

Bojan Prosenjak, Filozofski fakultet Sveučilišta u Zagrebu, Hrvatska

Sandra Marman, Filozofski fakultet Sveučilišta u Rijeci, Hrvatska

Učinak učestalosti provedbe ispita na izvedbu učenika hrvatskog kao stranog jezika

Constructing a valid, reliable, authentic, interactive, impactful and practical language test can be a daunting task, especially if one tries to harmonise it with the CEFR. When assessing students' performance on an achievement test, the question arises whether it is preferable to administer shorter tests more often, or one comprehensive test at the end of a teaching unit. The latter option corresponds to one research group that claims that students' performance is better when such tests are administered, but is in contrast to another group with the opposite results. However, a third group indicates that testing frequency does not have a significant effect on students' performance, nor is there a universally appropriate test length. Another possible effect on students' performance are their attitudes towards test length and testing frequency and their effect on performance. Therefore, the goal of this study is to compare the performance of students on progress achievement tests (PAT) and on a final comprehensive test (FCT), as well as to analyse their attitudes towards test length and testing frequency. The participants of the study were 14 Ukrainian students of Croatian as a Second Language at A1 level in Croatia. Our results show that, although the participants show their overall preference to taking PATs, which is in accordance with the results of Vaessen et al. (2017) and Öncül (2017), their performance was insignificantly better in an FCT than in PATs in all the sections apart from the writing production, which is in line with the research by Mines (2014) and Prosenjak (2020), and in part contrary to the results of Bangert-Drowns et al. (1991) and Gholami i Moghaddam (2013).

Keywords: achievement test, test performance, test length, testing frequency, attitudes towards test length and testing frequency

Izrada valjanog, pouzdanog, autentičnog, poticajnog, utjecajnog i praktičnog jezičnog ispita izazovan je zadatak, naročito ako ga se nastoji uskladiti sa ZEROJ-em. Kada je riječ o procjeni izvedbe učenika ispitom postignuća, postavlja se pitanje je li bolje češće provoditi kraće ispite ili na kraju određenog obrazovnog razdoblja provesti jedan sveobuhvatan ispit. Potonja je mogućnost u skladu s rezultatima jedne skupine istraživanja koja navode da ispitanici tada postižu bolje rezultate, no u suprotnosti je s rezultatima druge skupine. Međutim, treća ukazuje na to da učestalost provedbe ispita nema značajan učinak na izvedbu ispitanika te da ne postoji točna dužina ispita koja bi se smatrala prikladnom. Pritom su važni i stavovi ispitanika o dužini ispita i učestalosti ispitivanja koji mogu imati učinak na izvedbu. Stoga se ovim istraživanjem usporedila izvedba ispitanika na kraćim ispitima postignuća (KIP) i na sveobuhvatnom ispitivanju postignuća (SIP) te su se analizirali njihovi stavovi prema dužini ispita i učestalosti ispitivanja. U istraživanju je sudjelovalo 14 polaznika nastave hrvatskog kao stranog jezika kojima je materinski jezik ukrajinski i koji pohađaju nastavu na razini A1 u Hrvatskoj. Rezultati ovog istraživanja pokazali su da iako su sudionici skloniji polaganju KIP-ova, što je u skladu s rezultatima Vaessena i sur. (2017) te Öncül (2017), njihova je izvedba bila neznatno bolja na SIP-u nego na KIP-ovima u svim dijelovima osim u dijelu pisane proizvodnje, što je u skladu s istraživanjima Mines (2014) i Prosenjaka (2020), a djelomično u suprotnosti s rezultatima Bangert-Drowns i sur. (1991) te Gholami i Moghaddama (2013).

Ključne riječi: ispit postignuća, izvedba na ispitu, dužina ispita, učestalost ispitivanja, stavovi o dužini i učestalosti ispitivanja



1. Uvod

Istraživanja provjere jezičnog znanja učenika sve su brojnija, a uz njih se vežu i različiti pojmovi u hrvatskom jeziku, primjerice ocjenjivanje, procjenjivanje, provjeravanje, ispitivanje, mjerenje, vrjednovanje. U kontekstu procjenjivanja i vrjednovanja u hrvatskom jeziku Jelaska i Cvikić (2008) raspravljaju o nazivima procjenjivanje, vrjednovanje i provjeravanje, pri čemu se naziv procjenjivanje (eng. *assessment*) odnosi na „davanje suda o korisnikovom jezičnom znanju i umijeću“, a osim na temelju jezičnog ispita korisnika stranog jezika može se procjenjivati promatranjem u različitim jezičnim djelatnostima (ibid., 124). Tomu treba dodati da se korisnik ne procjenjuje na temelju očekivanja, nego prema tomu kako i koliko zna. Vrjednovanje se (eng. *evaluation*) također odnosi na davanje suda o korisnikovu jezičnom znanju i umijeću, ali i o uspješnosti programa koji je korisnik završio, pa ono može „uključivati i samu primjerenost načina procjenjivanja u odnosu na okolnosti ili razvijanje njegova znanja, smislenosti toga procjenjivanja i slično“ (ibid.). Provjeravanje se, pak, odnosi na procjenjivanje, odnosno davanje suda o korisnikovom znanju, na temelju očekivanja. U vezi s tim, Jelaska i Cvikić (ibid., 22) govore o engleskom nazivu *achievement assessment* jer se provjeravanjem može provjeriti je li korisnik ostvario očekivano postignuće.

U ovom će se radu pojam procjenjivanje upotrebljavati u užem smislu zato što ispiti postignuća procjenjuju postignuće ispitanika, što je u skladu s Bagarić Medve i Škarica (2023), koje za ovaj vid procjenjivanja kažu da se primjenjuje za procjenu razine znanja i sposobnosti ispitanika s pomoću ljestvica za procjenjivanje te se isprepliće s pojmom ispitivanja (eng. *testing*; usp. Kramer, 2013). Nasuprot tomu, pod procjenjivanjem u širem smislu *Zajednički europski referentni okvir za jezike*, odnosno ZEROJ (Council of Europe, 2001) podrazumijeva procjenjivanje jezičnog umijeća (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.

U kontekstu hrvatskog (kao inog) jezika istraživanja procjenjivanja i vrjednovanja znanja učenika usmjerena su na izradu jezičnih ispita (Ferbežar & Požgaj-Hadži, 2008), standardiziranih razredbenih ispita (Cvikić & Bošnjak, 2008), službenu provjeru poznavanja inojezičnog hrvatskog (Udier & Jelaska, 2008) te na sustave ocjenjivanja (Novak-Milić & Barbaroša-Šikić, 2008) i vrjednovanje statističkom analizom (Valdevit & Jelaska, 2008). Međutim, nastavnici se jezika svakodnevno suočavaju s problemom kako objektivno i učinkovito procjenjivati postignuće svojih učenika, pa tako i oni koji poučavaju hrvatskom kao stranom jeziku. Upravo su u tu svrhu dosad primijenjeni različiti ispiti, a oni koji se primjenjuju kako bi se utvrdilo u kojoj su mjeri učenici ispunili ciljeve koje im je postavio određeni tečaj, odnosno koliko su dobro ovladali određenom vještinom ili gradivom, nazivaju se ispitima postignuća (usp. Anastasi, 1976; Hughes, 1989; Bagarić Medve, 2011, 2023).

2. Teorijske postavke

Pri izradi svakog jezičnog ispita valja se voditi njegovim temeljnim svojstvom – korisnosti, odnosno primjenom za koju je određeni ispit namijenjen, tj. krovnim svojstvom koje u sebi

objedinjuje šest zasebnih svojstava, a to su pouzdanost, valjanost konstrukta, prirodnost ili autentičnost, poticajnost, utjecajnost te izvedivost ili praktičnost (Bachman & Palmer, 1996)¹. Iako krovno svojstvo korisnosti ispita u sebi objedinjuje ovih šest 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. Također, na valjanost ne treba gledati kao na opće svojstvo ispita niti pretpostavljati da može biti potpuna, već kao na stupanj do kojeg dokazi i teorija podupiru tumačenja rezultata ispita koja proizlaze iz namijenjenih primjena ispita, kao što ističe Shepard (2016) oblikujući najnoviju definiciju valjanosti.

Nadalje, autentičnost ili prirodnost ispita predstavlja „jamstvo da ispitni zadatci odgovaraju jezičnim zadatcima izvan samog ispita i da se rezultat ostvaren na ispitu može proširiti na rezultat i njegovo tumačenje u ciljnoj primjeni jezika“ (Bachman, 2005, 26). Poticajnost ispita, prema Fulcheru i Davidsonu (2007), podrazumijeva stupanj do kojeg su pojedinačna obilježja ispitanika, poput jezične sposobnosti, pozadinskog znanja, motivacije i dr., uključena prilikom polaganja ispita. Utjecajnost se, pak, definira kao „stupanj kojim ispit potiče ispitanikove sposobnosti koje želimo mjeriti“ (Udier & Jelaska, 2008, 250), a prema Bachmanu i Palmeru (1996) i Bachmanu (2005), ona uključuje povratni učinak, odnosno učinak na poučavanje, kao i načine na koje primjena ispita utječe na pojedince (ispitanike, korisnike ispita i donositelje odluka, među kojima su i nastavnici, ali i šire – na razredne i poslovne kolege ispitanika, buduće poslodavce ispitanika i dr.), institucije, obrazovni sustav i društvo u cjelini. Naposljetku, Fulcher i Davidson (2007) navode da se izvedivost bavi provedbom ispita, a ne značenjem ispitnih rezultata.

Pri izradi jezičnih ispita, nije dovoljno uzeti u obzir samo njihova svojstva, već se također postavlja pitanje ima li dužina tih ispita te učestalost njihove provedbe određeni učinak na procjenjivanje, odnosno na postignuće ispitanika. Upravo su se zato provodila različita istraživanja kako bi se utvrdilo može li se svrha ispitivanja postići provedbom ispita različitih dužina i različitom učestalosti. Kao što Burton (2006) zaključuje u svojem pregledu literature, ne postoji određena dužina ispita koja se može smatrati prikladnom. Međutim, prema istraživanju Ackermana i Kanfer (2009) te Jensen i sur. (2013), izvedba ispitanika na dužim ispitima bila je bolja, iako je primijećen sve veći subjektivni kognitivni umor kako se dužina ispita povećavala. Ti su rezultati također pokazali da su razlike među osobnostima učenika, interesima, motivacijama, stilovima učenja i strategijama polaganja ispita značajnije od dužine ispita ili vremena provedenog na ispitu. Wells i Wollack (2003), čiji su rezultati pokazali da duži ispiti dovode do veće pouzdanosti, savjetuju da se prije razmatranja o produženju ispita u obzir moraju uzeti vremensko ograničenje i mogući umor ispitanika.

Mogući učinak dužine ispita na izvedbu ispitanika manje se istraživao od učinka učestalosti ispitivanja. Naime, mnoga su se istraživanja usredotočila na formativno procjenjivanje, za koje Bagarić i Škarica (2023, 456) 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

¹ Usp. McNamara, 2000; Bachman, 2005; McNamara & Roever, 2006; Fulcher & Davidson, 2007; Udier & Jelaska, 2008; Kraš & Miličević, 2015; Shepard, 2016; Chapelle & Lee, 2021.

napredovanju i potencijalnim problemima te načinima kako ih riješiti i unaprijediti svoje učenje.²

Premda se u našem istraživanju ne ispituje učinak ovog načina ispitivanja na izvedbu ispitanika, već je težište stavljeno na usporedbu rezultata ispitanika na jednakom ispitu postignuća podijeljenom na dijelove i na sveobuhvatnom, valja prvo izdvojiti neka od istraživanja vezanih za učestalost i dužinu provedbe ispita.

Istraživanja koja su proveli Mines (2014) i Prosenjak (2020) ukazuju na to da učestalost ispitivanja ima malen utjecaj na prosječnu ili konačnu ocjenu učenika. Međutim, rezultati do kojih su došli Gholami i Moghaddam (2013) pokazali su da je izvedba učenika koji su polagali tjedne ispite u sklopu formativnog procjenjivanja bila bolja na završnom ispitu postignuća od onih koji su polagali samo jedan kolokvij, a Bangert-Drowns i sur. (1991) tvrdili su da je učestalije ispitivanje dovelo do bolje izvedbe prema kriterijskom procjenjivanju, ali ne i do poboljšanja njihova postignuća. Naime, kriterijskim procjenjivanjem „jezična izvedba ispitanika/učenika procjenjuje [se] s obzirom na unaprijed određena jezična znanja i sposobnosti koje se zadatkom ispituju i koje su jasno opisane za svaku razinu“ (Bagarić Medve, 2023, 467).³

Osim toga, rezultati pilot-istraživanja Vaessena i sur. (2017) pokazali su da većina studenata učestalo procjenjivanje vidi kao motivaciju za učenje, ali da bi taj učinak učestalosti mogao biti manji u slučaju da se oblici ispita na kraju svake cjeline i završnog ispita razlikuju. Nadalje, rezultati istraživanja koje je provela Öncül (2017) pokazuju da ispitanici smatraju da, iako učestali ispiti imaju pozitivan učinak na njihovo učenje, na upamćivanje naučenog i na vještine samopraćenja, ipak ne poboljšavaju redovito pohađanje nastave, navike učenja i izvedbu na ispitu, niti smanjuju ispitnu anksioznost, koju je definirao veći broj znanstvenika (usp. Zeidner, 1998; Živčić-Bećirević, 2003; Urbina, 2004; Erceg Jugović & Lauri Korajlija, 2012; Aliakbari & Gheitasi, 2017; Cheng & Zheng, 2021).

Naposljetku, iako ispitanici imaju pozitivniji stav prema predmetu i nastavniku ako se ispiti provode često (Bangert-Drowns & sur., 1991), a usto navode i smanjenu razinu ispitne anksioznosti (Vaessen & sur., 2017), istraživači bi trebali uzeti u obzir učinak učestalosti ispitivanja na učenje, a nastavnici bi shodno tomu mogli planirati ispitivanja tako da budu prikladna za učenike umjesto da se usredotočuju na samu učestalost ispitivanja (ibid.). S obzirom na djelomično različite rezultate dosadašnjih istraživanja, svrha je ovog rada ispitati i utvrditi učinak triju kraćih ispita tijekom semestra i jednog sveobuhvatnog na kraju semestra na izvedbu dviju skupina ispitanika.

3. Metodologija istraživanja

Kako bi se dobio uvid u učinak dužine i učestalosti procjenjivanja ispitima postignuća na izvedbu učenika hrvatskog kao stranog jezika, naše je istraživanje kao svoj glavni cilj imalo ispitati učinak kraćih i češćih pisanih ispita postignuća hrvatskog kao stranog jezika na izvedbu učenika. Nadalje, njegovi su pojedinačni ciljevi sljedeći:

² Usp. Bachman, 1990; Brown & Abeywickrama, 2010; Fulcher, 2010; Dobson, 2013; Woodin & Müller-Hartmann, 2013; Green, 2014; Fox, 2017; Poehner & dr., 2017.

³ Usp. Brown & Abeywickrama, 2010; Hughes & Hughes, 2020.

1. usporediti rezultate učenika na kraćim ispitima postignuća (KIP) i sveobuhvatnom ispitu postignuća (SIP);
2. ispitati stavove učenika o dužini ispita postignuća i učestalosti ispitivanja.

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

- IP1 Hoće li izvedba učenika biti bolja ako češće polažu KIP-ove tijekom jednog semestra ili ako jednom polažu jedan SIP na kraju semestra?
- IP2 Je li izvedba učenika na KIP-ovima i SIP-u povezana s njihovim stavovima o dužini ispita postignuća i učestalosti ispitivanja?

3.1 Sudionici

U ovom je istraživanju sudjelovalo četrnaestero učenika hrvatskog kao stranog jezika na razini A1 prema ZEROJ-u, u dobi od 17 do 50 godina, od kojih dvanaest žena i dvojica muškaraca, a kojima je materinski jezik ukrajinski.⁴

Za potrebe ovog istraživanja odabran je prigodni neprobabilistički uzorak, a to je i najčešći uzorak u istraživanjima stranog jezika. Prigodni uzorak odabire se jer zadovoljava određene kriterije koje zahtijeva istraživanje, poput dostupnosti određene skupine ispitanika u određeno vrijeme (usp. Mackey & Gass, 2012, 81). Usto, svi su sudionici bili učenici istog nastavnika, koji je sudjelovao u provedbi ispita i procjenjivanju izvedbe sudionika, no ne i u izradi ispita postignuća, koji su izradili istraživači.

3.2 Prikupljanje podataka

Prikupljanje podataka odvijalo se u dvjema skupinama u koje su sudionici stavljeni nasumično na početku semestra. Skupina A obuhvaćala je sedmero sudionika koji su polagali tri kraća ispita postignuća (KIP) tijekom jednog semestra u akademskoj godini 2023./2024., u međusobnom razmaku od otprilike mjesec dana. Skupina B od također sedmero sudionika polagala je jedan sveobuhvatan ispit postignuća (SIP) na kraju semestra.

Prije početka istraživanja sudionicima je objašnjena svrha istraživanja. Istaknuto je da će njihovi rezultati biti prikazani anonimno i da će se njihovi odgovori primijeniti samo u svrhu ovog istraživanja. Sudionicima je objašnjeno da je njihovo sudjelovanje dobrovoljno i da mogu odustati u bilo kojem trenutku. Uz usmena pojašnjenja, sudionici su dobili i pisani obaviješteni pristanak te su svojim potpisom pristali sudjelovati u istraživanju.

3.3 Instrumenti

Kako bi se odgovorilo na postavljena istraživačka pitanja, u istraživanju su korištena dva instrumenta – jezični ispit postignuća i upitnik o stavovima sudionika o dužini ispita postignuća i učestalosti ispitivanja. U nastavku se donosi opis instrumenata.

⁴ Svi sudionici došli su u Hrvatsku zbog rata u Ukrajini te su do ovog istraživanja u njoj živjeli od četiri do osamnaest mjeseci i pokazali su visoku motivaciju za učenje hrvatskog jezika. Tomu svjedoče odgovori na pitanje vezano za razlog učenja hrvatskoga jezika – posao i ostanak u Hrvatskoj (72 %) te sporazumijevanje na hrvatskom jeziku (28 %).

3.3.1 Jezični ispiti postignuća

Jezične ispite postignuća osmislili su autori ovog istraživanja, a slijedili su vrste zadataka koji se primjenjuju na nastavi, istovjetne onima u nastavnim materijalima kojima se koriste sudionici kako bi ih mogli uvježbavati kod kuće i pripremiti se za izvedbu na ispitima. Uzimajući u obzir činjenicu da valjanost nikad nije potpuna, već je na određenom stupnju⁵, može se tvrditi da je valjanost naših ispita na visoku stupnju jer se na temelju rezultata ostvarenih na njima može doći do zaključaka o onim jezičnim vještinama sudionika koje su ovi ispiti mjerili – razumijevanje teksta čitanjem, razumijevanje i primjena vokabulara i gramatike te pisana proizvodnja (Davies & Elder, 2005). Nadalje, nastojalo se slijediti načelo autentičnosti ispita pri izradi ispitnih zadataka koji se odnose na primjenu hrvatskog kao stranog jezika na razini A1. Usto, ispiti su bili poticajni jer su teme i ispitivane jezične sposobnosti bile raznovrsne s obzirom na dob i interese naših sudionika, utjecajni jer je njihov povratni učinak od koristi svim dionicima ovog procesa procjenjivanja (učenicima, nastavnicima i ustanovi na kojoj se provodi nastava) te izvedivi jer su se mogli provesti tijekom jednog semestra.

SIP se sastojao od šesnaest ispitnih čestica podijeljenih u tri dijela – jedna čestica u dijelu razumijevanja teksta čitanjem, četrnaest čestica u dijelu razumijevanja i primjene vokabulara i gramatike te jedna čestica u dijelu pisane proizvodnje. KIP-ovi su sadržavali iste zadatke iz SIP-a, samo što su dvije ispitne čestice, one iz razumijevanja teksta čitanjem i pisane proizvodnje bile podijeljene na tri dijela (po jedan za svaki KIP), a četrnaest čestica koje su ispitivale razumijevanje i primjenu vokabulara i gramatike bile su približno ravnomjerno raspoređene u trima KIP-ovima (tablica 1). Zadatci koji su se odnosili na razumijevanje i primjenu vokabulara i gramatike bili su poluotvorenog tipa – zadatci dopunjavanja i kratkih odgovora.

Pri izradi ispita slijedili smo dosadašnje spoznaje prema kojima bi učinak učestalosti mogao biti manji u slučaju da se oblici ispita na kraju svake cjeline i završnog ispita razlikuju (Vaessen i sur., 2017). Usto, trajanje svakog KIP-a bilo je do trideset minuta, a SIP-a 90 minuta, odnosno zbroju trajanja triju KIP-ova (tablica 1).

SIP	KIP	Dio ispita / vještina	Broj ispitne čestice	Broj bodova
1	1	Razumijevanje teksta čitanjem	1 (1. dio)	2
	2		1 (2. dio)	2
	3		1 (3. dio)	2
	1	Razumijevanje i primjena vokabulara i gramatike	2	6
	1		3	5
	1		4	5
	1		5	5
	1		6	5
	2		7	5
	2		8	5
	2		9	5
	2		10	5
	3		11	5

⁵ Usp. 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; Shepard, 2016; Chapelle, 2021; Green & Fulcher, 2021.

	3		12	5
	3		13	5
	3		14	6
	3		15	4
	1	Pisana proizvodnja	16 (1. dio)	5
	2		16 (2. dio)	5
	3		16 (3. dio)	10

Tablica 1: Struktura kraćih ispita postignuća (KIP) i sveobuhvatnog ispita postignuća (SIP)

Iako su još Hughes (1989) i Bachman (1990) tvrdili da je bolje usredotočiti se na izravno ispitivanje, što znači da „zahtjevi zadataka odražavaju u najvećoj mogućoj mjeri jezičnu uporabu u autentičnim lingvističkim, situacijskim, kulturnim i afektivnim situacijama (Bagarić Medve, 2023, 467), zadatci u dijelu razumijevanja i primjene vokabulara i gramatike u našim ispitima ispitivali su vještine neizravno zbog vremenskog ograničenja i praktičnosti provedbe ispita. U dijelovima razumijevanja teksta čitanjem i pisane proizvodnje vještine su se sudionika ispitivale izravno jer su sudionici odgovarali na pitanja povezana s danim tekstom, odnosno pisali su poluvođeno kraći tekst prema zadanim nalogima. Rezultati izvedbe sudionika izraženi su u postocima. U svim zadacima, osim zadatka pisane proizvodnje, primijenjeno je procjenjivanje zasebnih čestica, a svaki je točan odgovor donosio jedan bod, dok je u prethodno navedenom zadatku primijenjeno integrativno procjenjivanje, što znači da su sudionici kombinirali različite jezične sastavnice kako bi riješili zadatak (ibid.). Osim toga, ispiti primijenjeni u našem istraživanju oni su koji upućuju na obilježje (eng. *criterion-referenced test*), što znači da se „rezultati tumače s obzirom na postignut stupanj sposobnosti ili sam sadržaj, a ne s obzirom na rezultate ostalih ispitanika“ (Cvikić & Bošnjak, 2008, 185). Nadalje, većina zadataka bodovana je objektivno, što znači da prosudba procjenjivača nije bila potrebna tijekom bodovanja (Hughes, 1989, 19), osim zadatka pisane proizvodnje, u kojem je bila potrebna prosudba procjenjivača.

3.3.2 Upitnik

Osim postignuća sudionika na jezičnom ispitu bilo je važno ispitati i njihove stavove s obzirom na to da je istraživanje Öncül (2017) pokazalo da ispitanici smatraju da učestali ispiti ne poboljšavaju redovito pohađanje nastave, navike učenja i izvedbu na ispitu, iako imaju pozitivan učinak na njihovo učenje, na upamćivanje naučenog i na vještine samopraćenja. Stoga je drugi instrument našeg istraživanja bio upitnik o stavovima sudionika o dužini ispita postignuća i učestalosti ispitivanja (dalje u tekstu: Upitnik) koji su sudionici ispunili po završetku provedbe ispitâ – skupina A nakon rješavanja trećeg KIP-a, a skupina B nakon SIP-a.

Upitnik je bio na hrvatskom jeziku, a nastavnik koji je nadgledao ispit također je pomogao učenicima pri razumijevanju čestica Upitnika. Sastojao se od šest čestica:

1. Koliko godina imate?
2. Koji je Vaš spol?
3. Koji je Vaš razlog dolaska u Hrvatsku?
4. Koji je Vaš razlog učenja hrvatskog jezika?
5. Koliko ste dugo ukupno boravili u Hrvatskoj prije početka učenja hrvatskog jezika?
6. Je li Vam draže polagati više kraćih ispita tijekom semestra ili jedan sveobuhvatan na kraju semestra? Zašto?

4. Rezultati i rasprava

U ovom su poglavlju predstavljeni rezultati istraživanja o kojima se raspravlja s obzirom na teorijske spoznaje i druga slična istraživanja navedena u uvodu ovog rada. Nakon provedbe i procjenjivanja kraćih ispita postignuća (KIP) i sveobuhvatnog ispita postignuća (SIP) te analize odgovora na Upitniku, došlo se do sljedećih rezultata kako bi se odgovorilo na prvo istraživačko pitanje – Hoće li izvedba učenika biti bolja ako češće polažu KIP-ove tijekom jednog semestra ili ako jednom polažu jedan SIP na kraju semestra?

Kao što se može vidjeti u tablici 2, na trima KIP-ovima srednje vrijednosti izvedbe sudionika skupine A bile su izrazito visoke u dijelu razumijevanja teksta čitanjem, iznadprosječne u dijelu razumijevanja i primjene vokabulara i gramatike, a u dijelu pisane proizvodnje bile su ispodprosječne na prvom i trećem KIP-u, dok su na drugom bile iznadprosječne. Iz iste je tablice vidljivo da su na SIP-u srednje vrijednosti izvedbe sudionika skupine B na najvišoj razini u dijelu razumijevanja teksta čitanjem, iznadprosječne u dijelu razumijevanja i primjene vokabulara i gramatike, a ispodprosječne u dijelu pisane proizvodnje. Ipak, treba uzeti u obzir malen uzorak sudionika istraživanja, pa je na temelju toga teško poopćiti zaključke.

Ispit	Dio ispita	Najmanja vrijednost	Najveća vrijednost	<i>M</i>	<i>SD</i>	Asimetričnost distribucije ^a	Spljoštenost distribucije ^b
KIP 1	Razumijevanje teksta čitanjem	1,00	2,00	1,86	0,38	-2,646	7,000
	Razumijevanje i primjena vokabulara i gramatike	14,50	24,00	18,86	3,68	0,179	-1,745
	Pisana proizvodnja	0,00	3,50	2,00	1,26	-0,439	-0,909
	Ukupno	16,50	29,50	22,71	4,77	-0,022	-0,909
KIP 2	Razumijevanje teksta čitanjem	1,00	2,00	1,71	0,49	-1,230	-0,840
	Razumijevanje i primjena vokabulara i gramatike	11,50	17,50	15,07	1,99	-0,749	0,953
	Pisana proizvodnja	2,00	5,00	3,43	0,98	0,277	0,042
	Ukupno	16,50	22,00	20,21	2,14	-1,148	-0,089
KIP 3	Razumijevanje teksta čitanjem	2,00	2,00	2,00	0,00	0,000	0,000
	Razumijevanje i primjena vokabulara i gramatike	14,00	23,00	18,50	2,74	-0,017	1,463
	Pisana proizvodnja	2,00	8,00	4,36	1,93	1,106	1,730
	Ukupno	18,00	33,00	24,86	4,55	0,539	1,818
SIP	Razumijevanje teksta čitanjem	6,00	6,00	6,00	0,00	0,000	0,000

Razumijevanje i primjena vokabulara i gramatike	45,50	69,00	55,43	10,22	0,429	-2,405
Pisana proizvodnja	0,00	13,00	7,14	4,49	-0,441	-0,445
Ukupno	54,50	88,00	68,57	14,02	0,467	-2,196

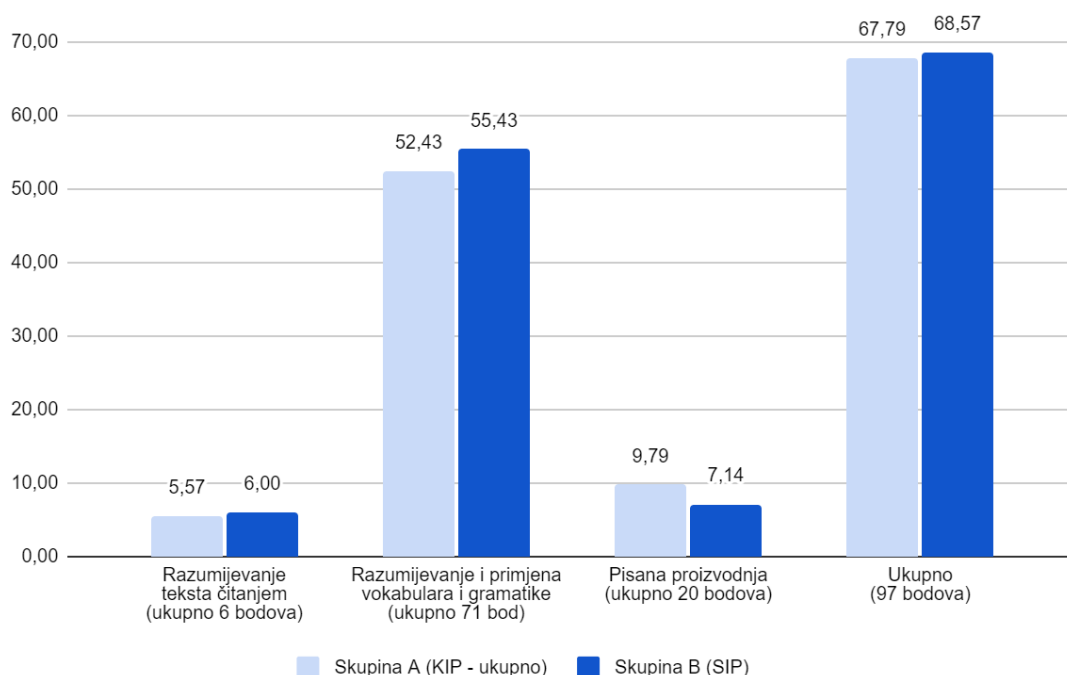
^a Standardna pogreška mjere asimetričnosti distribucije iznosi 0,794.

^b Standardna pogreška mjere spljoštenosti distribucije iznosi 1,587.

Tablica 2: Deskriptivna statistika izvedbe sudionika na KIP-ovima i na SIP-u (N = 7)

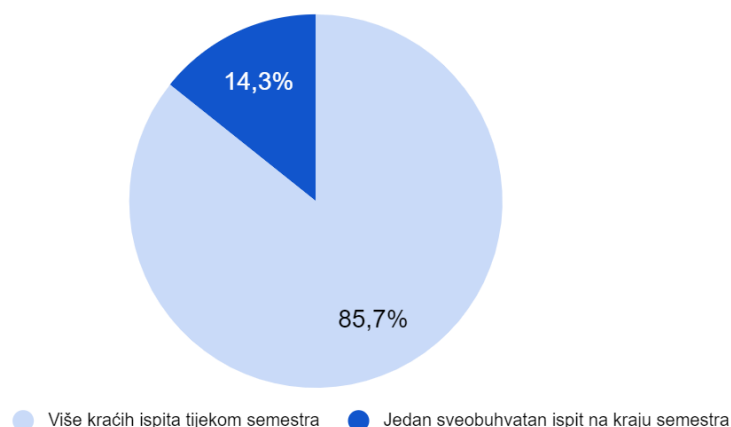
Kad se usporede srednje vrijednosti izvedbe sudionika skupina A i B (slika 1), i to na svakom od triju dijelova ispita te ukupni rezultati, razvidno je da su srednje vrijednosti nešto niže na KIP-ovima ($M = 5,57$) u odnosu na SIP ($M = 6,00$) u dijelu razumijevanja teksta čitanjem. Iako razlika postoji, može se reći da su sudionici u objema skupinama ostvarili približno jednak rezultat te da je razlika u srednjim vrijednostima relativno malena. Rezultati također pokazuju da su srednje vrijednosti nešto niže na KIP-ovima ($M = 52,43$) u odnosu na SIP ($M = 55,43$) i u dijelu razumijevanja i primjene vokabulara i gramatike. Premda je uzorak ispitanika ovog istraživanja vrlo malen i svrha rada nikako nije uopćivanje dobivenih rezultata, do sličnih nalaza došao je Prosenjak (2020). Naime, njegovo je istraživanje pokazalo da je postignuće učenika slično u svim vještinama, bez obzira na to jesu li učenici tijekom semestra bili izloženi češćim i kraćim testovima ili rjeđim i duljim. Suprotno tome, a kao što se može vidjeti na slici 1, srednje su vrijednosti na KIP-ovima ($M = 9,79$) zamjetno više na dijelu pisane proizvodnje u odnosu na SIP ($M = 7,14$).

Što se tiče srednjih vrijednosti za ukupnu izvedbu sudionika na ovim dvama ispitima, vidljivo je da su one neznatno više na SIP-u nego na KIP-ovima. S obzirom na navedene rezultate, može se opaziti da je razlika u konačnim rezultatima učenika na KIP-ovima i na SIP-u minimalna. Takav rezultat u skladu je s nekim dosadašnjim istraživanjima koja pokazuju da učestalost provedbe ispita ima malen utjecaj na konačnu ocjenu učenika (npr. Mines, 2014), ili da ne postoji određena dužina ispita koja se može smatrati prikladnom (Burton, 2006). Osim toga, rezultati ovog istraživanja djelomično su u skladu s rezultatima drugih istraživanja koja su pokazala da je učestalije ispitivanje dovelo do bolje izvedbe (npr. Gholami & Moghaddam, 2013), jer se u našem istraživanju pokazalo da su srednje vrijednosti na KIP-ovima zanemarivo više samo za dio pisane proizvodnje. Moguće obrazloženje te razlike vidimo u tome da sudionici na KIP-ovima moraju napisati kraći poluvođeni tekst koji slijedi sadržaj obrađen na nastavi do trenutka pisanja pojedinog KIP-a, dok taj zadatak u SIP-u sadrži više uputa koje se odnose na sadržaj obrađen u cijelom semestru te se od sudionika očekuje i duži tekst.



Slika 1: Usporedba izvedbe sudionika na KIP-ovima (N = 7) i na SIP-u (N = 7)

Kako bi se odgovorilo na drugo istraživačko pitanje – Je li izvedba učenika na KIP-ovima i SIP-u povezana s njihovim stavovima o dužini ispita postignuća i učestalosti ispitivanja? – analizirali su se odgovori sudionika na Upitniku. Kao što se može vidjeti na slici 2, većina sudionika navela je da više voli polagati nekoliko kraćih ispita tijekom semestra, nego jedan sveobuhvatan na kraju semestra, što preferira samo dvoje sudionika iz skupine B.



Slika 2: Usporedba preferencija prema dužini i učestalosti ispita (N = 14)

Iako su sudionici ovog istraživanja skloniji polaganju nekoliko KIP-ova tijekom semestra, rezultati su pokazali da su gotovo u svim dijelovima ispita, osim dijela pisane proizvodnje, uspješniji bili oni sudionici koji su pisali jedan SIP. Ustanovljeno je da su dva sudionika koja su navela da više vole polagati jedan SIP na kraju semestra svoj odgovor obrazložila činjenicom da je takvo

polaganje ispita „lakše” te da „moraju učiti samo jednom u semestru”. S druge strane, nastavnik koji je provodio ispite sa sudionicima ovog istraživanja, a koji je ujedno procjenjivač ispita, primijetio je da su učenici koji su pisali KIP-ove pristupali ispitima s povišenom razinom straha i anksioznosti. Kako bi se dobio uvid u mogući učinak učestalosti ispitivanja i dužine ispita na razinu ispitne anksioznosti, valjalo bi provesti dodatna istraživanja budući da su odgovori sudionika ovog istraživanja i iskustvo njihova nastavnika u suprotnosti s nekim istraživanjima koja su utvrdila da učenici koji pišu više kraćih ispita imaju smanjenu razinu ispitne anksioznosti (Vaessen & sur., 2017).

5. Zaključak

Svrha je ovog istraživanja bila istražiti učinak učestalosti provedbe ispita na postignuće učenika hrvatskog kao stranog jezika i to usporedbom rezultata učenika na kraćim ispitima postignuća (KIP) i sveobuhvatnom ispitu postignuća (SIP). Naši su rezultati pokazali da su srednje vrijednosti njihove izvedbe na jezičnim ispitima neznatno više na SIP-u nego na KIP-ovima. Prema tomu, može se zaključiti da je razlika u konačnim rezultatima učenika na KIP-ovima i na SIP-u relativno malena, a takav je rezultat u skladu s nekim dosadašnjim istraživanjima koja pokazuju da učestalost provedbe ispita ima malen utjecaj na konačnu izvedbu učenika (Mines, 2014; Prosenjak, 2020). Osim toga, u našem se istraživanju pokazalo da su srednje vrijednosti na KIP-ovima nezanemarivo više samo za dio pisane proizvodnje.

Nadalje, premda su sudionici skloniji polaganju KIP-ova, što su također zaključili Vaessen i sur. (2017) te Öncül (2017), rezultati su pokazali da su gotovo u svim dijelovima ispita, osim dijela pisane proizvodnje, uspješniji bili oni sudionici koji su pisali SIP. Ovi su naši rezultati u skladu s rezultatima Gholami i Moghaddam (2013), a djelomično u suprotnosti s rezultatima do kojih su došli Bangert-Drowns i sur. (1991). S obzirom na naše spoznaje, predlažemo da se u daljnjim istraživanjima ispita učinak učestalosti provedbe ispita na većem uzorku sudionika u svakoj skupini, a i na različitim razinama ovladanosti hrvatskim jezikom. Prijedlog je također da se s dvjema skupinama učenika provedu i KIP-ovi i SIP kako bi se utvrdilo postoji li razlika kod iste skupine učenika u pisanju ispita različite dužine koji se provode različitom učestalosti.

Smatramo da je naše istraživanje, iako provedeno na malenu uzorku, uz ostvarivanje svojeg zadanog cilja, također korisno svim dionicima u procesu procjenjivanja. Prvo, samim se ispitnicima, odnosno učenicima, mogu predstaviti rezultati koji ukazuju na to da izvedba na ispitu nije uvijek u skladu sa stavovima o dužini ispita i učestalosti ispitivanja. Zatim, nastavnici, odnosno autori ispita, mogu prilagoditi procjenjivanje preferencijama svojih učenika budući da jedna određena dužina ispita i jedna određena učestalost ispitivanja nemaju uvijek učinak na izvedbu na ispitu. Naposljetku, obrazovne ustanove mogu planirati i provoditi ispitivanje na način koji smatraju da im je najprikladniji zbog nepostojanja prethodno navedenog učinka.

6. Literatura

- Ackerman, P. L. & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, 15(2), 163–181.
- Aliakbari, M. & 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.
- Anastasi, A. (1976). *Psychological Testing*. Macmillan.
- Alderson, J. C. & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching*, 34, 213–236.
- Angoff, W. H. (1988). Validity: An Evolving Concept. U Wainer, H. i Braun, H. (ur.), *Test validity* (19–32). Lawrence Erlbaum.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. 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. & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- 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. & Pavičić Takač, V. (ur.), *Učenje i poučavanje stranih jezika* (461–500). Filozofski fakultet Osijek.
- Bagarić Medve, V. & Škarica, I. (2023). Vrednovanje u nastavi stranih jezika. U Bagarić Medve, V. & Pavičić Takač, V. (ur.), *Učenje i poučavanje stranih jezika* (451–459). Filozofski fakultet Osijek.
- Bangert-Drowns, R. L., Kulik J. A. & Kulik C. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, 85(2), 88–99.
- Brown, H. D. & Abeywickrama, P. (2010). *Language Assessment: Principles and Classroom Practices*. Pearson Education.
- Burton, R. F. (2006). Sampling knowledge and understanding: how long should a test be. *Assessment & Evaluation in Higher Education*, 31(5), 569–582.
- 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* (11–20). Routledge.
- Chapelle, C. A. & Lee, H. (2021). Understanding Argument-Based Validity in Language Testing. U Chapelle, C. A. & Voss, E. (ur.), *Validity Argument in Language Testing – Case Studies of Validation Research* (19–44). Cambridge University Press.
- Cheng, L. & Zheng, Y. (2021). Measuring Anxiety. U Winke, P. i Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (177–186). Routledge.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Cvikić, L. & Bošnjak, M. (2008). Teorijski okvir razredbenoga ispita inojezičnoga hrvatskoga. *Lahor*, 2(6), 183–197.
- Davies, A. & Elder, C. (2005). Validity and validation in language testing. U Hinkel, E. (ur.), *Handbook of research in second language teaching and learning* (795–814). Routledge.
- Dobson, A. (2013). Evaluation. U Byram, M. & Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning*, (233–239). Routledge.
- Erceg Jugović, I. & Lauri Korajlija, A. (2012). Povezanost ispitne anksioznosti s perfekcionizmom. *Psihologijske teme*, 21(2), 299–316.
- Ferbežar, I. & Požgaj-Hadži, V. (2008). Kako izraditi jezični test. *Lahor*, 2(6), 165–182.
- Fox, J. (2017). Using Portfolios for Assessment/Alternative Assessment. U Shohamy, E., Or, I. G. & May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (135–147). Springer.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon: Routledge.
- Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.
- Gholami, V. & Moghaddam, M. M. (2013). The Effect of Weekly Quizzes on Students' Final Achievement Score. *I. J. Modern Education and Computer Science*, 1, 36–41.
- Green, A. (2014). *Exploring Language Assessment and Testing*. Routledge.
- Green, A. & Fulcher, G. (2021). Test Design Cycle. U Winke, P. & Brunfaut, T. (ur.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (69–77). Routledge.

- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Hughes, A. & Hughes, J. (2020). *Testing for Language Teachers*. Cambridge University Press.
- Jelaska, Z. & Cvikić, L. (2008). Procjenjivanje i vrjednovanje u hrvatskome jeziku. *Lahor*, 1(5), 115–125.
- Jensen, J. L., Berry, D. A. & Kummer, T. A. (2013). Investigating the effects of exam length on performance and cognitive fatigue. *PLoS One*, 8(8); e70270. doi:10.1371/journal.pone.0070270
- Kramer, J. (2013). Assessment and testing. U Byram, M. & Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning* (52–57). Routledge.
- Kraš, T. & Miličević, M. (2015). *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika*. Filozofski fakultet Sveučilišta u Rijeci.
- Mackey, A. & Gass, S. M. (2012). *Research Methods in Second Language Acquisition: A Practical Guide*. John Wiley & Sons, Inc.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.
- Messick, S. (1989). Validity. U Linn, R. L. (ur.), *Educational measurement* (13–103). Macmillan.
- Milas, G. (2009). *Istraživačke metode u psihologiji i drugim društvenim znanostima*. 2. izd. Naklada Slap.
- Mines, R. O. Jr. (2014). *The impact of testing frequency and final exams on student performance*. Proceedings 2014 American Society of Engineering Education Annual Conference.
- Novak-Milić, J. & Barbaroša-Šikić, M. (2008). Različiti sustavi ocjenjivanja i usporedne ljestvice. *Lahor*, 2 (6), 198–209.
- Öncül, G. (2017). Frequent testing: What are the real impacts of frequent quizzes on students, teachers, and instruction? *Turkish Online Journal of English Language Teaching (TOJELT)*, 2(1), 1–19.
- Poehner, M. E., Davin, K. J. & Lantolf, J. P. (2017). Dynamic Assessment. U Shohamy, E., Or, I. G. & May, S. (ur.), *Encyclopedia of Language and Education: Language Testing and Assessment*, 3. izd. (243–256). Springer.
- 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.
- Shepard, L. A. (2016). Evaluating test validity: reprise and progress. *Assessment in Education: Principles, Policy & Practice*, 23(2). 268–280.
- Udier, S. L. & Jelaska, Z. (2008). Službena provjera poznavanja inojezičnoga hrvatskoga jezika. *Lahor*, 6(2). 237–255.
- Urbina, S. (2004). *Essentials of psychological testing*. John Wiley & Sons, Inc.
- Vaessen, B. E., van den Beemt, A., van de Watering, G., Meeuwen, L. W., van Lemmens, L. & den Brok, P. (2017). Students' perception of frequent assessments and its relation to motivation and grades in a statistics course: a pilot study. *Assessment & Evaluation in Higher Education*, 42(6), 872–886.
- Valdevit, M. & Jelaska, Z. (2008). Srednje vrijednosti u ocjenjivanju: procjenjivanje i vrjednovanje statističkom analizom. *Lahor*, 2(6), 210–236.
- Weir, C. J. (2005). *Language Testing and Validation*. Palgrave Macmillan.
- Wells, C. & Wollack J. (2003). *An instructor's guide to understanding test reliability*. Testing & Evaluation Services, University of Wisconsin. <https://testing.wisc.edu/Reliability.pdf> (pristupljeno 13. siječnja 2024).
- Woodin, J. & Müller-Hartmann, A. (2013). Task-based teaching and assessment. U Byram, M. & Hu, A. (ur.), *Routledge Encyclopedia of Language Teaching and Learning* (687–692). Routledge.
- Zeidner, M. (1998). *Test Anxiety – The State of the Art*. Kluwer Academic Publishers.
- Živčić-Bećirević, I. (2003). Uloga automatskih misli i ispitne anksioznosti u uspjehu studenata. *Društvena istraživanja: časopis za opća društvena pitanja*, 5(67), 703–720.