,50	10,50	22,75					
100.	2	2	5,00	11,00	16,75	8,50	4,00	3,50	16,00	8,50	4,50	4,50	17,50	32,75	5,00	12,00	13,00	4,50	4,00	4,00	12,50	4,50	4,50	4,50	13,50	30,00					
101.	2	1	3,00	11,00	11,50	5,50	4,00	2,00	11,50	5,50	3,00	3,00	11,50	25,50	3,00	9,00	9,75	3,50	3,00	3,00	9,50	3,00	4,50	2,50	10,00	21,75					
102.	2	2	4,00	11,00	13,25	6,50	5,00	1,50	13,00	6,00	5,50	2,00	13,50	28,25	5,00	11,00	15,50	7,00	4,50	4,00	15,50	7,00	5,50	3,00	15,50	31,50					
103.	2	1	3,00	11,00	12,00	6,00	3,00	3,50	12,50	5,00	3,50	3,00	11,50	26,00	3,00	13,00	10,00	4,50	3,00	3,00	10,50	4,00	3,50	2,00	9,50	26,00					
104.	2	2	2,00	9,00	7,00	3,50	2,50	2,00	8,00	4,50	1,00	0,50	6,00	18,00	3,00	9,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	12,00					
105.	1	2	1,00	10,00	10,75	3,50	3,50	3,50	10,50	3,50	4,00	3,50	11,00	21,75	2,00	9,00	8,25	2,00	3,00	3,50	8,50	1,50	4,00	2,50	8,00	19,25					
106.	2	2	5,00	9,00	17,00	8,00	6,00	3,00	17,00	7,00	6,00	4,00	17,00	31,00	2,00	11,00	15,25	7,50	5,00	4,00	16,50	6,00	4,50	3,50	14,00	28,25					
107.	2	2	3,00	11,00	12,00	6,00	3,50	3,50	13,00	5,50	2,50	3,00	11,00	26,00	4,00	14,00	7,75	3,00	3,00	2,50	8,50	3,00	2,00	2,00	7,00	25,75					
108.	2	1	3,00	11,00	11,50	6,00	3,00	3,00	12,00	5,00	3,50	2,50	11,00	25,50	2,00	12,00	9,75	3,50	4,00	3,00	10,50	3,00	3,00	3,00	9,00	23,75					
109.	1	2	2,00	9,00	10,00	3,50	4,00	2,50	10,00	3,00	4,00	3,00	10,00	21,00	4,00	10,00	8,75	3,50	2,50	3,00	9,00	4,00	2,00	2,50	8,50	22,75					
110.	2	1	5,00	13,00	8,75	1,50	5,00	2,00	8,50	1,00	5,00	3,00	9,00	26,75	2,00	8,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	10,00					
111.	1	4	3,00	11,00	12,25	5,00	4,50	3,50	13,00	4,50	4,00	3,00	11,50	26,25	2,00	13,00	9,25	4,50	3,00	2,50	10,00	3,00	3,00	2,50	8,50	24,25					
112.	1	2	4,00	12,00	10,25	2,00	5,00	4,00	11,00	1,00	5,00	3,50	9,50	26,25	4,00	13,00	7,25	1,50	2,50	3,50	7,50	1,50	2,50	3,00	7,00	24,25					
113.	2	10	3,00	10,00	6,25	1,50	3,00	1,50	6,00	2,00	2,50	2,00	6,50	19,25	1,00	11,00	12,75	3,50	6,00	4,00	13,50	4,00	5,50	2,50	12,00	24,75					
114.	2	10	5,00	10,00	10,75	4,00	4,00	2,50	10,50	4,00	4,50	2,50	11,00	25,75	5,00	12,00	11,00	3,50	5,50	3,50	12,50	2,50	5,00	2,00	9,50	28,00					
115.	1	1	2,00	3,00	6,50	1,00	3,00	2,50	6,50	1,50	3,00	2,00	6,50	11,50	2,00	10,00	7,25	3,00	2,00	3,00	8,00	2,50	2,50	1,50	6,50	19,25					
116.	2	10	4,00	10,00	8,75	4,50	2,00	2,00	8,50	4,00	3,00	2,00	9,00	22,75	1,00	11,00	12,00	5,50	4,50	2,00	12,00	5,50	4,50	2,00	12,00	24,00					
117.	2	1	5,00	13,00	13,50	6,00	4,00	3,50	13,50	6,00	4,50	3,00	13,50	31,50	2,00	10,00	9,75	2,50	4,50	3,00	10,00	2,50	5,00	2,00	9,50	21,75					
118.	2	1	3,00	14,00	13,25	4,50	5,00	4,50	14,00	3,50	5,00	4,00	12,50	30,25	2,00	9,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	11,00					
119.	1	1	5,00	11,00	15,25	8,50	3,00	3,00	14,50	8,00	4,00	4,00	16,00	31,25	5,00	15,00	12,25	6,50	3,00	2,50	12,00	6,00	3,00	3,50	12,50	32,25					
120.	1	4	3,00	15,00	14,00	6,50	5,00	3,00	14,50	6,50	4,00	3,00	13,50	32,00	4,00	13,00	13,00	5,00	4,50	4,00	13,50	4,00	5,50	3,00	12,50	30,00					
121.	2	1	4,00	14,00	11,75	6,50	3,00	2,50	12,00	6,50	3,00	2,00	11,50	29,75	5,00	13,00	13,50	7,00	4,50	3,00	14,50	5,50	5,00	2,00	12,50	31,50					
122.	2	1	3,00	8,00	11,50	5,50	3,00	2,50	11,00	6,00	3,00	3,00	12,00	22,50	3,00	8,00	11,50	4,50	3,00	4,00	11,50	5,00	3,00	3,50	11,50	22,50					

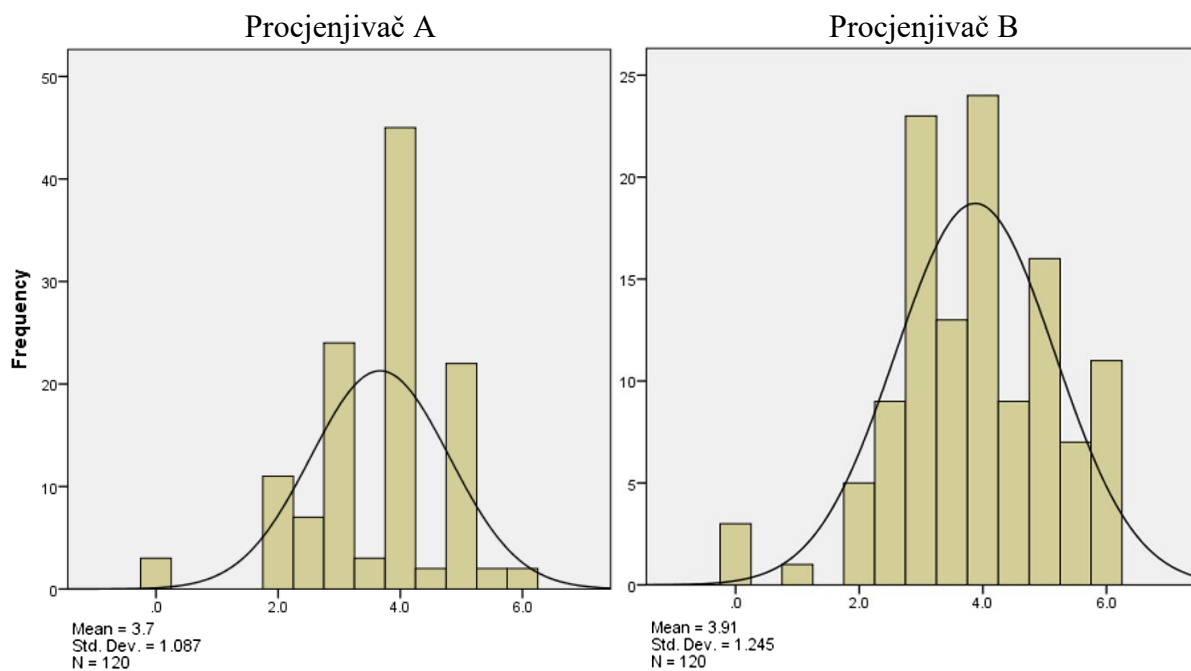
* Podatci isključeni iz daljnje obrade jer ovi sudionici na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Prilog 1.2. Histogrami procjenjivača A i B na korespondentnim mjerama

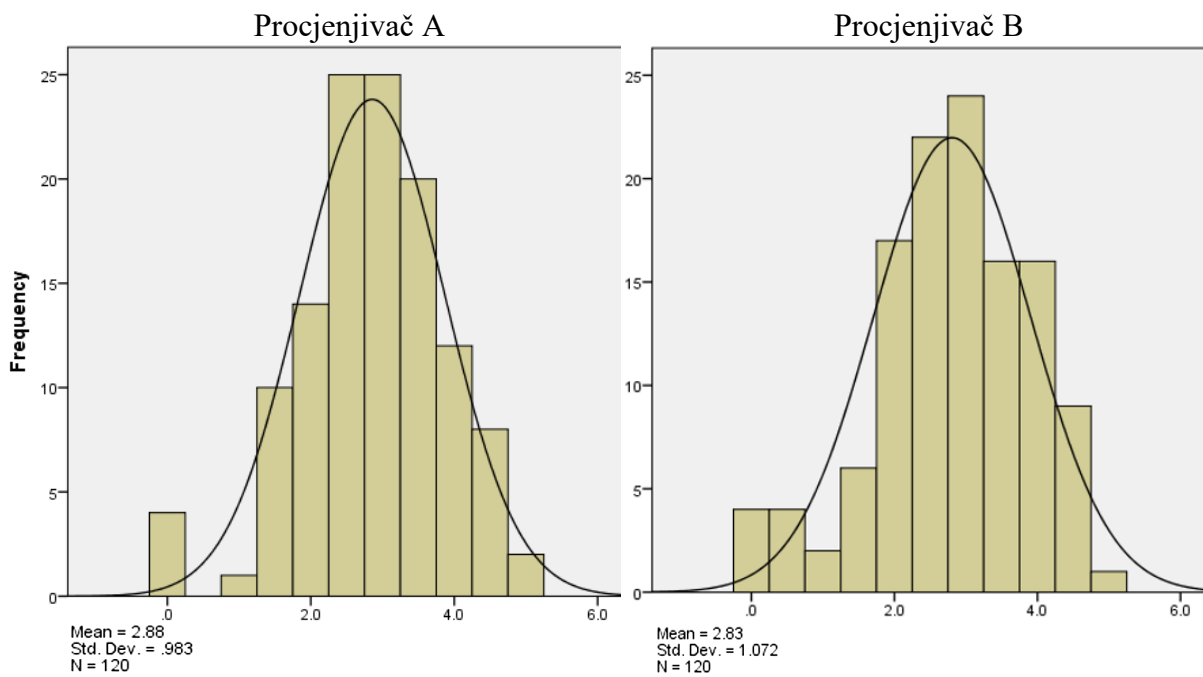
1. E-ispit – kategorija „oblik“



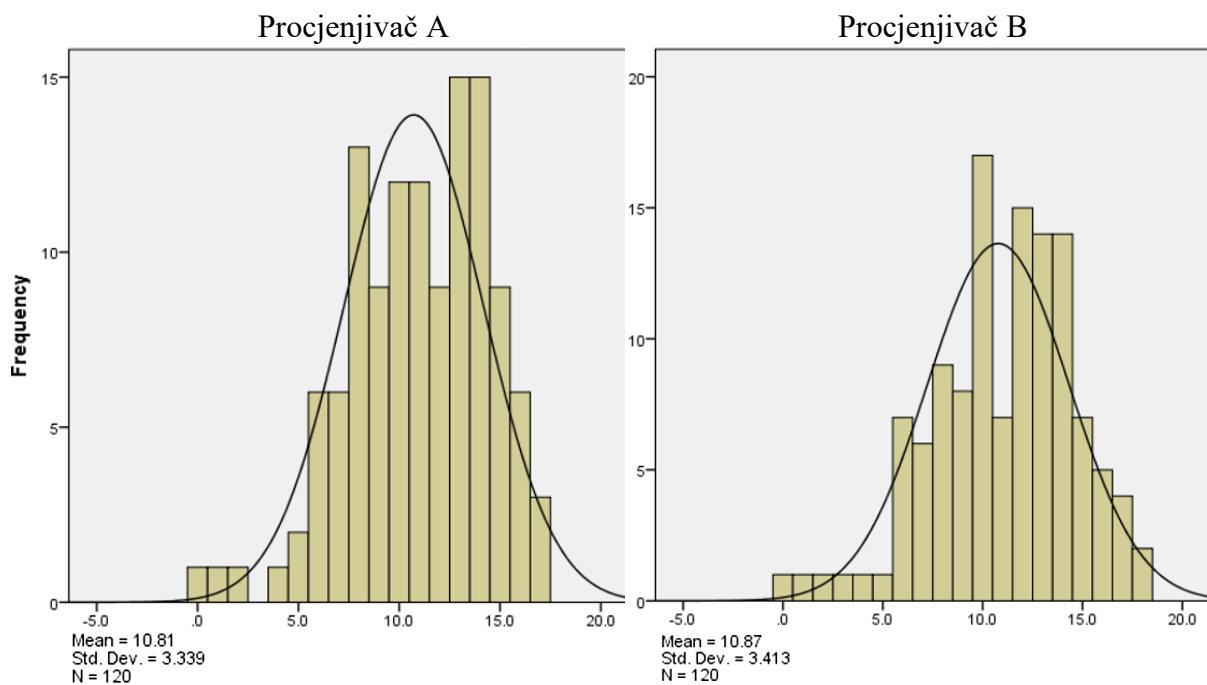
2. E-ispit – kategorija „sadržaj“



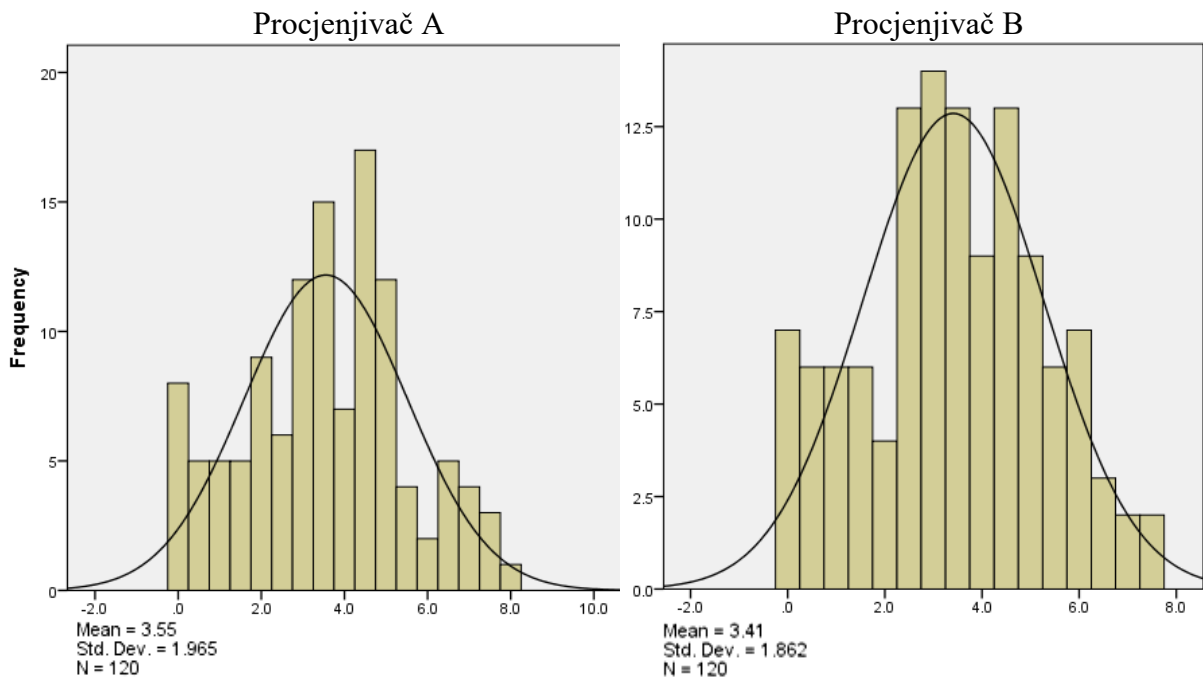
3. E-ispit – kategorija „jezik“



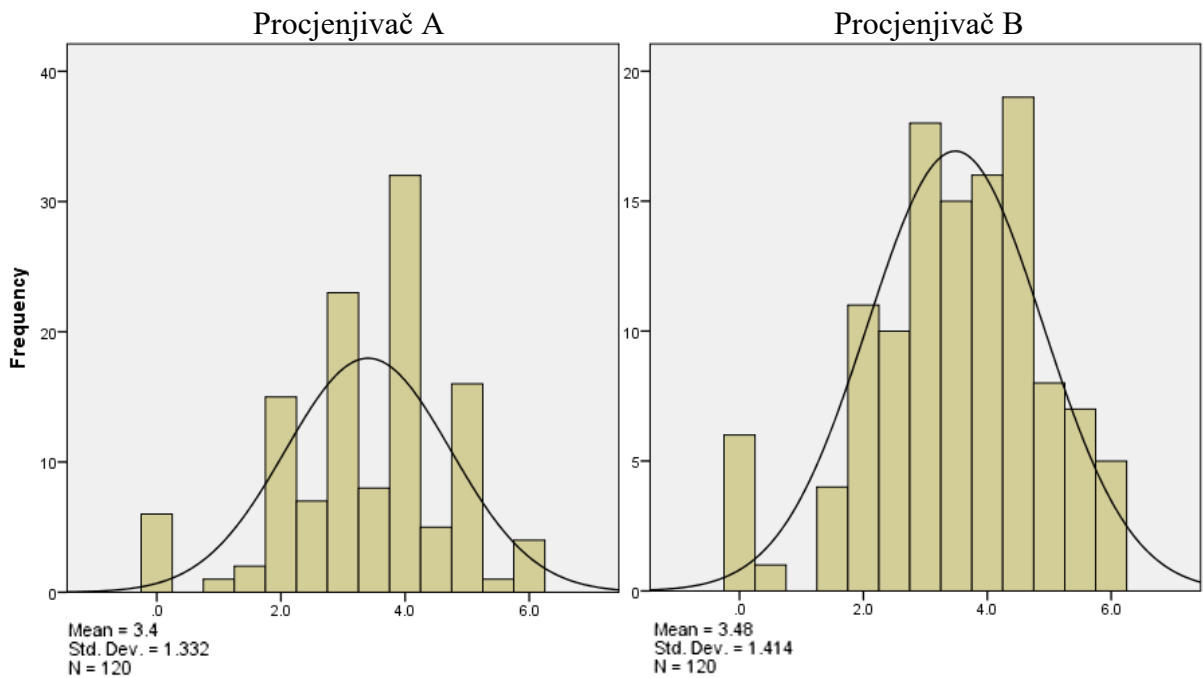
4. E-ispit – ukupno



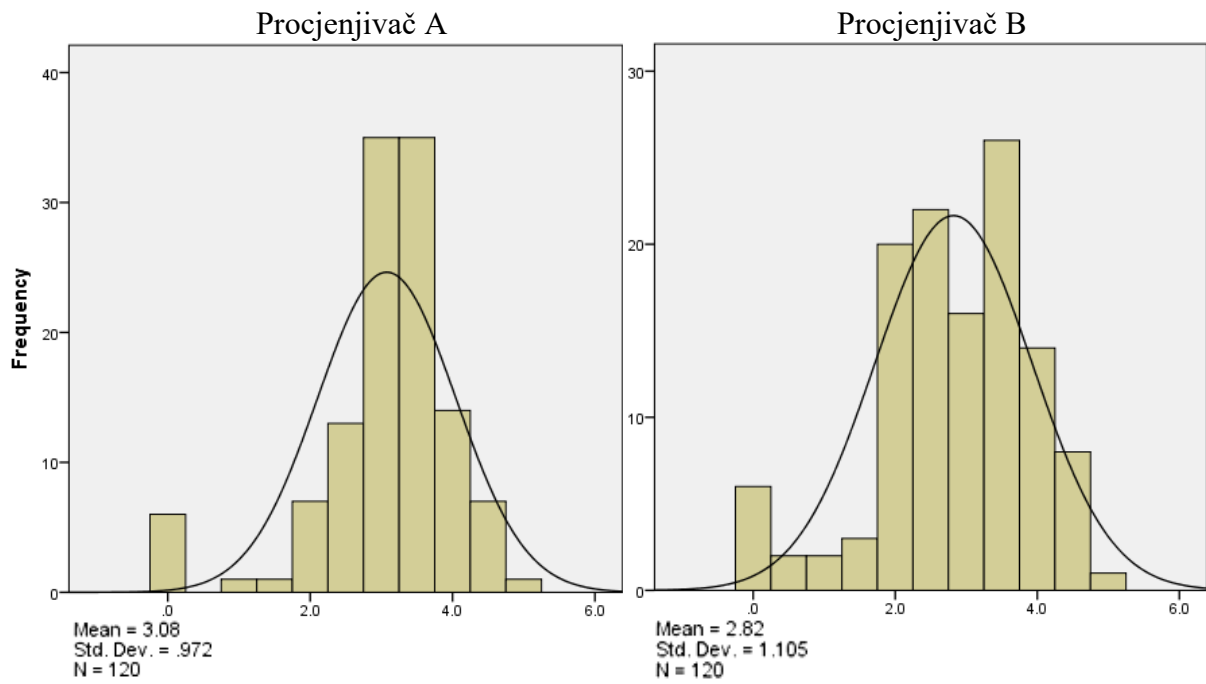
5. Ispit na papiru – kategorija „oblik“



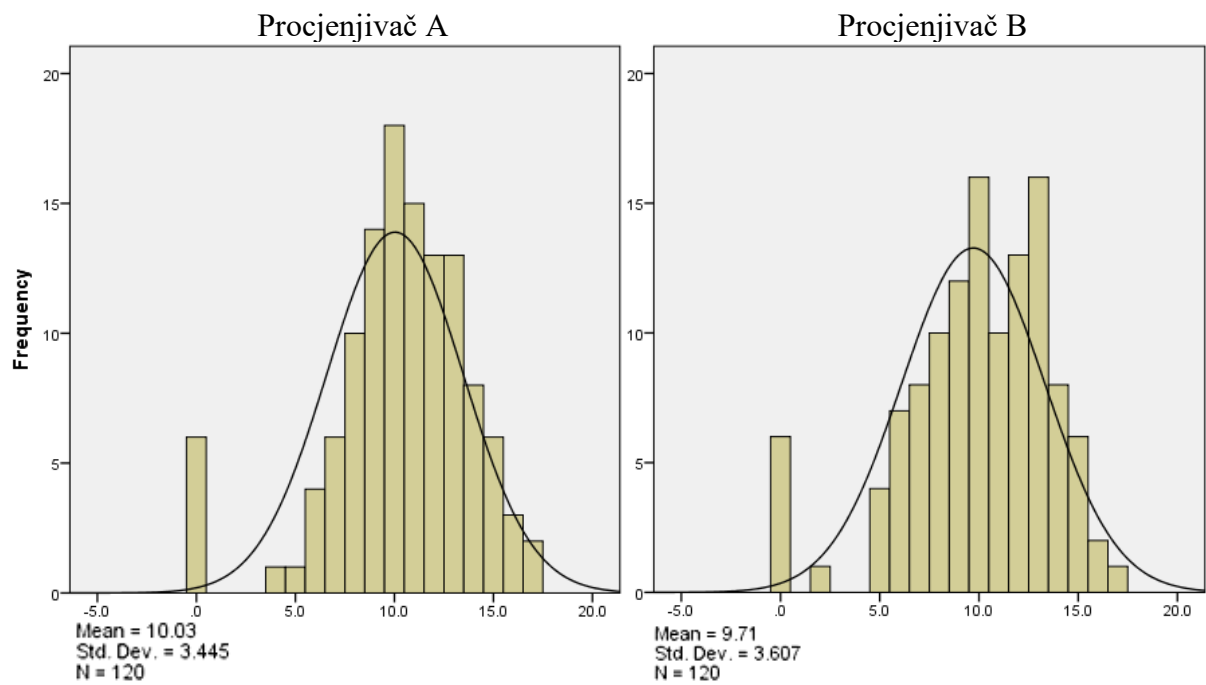
6. Ispit na papiru – kategorija „sadržaj“



7. Ispit na papiru – kategorija „jezik“



8. Ispit na papiru – ukupno



PRILOG 2. Upitnik o stavovima o e-ispitima Poslovnog engleskog jezika

(iz Googleova obrasca)

U ovom upitniku nalazi se 20 tvrdnja o ispitima Poslovnog engleskog jezika (PEJ).

Odaberite koliko se slažete s navedenom tvrdnjom – od 1 (*Nikako se ne slažem s tvrdnjom.*) do 6 (*Sasvim se slažem s tvrdnjom.*).

Odaberite odgovor na koji ste prvo pomislili (nemojte predugo razmišljati o svojim odgovorima). **NE BRINITE AKO VAM SE ODGOVORI BUDU ČINILI NEDOSLJEDNIMA.** Odaberite samo jedan odgovor za svaku tvrdnju. Čak i ako ni jedan odgovor nije točna slika Vašeg stava, molimo Vas da odaberete onaj koji mu je najbliži (važno je dati odgovore na sve tvrdnje).

1. Mislim da je dizajn e-ispita PEJ-a sve u svemu odgovarajuć.
2. Mislim da e-ispiti PEJ-a sveobuhvatno ispituju moju razinu poznavanja engleskog jezika.
3. Vjerujem da bodovi na e-ispitima PEJ-a uistinu odražavaju moju razinu poznavanja engleskog jezika.
4. Mislim da e-ispiti PEJ-a ispituju one jezične sposobnosti i jezične vještine koje trebam za studij i posao.
5. Mislim da zadatci u e-ispitima PEJ-a odražavaju primjenu jezika u stvarnom životu.
6. Lako mogu pristupiti informacijama o e-ispitima PEJ-a prije polaganja ispita.
7. Razumijem na koji se način u e-ispitima PEJ-a procjenjuju moje sposobnosti u engleskom jeziku.
8. Razumijem mjerila bodovanja u zadacima pisanja na e-ispitima PEJ-a.
9. Razumijem mjerila bodovanja u zadacima čitanja i poznavanja vokabulara na e-ispitima PEJ-a.
10. Razumijem mjerila bodovanja u zadacima govorenja na e-ispitima PEJ-a.
11. Upoznat(a) sam s internetskom stranicom, odnosno Googleovom učionicom u kojoj su svi podatci vezani za e-ispite PEJ-a.
12. Nastavnici mi daju dovoljno informacija o e-ispitima PEJ-a.
13. Zadovoljan/na sam načinom provedbe e-ispita PEJ-a.
14. Zadovoljan/sam uvjetima u kojima polažem e-ispite PEJ-a.
15. Smatram da nastavnici dobro nadgledaju e-ispite PEJ-a.
16. Poboljšao/la sam svoju razinu poznavanja engleskog jezika pripremajući se i polažući e-ispite PEJ-a.
17. Mislim da me e-ispiti PEJ-a potiču da provodim više vremena učeći engleski jezik.
18. Mislim da me e-ispiti PEJ-a potiču da više slušam emisije na engleskom jeziku na radiju, TV-u, internetu i dr.
19. Mislim da me e-ispiti PEJ-a potiču da više čitam tekstove na engleskom jeziku.
20. Vjerujem da me e-ispiti PEJ-a potiču da marljivije učim engleski jezik.

Prilog 2.1. Izjave iz upitnika o stavovima o ispitu The Fudan English Test (FET)

(Fan i Ji, 2014)

STATEMENTS IN THE SURVEY QUESTIONNAIRE

1. I think the design of the test is on the whole reasonable.
2. I think the FET is a comprehensive test of my English proficiency.
3. I believe the FET test score can truly reflect my English level.
4. I think the language abilities assessed in the FET are what I need for my studies and work.
5. I think the test items in the FET reflect real-life language use.
6. I can easily access test-related information prior to taking the test.
7. I understand how the FET assesses my English language abilities.
8. I understand the scoring criteria for the FET composition section.
9. I understand the scoring criteria for the FET speaking test.
10. I am familiar with the official webpage of the FET.
11. My English teachers have provided information about the test.
12. I am satisfied with the organization of the test.
13. I am satisfied with the examination hall for the FET written test.
14. I think the proctors have performed their duties well in the written test.
15. I think the examiners have performed their duties well in the speaking test.
16. I can perform to the best of my ability in the FET speaking test.
17. I can adjust myself well to computer-based speaking test.
18. I believe the FET speaking test can well assess my English speaking abilities.
19. Without an interlocutor in the FET speaking test, I feel much more relaxed.
20. I have improved my English level through preparing and taking the FET.
21. I think the FET has motivated me to spend more time learning English.
22. I think the FET has motivated me to listen to more English radio programs.
23. I think the FET has motivated me to read more English newspapers and magazines.
24. I believe the FET has pushed me to study harder in English.

Prilog 2.2. Odgovori sudionika na Upitniku o stavovima o e-ispitima PEJ-a

Sudionik	Čestica																			
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1	3	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	2	1	1	3
2	5	5	6	5	6	4	5	5	5	5	6	6	6	5	5	6	6	6	6	6
3	6	6	5	5	6	4	5	6	6	6	6	5	5	5	5	4	4	3	4	4
4	6	5	5	6	6	6	6	6	5	5	6	6	6	6	6	6	6	5	5	5
5	3	3	3	4	3	3	3	4	3	4	6	3	2	2	5	4	3	4	2	2
6	4	3	3	4	5	6	4	5	4	3	6	6	4	3	3	4	2	3	3	2
7	4	5	4	5	5	6	5	6	6	6	6	6	5	4	5	6	6	3	5	6
8	6	6	5	6	5	6	6	6	5	6	6	5	6	6	6	5	6	6	6	6
9*	4	6	4	4	5	5	6	6	6	6	6	5	4	4	5	2	2	3	2	2
10	3	4	3	5	4	6	5	4	4	6	6	6	2	1	6	4	3	4	3	3
11	5	6	5	5	4	5	5	5	5	5	6	5	4	4	5	5	5	3	3	5
12	3	4	3	2	2	5	4	4	3	3	6	5	2	1	3	2	2	2	2	3
13	6	5	4	5	5	6	6	6	6	6	6	6	5	6	6	5	6	4	4	4
14	5	4	5	6	5	4	6	4	6	2	6	4	3	4	6	6	3	1	1	2
15	5	4	3	4	3	6	6	6	6	6	6	5	5	5	6	5	6	4	3	6
16	5	5	5	4	3	6	4	4	4	3	6	4	4	3	4	4	3	1	1	4
17	4	5	5	5	3	5	4	5	5	5	6	6	5	4	5	6	4	4	5	5
18	6	5	5	6	6	6	5	5	5	5	6	6	6	6	6	6	6	4	4	6
19	5	6	4	6	6	6	6	5	6	4	6	6	3	3	6	6	6	3	3	6
20	5	5	6	5	6	6	5	5	4	5	6	6	6	6	6	5	6	6	6	5
21	5	5	6	6	6	6	4	3	3	3	6	6	6	4	6	6	6	6	6	6
22	4	5	3	5	4	3	5	4	5	5	6	4	2	4	5	4	4	3	3	2
23	5	5	6	5	4	6	6	6	6	4	6	6	6	6	6	6	6	3	2	5
24	6	6	6	6	5	6	6	6	6	4	5	5	6	6	6	6	5	6	4	5
25	6	6	6	6	4	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6
26	4	2	2	3	4	5	3	2	2	5	6	5	2	2	5	5	4	4	1	3
27	6	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	4	4	6
28	5	5	5	6	4	5	5	4	5	4	6	6	5	3	5	5	3	3	4	4
29	5	6	3	6	6	6	4	6	6	6	6	6	4	5	6	1	3	5	5	6
30	5	5	5	5	4	5	5	3	5	3	5	5	5	4	5	5	5	5	5	5
31	3	4	5	5	4	5	3	3	3	3	6	6	2	3	5	4	3	2	3	6
32	6	6	6	6	6	6	6	6	6	6	6	6	5	5	6	6	6	5	5	5
33	6	6	6	6	6	6	5	6	3	4	6	6	6	6	6	6	6	6	6	6
34	4	3	3	3	3	2	4	3	3	3	1	3	4	4	4	2	2	2	2	2
35	4	5	5	5	4	6	5	5	5	5	6	5	5	4	4	3	3	1	1	4
36	3	5	5	6	4	6	6	6	4	4	6	6	6	4	6	6	3	1	1	4
37	6	6	5	6	5	5	5	5	5	5	6	6	5	5	4	5	4	4	4	4
38	5	5	6	6	4	2	4	5	4	1	6	4	5	5	6	5	5	2	5	3
39*	5	3	3	6	6	6	3	3	3	3	6	6	2	3	6	4	1	1	1	1
40	6	5	5	6	4	6	6	6	6	6	6	6	6	6	6	5	5	3	3	5
41	4	3	4	2	2	1	5	6	6	4	4	5	3	3	5	5	2	1	1	2
42	5	4	4	5	3	5	5	5	4	4	4	4	4	4	5	5	5	2	2	4
43	6	6	6	5	3	6	6	6	6	6	6	6	5	3	4	5	5	1	4	6
44	4	6	6	5	6	6	6	3	4	4	6	5	5	5	6	5	6	3	4	6
45	6	6	6	6	3	4	6	4	5	2	5	4	5	6	6	6	6	1	3	6

Sudionik	Čestica																			
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
46	6	5	4	4	4	5	5	5	5	4	5	3	4	3	2	4	3	2	2	2
47	5	3	3	6	2	6	5	4	5	5	6	6	4	4	6	6	6	2	2	6
48	5	5	6	6	4	5	5	6	5	5	6	6	6	6	6	6	3	6	6	3
49	4	4	5	6	4	6	6	6	6	6	6	3	3	2	6	3	5	1	1	5
50	5	5	5	5	5	5	5	5	5	5	5	5	5	3	6	5	6	6	5	5
51	5	4	3	5	6	5	6	6	6	6	6	5	5	5	5	4	4	3	4	5
52	5	5	4	4	4	4	4	6	3	6	6	5	5	5	5	5	4	3	6	6
53	2	1	1	4	1	5	3	2	5	6	6	6	5	3	4	5	2	2	6	4
54	4	5	5	6	5	6	5	6	6	6	6	5	6	6	6	5	5	5	4	6
55	5	5	4	3	3	4	4	6	5	3	5	5	5	4	5	5	2	4	2	2
56	5	6	6	5	6	5	6	6	6	6	6	6	5	5	5	6	6	3	4	6
57	5	6	6	6	6	6	6	4	5	6	6	6	6	6	6	6	6	6	6	6
58	6	5	4	6	6	6	5	6	6	6	6	6	6	6	5	3	4	3	4	4
59	6	5	5	6	4	5	6	6	6	6	6	6	6	6	6	6	5	6	6	6
60	6	6	5	6	6	6	6	5	6	5	6	6	6	4	6	6	4	1	2	3
61	4	4	4	6	4	5	3	5	6	6	6	4	5	3	6	6	2	3	2	4
62	5	4	5	5	6	6	6	4	6	6	6	6	6	5	6	6	4	3	4	4
63	6	6	6	6	5	6	6	6	6	6	6	6	5	6	6	6	5	6	6	6
64	3	6	4	4	4	3	4	3	3	4	3	4	5	4	6	4	5	3	5	5
65	6	5	5	6	6	6	6	6	6	6	6	6	6	5	4	5	5	2	2	5
66	5	5	5	6	5	6	5	4	4	4	6	5	4	5	5	6	6	5	5	5
67	3	6	4	6	5	5	3	4	5	4	5	6	4	4	6	4	6	2	3	5
68	5	5	3	3	2	6	4	4	6	3	6	6	5	6	6	4	4	2	2	3
69	6	6	5	6	6	6	6	4	6	5	6	6	6	6	6	6	5	5	5	5
70	5	5	5	6	5	5	6	4	5	5	6	6	5	5	6	4	4	3	3	4
71	6	6	4	6	6	6	6	5	5	5	6	6	6	5	6	6	6	4	4	6
72	6	6	4	6	4	6	6	6	5	6	6	6	6	6	6	6	5	4	5	6
73	6	4	3	6	5	6	6	6	6	4	6	6	5	6	4	4	3	1	1	1
74	6	6	6	6	6	6	5	6	6	6	5	6	6	6	6	6	4	4	5	6
75	6	6	6	6	6	6	6	6	6	6	6	6	6	4	6	5	6	1	6	6
76	6	6	6	5	6	6	6	5	4	6	6	6	6	4	6	6	6	4	6	6
77	6	3	3	4	3	3	6	6	5	6	6	6	5	6	6	5	5	3	5	5
78	6	5	3	2	2	6	5	4	5	4	6	6	5	4	5	6	5	3	4	6
79	3	4	3	3	3	3	5	6	6	6	6	6	4	2	6	4	5	5	6	4
80	4	3	2	4	3	3	4	3	3	3	5	3	4	2	5	5	5	5	5	5
81	6	6	6	6	6	6	6	6	6	6	6	3	5	5	6	6	6	6	6	6
82	5	6	3	5	5	4	5	3	5	3	5	3	3	1	5	5	5	5	4	5
83	4	5	2	5	4	4	3	3	3	3	5	5	5	4	4	4	2	3	2	2
84	4	5	2	3	4	4	5	4	3	4	6	4	4	2	5	5	3	5	5	4
85	6	5	4	5	5	6	6	6	6	6	6	5	5	3	6	6	5	4	6	5
86	4	5	5	3	2	6	4	5	5	6	6	5	5	5	5	4	6	6	6	5
87	6	5	3	6	6	6	6	4	4	4	6	5	5	4	6	6	5	1	1	3
88	5	4	2	5	4	4	3	2	3	6	6	5	4	2	5	6	6	5	3	5
89	6	6	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
90	5	5	4	6	6	5	4	4	6	5	4	6	6	4	6	5	5	5	4	4
91	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
92	3	2	1	6	6	6	4	3	4	4	6	5	1	2	6	3	2	2	2	6
93	3	4	2	3	2	4	3	5	5	4	4	3	3	4	6	3	2	1	1	3

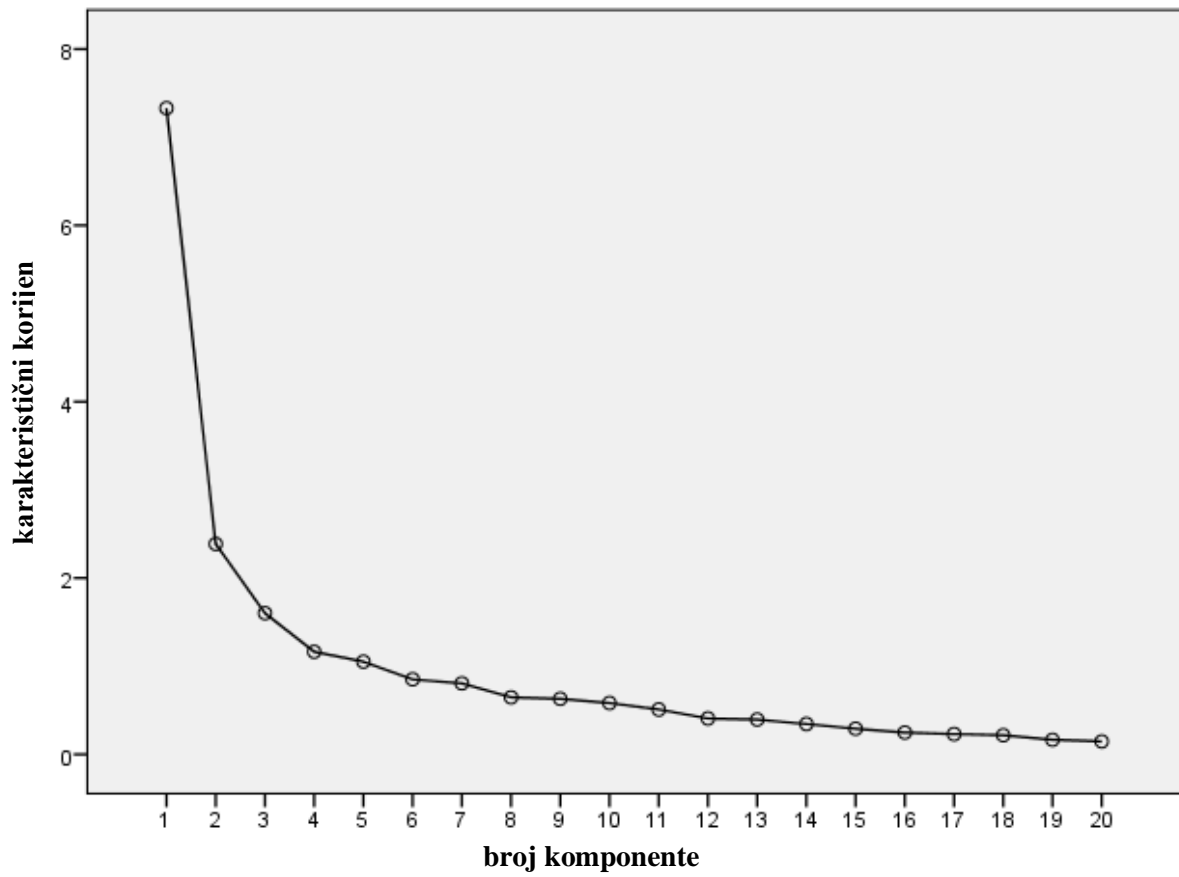
Sudionik	Čestica																			
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
94	5	5	5	5	4	4	3	6	6	6	6	5	4	3	6	5	5	5	5	6
95	6	6	6	6	5	4	5	6	6	4	6	6	6	5	6	6	4	5	6	5
96	6	5	4	6	5	6	5	4	5	4	6	5	5	4	5	6	6	6	6	6
97	4	4	3	3	2	1	3	2	2	1	3	2	1	3	5	2	4	3	4	3
98	4	6	3	5	4	3	3	4	4	4	6	4	4	5	6	2	2	2	2	4
99	4	5	3	6	6	6	6	6	6	6	6	4	2	2	6	3	1	3	3	1
100	5	4	4	6	5	6	6	4	5	5	6	5	4	5	6	6	5	2	4	4
101	6	6	4	6	6	6	6	6	4	6	6	6	4	4	6	6	6	6	5	6
102	5	5	4	6	6	6	5	5	4	5	6	4	5	6	6	6	6	3	3	3
103	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
104	2	3	3	3	3	4	4	5	3	3	5	4	3	4	5	4	3	4	5	4
105	4	4	3	3	3	5	5	5	4	4	6	4	3	4	4	3	5	4	5	5
106	6	6	5	6	5	3	5	6	6	6	6	4	5	2	6	1	1	1	1	1
107	5	5	4	5	6	5	5	5	4	5	6	5	6	6	6	6	1	6	2	1
108	5	6	6	6	5	5	6	6	5	6	6	6	4	4	6	6	5	5	5	5
109	5	5	1	4	4	6	5	3	5	5	6	6	6	6	6	3	3	1	1	6
110	6	6	5	6	6	1	6	6	6	6	6	5	5	4	5	4	5	4	5	4
111	6	6	5	6	5	6	5	6	4	6	6	5	3	6	4	6	5	2	2	3
112	4	1	1	5	5	6	6	6	6	6	6	6	3	3	1	1	1	1	1	1
113	5	6	4	6	6	6	6	6	6	6	6	6	5	4	6	6	6	4	4	6
114	5	5	2	6	3	5	6	5	6	6	6	6	4	1	2	3	1	1	1	2
115	5	5	4	6	5	6	5	6	6	6	6	6	5	6	6	6	6	5	5	5
116	5	6	4	6	6	6	5	4	6	1	1	6	6	3	4	6	1	1	1	1
117	4	5	5	5	5	6	5	5	4	4	5	6	5	5	6	4	4	4	4	5
118	5	4	4	5	6	6	6	6	6	6	6	6	6	6	6	6	5	3	4	4
119	4	5	5	5	6	6	4	6	5	4	6	3	4	4	2	5	2	5	5	5
120	6	5	5	4	4	6	6	6	6	6	6	6	6	6	6	4	2	1	1	2
121	6	6	6	6	6	6	6	6	6	6	6	6	5	4	6	6	6	6	6	6
122	4	5	5	5	6	6	5	6	6	6	6	6	6	3	6	4	4	6	5	6

* Podatci isključeni iz daljnje obrade jer ovi sudionici na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Prilog 2.3. Matrica komponentata latentne strukture Upitnika o stavovima o e-ispitima PEJ-a (N = 120)

Čestica	1	2	3	4	5	h ²
1. Mislim da je dizajn e-ispita PEJ-a sve u svemu odgovarajuć.	,695	-,138	-,292	,112	-,016	,600
2. Mislim da e-ispiti PEJ-a sveobuhvatno ispituju moju razinu poznavanja engleskog jezika.	,650	,074	-,432	,116	,226	,679
3. Vjerujem da bodovi na e-ispitima PEJ-a uistinu odražavaju moju razinu poznavanja engleskog jezika.	,695	,123	-,369	,122	,069	,654
4. Mislim da e-ispiti PEJ-a ispituju one jezične sposobnosti i jezične vještine koje trebam za studij i posao.	,653	-,207	-,182	-,298	,428	,774
5. Mislim da zadatci u e-ispitima PEJ-a odražavaju primjenu jezika u stvarnom životu.	,634	-,146	-,137	-,219	,495	,735
6. Lako mogu pristupiti informacijama o e-ispitima PEJ-a prije polaganja ispita.	,545	-,272	,285	-,499	-,013	,702
7. Razumijem na koji se način u e-ispitima PEJ-a procjenjuju moje sposobnosti u engleskom jeziku.	,668	-,402	-,065	,047	,023	,615
8. Razumijem mjerila bodovanja u zadacima pisanja na e-ispitima PEJ-a.	,581	-,426	,081	,519	-,007	,794
9. Razumijem mjerila bodovanja u zadacima čitanja i poznavanja vokabulara na e-ispitima PEJ-a.	,544	-,552	,038	,287	-,064	,689
10. Razumijem mjerila bodovanja u zadacima govorenja na e-ispitima PEJ-a.	,537	-,293	,564	,257	,068	,763
11. Upoznat(a) sam s internetskom stranicom, odnosno Googleovom učionicom u kojoj su svi podaci vezani za e-ispite PEJ-a.	,422	-,162	,599	-,096	,155	,596
12. Nastavnici mi daju dovoljno informacija o e-ispitima PEJ-a.	,578	-,227	,285	-,293	-,342	,669
13. Zadovoljan/na sam načinom provedbe e-ispita PEJ-a.	,730	-,064	-,213	-,026	-,411	,752
14. Zadovoljan/sam uvjetima u kojima polažem e-ispite PEJ-a.	,653	-,016	-,215	-,007	-,432	,660
15. Smatram da nastavnici dobro nadgledaju e-ispite PEJ-a.	,500	,284	-,069	-,090	-,058	,347
16. Poboljšao/la sam svoju razinu poznavanja engleskog jezika pripremajući se i polažući e-ispite PEJ-a.	,638	,268	-,106	-,277	-,174	,597
17. Mislim da me e-ispiti PEJ-a potiču da provodim više vremena učeći engleski jezik.	,658	,503	,072	-,047	-,023	,695
18. Mislim da me e-ispiti PEJ-a potiču da više slušam emisije na engleskom jeziku na radiju, TV-u, internetu i dr.	,483	,595	,210	,213	,153	,700
19. Mislim da me e-ispiti PEJ-a potiču da više čitam tekstove na engleskom jeziku.	,546	,613	,223	,292	,043	,811
20. Vjerujem da me e-ispiti PEJ-a potiču da marljivije učim engleski jezik.	,597	,506	,288	-,059	-,017	,700
karakteristični korijeni	7,330	2,386	1,601	1,164	1,051	
% objašnjene varijance	36,684	11,930	8,007	5,818	5,255	67,694

Prilog 2.4. *Scree plot* prikaz latentne strukture Upitnika o stavovima o e-ispitima PEJ-a (N = 120)



PRILOG 3. Upitnik o stavovima o ispitima općenito

U ovom dijelu upitnika nalazi se 17 tvrdnja vezanih za ispite.

Odaberite koliko se slažete s navedenom tvrdnjom – od 1 (*Nikako se ne slažem s tvrdnjom.*) do 5 (*Sasvim se slažem s tvrdnjom.*).

Odaberite odgovor na koji ste prvo pomislili (nemojte predugo razmišljati o svojim odgovorima). NE BRINITE AKO VAM SE ODGOVORI BUDU ČINILI NEDOSLJEDNIMA. Odaberite samo jedan odgovor za svaku tvrdnju. Čak i ako ni jedan odgovor nije točna slika Vašeg stava, molimo Vas da odaberete onaj koji mu je najbliži (važno je dati odgovore na sve tvrdnje).

STAVITE „X“ U ODGOVARAJUĆU KUĆICU.		1 Nikako se ne slažem	2 Ne slažem se	3 Niti se slažem, niti ne slažem	4 Slažem se	5 Sasvim se slažem
1.	Ispiti me motiviraju da više učim.					
2.	Ispiti su važni za razumijevanje gradiva.					
3.	Ispiti mi pomažu organizirati vrijeme za učenje.					
4.	Ispiti mi pomažu usredotočiti se na važne dijelove gradiva.					
5.	Ispiti mi pokazuju širu sliku onoga što učim.					
6.	Ispiti su korisni za određivanje moje razine postignuća.					
7.	Ispiti mi pomažu odrediti moje akademske slabe točke.					
8.	Tijekom ispita učim mnoge korisne stvari (npr. organizaciju, upravljanje vremenom, kako raditi brže itd.).					
9.	Za mene je polaganje ispita bolno iskustvo.					
10.	Znam kako se primjereno ponašati na ispitu.					
11.	Smatram da se ispiti ne bi trebali koristiti u procesu poučavanja.					
12.	Polaganje ispita povećava moje samopouzdanje.					
13.	Trudim se izbjegavati kolegije koji imaju mnogo ispita.					
14.	Uživam polagati ispite.					
15.	Ispiti mi pružaju priliku da pokažem svoje znanje i vještine koje sam naučio tijekom kolegija.					
16.	Više volim kolegije koji imaju mnogo ispita.					
17.	Općenito, smatram da su ispiti korisno iskustvo za mene.					

Prilog 3.1. Ljestvica stavova o ispitima

(Dodeen, 2008 u Muñoz, 2017)

Attitude toward Tests Scale (ATS). English version

Attitude toward Tests Scale—by Hamzeh Dodeen (Translated from Arabic to English by Hamzeh Dodeen)

No.	Item	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	Tests motivate me to study hard					
2	Tests are important to understand materials					
3	Tests help me organize studying time					
4	Tests help me focus on important materials					
5	Tests help me make a big picture of what I am studying					
6	Tests are useful to determine my achievement level					
7	Tests help me identify my academic weaknesses					
8	During tests, I learn many useful skills (e.g., organization, time management, working fast ...)					
9	For me, taking tests is a painful experience					
10	I know how to behave appropriately in tests					
11	I believe that tests should not be used in teaching					
12	Taking tests increases my self-confidence					
13	I try to avoid courses that have many tests					
14	I enjoy taking tests					
15	Tests give me opportunity to present my skills and knowledge I learned during the course					
16	I prefer courses that require many tests					
17	Overall, I think taking tests is a useful experience for me					

Prilog 3.2. Odgovori sudionika na Upitniku o stavovima o ispitima općenito

Sudionik	Čestica																
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1	5	3	2	4	4	1	4	4	2	5	4	2	2	2	4	2	4
2	4	3	4	4	3	3	4	2	2	5	4	3	2	3	4	3	3
3	5	4	3	3	2	2	3	4	5	4	4	3	3	1	2	2	3
4	5	3	3	4	2	3	3	3	2	5	4	4	3	4	3	2	3
5	3	4	4	4	2	3	3	4	4	5	3	2	2	2	4	3	2
6	5	5	4	4	5	5	5	5	2	5	1	4	2	4	4	3	4
7	5	5	4	5	3	3	5	4	3	5	3	4	2	3	4	4	4
8	4	4	5	3	4	4	4	5	2	4	2	4	2	3	4	1	4
9*	4	4	4	4	4	4	4	4	2	5	2	3	3	2	3	2	4
10	2	3	2	2	2	1	3	4	4	5	4	1	2	1	3	1	2
11	4	4	4	3	3	3	4	4	4	4	2	3	4	1	4	1	3
12	4	4	3	4	2	3	3	2	2	5	2	2	2	1	3	1	3
13	4	4	2	2	2	5	4	4	3	4	2	5	3	5	4	4	4
14	2	3	1	2	4	4	4	1	5	5	3	2	4	1	4	1	2
15	4	4	4	3	3	4	4	4	4	5	2	2	2	1	4	2	2
16	2	3	4	3	4	4	4	5	3	5	3	4	2	3	3	4	4
17	4	4	5	3	2	5	5	4	3	4	2	5	3	4	4	2	4
18	4	4	4	4	4	5	5	5	2	5	2	5	3	3	4	3	4
19	4	4	4	4	4	2	3	4	5	5	3	1	3	1	3	1	3
20	3	4	4	4	3	4	3	4	4	5	3	5	2	2	3	2	3
21	4	4	4	4	4	4	3	4	3	5	1	3	1	2	4	1	4
22	5	4	4	5	3	4	3	3	2	4	2	4	2	3	4	3	4
23	2	4	3	1	3	3	4	3	4	5	1	2	2	2	2	4	3
24	3	2	3	2	2	4	3	3	4	5	3	3	4	2	3	1	3
25	4	4	4	2	2	4	4	2	3	4	3	4	2	2	4	2	3
26	3	3	2	2	2	4	4	3	5	5	3	1	2	1	3	1	3
27	4	4	4	3	3	4	4	4	3	4	3	3	4	3	4	3	3
28	4	4	3	3	5	4	4	5	3	4	2	4	2	3	4	2	4
29	4	2	5	3	3	5	4	5	5	5	3	1	3	1	4	3	2
30	3	4	4	4	2	2	4	3	4	5	3	3	4	1	2	2	3
31	4	4	3	4	4	5	4	5	5	5	2	2	4	1	4	1	3
32	5	5	4	4	4	4	5	5	2	5	2	3	1	2	4	1	4
33	3	3	3	2	1	1	2	4	5	5	3	2	1	1	2	1	2
34	3	2	2	2	3	2	2	3	4	5	3	2	3	2	3	1	3
35	4	4	4	4	3	3	3	4	2	4	2	5	2	2	4	1	4
36	4	4	5	5	4	3	4	5	4	5	3	2	2	2	3	2	4
37	3	4	4	4	4	4	3	4	2	5	3	4	2	3	3	1	3
38	5	5	5	3	2	3	4	4	3	5	1	5	3	1	5	3	4
39*	5	5	5	5	4	3	2	4	3	5	4	3	4	4	3	4	3
40	2	4	5	3	1	1	4	5	2	4	3	4	3	1	4	2	4
41	4	4	2	2	2	4	2	4	4	4	2	1	4	2	2	2	2
42	2	1	3	3	2	1	3	2	4	5	4	2	3	2	3	2	2
43	2	3	1	1	3	4	4	3	3	5	3	2	3	1	3	3	2
44	4	2	4	2	2	2	4	2	3	5	4	2	2	1	3	1	2
45	2	1	1	2	2	5	1	4	5	5	5	1	3	2	1	1	1

Sudionik	Čestica																
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
46	5	5	5	5	4	4	3	4	2	5	2	4	2	4	4	3	4
47	2	4	4	3	4	1	2	5	5	5	5	1	3	1	2	4	2
48	4	2	2	2	2	1	4	1	3	5	3	5	3	2	4	3	3
49	5	4	4	4	4	3	3	4	3	5	3	2	1	2	4	1	3
50	4	4	3	4	3	4	4	2	5	5	3	4	3	1	3	1	3
51	4	3	4	4	3	2	3	4	5	4	3	2	3	2	3	1	2
52	5	4	5	5	3	4	4	5	1	5	3	4	3	4	4	4	4
53	2	3	4	3	1	2	3	4	3	5	2	3	4	1	2	2	3
54	4	3	4	4	4	2	3	5	3	5	4	2	1	3	3	2	3
55	5	4	4	3	2	2	2	4	2	5	2	3	4	3	2	1	2
56	5	4	3	4	4	4	3	5	2	5	2	5	1	3	4	3	1
57	5	4	5	5	5	3	5	5	4	5	3	1	2	2	3	4	4
58	3	4	3	4	3	2	4	4	3	5	4	4	2	2	3	3	2
59	2	2	2	1	2	1	3	4	3	5	5	2	2	1	3	1	2
60	4	3	1	1	2	3	5	3	4	2	3	1	1	2	4	1	3
61	3	4	4	3	3	3	4	5	2	4	3	3	2	2	4	1	4
62	4	4	2	3	4	3	4	5	3	5	3	4	3	1	3	4	4
63	4	5	4	4	4	5	4	3	2	5	2	4	2	3	4	2	3
64	2	4	3	3	2	4	4	5	3	5	2	3	2	3	4	1	3
65	5	3	5	5	5	5	5	3	4	5	1	4	3	4	5	3	5
66	3	4	4	4	4	5	3	4	2	5	2	3	3	3	4	2	4
67	3	3	4	4	3	4	4	4	5	5	3	3	1	2	3	1	3
68	4	4	2	3	4	4	5	4	3	5	2	4	2	2	4	2	3
69	3	4	5	4	4	2	3	4	3	5	3	5	2	3	5	2	3
70	4	4	4	3	5	4	4	3	2	4	3	3	2	3	4	1	4
71	3	4	3	4	4	5	5	4	3	5	1	3	2	3	4	3	4
72	4	4	4	3	3	4	4	4	3	5	3	4	4	1	4	2	3
73	5	2	5	4	1	3	5	5	4	5	3	5	1	2	4	4	3
74	3	3	4	4	3	2	4	4	2	4	2	5	2	2	4	2	3
75	4	4	4	4	2	3	2	3	1	5	2	4	2	2	3	1	2
76	3	2	1	3	2	5	2	4	5	5	5	1	3	1	1	4	5
77	4	2	3	2	2	2	4	3	3	5	4	4	5	2	4	2	3
78	4	2	2	3	1	1	1	2	4	5	3	3	2	1	2	1	2
79	4	3	2	4	3	2	4	4	4	5	1	2	1	2	4	3	3
80	2	4	3	3	3	4	2	4	4	4	2	2	4	3	4	2	3
81	3	1	4	1	5	5	5	5	4	5	3	2	1	1	5	3	3
82	1	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	3
83	2	4	4	2	2	3	2	3	1	4	3	2	2	1	4	1	3
84	4	4	4	4	4	2	4	4	5	5	3	3	3	3	3	3	3
85	4	3	4	4	5	5	3	4	5	5	4	2	4	1	4	2	2
86	5	5	3	5	5	4	4	5	4	5	2	3	3	1	5	1	4
87	3	2	3	1	1	1	1	4	4	5	5	5	2	1	3	1	1
88	3	4	4	5	3	4	2	3	3	5	2	3	3	2	4	2	2
89	4	4	5	5	5	5	5	5	2	5	1	5	1	3	5	3	5
90	5	4	4	5	3	2	4	5	2	5	2	4	3	3	4	2	4
91	5	5	5	5	3	4	3	3	4	5	3	4	4	3	3	2	2
92	3	4	3	3	4	5	4	4	4	5	4	1	2	1	3	1	3
93	4	2	3	4	2	1	2	4	3	5	3	2	3	1	4	2	3

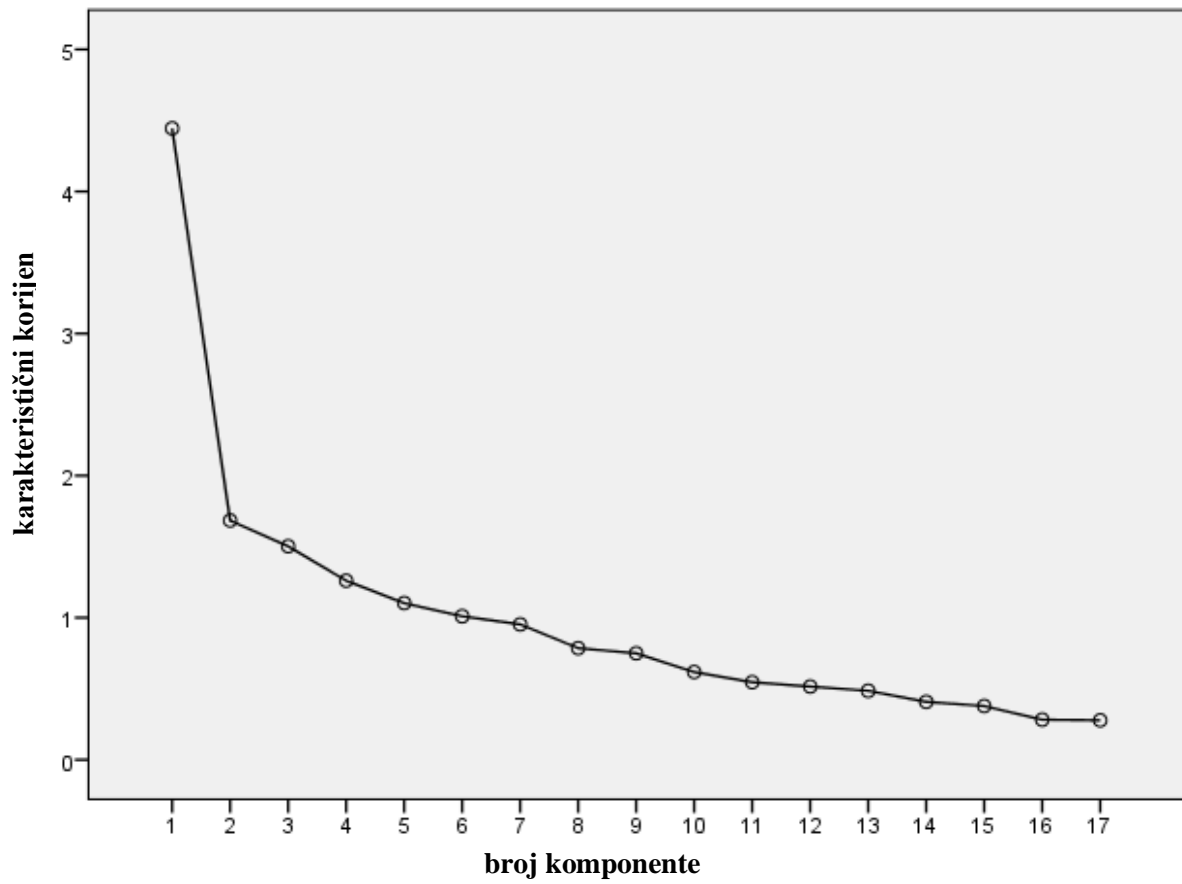
Sudionik	Čestica																
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
94	4	3	4	3	4	3	2	1	3	5	1	3	2	1	4	2	4
95	4	3	2	1	3	4	4	4	5	5	3	4	2	3	4	1	2
96	4	2	5	5	4	3	4	4	4	5	4	4	3	2	4	1	1
97	4	4	4	4	4	2	2	1	3	5	4	1	3	1	2	1	2
98	4	2	4	3	1	2	3	4	4	5	3	2	3	1	3	1	2
99	3	4	5	5	3	1	1	4	3	5	1	5	3	1	3	1	2
100	4	4	2	4	2	4	2	3	2	5	2	3	4	1	4	1	4
101	5	5	5	5	5	4	5	5	5	3	3	4	2	1	4	3	4
102	4	3	4	4	2	3	2	4	1	4	3	4	1	3	3	2	4
103	3	3	4	4	3	1	4	3	2	5	4	4	2	2	4	1	2
104	2	3	3	2	3	3	3	4	2	5	3	4	4	2	4	2	3
105	4	4	4	2	2	4	4	3	2	4	2	3	3	2	3	2	4
106	5	3	4	5	2	2	3	4	3	5	3	2	1	2	4	1	3
107	5	3	4	4	3	3	3	4	5	5	2	3	2	2	4	2	3
108	5	3	5	4	3	2	3	4	1	5	3	4	1	2	4	1	1
109	2	4	3	3	2	5	3	3	5	5	3	1	3	1	4	1	2
110	3	3	1	3	3	4	4	2	5	5	2	1	4	1	2	1	3
111	5	5	4	5	2	5	5	4	3	5	3	2	4	1	2	1	1
112	2	1	1	3	1	3	2	1	1	5	1	1	1	4	2	1	1
113	4	3	3	4	3	3	3	4	3	5	3	3	2	4	4	3	3
114	5	2	4	3	2	2	2	4	4	5	3	2	1	2	4	3	3
115	4	4	4	5	5	5	5	4	5	5	4	2	4	4	2	1	3
116	4	1	5	3	1	4	5	5	4	5	1	5	3	1	5	3	5
117	3	4	4	4	3	4	4	4	1	5	2	4	3	3	4	2	4
118	4	3	4	3	2	3	5	4	4	5	3	2	1	1	3	2	3
119	4	3	4	3	3	4	4	5	1	5	2	5	2	2	4	2	4
120	3	4	3	2	1	2	3	3	1	5	3	3	1	3	3	2	3
121	3	3	3	4	5	3	3	4	3	5	3	3	2	3	3	2	3
122	3	3	2	3	3	2	3	4	4	5	2	2	2	1	3	1	3

* Podatci isključeni iz daljnje obrade jer ovi sudionici na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Prilog 3.3. Matrica komponenata latentne strukture Upitnika o stavovima o ispitima općenito (N = 120)

Čestica	1	h ²
1. Ispiti me motiviraju da više učim.	,557	,461
2. Ispiti su važni za razumijevanje gradiva.	,576	,650
3. Ispiti mi pomažu organizirati vrijeme za učenje.	,607	,693
4. Ispiti mi pomažu usredotočiti se na važne dijelove gradiva.	,583	,788
5. Ispiti mi pokazuju širu sliku onoga što učim.	,533	,692
6. Ispiti su korisni za određivanje moje razine postignuća.	,445	,706
7. Ispiti mi pomažu odrediti moje akademske slabe točke.	,569	,603
8. Tijekom ispita učim mnoge korisne stvari (npr. organizaciju, upravljanje vremenom, kako raditi brže itd.).	,474	,543
10. Znam kako se primjereno ponašati na ispitu.	-,078	,712
12. Polaganje ispita povećava moje samopouzdanje.	,585	,702
14. Uživam polagati ispite.	,528	,654
15. Ispiti mi pružaju priliku da pokažem svoje znanje i vještine koje sam naučio tijekom kolegija.	,637	,497
16. Više volim kolegije koji imaju mnogo ispita.	,425	,651
17. Općenito, smatram da su ispiti korisno iskustvo za mene.	,650	,562
9. (rekodirano) Za mene je polaganje ispita bolno iskustvo.	,376	,733
11. (rekodirano) Smatram da se ispiti ne bi trebali koristiti u procesu poučavanja.	,518	,506
13. (rekodirano) Trudim se izbjegavati kolegije koji imaju mnogo ispita.	,133	,849
karakteristični korijeni	4,444	
% objašnjene varijance	26,142	

Prilog 3.4. *Scree plot* prikaz latentne strukture Upitnika o stavovima o ispitima općenito (N = 120)



PRILOG 4. Upitnik o stavovima o računalima

(iz Googleova obrasca)

U ovom dijelu upitnika nalazi se 53 tvrdnja povezanih s računalima.

Odaberite koji od pet mogućih odgovora koji se nalaze uz svaku tvrdnju najbolje opisuje Vaš stav o navedenoj tvrdnji.

Odaberite koliko se slažete s navedenom tvrdnjom – od 1 (*Nikako se ne slažem s tvrdnjom.*) do 5 (*Sasvim se slažem s tvrdnjom.*).

Odaberite odgovor na koji ste prvo pomislili (nemojte predugo razmišljati o svojim odgovorima). **NE BRINITE AKO VAM SE ODGOVORI BUDU ČINILI NEDOSLJEDNIMA.**

Odaberite samo jedan odgovor za svaku tvrdnju. Čak i ako ni jedan odgovor nije točna slika Vašeg stava, molimo Vas da odaberete onaj koji mu je najbliži (važno je dati odgovore na sve tvrdnje).

1. Uvijek osjećam tjeskobu kad se koristim računalom.
2. Poduzimam sve što mogu da se izbjegnem koristiti računalom.
3. Važno mi je znati koristiti se računalom.
4. Moj društveni život ponekad pati zbog mojih aktivnosti na računalu.
5. Teško mi je razumjeti računala.
6. Često osjećam da na računalo potrošim više novca nego što si mogu priuštiti.
7. Računala mi olakšavaju život.
8. Računala me straše.
9. Ponekad je došlo do svađe kod kuće zbog vremena koje provodim na računalu.
10. Često osjetim nalet uzbuđenja dok sam na računalu.
11. Izbjegavam računala što više mogu.
12. Smatram da sam ovisnik o računalu.
13. Osjećam da nemam kontrolu kad se koristim računalom.
14. Kad ljudi razgovaraju o računalima, ne obraćam baš pažnju na to što govore.
15. Računalni žargon zvuči mi glupo.
16. Ne bi mi smetalo da se nikad više ne koristim računalom.
17. Ponekad kasnim na dogovore zbog mojih aktivnosti na računalu.
18. Teško mi je tolerirati računala.
19. Računala me često naljute.
20. Sviđa mi se izazov koji predstavlja učenje kako se koristiti računalom.
21. Radije se koristim olovkom i papirom nego programom poput Worda kad pripremam završnu inačicu nekog pisanog rada.
22. Ne osjećam tjeskobu dok koristim računalo.
23. Ne razumijem zašto ljudi vole računala.
24. Trudim se da vrijeme koje provodim za računalom traje što je duže moguće.
25. Važno mi je da budem vješt/-a na računalu.
26. Ne brine me mogućnost neuspjeha kad radim neki zadatak na računalu.
27. Osjećam se sretno pri pomisli na korištenje računalom.
28. Računala su za mene previše stručna.

29. Kad vidim računalo, osjećam kako me privlači.
30. Nikad nisam pokušao/-la izbjeći koristiti se računalom.
31. Ne bih volio/voljela NE koristiti se računalom više od nekoliko dana.
32. Volim gledati dokumentarne emisije o računalima.
33. Nikad mi nije bilo loše od pomisli da se moram koristiti računalom.
34. Osjećam se moćno kad sam na računalu.
35. Računala mi nisu važna u životu.
36. Malen dio slobodna vremena provodim na računalu.
37. Nikad nisam koristio računalo kao bijeg od društvenog života.
38. Smiren/-a sam kad se koristim računalom.
39. Imam izraženu želju da što više vremena provodim koristeći se računalom.
40. Aktivnosti na računalu ponekad ometaju moje obveze (studiranje, posao i dr.).
41. Rijetko me brine da ću napraviti ogromnu grešku kad sam na računalu.
42. Kad se ne koristim računalom, često se osjećam nemirno.
43. Često se ne naspavam dovoljno zbog računala.
44. Moje misli često postanu zbrkane kad se trebam koristiti računalom.
45. Rijetko razmišljam o računalu kad se ne koristim njime.
46. Pokušao/-la sam smanjiti količinu vremena koje provodim na računalu, ali bezuspješno.
47. Nikad ne preskačem obroke zbog aktivnosti na računalu.
48. Osjećam više tjeskobe povezane s računalima nego što bih trebao/-la.
49. Lako mi je koristiti se računalom.
50. Moja mi tjeskoba povezana s računalima smeta.
51. Ponekad zaboravim važne stvari zbog svojeg zanimanja za računala.
52. Nikad me nije brinulo što se ne znam nositi s računalima.
53. Računala su preznanstvena za mene.

Vaše iskustvo s računalima

Koliko se često trenutačno koristite računalom? Molimo Vas, odaberite jedan odgovor:

- Svaki dan.
- Dvaput ili triput tjedno.
- Najmanje jednom tjedno.
- Rjeđe od jednom tjedno.

Imate li računalo kod kuće?

- Da.
- Ne.

Prilog 4.1. Istraživanje stavova o računalima

(Hewson i sur., 2007; Hewson i Charlton, 2019)

COMPUTING ATTITUDES STUDY

Consent

This questionnaire forms part of a study investigating students' attitudes to computers.

Your participation is voluntary, and you may discontinue at any time, without prejudice. Any data collected will be kept strictly confidential.

I give my informed consent to participate in this study of attitudes to computers. I have read and understand this consent section.

Signature:

Student ID Number: _____

Demographic Details

Age: _____ Gender: _____

Have either of your parents ever studied for a degree? Yes / No (please circle as appropriate)

Which of these ethnic groups do you belong to? Please circle the appropriate group:

1. White - British	6. Other Black Background	11. Other Asian Background	16. Other Ethnic Background
2. White – Irish	7. Asian or Asian British – Indian	12. Mixed – White and Black Caribbean	17. Not Known
3. Other White Background	8. Asian or Asian British – Pakistani	13. Mixed – White and Black African	18. Information Refused
4. Black or Black British Caribbean	9. Asian or Asian British – Bangladeshi	14. Mixed – White and Asian	
5. Black or Black British African	10. Chinese	15. Other Mixed Background	

PLEASE TURN THE PAGE

Computing Attitudes

On this and the following pages you will find a number of statements concerning computers.

Circle which of the five possible responses, given at the side of each question, best sums up your attitude as far as the statement is concerned.

Circle SA if you Strongly Agree with the statement.

Circle A if you just about Agree with the statement.

Circle N if you neither agree nor disagree with the statement.

Circle D if you just about Disagree with the statement.

Circle SD if you Strongly Disagree with the statement.

Just indicate your initial thoughts (do not ponder too long over any of your responses). DO NOT WORRY IF YOUR ANSWERS SEEM INCONSISTENT.

Circle only one choice per statement. Even if none of the choices is an exact reflection of your attitude, please indicate that choice which comes nearest (it is important that all statements are answered).

1. I always feel anxious when using computers SA A N D SD
2. I go out of my way to avoid using computers SA A N D SD
3. It is important for me to be able to use computers..... SA A N D SD
4. My social life has sometimes suffered because of my computing activities SA A N D SD
5. I have problems in understanding computers SA A N D SD
6. I often feel that I spend more money than I can afford on computing SA A N D SD
7. Computers make my life easier SA A N D SD
8. I find computers threatening SA A N D SD
9. Arguments have sometimes arisen at home because of the time I spend on computing activities SA A N D SD
10. I often experience a buzz of excitement while computing SA A N D SD
11. The less I have to do with computers the better SA A N D SD
12. I think that I am addicted to computing SA A N D SD
13. I don't feel in control when I use a computer SA A N D SD
14. I pay little attention when people talk about computers SA A N D SD
15. Computer jargon sounds stupid to me SA A N D SD
16. It would not matter to me if I never used a computer again SA A N D SD
17. I am sometimes late for engagements because of my computing activities SA A N D SD
18. I find computers hard to tolerate SA A N D SD
19. I often get irritated with computers SA A N D SD

PLEASE TURN THE PAGE

20. I like the challenge which learning to use computers presents SA A N D SD
21. I prefer to use a pen and paper, rather than a word processor, when preparing the
final copy of a piece of writing SA A N D SD
22. I do not feel anxious about using computers SA A N D SD
23. I can't understand why people like computers SA A N D SD
24. I try to make my sessions with computers last as long as possible SA A N D SD
25. It is important for me to be good at computing SA A N D SD
26. The possibility of failing at a computing task does not worry me SA A N D SD
27. I feel happy at the thought of using a computer SA A N D SD
28. Computing is too technical for me SA A N D SD
29. When I see a computer I feel drawn towards it SA A N D SD
30. I have never tried to avoid using computers SA A N D SD
31. I would hate to go without using a computer for more than a few days SA A N D SD
32. I like to watch documentaries about computers on television SA A N D SD
33. I have never felt ill at the thought of having to use a computer SA A N D SD
34. I feel a sense of power when I am computing SA A N D SD
35. Computing is unimportant in my life SA A N D SD
36. I spend little of my spare time computing SA A N D SD
37. I have never used computing as an escape from socialising SA A N D SD
38. I am calm when using a computer SA A N D SD
39. I tend to want to spend increasing amounts of time using computers SA A N D SD
40. Computing activities have sometimes interfered with my work SA A N D SD
41. I seldom worry about making a catastrophic mistake when I am computing SA A N D SD
42. When I am not using a computer, I often feel agitated SA A N D SD
43. I often fail to get enough sleep because of my computing activities SA A N D SD
44. My thoughts often become jumbled when I have to use a computer SA A N D SD
45. I rarely think about computing when I am not using a computer SA A N D SD
46. I have made unsuccessful attempts to reduce the time I spend computing SA A N D SD
47. I never miss meals because of my computing activities SA A N D SD
48. I am more anxious about computers than I should be SA A N D SD

PLEASE TURN THE PAGE

49. It is easy for me to use computers SA A N D SD
50. My anxiety about computers bothers me SA A N D SD
51. I sometimes neglect important things because of an interest in computers SA A N D SD
52. I have never been worried about an inability to cope with computers SA A N D SD
53. Computers are too scientific for me SA A N D SD

Finally, please answer the following two questions relating to your computing experience:

i) How often do you currently use a computer? Please circle the appropriate response:

- | | | | |
|----------|--------------------------------|---------------------------|----------------------------|
| 1 | 2 | 3 | 4 |
| Everyday | Two or three times
per week | At least once
per week | Less than once
per week |

ii) Do you have a computer at home? Yes / No (please circle as appropriate)

THANKS FOR YOUR HELP.
PLEASE CHECK THAT YOU HAVE RESPONDED TO ALL THE STATEMENTS AND QUESTIONS

Prilog 4.2. Odgovori sudionika na upitniku o stavovima o računalima

1. dio – čestice 1-27/53

Sudionik	Čestica																										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
1	3	1	5	1	1	1	5	1	1	1	1	1	1	4	3	1	1	1	5	5	1	1	1	1	5	1	3
2	3	1	5	2	2	2	4	1	2	1	2	2	1	1	1	3	1	1	2	5	2	4	1	2	5	4	3
3	4	1	5	5	2	1	4	1	1	3	1	5	1	2	1	1	1	3	4	3	5	4	1	5	5	1	3
4	2	1	5	4	1	1	5	1	3	4	1	2	1	3	2	2	2	2	3	5	1	5	3	3	5	3	3
5	2	2	5	4	1	3	4	2	1	2	3	2	2	2	3	1	1	3	2	3	3	4	3	2	5	2	2
6	2	1	5	2	1	3	4	3	4	2	3	1	2	2	3	2	2	3	2	4	2	4	2	2	5	3	3
7	3	5	5	5	1	1	4	2	1	2	4	1	1	1	3	3	1	2	2	3	4	2	2	1	5	4	1
8	2	1	5	4	1	1	5	1	1	2	1	1	1	3	2	1	1	1	2	4	1	4	1	1	5	4	3
9*	1	3	5	4	1	1	4	1	4	1	1	2	1	3	3	1	4	1	3	1	1	5	1	1	5	2	2
10	2	1	5	5	2	2	5	1	3	3	1	1	1	2	3	2	4	1	2	2	2	5	1	1	5	4	4
11	2	1	5	3	1	2	5	2	1	1	1	2	1	4	4	2	3	1	2	2	3	5	1	1	5	3	3
12	3	4	5	2	3	1	3	1	2	1	4	1	1	5	4	5	2	1	1	1	3	5	3	1	4	3	2
13	2	1	5	1	1	1	4	1	1	1	1	1	1	4	3	3	1	1	1	4	3	5	1	1	4	3	1
14	2	2	5	2	3	1	4	3	3	1	2	2	3	3	4	1	4	1	4	5	1	4	1	5	5	4	2
15	1	1	5	1	1	1	3	1	2	1	1	1	1	4	3	2	1	1	2	3	2	5	3	1	4	3	2
16	1	1	4	3	2	1	4	1	2	2	1	4	3	2	2	1	2	2	3	2	3	4	2	3	4	3	3
17	4	3	5	2	4	2	3	3	1	1	3	1	1	5	5	2	1	1	3	4	3	2	3	1	5	5	2
18	1	1	5	1	2	1	5	1	1	3	1	1	2	2	3	1	1	1	1	4	3	5	2	2	5	4	3
19	2	1	5	5	1	1	5	1	4	2	1	5	3	2	2	1	5	1	5	5	1	2	1	2	5	1	3
20	3	1	5	4	1	3	4	1	2	2	1	1	1	3	3	1	3	3	4	3	2	2	1	2	4	2	3
21	1	1	5	4	1	1	5	1	3	2	1	3	1	1	2	1	1	1	1	5	1	5	1	3	5	5	5
22	1	1	4	1	2	1	3	1	2	2	1	1	1	4	3	2	1	1	2	2	3	5	3	2	4	3	3
23	2	2	5	2	2	1	3	3	1	1	3	1	1	1	2	5	1	2	3	2	5	3	2	1	5	3	1
24	3	4	5	5	3	3	3	3	5	3	4	1	1	4	4	2	2	2	4	3	3	4	4	2	5	3	2
25	2	1	5	3	2	1	5	1	1	4	1	1	1	2	2	4	3	2	2	4	2	4	1	1	5	1	4
26	2	1	5	1	2	1	5	3	1	1	2	1	1	2	3	5	1	1	3	2	5	3	2	3	5	2	1

Sudionik	Čestica																										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
27	4	3	5	4	1	2	5	2	2	1	4	1	1	4	2	4	1	3	2	4	2	4	3	1	5	3	3
28	2	3	5	1	2	1	4	2	2	1	4	1	1	3	3	3	1	2	2	2	3	2	2	1	5	3	2
29	1	1	5	5	1	1	4	1	1	1	3	2	1	5	5	4	4	2	5	1	5	5	3	1	3	1	3
30	2	2	5	4	2	3	5	2	3	2	2	3	2	2	2	2	4	2	2	3	1	4	2	3	5	2	3
31	2	4	5	1	2	3	4	1	4	1	3	2	1	1	3	3	3	3	2	4	4	3	3	1	5	1	3
32	1	1	5	3	1	1	4	1	4	2	2	1	1	1	1	1	1	1	1	4	5	5	1	1	5	3	4
33	2	1	5	3	1	4	5	1	1	1	1	1	1	2	1	3	3	1	1	4	3	1	1	1	5	3	3
34	2	2	3	1	2	1	2	2	1	1	2	1	1	3	3	3	1	2	1	1	3	5	4	1	2	2	1
35	1	1	5	2	2	1	5	1	3	3	1	4	4	3	2	1	4	1	4	3	1	5	1	2	3	2	3
36	3	2	5	4	2	4	5	1	4	3	2	5	3	3	4	2	1	1	4	2	5	5	3	1	4	2	2
37	1	1	5	1	1	1	5	1	1	3	3	1	1	2	2	3	1	1	1	4	1	5	1	1	5	2	2
38	4	4	5	5	1	1	5	1	1	4	1	4	4	1	2	1	5	4	4	2	5	3	1	2	5	5	4
39*	3	3	5	4	4	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	5	3	4	4	5	4	4
40	1	1	1	4	2	3	4	3	1	1	3	2	1	3	3	5	1	1	1	4	2	5	3	1	5	5	4
41	4	3	4	5	2	2	3	2	2	2	4	1	2	2	4	3	2	2	5	2	5	3	3	2	4	2	2
42	2	4	5	5	2	2	4	1	1	2	3	1	1	3	3	2	3	3	3	2	2	4	2	1	4	2	2
43	1	1	5	3	2	4	4	1	4	1	1	2	1	4	3	1	4	1	1	3	4	4	1	1	5	2	2
44	2	2	5	3	2	1	4	1	2	2	1	2	1	3	1	4	2	1	2	2	3	4	2	1	5	1	1
45	1	1	5	1	1	1	5	1	2	2	2	1	1	4	2	4	1	1	1	5	2	5	4	1	5	5	1
46	1	1	5	2	1	1	3	1	1	3	1	1	1	1	2	1	2	3	3	3	3	5	2	1	5	2	3
47	3	2	5	1	2	4	5	1	5	1	2	1	1	5	3	2	4	1	5	5	5	5	2	3	5	1	4
48	1	1	5	2	2	2	4	1	3	1	4	1	1	2	2	1	1	2	2	4	5	5	3	1	4	2	2
49	1	1	5	2	1	1	5	1	2	1	1	1	1	1	1	4	1	1	1	4	3	5	2	1	5	2	2
50	3	1	5	4	3	3	4	4	1	1	3	2	1	2	3	2	1	1	3	2	1	3	3	1	4	2	2
51	1	1	4	1	3	1	5	2	1	2	2	2	2	3	3	2	1	1	1	3	2	4	3	2	3	3	2
52	2	2	5	5	3	2	3	1	3	3	4	2	2	3	2	2	2	2	2	3	4	4	3	4	4	2	4
53	1	3	5	5	4	4	5	2	4	1	4	1	4	1	5	5	1	4	5	2	5	5	2	1	4	4	4
54	1	1	5	3	1	2	5	1	3	3	1	2	2	2	2	1	3	1	1	4	2	5	1	3	5	5	4
55	1	2	5	2	2	1	4	2	1	1	2	2	1	2	2	1	2	1	2	1	2	4	2	2	5	3	3
56	1	1	5	4	2	1	4	1	2	1	1	1	1	2	3	1	3	2	3	1	3	5	2	3	5	3	3

Sudionik	Čestica																											
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	
57	1	1	5	4	1	1	4	1	1	1	1	3	3	3	3	1	1	1	1	2	4	5	2	1	5	2	3	
58	4	1	5	3	5	3	4	3	1	3	1	2	1	2	2	3	3	1	3	3	1	5	2	2	5	2	3	
59	2	1	5	1	1	2	4	1	3	4	2	2	2	1	1	1	1	1	3	4	2	4	1	2	5	1	5	
60	1	1	5	5	2	1	5	1	2	1	2	1	1	1	2	1	3	1	2	1	5	4	1	1	5	3	1	
61	3	2	5	2	2	2	5	2	1	1	1	1	1	3	4	1	4	1	2	4	3	5	2	2	5	3	3	
62	2	1	5	4	3	4	5	4	1	2	1	1	2	4	4	1	4	3	3	3	2	4	3	1	5	2	2	
63	1	1	5	1	1	1	5	1	2	2	1	1	1	1	1	1	1	1	1	5	1	5	1	2	5	3	4	
64	2	2	5	1	1	3	4	1	2	3	2	1	2	3	4	2	3	2	3	3	1	5	2	1	5	2	3	
65	1	1	5	3	1	1	5	1	1	3	1	2	1	2	1	1	2	1	1	3	1	5	1	2	5	1	4	
66	1	1	5	2	1	2	5	1	2	1	1	1	1	3	1	1	1	1	3	3	3	5	1	2	5	3	2	
67	3	1	5	3	2	1	5	1	2	1	1	3	2	2	2	1	3	3	2	1	3	5	2	3	5	3	3	
68	1	1	5	2	1	1	4	1	1	2	2	1	1	4	3	5	1	2	4	2	3	5	2	1	5	4	3	
69	1	1	5	2	1	1	5	1	1	1	3	1	1	4	3	2	1	1	1	5	1	5	1	1	5	4	3	
70	1	1	5	2	4	1	3	1	1	1	2	1	1	3	2	3	1	1	1	3	3	5	1	1	4	3	2	
71	4	2	5	1	4	1	4	2	1	2	2	1	2	4	3	4	1	1	3	3	4	3	1	2	4	2	3	
72	1	1	5	1	2	1	3	1	1	1	2	1	1	2	3	2	1	1	3	3	2	5	3	2	3	3	2	
73	1	1	5	1	2	1	5	1	1	1	1	1	1	5	3	1	2	1	3	1	5	5	2	2	5	5	4	
74	1	1	5	2	1	1	5	1	2	1	3	1	1	3	1	1	1	1	1	5	1	5	3	1	5	5	4	
75	1	1	5	1	1	1	5	1	3	1	3	3	1	1	1	5	1	1	3	3	1	5	1	1	5	3	2	
76	3	1	5	3	3	1	5	1	1	4	1	1	1	3	2	1	1	2	2	5	2	1	2	2	4	1	2	
77	1	2	5	5	1	1	5	1	2	1	3	1	1	3	2	2	2	2	1	2	3	2	5	1	1	5	4	3
78	1	1	5	3	2	3	5	1	3	1	2	4	4	3	2	1	1	1	4	3	1	5	2	1	5	3	3	
79	2	1	5	2	1	1	5	1	3	4	1	1	1	1	3	1	1	1	1	4	4	5	3	1	5	5	3	
80	2	2	5	4	4	2	4	2	2	3	3	1	1	3	3	1	2	3	4	4	1	3	3	2	5	1	2	
81	3	3	5	2	1	1	5	1	2	1	3	5	5	3	3	1	3	1	3	2	1	4	1	1	5	2	2	
82	1	1	5	1	1	1	5	1	1	1	3	3	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3
83	2	1	5	3	2	3	5	1	2	4	2	3	2	3	3	1	2	1	3	2	2	4	1	3	5	3	4	
84	3	2	5	4	2	3	4	2	3	1	3	3	3	4	4	3	4	3	4	3	4	3	2	2	5	2	2	
85	2	1	5	4	2	3	4	1	3	3	2	4	1	3	3	1	2	1	2	4	2	1	2	2	5	4	3	
86	3	1	5	4	2	4	5	3	5	5	3	5	5	2	5	5	5	4	5	5	4	5	2	4	5	1	4	

Sudionik	Čestica																										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
87	3	1	5	5	2	2	5	1	5	5	1	1	1	1	5	1	4	1	1	5	1	5	1	2	5	1	3
88	3	1	5	5	1	3	5	1	4	4	1	4	2	2	1	3	2	2	2	1	1	4	1	3	5	5	2
89	3	1	5	1	1	1	5	1	1	2	3	1	1	1	5	3	1	3	3	5	5	4	3	1	5	5	2
90	4	3	5	5	2	4	4	3	1	1	4	1	3	3	2	2	2	3	3	3	3	3	3	2	5	3	2
91	5	1	5	3	1	2	5	1	1	1	1	2	1	2	3	1	3	2	2	3	3	5	3	3	5	3	2
92	2	1	5	1	2	3	4	1	5	5	3	1	5	3	2	3	1	3	5	3	5	2	2	5	5	4	3
93	1	1	5	1	1	1	4	1	1	3	3	1	1	2	3	4	1	1	1	2	1	5	2	2	4	2	2
94	2	2	4	1	3	1	4	2	1	1	2	2	3	2	4	2	1	2	3	2	1	5	2	2	3	3	3
95	3	3	5	4	3	2	4	4	4	3	3	2	2	4	3	2	3	2	3	2	5	3	3	3	4	1	2
96	4	4	5	2	4	2	4	3	4	2	3	2	1	2	3	3	1	3	3	3	4	3	3	1	4	3	2
97	3	3	5	5	2	2	3	1	1	1	1	1	1	2	4	4	5	4	4	2	3	3	3	1	5	5	1
98	1	1	5	1	1	1	5	1	3	1	4	1	1	2	3	1	1	1	1	3	5	5	2	1	5	3	1
99	4	1	5	3	1	1	5	1	1	1	4	2	1	4	3	2	1	1	1	2	3	4	1	1	5	3	2
100	1	3	5	1	2	1	5	1	1	1	2	1	1	4	2	3	1	1	3	1	4	5	2	1	5	3	2
101	3	2	5	4	2	3	5	3	5	2	3	1	1	1	3	3	1	1	3	4	2	3	1	1	5	5	3
102	3	5	4	1	3	3	1	1	2	3	1	1	1	2	3	3	1	3	3	3	3	2	2	1	4	2	2
103	1	1	5	5	1	4	5	1	1	1	1	1	1	1	1	1	1	1	1	5	1	5	1	1	5	5	5
104	1	1	5	3	1	1	4	1	5	2	1	5	3	2	3	1	1	1	2	2	5	5	1	2	5	3	2
105	1	1	5	1	1	3	4	1	4	3	1	1	2	3	3	1	1	1	2	2	2	5	2	2	4	4	3
106	3	3	5	5	2	4	3	3	3	2	4	1	1	5	3	4	5	2	4	2	5	3	2	1	5	1	1
107	5	5	5	3	2	4	2	3	2	1	5	1	1	5	3	5	1	3	4	1	5	3	3	1	5	1	1
108	1	1	5	3	2	1	4	1	2	3	1	1	1	2	1	4	1	4	2	3	4	5	2	3	5	1	3
109	2	1	5	1	3	1	5	1	3	2	3	1	1	5	1	3	1	1	1	1	3	5	1	1	5	1	3
110	5	3	5	3	4	1	3	4	5	2	3	3	1	3	5	2	1	3	5	1	4	2	2	1	4	1	1
111	1	1	5	1	1	1	5	1	4	3	1	1	1	1	1	1	1	1	1	3	2	5	1	1	5	5	4
112	1	1	5	5	1	1	3	1	5	3	1	1	1	1	1	5	1	1	1	1	1	1	1	1	5	5	3
113	2	1	5	2	3	1	3	1	1	1	3	1	1	1	2	1	1	1	1	3	4	2	1	1	3	3	2
114	2	1	5	3	1	1	5	1	1	3	1	1	1	3	3	1	1	1	2	3	5	5	3	1	5	4	3
115	1	1	5	2	2	1	5	1	2	1	1	2	1	3	2	1	1	2	3	4	2	5	1	3	5	2	2
116	1	3	5	5	3	4	4	1	5	1	2	1	2	5	4	5	4	3	4	5	5	5	3	1	5	3	3

Sudionik	Čestica																										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
117	1	1	5	3	2	2	4	1	3	2	1	1	1	2	2	1	2	1	3	4	3	5	3	2	4	3	2
118	3	3	4	4	2	2	4	2	1	1	3	2	1	3	3	3	3	2	2	3	4	3	3	1	3	2	2
119	4	1	5	5	2	4	5	1	4	5	1	2	2	1	2	1	4	4	3	4	2	4	1	2	4	2	4
120	1	1	5	4	1	2	3	1	1	1	3	1	1	4	3	1	4	2	1	2	5	5	2	1	5	4	1
121	2	1	1	3	3	1	5	2	2	1	1	2	3	4	3	2	1	3	4	5	1	4	1	2	5	2	4
122	1	1	5	3	2	2	3	1	5	3	3	3	1	1	2	1	2	2	2	3	4	4	4	2	3	3	2

2. dio – čestice 28-53/53

Sudionik	Čestica																										
	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53.	
1	1	1	1	1	2	5	1	1	5	3	2	1	1	3	5	1	5	5	1	5	4	5	1	1	5	1	
2	2	1	5	1	4	4	3	3	2	4	3	2	3	4	2	1	2	4	2	5	2	4	1	2	3	1	
3	1	1	5	5	4	5	1	1	1	5	3	3	1	3	4	5	4	2	3	2	2	5	4	5	3	1	
4	2	3	3	2	4	4	3	2	2	5	5	2	2	5	1	1	1	4	2	5	2	5	2	1	3	2	
5	2	1	3	2	2	4	3	2	4	3	2	2	3	4	2	3	1	2	2	4	3	5	2	1	3	1	
6	2	3	3	3	2	5	4	2	3	3	5	2	3	4	2	2	2	4	2	4	2	5	2	2	3	2	
7	1	1	2	2	1	3	3	4	4	1	5	1	2	1	1	2	1	1	2	1	1	5	1	1	5	1	
8	1	2	5	4	3	4	1	1	3	5	4	2	2	4	1	2	1	1	2	5	1	4	1	1	1	1	
9*	1	1	3	4	3	4	3	2	3	5	4	2	4	3	1	4	1	4	2	5	1	5	1	2	1	1	
10	2	2	4	3	1	4	3	2	2	2	3	1	3	4	1	1	2	3	2	5	1	4	1	3	3	2	
11	2	1	4	2	2	1	1	3	1	2	4	1	5	2	2	5	2	5	3	4	1	4	2	2	2	2	
12	2	1	4	2	1	1	1	4	5	5	1	1	1	3	1	1	1	5	1	5	1	4	1	2	1	2	
13	2	1	5	3	1	5	1	4	5	5	3	1	1	3	2	1	1	5	1	5	1	4	1	1	4	2	
14	4	1	4	5	3	1	2	1	1	2	3	2	4	3	1	5	3	2	1	5	3	2	1	4	2	4	
15	1	1	5	4	1	5	1	5	4	4	5	1	2	1	1	1	1	1	4	1	5	1	1	1	1	1	
16	2	2	4	4	2	4	3	3	2	3	2	2	4	3	2	4	2	2	3	4	1	4	1	3	4	2	
17	2	1	2	2	1	2	1	5	5	5	3	1	1	5	1	1	4	5	1	5	2	3	2	1	2	2	

Sudionik	Čestica																									
	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53.
18	2	3	4	3	2	4	3	1	2	3	4	3	2	4	2	1	1	3	1	5	1	5	3	1	3	3
19	1	2	2	3	4	4	4	1	2	1	2	2	1	5	1	3	1	5	5	3	2	5	2	5	1	1
20	2	2	3	3	1	5	4	3	2	4	2	2	2	2	1	3	3	3	3	2	2	3	1	2	2	3
21	1	1	1	4	3	5	3	1	1	5	3	1	3	3	1	1	1	4	1	3	1	5	1	1	1	1
22	2	1	4	3	2	5	3	3	4	5	4	1	2	4	1	2	2	4	1	5	2	4	1	2	3	2
23	2	1	1	1	1	3	1	4	5	5	3	1	1	3	3	2	2	5	1	5	4	3	2	1	3	3
24	2	1	2	1	1	2	1	4	1	5	3	1	3	3	1	3	4	4	2	5	4	3	4	2	2	2
25	2	2	5	1	2	5	2	4	3	2	4	2	4	4	2	2	2	4	2	5	2	5	1	1	2	2
26	1	1	3	3	1	5	1	4	3	4	4	1	1	3	1	1	1	5	1	5	3	3	3	1	3	2
27	1	1	3	2	3	4	3	4	5	5	4	2	1	5	1	1	1	5	1	5	2	5	1	1	3	1
28	2	1	1	2	1	3	1	4	4	5	3	2	3	4	2	2	3	4	2	4	2	4	2	2	3	2
29	1	1	3	1	1	3	1	3	5	5	3	1	4	5	1	4	4	5	1	5	3	4	3	1	1	3
30	1	3	4	2	3	4	3	2	3	3	4	2	2	4	2	4	2	3	3	5	1	4	1	2	2	2
31	4	1	4	2	3	4	2	2	5	4	3	1	5	3	1	5	3	5	4	5	3	3	3	2	1	3
32	1	3	4	4	4	5	3	1	4	5	4	2	2	4	1	1	1	5	1	5	1	4	1	1	4	1
33	1	1	1	1	1	5	3	3	3	3	3	4	3	4	2	4	2	4	2	2	3	3	3	4	2	4
34	2	1	3	1	1	2	1	5	4	4	2	1	1	5	1	1	2	5	1	4	2	3	1	1	4	2
35	2	4	2	5	2	3	3	1	1	2	4	3	3	5	1	2	2	4	3	4	2	4	2	2	2	2
36	3	2	1	3	1	4	3	2	2	2	3	2	5	3	2	5	2	3	5	2	3	4	3	2	3	2
37	2	1	2	1	1	2	2	2	1	5	5	1	1	4	1	1	1	5	1	5	1	5	1	1	1	1
38	2	4	5	5	5	3	4	1	1	4	3	2	5	4	3	5	2	5	1	5	3	5	3	5	1	2
39*	4	3	4	4	4	2	4	4	3	4	3	4	5	4	4	5	4	2	5	5	5	5	3	4	3	4
40	2	1	5	1	2	5	1	4	1	2	3	1	1	2	2	1	3	5	3	5	1	5	1	1	5	1
41	2	2	4	1	1	4	2	3	2	2	4	1	2	2	1	2	2	4	2	5	2	4	2	1	4	2
42	3	1	3	2	1	2	3	3	2	4	3	2	2	2	3	2	3	5	2	5	2	4	2	1	2	2
43	3	1	3	4	1	2	3	3	4	4	3	1	4	3	1	1	2	5	3	5	2	4	2	1	3	3
44	1	1	2	2	1	3	1	4	4	2	2	1	2	4	2	2	2	4	2	5	2	3	4	2	2	1
45	1	3	3	1	3	5	2	5	3	1	5	1	1	5	1	2	1	5	1	5	1	5	1	1	1	1
46	1	1	5	2	4	5	3	5	3	5	4	1	3	3	1	2	1	5	1	1	1	5	1	1	1	1
47	1	1	4	3	1	2	4	1	1	1	4	2	3	5	2	5	1	5	1	1	1	5	1	1	1	1

Sudionik	Čestica																									
	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53.
48	3	1	2	2	3	3	3	4	4	4	3	1	4	4	2	2	2	5	2	5	2	5	1	2	4	2
49	1	1	5	1	5	5	2	3	1	2	5	2	2	5	1	1	1	5	1	5	1	5	1	1	3	1
50	2	1	3	1	1	3	2	4	4	3	3	1	1	3	1	1	1	4	2	5	3	3	1	1	2	3
51	2	1	4	4	1	3	1	4	3	4	4	2	4	4	3	4	1	5	2	4	2	4	2	1	4	4
52	2	1	3	1	1	2	1	5	4	4	5	1	4	1	1	1	1	3	3	4	2	4	1	3	3	3
53	2	1	1	1	4	1	3	5	5	2	5	1	1	2	1	1	3	5	2	4	5	4	1	1	4	1
54	1	3	5	1	2	5	4	1	3	4	4	3	2	3	2	1	2	4	3	5	2	5	1	3	5	1
55	2	2	4	4	1	4	2	4	2	4	3	2	5	2	3	2	2	4	2	5	2	5	2	2	2	2
56	2	1	5	5	5	4	1	2	4	5	5	3	3	3	1	1	1	3	1	4	1	4	1	1	2	2
57	1	1	1	5	1	5	1	1	1	3	4	2	3	5	2	3	2	5	1	5	3	4	3	1	1	1
58	3	1	2	3	3	3	2	2	2	3	4	2	3	3	2	2	1	5	2	2	1	5	1	2	1	1
59	1	5	3	3	3	5	5	1	1	1	3	4	3	5	4	5	1	3	5	4	3	5	2	3	1	1
60	1	1	3	1	1	2	1	1	4	5	3	1	5	2	1	1	1	5	5	2	2	3	1	1	1	1
61	2	3	3	2	3	4	2	3	3	3	4	1	4	4	1	4	1	2	1	5	2	4	1	1	3	2
62	3	3	1	4	3	2	4	2	3	4	3	3	2	1	2	1	2	5	2	5	2	4	3	3	2	2
63	1	1	5	3	4	4	2	1	3	4	5	4	3	4	1	1	1	3	1	2	1	5	1	2	1	1
64	2	3	3	3	4	3	4	2	3	4	3	3	2	3	3	3	1	5	3	5	1	5	4	1	2	1
65	1	3	5	4	1	5	1	1	2	3	5	3	3	5	1	3	2	4	1	4	1	5	1	2	1	1
66	1	1	5	4	3	5	3	1	3	3	3	2	3	2	1	3	1	4	3	4	1	5	2	3	4	1
67	1	3	4	4	1	5	3	1	2	2	3	3	4	4	1	3	1	4	4	5	3	5	1	3	1	2
68	1	2	4	1	1	4	2	4	3	5	3	1	3	4	1	4	4	5	1	5	3	5	1	1	1	1
69	1	1	5	3	3	5	3	3	4	5	5	1	1	5	1	1	1	5	1	2	1	5	1	3	1	1
70	3	1	2	1	2	5	1	3	3	5	3	1	1	1	1	1	1	2	1	1	5	1	1	5	2	
71	3	1	4	1	2	2	3	3	5	5	3	2	1	2	1	1	3	5	1	5	2	3	3	1	3	1
72	1	1	1	5	1	1	1	3	3	4	3	1	2	3	1	1	1	1	1	5	1	3	1	1	3	4
73	1	1	5	5	1	5	5	1	1	1	5	2	3	3	1	1	1	5	1	5	1	5	1	1	5	1
74	1	1	4	5	2	5	3	2	3	3	5	2	1	5	1	1	1	5	1	5	1	5	1	1	3	1
75	2	1	3	1	1	5	1	3	5	5	3	1	3	3	3	1	1	1	3	3	1	5	1	1	1	1
76	5	2	1	5	2	5	5	1	3	3	5	5	2	3	1	1	1	1	1	5	1	5	1	1	3	1
77	1	2	2	3	1	5	3	3	4	4	3	2	1	5	1	1	1	5	1	3	1	5	2	1	4	1

Sudionik	Čestica																									
	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53.
78	2	1	4	2	1	2	3	2	4	2	4	1	4	3	1	2	1	5	2	5	1	4	1	1	3	2
79	1	1	4	3	4	3	4	2	3	5	4	2	2	2	1	1	1	4	1	5	1	5	1	1	1	1
80	3	1	3	2	1	2	2	1	3	3	2	2	4	1	3	4	4	4	3	2	2	4	3	2	3	2
81	2	1	1	1	3	4	1	3	2	1	3	1	5	1	3	5	1	3	4	3	1	5	1	5	1	1
82	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
83	2	1	3	2	1	4	2	1	1	2	4	3	5	4	2	3	2	2	4	5	1	5	1	2	1	1
84	3	2	2	2	4	3	2	3	2	2	3	2	2	2	3	3	3	4	4	5	4	4	4	3	2	3
85	2	2	2	2	4	5	3	1	2	3	3	2	4	3	1	1	1	3	3	5	1	4	1	2	3	1
86	1	4	1	1	1	5	3	1	1	1	4	2	5	1	4	5	5	5	5	1	4	4	2	5	3	3
87	2	4	5	5	1	5	5	1	2	5	4	3	5	5	1	3	5	1	2	2	2	5	1	1	3	1
88	1	4	3	1	3	4	4	1	1	2	3	2	5	5	1	5	5	3	2	5	4	4	4	1	1	1
89	4	1	1	3	3	4	3	4	3	5	5	1	1	5	2	5	3	5	1	1	1	5	5	1	1	3
90	4	2	2	2	2	3	2	2	3	4	3	2	4	2	2	3	3	5	2	5	2	3	4	2	1	3
91	2	5	3	2	1	2	3	3	2	3	4	1	3	3	1	5	5	3	3	3	3	5	2	2	2	2
92	4	1	5	3	2	3	4	3	5	5	3	3	5	4	3	5	5	3	1	5	3	4	3	3	2	2
93	2	1	5	1	1	5	1	2	4	4	3	1	3	4	1	1	1	5	1	5	1	4	1	1	5	1
94	3	1	5	3	1	2	4	2	2	2	3	3	2	5	3	2	2	3	2	5	3	4	3	2	2	2
95	3	2	3	3	2	1	2	2	2	4	3	3	3	3	4	3	4	3	4	2	5	3	4	2	3	3
96	3	1	2	3	1	3	2	3	4	4	3	2	3	2	2	1	2	4	4	3	4	3	4	1	3	2
97	3	1	2	1	1	2	1	5	5	5	2	1	1	2	1	1	1	1	1	1	3	3	3	1	1	2
98	2	1	3	3	1	5	1	1	5	5	5	1	5	5	1	1	1	1	1	5	1	5	1	1	5	1
99	1	1	5	3	1	5	1	4	4	3	5	1	4	5	1	1	1	2	3	1	3	5	1	2	5	1
100	2	1	5	2	1	5	1	4	5	5	4	1	3	3	1	1	1	5	1	5	1	4	1	1	4	2
101	1	3	5	3	2	3	3	2	2	5	3	3	3	1	1	5	3	5	1	2	5	4	5	3	2	1
102	3	1	1	1	1	1	3	2	2	3	2	1	2	2	3	4	3	5	1	5	3	3	2	1	2	2
103	1	1	5	5	1	5	1	3	1	5	5	1	5	5	1	1	1	5	1	5	1	5	1	1	3	1
104	1	2	4	2	1	1	1	1	1	5	4	1	3	2	2	3	1	3	5	5	1	4	1	1	1	1
105	2	2	3	3	4	4	3	2	3	4	3	2	2	4	1	1	2	4	1	5	2	5	1	1	3	2
106	2	1	3	2	2	3	4	3	3	5	1	1	4	3	1	4	1	5	3	3	1	4	3	1	4	2
107	2	1	2	1	2	3	3	3	3	3	3	1	5	3	3	5	4	5	3	5	3	4	3	4	3	2

Sudionik	Čestica																									
	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53.
108	1	3	3	1	1	4	3	1	1	2	3	2	2	3	1	1	1	4	1	5	1	4	1	1	4	1
109	5	1	5	1	1	5	3	4	4	1	5	2	5	3	1	1	1	5	1	5	1	5	1	3	3	1
110	3	1	1	1	1	1	1	5	5	5	1	1	1	1	1	1	3	5	1	5	4	1	4	1	2	5
111	1	2	2	5	5	5	1	3	3	5	5	1	5	5	1	1	1	4	1	2	1	5	1	4	5	1
112	1	1	1	1	1	3	1	3	3	3	3	1	1	1	1	1	1	1	1	1	1	5	1	1	1	1
113	2	1	1	3	2	3	3	2	4	5	3	1	1	4	3	1	2	3	3	5	1	3	3	1	3	3
114	1	3	5	4	2	5	4	5	5	5	3	1	1	3	2	3	1	1	1	3	1	5	1	1	1	1
115	2	2	2	4	2	2	2	4	3	2	3	2	3	2	1	2	2	4	2	3	1	5	1	2	2	1
116	2	1	1	1	1	1	3	5	5	5	3	1	2	5	1	1	1	1	1	5	4	3	3	1	2	1
117	1	1	3	2	1	2	1	3	5	5	3	2	3	2	1	3	2	5	2	4	1	5	1	1	1	1
118	2	1	1	1	1	2	1	3	5	1	5	1	2	2	1	1	3	5	2	5	3	3	2	2	3	1
119	3	3	3	5	3	4	3	1	1	5	4	3	4	3	2	5	3	2	1	1	3	4	4	3	3	2
120	2	1	2	2	1	5	1	3	5	5	4	1	3	5	1	1	1	5	2	3	2	5	2	2	1	1
121	2	1	3	5	1	4	3	4	3	2	3	3	3	3	3	3	2	4	3	5	3	4	3	2	3	2
122	2	2	2	2	1	1	1	2	4	5	2	5	3	3	4	4	2	4	4	4	4	3	3	2	3	3

* Podatci isključeni iz daljnje obrade jer ovi sudionici na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Prilog 4.3. Matrica komponentata latentne strukture Upitnika o stavovima o računalima (N = 120)

Čestica	1	2	h ²
1. Uvijek osjećam tjeskobu kad se koristim računalom. ¹	,551	,237	,360
2. Poduzimam sve što mogu da se izbjegnem koristiti računalom. ¹	,642	-,129	,429
3. Važno mi je znati koristiti se računalom.	-,091	,103	,019
4. Moj društveni život ponekad pati zbog mojih aktivnosti na računalu.	,224	,295	,137
5. Teško mi je razumjeti računala. ¹	,566	-,078	,327
6. Često osjećam da na računalo potrošim više novca nego što si mogu priuštiti.	,396	,329	,265
7.* Računala mi olakšavaju život. ¹	,520	-,368	,405
8. Računala me straše. ¹	,610	-,021	,372
9. Ponekad je došlo do svađe kod kuće zbog vremena koje provodim na računalu. ²	,163	,381	,172
10. Često osjetim nalet uzbuđenja dok sam na računalu. ²	-,096	,551	,313
11. Izbjegavam računala što više mogu. ¹	,507	-,264	,327
12. Smatram da sam ovisnik o računalu. ²	,085	,597	,364
13. Osjećam da nemam kontrolu kad se koristim računalom. ²	,195	,564	,356
14. Kad ljudi razgovaraju o računalima, ne obraćam baš pažnju na to što govore.	,243	-,247	,120
15. Računalni žargon zvuči mi glupo. ¹	,550	-,065	,306
16. Ne bi mi smetalo da se nikad više ne koristim računalom.	,378	-,350	,265
17. Ponekad kasnim na dogovore zbog mojih aktivnosti na računalu. ²	,227	,467	,269
18. Teško mi je tolerirati računala. ¹	,538	,286	,371
19. Računala me često naljute. ¹	,573	,352	,452
20. Sviđa mi se izazov koji predstavlja učenje kako se koristiti računalom.	-,298	,212	,134
21. Radije se koristim olovkom i papirom nego programom poput Worda kad pripremam završnu inačicu nekog pisanog rada. ¹	,400	-,123	,175
22.* Ne osjećam tjeskobu dok koristim računalu. ¹	,482	,091	,240
23. Ne razumijem zašto ljudi vole računala. ¹	,440	-,300	,283
24. Trudim se da vrijeme koje provodim za računalom traje što je duže moguće. ²	,006	,575	,330
25.* Važno mi je da budem vješt/-a na računalu. ¹	,309	-,201	,136
26.* Ne brine me mogućnost neuspjeha kad radim neki zadatak na računalu. ¹	,314	,184	,132
27.* Osjećam se sretno pri pomisli na korištenje računalom.	,453	-,449	,407
28. Računala su za mene previše stručna. ¹	,473	,058	,227
29. Kad vidim računalo, osjećam kako me privlači. ²	-,169	,609	,400
30.* Nikad nisam pokušao/-la izbjeći koristiti se računalom. ¹	,462	-,021	,214
31.* Ne bih volio/voljela NE koristiti se računalom više od nekoliko dana.	,383	-,317	,247
32. Volim gledati dokumentarne emisije o računalima.	-,282	,309	,176
33.* Nikad mi nije bilo loše od pomisli da se moram koristiti računalom. ¹	,619	-,073	,388
34. Osjećam se moćno kad sam na računalu. ²	-,123	,493	,258
35.* Računala mi nisu važna u životu. ²	-,314	,617	,479
36.* Malen dio slobodna vremena provodim na računalu. ²	-,220	,617	,429
37.* Nikad nisam koristio računalo kao bijeg od društvenog života. ²	-,017	,343	,118
38.* Smiren/-a sam kad se koristim računalom. ¹	,525	,045	,277
39. Imam izraženu želju da što više vremena provodim koristeći se računalom. ²	-,145	,582	,360
40. Aktivnosti na računalu ponekad ometaju moje obveze (studiranje, posao i dr.). ²	-,013	,460	,211

41.* Rijetko me brine da ću napraviti ogromnu grešku kad sam na računalu. ¹	,404	,003	,163
42. Kad se ne koristim računalom, često se osjećam nemirno. ²	,299	,404	,253
43. Često se ne naspavam dovoljno zbog računala. ²	,291	,649	,505
44. Moje misli često postanu zbrkane kad se trebam koristiti računalom. ¹	,516	,352	,390
45. Rijetko razmišljam o računalu kad se ne koristim njime.	,131	-,140	,037
46. Pokušao/-la sam smanjiti količinu vremena koje provodim na računalu, ali bezuspješno. ²	,235	,501	,306
47.* Nikad ne preskačem obroke zbog aktivnosti na računalu. ²	,010	-,209	,044
48. Osjećam više tjeskobe povezane s računalima nego što bih trebao/-la. ¹	,652	,273	,500
49.* Lako mi je koristiti se računalom. ¹	,699	-,121	,504
50. Moja mi tjeskoba povezana s računalima smeta. ¹	,578	,289	,418
51. Ponekad zaboravim važne stvari zbog svojeg zanimanja za računala. ²	,111	,624	,402
52. Nikad me nije brinulo što se ne znam nositi s računalima. ²	-,015	-,198	,039
53. Računala su preznanstvena za mene. ¹	,565	,048	,321
karakteristični korijeni	8,269	6,866	
% objašnjene varijance	15,602	12,954	28,556

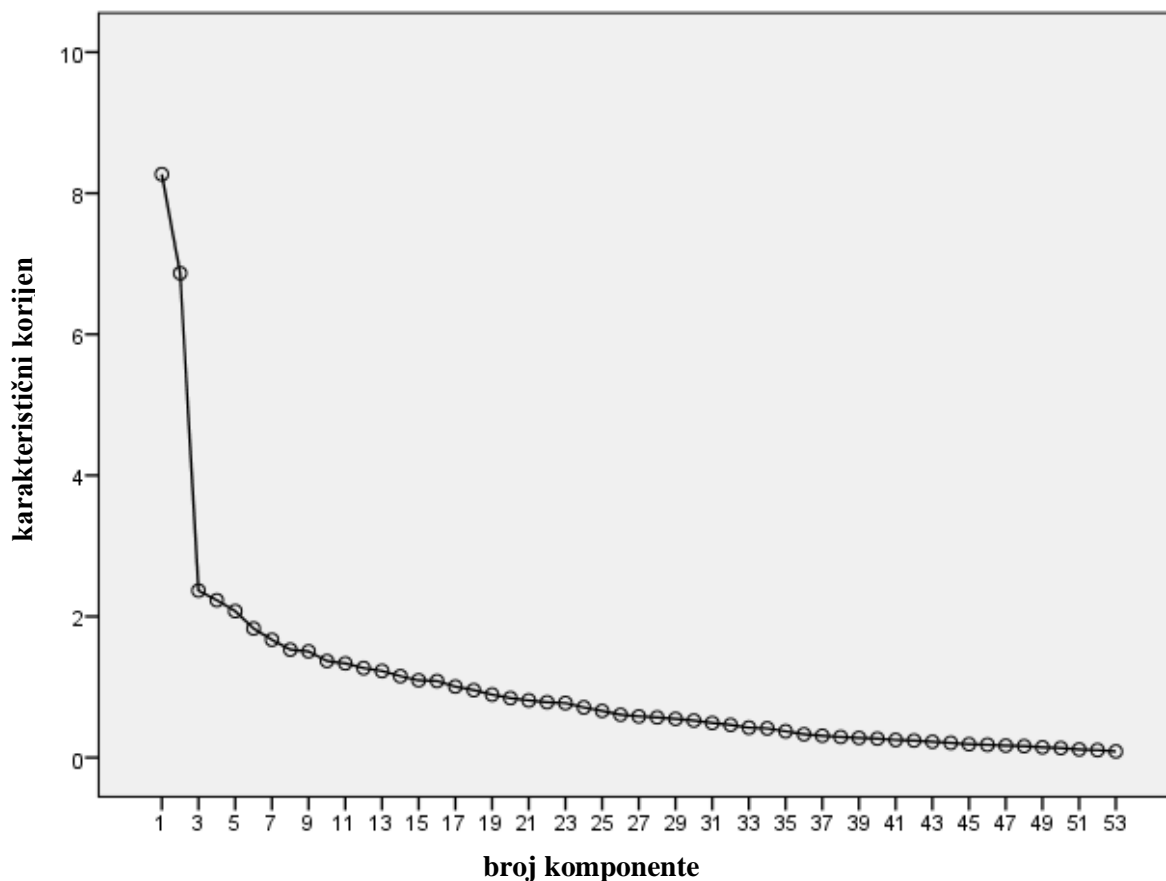
abe Čestice isključene iz daljnje obrade.

* Rekodirane čestice.

¹ Čestice koje pripadaju 1. faktoru.

² Čestice koje pripadaju 2. faktoru.

Prilog 4.4. Scree plot prikaz latentne strukture Upitnika o stavovima o računalima (N = 120)



PRILOG 5. Upitnik o razini ispitne anksioznosti

U ovom dijelu upitnika nalazi se 17 tvrdnja vezanih za ispite.

Odaberite koliko se slažete s navedenom tvrdnjom – od 1 (*Nikako se ne slažem s tvrdnjom.*) do 5 (*Sasvim se slažem s tvrdnjom.*).

Odaberite odgovor na koji ste prvo pomislili (nemojte predugo razmišljati o svojim odgovorima). NE BRINITE AKO VAM SE ODGOVORI BUDU ČINILI NEDOSLJEDNIMA.

Odaberite samo jedan odgovor za svaku tvrdnju. Čak i ako ni jedan odgovor nije točna slika Vašeg stava, molimo Vas da odaberete onaj koji mu je najbliži (važno je dati odgovore na sve tvrdnje).

STAVITE „X“ U ODGOVARAJUĆU KUĆICU.

	1 Nikako se ne slažem	2 Ne slažem se	3 Niti se slažem, niti ne slažem	4 Slažem se	5 Sasvim se slažem
1. Što sam bliže ispitu, teže mi je usredotočiti se na gradivo.					
2. Kad učim za ispit, brinem se da se neću sjetiti gradiva na samom ispitu.					
3. Za vrijeme važnih ispita smatram da mi ide loše i da ću pasti.					
4. Gubim koncentraciju na važnim ispitima i ne mogu se sjetiti gradiva koje sam znao prije ispita.					
5. Sjetim se odgovora na pitanja iz ispita tek nakon što završi.					
6. Toliko se brinem prije važnog ispita da sam toliko izmoren(a) da ne mogu dati sve od sebe na ispitu.					
7. Nisam dobre volje niti sam sav svoj / sva svoja dok polažem važne ispite.					
8. Primijetim da mi nekad misli odlutaju dok polažem važne ispite.					
9. Nakon ispita brinem se o tomu jesam li dao/dala sve od sebe.					
10. Borim se s pisanim zadacima ili ih izbjegavam jer smatram da štogod napravio/napravila, to neće biti dovoljno dobro. Želim da to bude savršeno.					

Prilog 5.1. Ljestvica ispitne anksioznosti Westside

(Driscoll, 2007 u Aliakbari i Gheitasi, 2017)

Westside Test Anxiety Scale

Rate how true each of the following is of you, from extremely or always true, to not at all or never true. Use the following 5 point scale.

- | 5 | 4 | 3 | 2 | 1 |
|-----------------------------|---------------------------|---------------------------------|----------------------------|-----------------------------|
| extremely or
always true | highly or
usually true | moderately or
sometimes true | slightly or
seldom true | not at all or
never true |
- ___ 1) The closer I am to a major exam, the harder it is for me to concentrate on the material.
- ___ 2) When I study, I worry that I will not remember the material on the exam.
- ___ 3) During important exams, I think that I am doing awful or that I may fail.
- ___ 4) I lose focus on important exams, and I cannot remember material that I knew before the exam.
- ___ 5) I finally remember the answer to exam questions after the exam is already over.
- ___ 6) I worry so much before a major exam that I am too worn out to do my best on the exam.
- ___ 7) I feel out of sorts or not really myself when I take important exams.
- ___ 8) I find that my mind sometimes wanders when I am taking important exams.
- ___ 9) After an exam, I worry about whether I did well enough.
- ___ 10) I struggle with writing assignments, or avoid them as long as I can. I feel that whatever I do will not be good enough.
- _____ Sum of the 10 questions
- _____ Divide the sum by 10. This is your Test Anxiety score.

What does your test anxiety score mean?

- 1.0—1.9 Comfortably low test anxiety
- 2.0—2.5 Normal or average test anxiety
- 2.5—2.9 High normal test anxiety
- 3.0—3.4 Moderately high (some items rated 4=high)
- 3.5—3.9 High test anxiety (half or more of the items rated 4=high)
- 4.0—5.0 Extremely high anxiety (items rated 4=high and 5=extreme)

© 2004 by Richard Driscoll, Ph.D.

Prilog 5.2. Odgovori sudionika na upitniku o razini ispitne anksioznosti

Sudionik	Čestica									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1	5	3	2	4	4	1	4	4	2	5
2	4	3	4	4	3	3	4	2	2	5
3	5	4	3	3	2	2	3	4	5	4
4	5	3	3	4	2	3	3	3	2	5
5	3	4	4	4	2	3	3	4	4	5
6	5	5	4	4	5	5	5	5	2	5
7	5	5	4	5	3	3	5	4	3	5
8	4	4	5	3	4	4	4	5	2	4
9*	4	4	4	4	4	4	4	4	2	5
10	2	3	2	2	2	1	3	4	4	5
11	4	4	4	3	3	3	4	4	4	4
12	4	4	3	4	2	3	3	2	2	5
13	4	4	2	2	2	5	4	4	3	4
14	2	3	1	2	4	4	4	1	5	5
15	4	4	4	3	3	4	4	4	4	5
16	2	3	4	3	4	4	4	5	3	5
17	4	4	5	3	2	5	5	4	3	4
18	4	4	4	4	4	5	5	5	2	5
19	4	4	4	4	4	2	3	4	5	5
20	3	4	4	4	3	4	3	4	4	5
21	4	4	4	4	4	4	3	4	3	5
22	5	4	4	5	3	4	3	3	2	4
23	2	4	3	1	3	3	4	3	4	5
24	3	2	3	2	2	4	3	3	4	5
25	4	4	4	2	2	4	4	2	3	4
26	3	3	2	2	2	4	4	3	5	5
27	4	4	4	3	3	4	4	4	3	4
28	4	4	3	3	5	4	4	5	3	4
29	4	2	5	3	3	5	4	5	5	5
30	3	4	4	4	2	2	4	3	4	5
31	4	4	3	4	4	5	4	5	5	5
32	5	5	4	4	4	4	5	5	2	5
33	3	3	3	2	1	1	2	4	5	5
34	3	2	2	2	3	2	2	3	4	5
35	4	4	4	4	3	3	3	4	2	4
36	4	4	5	5	4	3	4	5	4	5
37	3	4	4	4	4	4	3	4	2	5
38	5	5	5	3	2	3	4	4	3	5
39*	5	5	5	5	4	3	2	4	3	5
40	2	4	5	3	1	1	4	5	2	4
41	4	4	2	2	2	4	2	4	4	4
42	2	1	3	3	2	1	3	2	4	5
43	2	3	1	1	3	4	4	3	3	5
44	4	2	4	2	2	2	4	2	3	5
45	2	1	1	2	2	5	1	4	5	5

Sudionik	Čestica									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
46	5	5	5	5	4	4	3	4	2	5
47	2	4	4	3	4	1	2	5	5	5
48	4	2	2	2	2	1	4	1	3	5
49	5	4	4	4	4	3	3	4	3	5
50	4	4	3	4	3	4	4	2	5	5
51	4	3	4	4	3	2	3	4	5	4
52	5	4	5	5	3	4	4	5	1	5
53	2	3	4	3	1	2	3	4	3	5
54	4	3	4	4	4	2	3	5	3	5
55	5	4	4	3	2	2	2	4	2	5
56	5	4	3	4	4	4	3	5	2	5
57	5	4	5	5	5	3	5	5	4	5
58	3	4	3	4	3	2	4	4	3	5
59	2	2	2	1	2	1	3	4	3	5
60	4	3	1	1	2	3	5	3	4	2
61	3	4	4	3	3	3	4	5	2	4
62	4	4	2	3	4	3	4	5	3	5
63	4	5	4	4	4	5	4	3	2	5
64	2	4	3	3	2	4	4	5	3	5
65	5	3	5	5	5	5	5	3	4	5
66	3	4	4	4	4	5	3	4	2	5
67	3	3	4	4	3	4	4	4	5	5
68	4	4	2	3	4	4	5	4	3	5
69	3	4	5	4	4	2	3	4	3	5
70	4	4	4	3	5	4	4	3	2	4
71	3	4	3	4	4	5	5	4	3	5
72	4	4	4	3	3	4	4	4	3	5
73	5	2	5	4	1	3	5	5	4	5
74	3	3	4	4	3	2	4	4	2	4
75	4	4	4	4	2	3	2	3	1	5
76	3	2	1	3	2	5	2	4	5	5
77	4	2	3	2	2	2	4	3	3	5
78	4	2	2	3	1	1	1	2	4	5
79	4	3	2	4	3	2	4	4	4	5
80	2	4	3	3	3	4	2	4	4	4
81	3	1	4	1	5	5	5	5	4	5
82	1	3	3	3	3	3	3	3	3	3
83	2	4	4	2	2	3	2	3	1	4
84	4	4	4	4	4	2	4	4	5	5
85	4	3	4	4	5	5	3	4	5	5
86	5	5	3	5	5	4	4	5	4	5
87	3	2	3	1	1	1	1	4	4	5
88	3	4	4	5	3	4	2	3	3	5
89	4	4	5	5	5	5	5	5	2	5
90	5	4	4	5	3	2	4	5	2	5
91	5	5	5	5	3	4	3	3	4	5
92	3	4	3	3	4	5	4	4	4	5
93	4	2	3	4	2	1	2	4	3	5

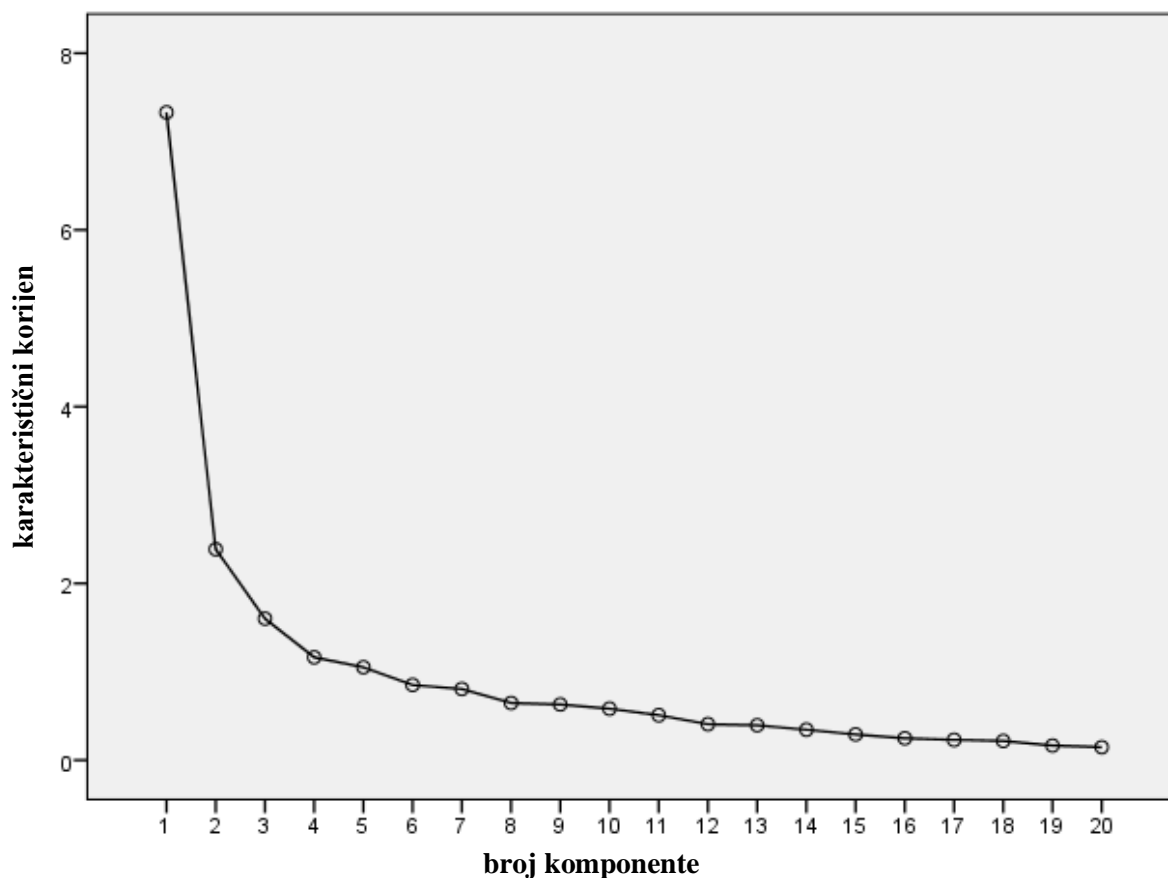
Sudionik	Čestica									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
94	4	3	4	3	4	3	2	1	3	5
95	4	3	2	1	3	4	4	4	5	5
96	4	2	5	5	4	3	4	4	4	5
97	4	4	4	4	4	2	2	1	3	5
98	4	2	4	3	1	2	3	4	4	5
99	3	4	5	5	3	1	1	4	3	5
100	4	4	2	4	2	4	2	3	2	5
101	5	5	5	5	5	4	5	5	5	3
102	4	3	4	4	2	3	2	4	1	4
103	3	3	4	4	3	1	4	3	2	5
104	2	3	3	2	3	3	3	4	2	5
105	4	4	4	2	2	4	4	3	2	4
106	5	3	4	5	2	2	3	4	3	5
107	5	3	4	4	3	3	3	4	5	5
108	5	3	5	4	3	2	3	4	1	5
109	2	4	3	3	2	5	3	3	5	5
110	3	3	1	3	3	4	4	2	5	5
111	5	5	4	5	2	5	5	4	3	5
112	2	1	1	3	1	3	2	1	1	5
113	4	3	3	4	3	3	3	4	3	5
114	5	2	4	3	2	2	2	4	4	5
115	4	4	4	5	5	5	5	4	5	5
116	4	1	5	3	1	4	5	5	4	5
117	3	4	4	4	3	4	4	4	1	5
118	4	3	4	3	2	3	5	4	4	5
119	4	3	4	3	3	4	4	5	1	5
120	3	4	3	2	1	2	3	3	1	5
121	3	3	3	4	5	3	3	4	3	5
122	3	3	2	3	3	2	3	4	4	5

* Podatci isključeni iz daljnje obrade jer ovi sudionici na svim varijablama pisanja u sklopu e-ispita ostvaruju rezultat 0. Razumno je za pretpostaviti da je riječ o nemotiviranim sudionicima, pa bi uključivanje njihovih rezultata dovelo do nerealna odstupanja rezultata od stvarnih vrijednosti.

Prilog 5.3. Matrica komponenata latentne strukture Upitnika o razini ispitne anksioznosti (N = 120)

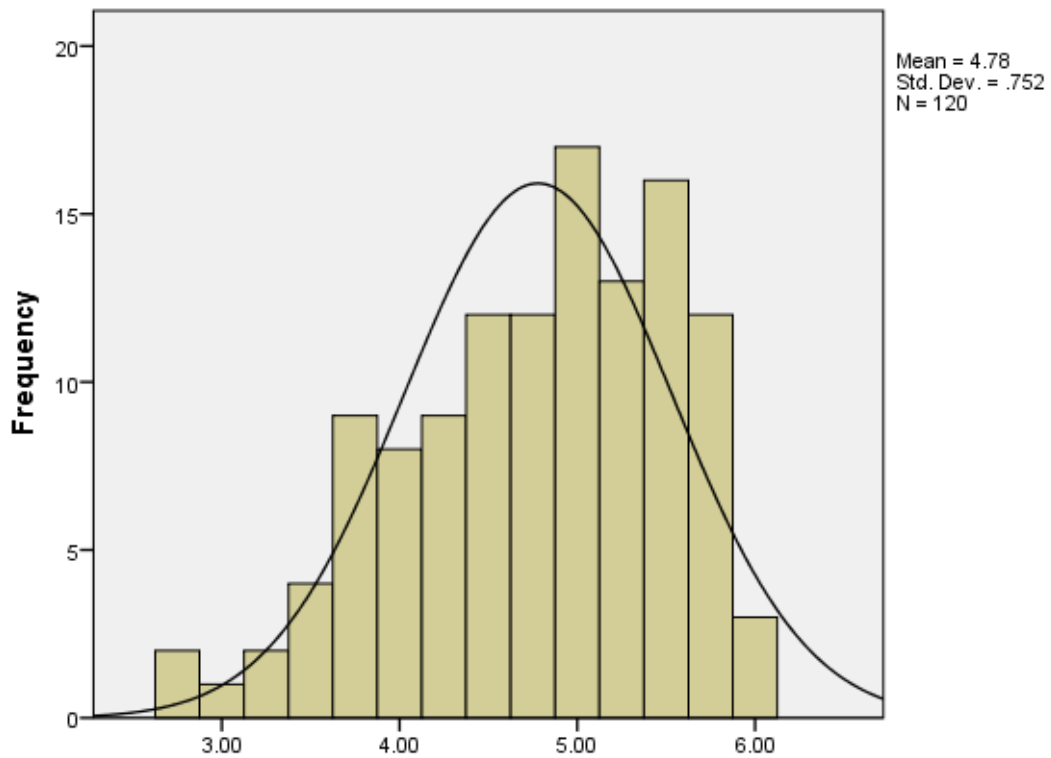
Čestica	1	h ²
1. Što sam bliže ispitu, teže mi je usredotočiti se na gradivo.	,534	,304
2. Kad učim za ispit, brinem se da se neću sjetiti gradiva na samom ispitu.	,721	,541
3. Za vrijeme važnih ispita smatram da mi ide loše i da ću pasti.	,787	,622
4. Gubim koncentraciju na važnim ispitima i ne mogu se sjetiti gradiva koje sam znao prije ispita.	,848	,758
5. Sjetim se odgovora na pitanja iz ispita tek nakon što završi.	,681	,490
6. Toliko se brinem prije važnog ispita da sam toliko izmoren(a) da ne mogu dati sve od sebe na ispitu.	,848	,724
7. Nisam dobre volje niti sam sav svoj / sva svoja dok polažem važne ispite.	,710	,587
8. Primijetim da mi nekad misli odlutaju dok polažem važne ispite.	,407	,472
9. Nakon ispita brinem se o tomu jesam li dao/dala sve od sebe.	,472	,810
10. Borim se s pisanim zadacima ili ih izbjegavam jer smatram da štogod napravio/napravila, to neće biti dovoljno dobro. Želim da to bude savršeno.	,677	,562
karakteristični korijeni	4,676	
% objašnjene varijance	46,758	

Prilog 5.4. Scree plot prikaz latentne strukture Upitnika o razini ispitne anksioznosti (N = 120)

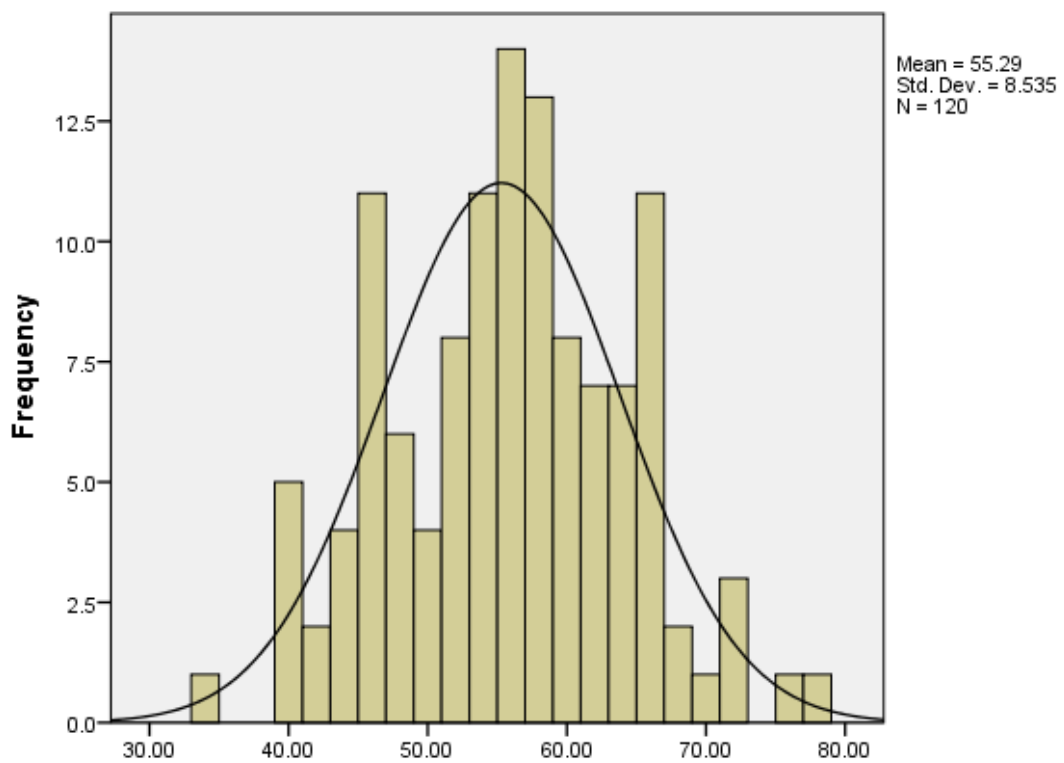


PRILOG 6. Histogrami distribucija kompozitnih varijabla

1. Rezultati upitnika o stavovima o e-ispitima Poslovnog engleskog jezika

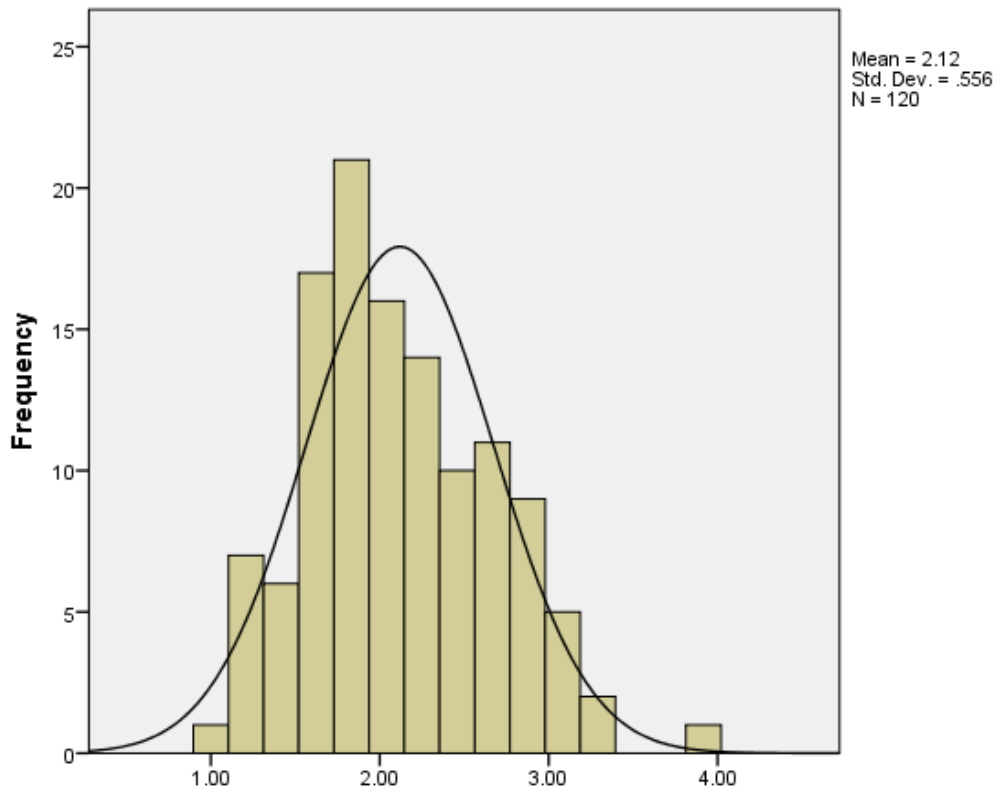


2. Rezultati upitnika o stavovima o ispitima općenito

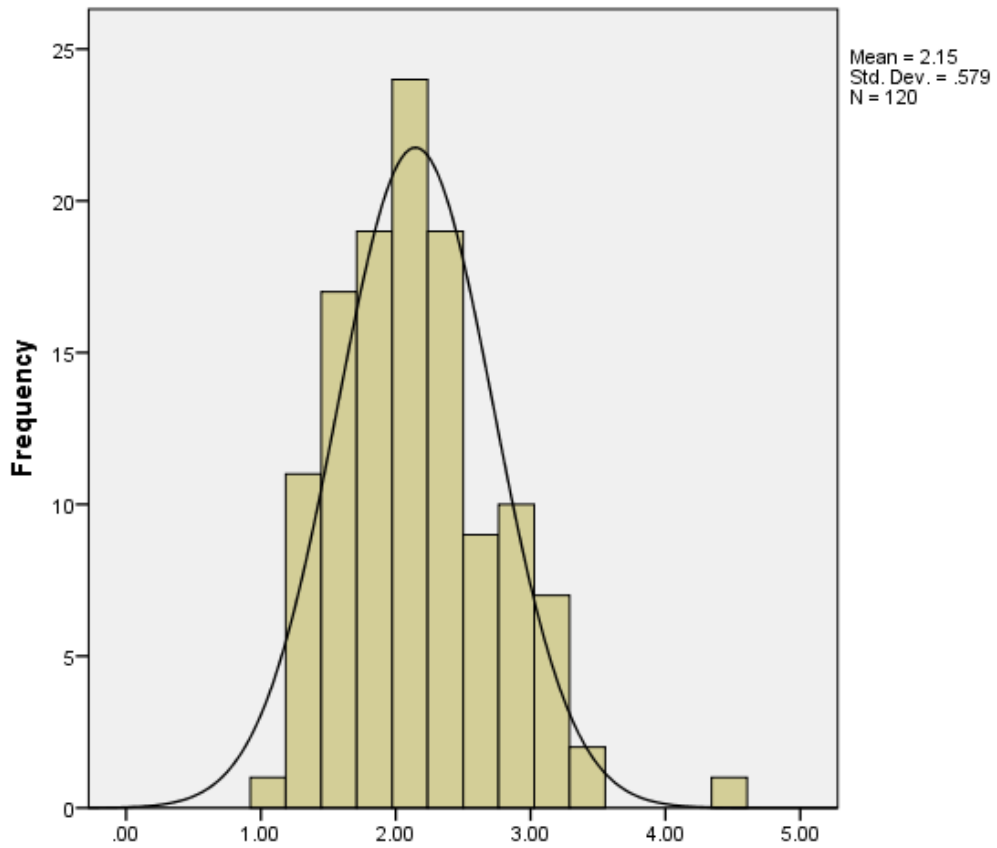


3. Rezultati upitnika o stavovima o računalima

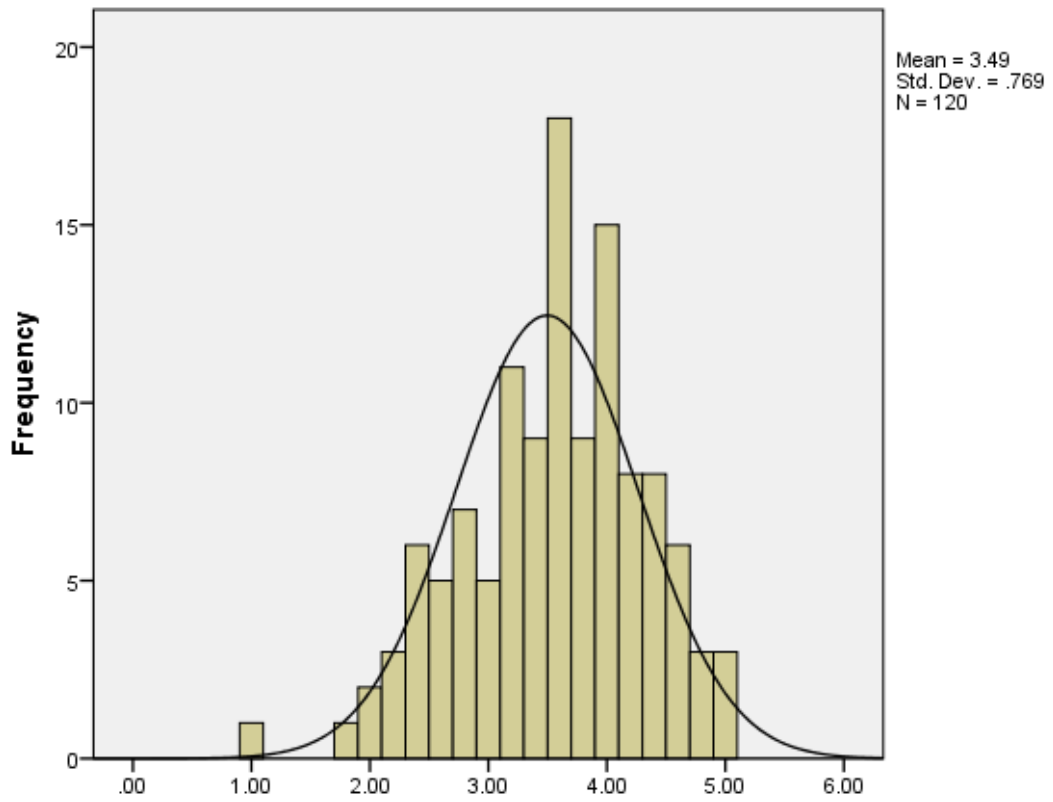
a. Faktor „anksioznost vezana za računala“



b. Faktor „ovisnost o računalima“



4. Rezultati upitnika o razini ispitne anksioznosti



PRILOG 7. Glosar

ability – sposobnost (kao sastavnica jezične kompetencije)

achievement assessment – procjenjivanje postignuća

achievement tests – ispiti postignuća

alternate form tests – ispiti zamjenjiva oblika

aptitude tests – ispiti sposobnosti

assessment – procjenjivanje

assessment as learning – procjenjivanje kao učenje

assessment for learning – procjenjivanje za učenje

assessment of learning – procjenjivanje naučenoga

assessment use argument – potkrjepa procjenjivanja

assessment utilization argument – potkrjepa primjene procjenjivanja

assessment validity argument – potkrjepa valjanosti procjenjivanja

authenticity – prirodnost (ispita) / autentičnost (ispita)

automated writing evaluation – automatsko procjenjivanje pisane proizvodnje

automatic scoring – automatsko bodovanje

backing – potpora (u potkrjepi procjenjivanja)

backwash (ili *washback*) – povratni učinak

band descriptor (ili *rubric, scale, scoring rubric, rating scale, level descriptor*) – ljestvica za procjenjivanje

behavioural intention (of a test use) – bihevioralna namjera (primjene ispita)

Business English – poslovni engleski

claim – tvrdnja (u potkrjepi procjenjivanja)

cognitive test anxiety – kognitivna ispitna anksioznost, kognitivna sastavnica ispitne anksioznosti

collaboration – suradnja (sustručnjaka)

competence – kompetencija (tj. ukupnost jezičnih sposobnosti)

comparable (test) form – usporediv oblik (ispita)

computer-adaptive testing – prilagodljivo ispitivanje na računalu

computer anxiety – računalna anksioznost

computer-assisted language assessment – računalno potpomognuto jezično procjenjivanje

computer-assisted language learning – računalno potpomognuto učenje jezika
computer-assisted language testing – računalno potpomognuto jezično ispitivanje
computer engagement – povezanost s računalima
computerized tests / computer-based tests – ispiti na računalu
concurrent validity – usporedna valjanost
consequential validity – posljedična valjanost
construct irrelevance – beznačajnost konstrukta
construct under-representation – nedovoljna zastupljenost konstrukta
construct validation – utvrđivanje valjanosti konstrukta
construct validity – konstruktna valjanost
content validity – sadržajna valjanost
context validity – kontekstna valjanost
continuous assessment – kontinuirano procjenjivanje
convergent validity – konvergentna valjanost
cooperation – savjetovanje (među sustručnjacima)
criterion-referenced assessment – kriterijsko procjenjivanje
criterion-related validity – kriterijska valjanost
data – podatak, podatci (u potkrjepi procjenjivanja)
debilitating anxiety – otežavajuća anksioznost
descriptive part (of assessment use argument) – opisni dio (potkrjepe procjenjivanja)
descriptors – opisnici
diagnostic assessment – dijagnostičko procjenjivanje
diagnostic tests – dijagnostički ispiti
discourse feedback – povratne obavijesti o diskursnim oznakama
discriminant validity – divergentna valjanost
dynamic assessment – dinamično procjenjivanje
educational assessment – obrazovno procjenjivanje
emergency remote teaching – nastava na daljinu iz krajnje nužde
emotionality – emocionalnost, emocionalna sastavnica ispitne anksioznosti
English as a Foreign Language – engleski kao strani jezik
English as a Second Language – engleski kao drugi jezik
English for Academic Purposes – engleski za akademske namjene

English for Business Purposes – engleski za poslovne namjene
English for Legal Purposes – engleski za pravne namjene
English for Medical Purposes – engleski u medicini
English for Occupational Purposes – engleski za profesionalne namjene
English for Science and Technology – engleski u znanosti i tehnologiji
English for Specific Purposes (ESP) – engleski kao jezik struke (EJS) / engleski za posebne namjene (EPN)
English for Vocational Purposes – engleski za strukovne namjene
equivalent (test) form – istovrijedan oblik (ispita)
error feedback – povratne obavijesti o greškama
evaluation – vrjednovanje
experimental validity – eksperimentalna valjanost
expert speaker – stručni govornik
external assessment – vanjsko procjenjivanje
extrapolation – proširivanje
eye-tracking method – metoda praćenja pokreta očiju
face validity – pojavna valjanost
facilitating anxiety – olakšavajuća anksioznost
feasibility (ili practicality) – izvedivost (ili praktičnost) (ispita)
formal assessment – formalno procjenjivanje
formative assessment – formativno procjenjivanje
gatekeepers – nadležne osobe (u institucijama u kojima se provodi ispitivanje)
General English – opći engleski
generalizable (test) – uopćivajuć (ispit)
generalization – uopćivanje
impact – utjecajnost (ispita)
impartial (test) – nepristran (ispit)
informal assessment – neformalno procjenjivanje
intended consequences – namijenjene posljedice (jedno od jamstava potkrjepe procjenjivanja)
intended use (of a test) – namijenjena primjena (ispita)
interactiveness – poticajnost (ispita)
internal assessment – unutarnje procjenjivanje

interpretive argument – potkrjepa tumačenja

interpretivist validity – interpretivistička valjanost

Intraclass Correlation Coefficient (ICC) – unutarklasni koeficijent korelacije

intrinsic validity – unutarnja valjanost

item response theory – teorija odgovora na zadatke

language assessment – jezično procjenjivanje

Language for Specific Purposes – jezik struke (ostali nazivi u primjeni: jezik za posebne namjene, strani jezik u struci, strani jezik struke i sl.)

language learning aptitude – sposobnost učenja jezika

language proficiency – jezično umijeće

language testing – jezično ispitivanje

Legal English – pravni engleski

level descriptor (ili *rubric*, *scale*, *scoring rubric*, *rating scale*, *band descriptor*) – ljestvica za procjenjivanje

meaningful (test) – smislen (ispit)

measurement – mjerenje

missing value analysis – analiza podataka koji nedostaju

native speaker – izvorni govornik

natural language processing – obrada prirodnog jezika

non-standardized assessment – nestandardizirano procjenjivanje

norm-referenced assessment – normativno procjenjivanje

outcome (of a test) – ishod (ispita)

outliers – odstupajući podatci

parallel (test) form – podudaran oblik (ispita)

peer assessment – vršnjačko procjenjivanje

perceived ease of use (of a test) – doživljena lakoća primjene (ispita)

perceived usefulness (of a test) – doživljena korisnost (ispita)

performance – izvedba (na ispitu)

performance assessment – procjenjivanje izvedbe

plausible rival hypothesis – hipoteza mogućih suparnika

positivist validity – pozitivistička valjanost

practicality (ili *feasibility*) – praktičnost (ili izvedivost) (ispita)

pragmatic validity – pragmatična valjanost

predictive validity – prediktivna valjanost

prescriptive part (of assessment use argument) – propisni dio (potkrjepe procjenjivanja)

Pre-Vocational English – engleski u pripremi za (određenu) struku

procrastination – odgađanje

proficiency – umijeće

proficiency assessment – procjenjivanje umijeća

proficiency tests – ispiti (jezičnog) umijeća

prognostic assessment – prognostičko procjenjivanje

quality (of a test outcome) – svojstvo (ishoda ispita)

rating criterion (ili scoring criterion) – mjerilo bodovanja

rating scale (ili rubric, scale, scoring rubric, band descriptor, level descriptor) – ljestvica za procjenjivanje

rebuttal – opovrgnuće (u potkrjepi procjenjivanja)

rebuttal data – dokazi opovrgnuća (u potkrjepi procjenjivanja)

relevance – značajnost (jedno od jamstava potkrjepe procjenjivanja)

relevant (test) – značajan (ispit)

reliability – pouzdanost

rubric (ili scale, scoring rubric, rating scale, band descriptor, level descriptor) – ljestvica za procjenjivanje

scale (ili rubric, scoring rubric, rating scale, band descriptor, level descriptor) – ljestvica za procjenjivanje

scoring – bodovanje

scoring criterion (ili rating criterion) – mjerilo bodovanja

scoring rubric (ili rubric, scale, rating scale, band descriptor, level descriptor) – ljestvica za procjenjivanje

second language – drugi jezik (jezik koji obuhvaća i drugi jezik – onaj koji se usvaja u sredini u kojoj se govori, i strani jezik – onaj koji se uči izvan takve sredine, obično u učionici)

self-assessment – samoprocjenjivanje

specifications – ispitna specifikacija

standardized assessment – standardizirano procjenjivanje

sufficiency – dostatnost (jedno od jamstava potkrjepe procjenjivanja)

summative assessment – sumativno procjenjivanje

target language use domain – ciljna primjena jezika

team teaching – timsko poučavanje (koje zajedno rade nastavnik i stručnjak uključen u određeni korak nastave i/ili procjenjivanja)

test – ispit

testing – ispitivanje

test usefulness – korisnost ispita

unitary competence hypothesis / unitary factor hypothesis – hipoteza jedinstvene jezične kompetencije

utility – upotrebljivost (jedno od jamstava potkrjepe procjenjivanja)

utilization – primjena

validation – utvrđivanje valjanosti

validity argument – potkrjepa valjanosti

variance inflation factor – VIF-pokazatelj

Vocational English – strukovni engleski

warrant – jamstvo (u potkrjepi procjenjivanja)

washback (ili *backwash*) – povratni učinak

washback hypothesis – hipoteza povratnog učinka

web-based testing – e-ispitivanje

web-based tests / Internet-based tests – e-ispiti

ŽIVOTOPIS

Bojan Prosenjak rođen je u Koprivnici gdje počinje svoje obrazovanje, a srednju školu završava u Velikoj Britaniji, nakon čega upisuje Filozofski fakultet Sveučilišta u Zagrebu te 2011. stječe zvanje profesora engleskog i francuskog jezika i književnosti. Zatim odlazi u Bruxelles gdje godinu dana radi u Vijeću EU-a kao asistent pravnicima-lingvistima.

Tijekom i nakon studija radi kao nastavnik engleskog i francuskog jezika te hrvatskog kao stranog jezika u školama stranih jezika, Privatnoj umjetničkoj gimnaziji i Francuskoj školi u Zagrebu. Ocjenjivač je i koordinator državne mature iz engleskog jezika. Također radi kao honorarni suradnik na izradi operativnih planova i nastavnih jedinica za udžbenike iz engleskog jezika te na izradi materijala za pripremu za državnu maturu iz engleskog jezika.

Nakon toga otvara svoj obrt za jezično usavršavanje i prevođenje Labos, imenovan je stalnim sudskim tumačem za engleski i francuski jezik te se bavi lektoriranjem i prevođenjem knjiga, tekstova i dokumenata iz raznih područja. Općem i poslovnom engleskom, francuskom i hrvatskom kao stranom jeziku poučava učenike svih uzrasta, pojedinačno i skupno, uživo i putem interneta, te ih priprema i za razne međunarodne certifikate. Usto upisuje poslijediplomski doktorski studij Glotodidaktike na Filozofskom fakultetu u Zagrebu, a radi i kao vanjski suradnik na Ekonomskom fakultetu u Zagrebu kao predavač Poslovnog engleskog jezika.

Popis objavljenih radova:

Prosenjak, B. i Lučev, I. (2020). The impact of peer assessment on the attainment level of oral presentations skills. *Language Value*, 12(1):30-55. DOI: 10.6035/LanguageV.2020.12.3

Prosenjak, B. (2020). How to construct your test: The influence of test length on students' achievement. *Competences of the 21st century: Certification of language proficiency*, 80-92. DOI: 10.31338/uw.9788323546917.

Jukić, S. i Prosenjak, B. (2019). Kako govornici slavenskih jezika rabe naveske u hrvatskome jeziku? *Komparativnoslavističke lingvokulturalne teme*, 151-164.

Vrhovac, Y., Seljan, S., Mencer Saluzzo, M. i Prosenjak, B. (2007). Une expérience interculturelle en tandem par Internet. *SRAZ*, 52, 301-315.

Prosenjak, B. (2007). Europski jezični portfolio u O. Š. I. Kršnjavi. *Strani jezici*, 36/1, 55-62.