

KAKOVOST V JEZIKOVNEM TESTIRANJU: MISIJA NEMOGOČE?

V prispevku bosta na kratko predstavljena javno veljavni izobraževalni program za odrasle Slovenščina kot drugi in tuji jezik, na katerem od leta 2015 temelji sistem preverjanja in certificiranja znanja slovenščine kot drugega in tujega jezika, in kontekst, v katerem je program nastal. Ob tem bodo kritično predstavljene in izpostavljene ključne strokovne dileme, še posebej tiste, ki se nanašajo na z njim povezano jezikovno testiranje. Sistem preverjanja znanja slovenščine kot drugega in tujega jezika je bil z novim programom razvit po strogih mednarodnih standardih kakovosti. Prispevek podrobneje predstavlja tri najzahtevnejše, in sicer so to primerljivost različic izpitolov, zanesljivost ocenjevanja in analiza podatkov.

Ključne besede: slovenščina kot drugi in tuji jezik, izobraževalni jezikovni programi, Skupni evropski jezikovni okvir, jezikovno testiranje, standardi kakovosti

0 Uvod

Povod za ta prispevek je – ob 100-letnici Univerze v Ljubljani – 25. obletnica uradne ustanovitve programa Izpitni center, ki deluje v okviru Centra za slovenščino kot drugi in tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani. Povod je torej povezan s formalnimi okoliščinami, razlogi pa so vsebinski. Vsaj dva lahko navedemo: prvi je sprejem javno veljavnega izobraževalnega programa za odrasle *Slovenščina kot drugi in tuji jezik*, na katerem temeljijo izpiti iz znanja slovenščine in ki strokovni javnosti še ni bil celovito predstavljen. Drugi pa je mednarodni znak kakovosti, ki so ga izpiti iz znanja slovenščine dobili po zahtevnem postopku zunanje evalvacije in tako postali primerljivi z izpiti, ki jih razvijajo in izvajajo velike institucije (*Goethejev inštitut, Cambridge English,*

Österreichisches Sprachdiplom, Università per Stranieri di Perugia idr.). Od obojega je minilo dovolj časa za kritičen razmislek in presojo o tem, kaj smo s tem pridobili in katere izzive prinašata.

1 Prelomnost programa *Slovenščina kot drugi in tuji jezik*

Leta 2011 smo dobili slovenski prevod dokumenta *Skupni evropski jezikovni okvir: učenje, poučevanje, ocenjevanje* (SEJO). Dokument, ki ga je leta 2001 izdal Svet Evrope, je bil razvit, poenostavljeno, z namenom doseči transparentnost in usklajenost pri učenju jezikov ter preverjanju in certificiranju jezikovnega znanja v različnih izobraževalnih kontekstih v Evropi (SEJO 2011: 27).¹

SEJO kot »odprt in prožen« dokument (SEJO 2011: 29) opisuje jezikovne zmožnosti na šestih temeljnih ravneh² in je bil zasnovan s ciljem, da bi bilo mogoče za vsakega govorca in govorko izdelati njegov oz. njen jezikovni profil. Kaj to pomeni? Običajno govorniki in govorki posamezne jezikovne dejavnosti obvladamo na različnih ravneh, v tujem jeziku na primer je pogosto naše razumevanje besedil na višji ravni kot zmožnost pisanja, lahko smo zelo učinkoviti pri pridobivanju informacij, veliko manjša pa je naša zmožnost pogajanja in podobno. Možnost takšnega jezikovnega profiliranja je morda komu bolj znana iz (samo)evalvacijskih evropskih jezikovnih listovnikov, ki so na voljo za različna starostna obdobja.³

Na SEJO je mogoče gledati kot na nedokončano delo (Skela 2011: 8). A kljub temu je večina institucij v EU, ki se ukvarjajo z jezikovnim testiranjem,⁴ do leta 2011 že uskladila svoje sisteme preverjanja in certificiranja jezikovnega znanja s tem dokumentom.⁵ V Izpitnem centru (IC) smo s temi postopki začeli v letu 2010, najprej s primerjavo opisnikov v tedaj veljavnem izobraževalnem programu *Slovenščina za tujce* in na njem temelječih izpitov iz znanja slovenščine na osnovni ravni⁶ z opisniki SEJO (podrobneje o opisu postopkov Ferbežar idr. 2014). Med letoma 2012 in 2014 je bil nato razvit nov izobraževalni program *Slovenščina kot drugi in tuji jezik* (SDTJ), ki je upošteval tako načela SEJO in opisnike v njem kot tudi dolgoletne izkušnje Centra za slovenščino s poučevanjem in preverjanjem znanja

¹ O slovenskih izkušnjah s SEJO gl. Lah 2014.

² Leta 2018 je bilo izvirnih šest ravni dopoljenih z opisniki ravni pod A1 in t. i. vmesnih ravni (po novem tako za A2, B1 in B2 obstajajo opisniki na spodnji in zgornji ravni), tako da je ravni skupno deset. <<https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>>. Dostop 24. 7. 2019.

³ Za slovenščino gl. <http://www.mizs.gov.si/si/delovna_podrocja/urad_za_razvoj_in_kakovost_izobrazevanja/jezikovno_izobrazevanje/evropski_jezikovni_portfolioevropski_jezikovni_listovnik_ejl/>. Dostop 24. 7. 2019.

⁴ Gl. spletno stran <www.alte.org>. Dostop 23. 7. 2019.

⁵ Za umestitev nacionalnih izpitov iz tujih jezikov na SEJO je poskrbel Državni izpitni center: <https://www.ric.si/ostalo/umestitve_izpitov_iz_tujih_jezikov_v_sejo/>. Dostop 24. 7. 2019.

⁶ Gl. letno poročilo Centra za slovenščino kot drugi in tuji jezik 2010 (<https://centerslo.si/wp-content/uploads/2015/10/LP_2010.pdf>). Dostop 24. 7. 2019).

SDTJ. Program je bil potrjen na Strokovnem svetu za izobraževanje odraslih in je z odredbo ministra za šolstvo v veljavo stopil septembra 2015.⁷

Program je zasnovan stopenjsko in upošteva tujega govorca in raven njegove samostojnosti pri sporazumevanju v slovenščini kot nepravem jeziku. Jezikovna zmožnost je na posameznih stopnjah opredeljena v obliki pozitivnih trditev (kaj govorec v slovenščini zna in zmore opraviti), tako kot to velja za SEJO. Program ima tri stopnje, njihovo primerljivost s SEJO kaže spodnja preglednica.

SDTJ	SEJO
/	A1 osnovni uporabnik: vstopna raven
Osnovna raven	A2 osnovni uporabnik: vmesna raven
	B1 samostojni uporabnik: raven sporazumevalnega praga
Višja raven	B2 samostojni uporabnik: višja raven
Raven odličnosti	C1 učinkoviti uporabnik: raven učinkovitosti
	C2 učinkoviti uporabnik: raven mojstrstva

Preglednica 1: Program SDTJ in SEJO.

Program SDTJ ni radikalno drugačen od prejšnjega programa *Slovenščina za tujce*, ki je že opredeljeval tri ravni. Njegova prelomnost je v usklajenosti s SEJO in kompleksno zasnovanem sistemu preverjanja znanja SDTJ: ta prinaša možnost jezikovnega profiliranja – jezikovni testi oz. izpiti⁸ na osnovni ravni in ravni odličnosti so namreč zasnovani progresivno (od A2 do B1 oz. od C1 do C2)⁹ – in tako bolj realno kaže posameznikovo jezikovno zmožnost v slovenščini.

V programu SDTJ opisani jezik je t. i. splošni jezik. To dejstvo odpira vprašanje, komu je v svoji splošnosti (en program za vse) v resnici namenjen, to pa je tudi problem SEJO kot univerzalnega dokumenta. Če že takšna splošnost ni problematična na osnovni ravni – ta opisuje raven jezikovne zmožnosti, ki jo za preživetje in osnovno samostojno delovanje v nekem jezikovnem okolju potrebujeta slehernik in slehernica – pa o tem kaže razmišljati ob opisovanju jezikovne zmožnosti na višji ravni in ravni odličnosti, kjer se govorke in govorci tudi jezikovno začemo specializirati (npr. za poklicne potrebe). Tu pa se odpira tudi vprašanje primernosti opisnikov: ali je o jezikovni zmožnosti, kakršna je pripisana uporabniku na ravni učinkovitosti in opisana zanj, sploh mogoče govoriti brez upoštevanja njegove

⁷ Uradni list RS, št. 6/15.

⁸ Besedi *izpit* in *test* sta v tem besedilu uporabljeni kot sopomenki; to velja tudi za zvezi *jezikovno testiranje* in *preverjanje jezikovnega znanja*.

⁹ Tako na podlagi doseženega rezultata na izpitu udeleženec oz. udeleženka dobiva potrdilo, na kateri ravni je njuna zmožnost razumevanja ob branju in poslušanju ter zmožnost pisnega in govornega sporazumevanja.

(visoke) izobrazbenosti? Ali torej tuji govorec, katerega izobrazba ni na (dovolj) visoki stopnji, glede na te opisnike pri sporazumevanju sploh lahko »postane« učinkovit? Vrnimo se k programu SDTJ: splošnost jezika velja tudi za preverjanje znanja, ki ga program določa v Izpitnem katalogu. Preverjanje znanja je po priporočilih SEJO operacionalizirano z jezikovnimi opravili, kot so na primer iskanje informacij ter primerjava izdelkov ali storitev na osnovni ravni, delanje zapiskov ob poslušanju na višji ravni in povzemanje informacij iz več besedil na ravni odličnosti.¹⁰ Posamezna opravila se izvajajo ob besedilih; avtorji izpitnih gradiv imajo natančne smernice za njihovo izbiro in izdelavo vprašanj ob njih. Smernice oz. izpitne specifikacije so bile pripravljene na podlagi SEJO, učnih gradiv in dolgoletnih izkušenj s poučevanjem in testiranjem SDTJ ter analize sporazumevalnih praks in potreb udeležencev in udeleženk izpitov (z vprašalnikom), torej tako »od zgoraj navzdol« kot »od spodaj navzgor«, kar je pri razvoju jezikovnih programov pri nas novost.¹¹

Prav z vključevanjem jezikovnih opravil poskušamo v čim večji meri zadostiti zahtevi po avtentičnosti, kolikor jo narava jezikovnega testa sploh dopušča: testi so namreč sami po sebi zunaj vsakršnega realnega sporazumevalnega konteksta in so kot taki neavtentični – avtentični pa so v pedagoškem kontekstu (Ferbežar 2016: 492).

Pri razvoju sistema preverjanja znanja SDTJ smo sledili tudi mednarodnim smernicam kakovostnega jezikovnega testiranja. To je pomenilo upoštevanje 17 standardov kakovosti, kot jih je opredelilo vplivno združenje evropskih jezikovnih testatorjev ALTE.¹² Ti standardi obsegajo pet sklopov oz. faz v procesu jezikovnega testiranja, in sicer

- 1) pripravo izpitnih gradiv (teoretični model, jasen namen izpita in ciljna publika, primerljivost različnih verzij izpitov na isti ravni, usposobljenost avtorjev gradiv, dokazljivost povezave s SEJO),
- 2) organizacijo in izvajanje izpitov (izvajalci in njihovo usposabljanje, zaupnost gradiva in njegova varna dostava, varovanje osebnih podatkov, podpora kandidatom s posebnimi potrebami),
- 3) točkovanje in ocenjevanje (natančnost in zanesljivost ocenjevanja, spremljanje ocenjevalcev),
- 4) analizo testov (zbiranje podatkov in analiza rezultatov za ugotavljanje težavnosti, občutljivosti, zanesljivosti testov) in

¹⁰ Gl. vzorčne teste na spletni strani Izpitnega centra Centra za slovenščino kot drugi in tuji jezik (IC) <<https://centerslo.si/izpiti/>>. Dostop 24. 7. 2019.

¹¹ Pridobivanje povratnih informacij udeležencev in udeleženk izpitov je običajna praksa tudi v tujini, v IC pa smo jih želeli vključiti tudi v druge faze testiranja. Tako je na primer v Svet IC, ki ga sestavljajo predstavniki različnih deležnikov in ki nadzoruje delo IC, imenovan tudi predstavnik udeleženk in udeležencev izpitov. Te prakse primerljive institucije v EU ne poznajo.

¹² IC v tem združenju že od leta 2000 zelo aktivno predstavlja Univerzo v Ljubljani, zadnjih deset let ima svojo predstavnico tudi v njegovih najvišjih organih.

- 5) komunikacijo z uporabniki (informiranje uporabnikov, posredovanje rezultatov in njihova interpretacija).¹³

Sklenemo lahko, da je program SDTJ zasnovan in se uresničuje v skladu z mednarodnimi standardi in razvojem stroke in je kot tak tudi mednarodno primerljiv. To dokazuje tudi znak kakovosti, ki je bil po strogi zunanji evalvaciji združenja ALTE septembra 2017 za pet let dodeljen izpitom iz znanja SDTJ na vseh treh ravneh.¹⁴

2 Trije najzahtevnejši standardi

V nadaljevanju se bomo osredotočili na tiste standarde, ki predstavljajo največji izziv tako pri razvoju sistema preverjanja znanja SDTJ kot pri njegovem vzdrževanju. To so primerljivost različnih verzij izpitov, zanesljivost ocenjevanja in analiza podatkov.

2.1 Primerljivost različnih verzij izpitov

Gre za enega izmed standardov iz prvega sklopa, povezan je z razvojem izpitnih gradiv oz. s tem, kako je posamični test sestavljen, katera opravila oz. naloge vsebuje, kakšna (kako dolga in jezikovno kompleksna) so izhodiščna besedila in kako zahtevna so vprašanja ob njih. Vsak test, četudi na isti ravni, je nov, sestavljajo ga različna besedila in vprašanja, verzije torej niso nikoli povsem enake. Primerljivost med njimi dosegamo z različnimi ukrepi, tu navajamo tri glavne.

Specifikacija izpita. Ta avtorjem in avtoricam izpitnih gradiv prinaša natančna navodila o tem, kaj se v testu meri (npr. branje, pisanje) in s katerimi opravili (npr. z iskanjem informacij in argumentov, z delanjem zapiskov ob poslušanju), kako dolgo naj bo posamezno besedilo, ki je izhodišče za opravilo, kako kompleksne naj bodo strukture v njem, kakšno naj bo besedišče, koliko vprašanj naj bo ob besedilu, na kateri ravni naj bodo in po čem naj sprašujejo (npr. po podatkih, po stališčih, po odnosu med ljudmi ipd.); specifikacija ima tudi seznam nezaželenih tem, da se njihovi avtorji v čim večji meri lahko izognejo pristranskosti (npr. glede na spol, izvor, prvi jezik). Vsaka raven ima svojo specifikacijo.

Sidranje. V vsakem novem testu je nekaj nalog, ki so bile vključene v katero od prejšnjih verzij in so se pokazale kot statistično ustrezne; take naloge predstavljajo t. i. sidra in so eno od zagotovil stabilnosti testov. Sidranje zahteva vzpostavitev t. i. banke nalog, to je zbirke vseh nalog, ki so jih vsebovali pretekli testi, z vsemi

¹³ <https://www.alte.org/resources/Documents/minimum_standards_en.pdf>, dostop 24. 7. 2019; načela v slovenskem prevodu: <https://centerslo.si/wp-content/uploads/2017/05/Minimalni-standardi_za-splet-2017.pdf>, dostop 24. 7. 2019.

¹⁴ Zunanja evalvacija zahteva natančen opis vseh 17 načel in predložitve dokazov (v obliki dokumentov, poročil, statističnih podatkov), kako se uresničujejo v praksi. Zunanja evalvacija vsebuje priporočila za izboljšave in ob ponovnem preverjanju po petih letih naj bi bila ob ponovni evalvaciji ta priporočila izpolnjena.

statističnimi podatki, ki povedo, ali je posamezna naloga za sidro primerna ali ne (gl. tudi v nadaljevanju).

Določitev mejne vrednosti in redno preverjanje, ali je ustrezno postavljena. Mejna vrednost je v programu SDTJ postavljena na 60 %. Izpit torej udeleženka ali udeleženec opravita, če dosežeta ta odstotek. Mejna vrednost za branje in poslušanje je bila postavljena z uporabo verjetnosti odziva v vrednosti 0,67. Uporabljena je bila t. i. metoda zvezka: posamezna vprašanja so bila glede na analizo rezultatov po klasični testni teoriji razvrščena od najlažjega do najtežjega. Deset strokovnjakov se je najprej seznanilo z ustrezno ravno in se dogovorilo o sprejemljivem govorcu, to je govorcu, katerega jezikovna zmožnost v slovenščini »ravno že« zadostuje za določeno raven oz. je zanjo sprejemljiva. Nato so v zvezku z vprašanji postavili mejo, in sicer so označili tisto vprašanje, na katero bi po njihovem mnenju dva od treh takšnih govorcev še lahko odgovorila (nadaljnjih vprašanj pa niso več brali). Postopek je bil izveden po priporočilih, opisanih v priložniku za uvrščanje izpitov na ravni SEJO (Manual 2009), za vsako raven posebej (gl. tudi Ferbežar idr. 2014).

Za pisanje in govorjenje smo določili za posamezno raven »tipično« produkcijo. Gre za vzorčno produkcijo oz. merilo (angl. *benchmark*), tj. govorno in pisno produkcijo takšnega uporabnika slovenščine, ki izkazuje že sprejemljivo jezikovno zmožnost za določeno raven oz. za izvedbo določenega opravila na tej ravni (o konceptu minimalno sprejemljivega uporabnika gl. Manual 2009: 62; tudi Ferbežar idr. 2014: 281, Ferbežar 2016: 494, op. 11).

Če se za trenutek pomudimo pri prej omenjenem jezikovnem profiliranju: udeleženke in udeleženci izpita na osnovni ravni izpit opravijo, če je njihov rezultat pri vsakem od štirih podtestov (Branje, Poslušanje, Pisanje, Govorjenje) najmanj 60 %; če pa dosežejo 85-odstotni rezultat ali več, to pomeni, da je njihova jezikovna zmožnost v slovenščini pri posameznem podtestu na ravni B1 SEJO. Pri tem izpitu sta bili torej postavljeni dve mejni vrednosti: za A2 in za B1, kot je razvidno tudi iz Preglednice 1. To je bilo mogoče, ker je izpit na osnovni ravni zasnovan progresivno. Podobno velja za izpit na ravni odličnosti, izpit na višji ravni pa jezikovno zmožnost meri na le eni ravni (B2 SEJO), tudi mejna vrednost je ena.

Preverjanje mejne vrednosti pomeni, da se od časa do časa vsi opisani postopki ponovijo. To je s stališča vzdrževanja sistema precej zamudno in drago, saj mora biti, če naj bodo postopki tudi relevantni, v vsakega od njih vključenih najmanj po 10 strokovnjakov (po možnosti različnih profilov, ne le s področja SDTJ). Vendar pa jih je zaradi primerljivosti različnih verzij testov in s tem tudi poštenosti oz. enakovredne obravnave udeležencev izpitov treba od časa do časa ponoviti.

2.2 Zanesljivost ocenjevanja

Eden izmed najpomembnejših standardov za zagotavljanje kakovosti jezikovnega testiranja je zanesljivost ocenjevanja. Statistična analiza kaže, da imajo izpiti na

osnovni ravni visok indeks zanesljivosti (Cronbachov koeficient alfa je 0,93 ali več) in so pri oceni zanesljivosti tudi precej stabilni. Izpit na višji ravni ima indeks zanesljivosti nekoliko nižji (0,86) – prvič zato, ker je zanesljivost na višjih ravneh znanja na splošno težje doseči,¹⁵ drugič pa tudi zaradi majhnega števila kandidatov (od 15 do 45 na posameznem roku). Raven odličnosti pa tako zaradi formata izpita (integrirani izpit, pri katerem se receptivni dejavnosti povezujeta s produktivnima)¹⁶ kot zaradi premajhnega števila kandidatov za zdaj ne dopušča kvantitativnih analiz. Zato skušamo zanesljivost pri tem izpitu med drugim dosegati z dvojnimi ali trojnimi ocenjevanjem po natančno določenih merilih in z omejenim številom ocenjevalk in ocenjevalcev.

V nadaljevanju predstavljamo ukrepe, ki smo jih v IC sprejeli za osnovno raven. Več ukrepov zahteva dejstvo, da se izpit na osnovni ravni izvaja na 14 institucijah po vsej Sloveniji in je v ocenjevanje vključenih približno 90 ocenjevalk in ocenjevalcev, ki morajo biti pri ocenjevanju čim bolj enotni. Ti ukrepi so

- 1) uvodno usposabljanje izvajalcev,
- 2) navodila za točkovanje in ocenjevanje,
- 3) standardizacija pred ocenjevanjem oz. t. i. umerjanje na podlagi vzorčnih pisnih in govornih produkcij,
- 4) točkovalne šablone, ocenjevalne lestvice, dvojno ocenjevanje pisne in govorne produkcije, centralno ocenjevanje pisne produkcije,
- 5) standardizacijski seminarji,
- 6) spremljanje ocenjevalcev (raziskave notranje skladnosti ocenjevanja, tj. pri istem ocenjevalcu, in ujemanja med ocenjevalci, sprotno spremljanje ocenjevalcev).

Če na kratko povzamemo, je namen vseh teh ukrepov enotno razumevanje in interpretacija ocenjevalnih meril, še posebej pri produktivnih delih izpita, tj. pisanju in govorjenju, kjer gre za subjektivno presojo ocenjevalk in ocenjevalcev, končni cilj pa je poštena obravnava udeleženk in udeležencev izpitov.

Bralkam in bralcem za boljšo prestavo tu natančneje predstavljamo standardizacijo oziroma umerjanje pred ocenjevanjem (ukrep 3), ki je manj poznano in ga le redko kot obvezen ukrep pred vsakim ocenjevanjem poznajo tudi drugod. Gre za to, da si ocenjevalke in ocenjevalci pred začetkom ocenjevanja govorne produkcije (torej pred ustnim izpitom) natančno pogledajo vzorčno produkcijo, tj. videoposnetek sprejemljivega uporabnika, ki pri govorjenju ravno že dosega raven A2. Pred vsakim ocenjevanjem pisanja pa se seznanijo z vzorčnimi pisnimi produkcijami na ravni A2, pod njo in nad njo. Takšen postopek nekako uravnava strogost ocenjevalcev, saj »merilo« ni prvi udeleženec izpita (oz. tista produkcija, ki se jo ocenjuje najprej),

¹⁵ V tem primeru je skupina oseb, ki se udeležuje izpitov, bolj homogena (npr. po izobrazbi); večja ko je podobnost med njimi, težje jih je razločevati med seboj, kar pa pomeni tudi nižjo zanesljivost.

¹⁶ Gl. vzorčni test za raven odličnosti <<https://centerslo.si/izpiti/izpiti-iz-znanja-slovenscine/izpit-na-ravni-odlicnosti/vzorcni-test/>>. Dostop 24. 7. 2019.

ampak nek za posamezno raven tipični govorec oz. njegova produkcija. Če bi bila namreč prva produkcija, ki se jo ocenjuje, nad pričakovano ravno, bi to lahko imelo za posledico strožje ocenjevanje vseh naslednjih produkcij – in nasprotno (če bi bila pod mejo pričakovanega, bi bilo seveda to za udeležence in udeleženke izpitov mnogo manj usodno).

Tu se odpira vprašanje tipičnega govorca oz. za posamezno raven tipične produkcije. Pri sporazumevanju težko govorimo o tipičnosti, sleherni govorec ima svoje posebnosti, ki so odvisne od njegovega prvega jezika, izobrazbe, učnega stila in učnih praks, socialnih in drugih okoliščin idr. Zato je ustrezno merilo oz. vzorčno produkcijo – še posebej pri govorjenju – izjemno težko najti. Za to morajo biti izpolnjeni najrazličnejši pogoji: poleg tega, da govorka ali govorec ravno že ustreza ravni, za katero potrebujemo merilo (o konceptu že sprejemljivega govorca gl. pogl. 2.1), mora biti dovolj reprezentativen za ciljno skupino. Najštevilčnejšo, in s tem najbolj tipično skupino v primeru izpitov iz znanja SDTJ predstavljajo govorce južnoslovanskih jezikov.¹⁷ Produkcija izbranega govorca mora biti poleg tega enakomerna (angl. *flat*), kar pomeni, da so vse ocenjevane kategorije (besedišče, jezikovne strukture, izgovor itd.) na približno enaki ravni, nenazadnje pa mora govorec govoriti dovolj razločno, da se ga na posnetku lahko razume. In navsezadnje: tak govorec mora dovoliti, da se ga posname in njegov posnetek uporabi za izobraževalne namene. Merila potrebujemo za vse ravni. Poleg predstavnika prevladujoče skupine govorcev pa vsaj za osnovno raven, kjer je zaradi velikega števila ocenjevalcev enotnost mnogo težje doseči, potrebujemo tudi produkcijo govorcev kakih bolj oddaljenih jezikov. Za IC vsekakor velik izziv.

Vse produkcije se ocenjujejo dvojno, s tem – ob dejstvu, da gre za subjektivno presojo ocenjevalcev – skušamo zagotoviti vsaj približno objektivnost: govorne produkcije se ocenjujejo pri izvajalcu, in sicer celostno in po analitični lestvici, oba ocenjevalca se morata za končno oceno uskladiti. Pisne produkcije se ocenjujejo analitično, in sicer enkrat pri zunanjem izvajalcu, drugič v IC (če je izvajalec IC, obakrat v IC), v primeru razhajanj med ocenami IC poskrbi za tretjo oceno. V IC tehtamo možnost celostnega ocenjevanja tudi pisne produkcije. To se namreč, vsaj na tako nizki ravni, kot je A2, zdi mnogo bolj življenjsko. Z analitično lestvico se namreč pisno in govorno produkcijo vrednoti po jezikoslovnih kategorijah (npr. besedišče, jezikovna pravilnost, izgovor), a na ravni A2 je smiselno dati prednost izpolnitvi naloge in sporočilu, ne pa toliko temu, kako jezikovno pravilno je to sporočilo posredovano (prednost naj bi torej imela vsebina pred obliko; gl. tudi Ferbežar 2016).

Kljub nenehni skrbi za skladnost ocenjevanja (angl. *intra-rater agreement*; posamezni ocenjevalec naj bi v različnih časovnih intervalih isto produkcijo ocenil enako) in ujemanje med ocenjevalci (angl. *inter-rater agreement*; vsi ocenjevalci enako ocenijo isto produkcijo) pa prihaja do razhajanj – kot že rečeno, produkcijo je za zdaj

¹⁷ Gl. letna poročila Centra za slovenščino kot drugi in tuji jezik, <<https://centerslo.si/knjige/letna-porocila/>>. Dostop 24. 7. 2019.

še težko vrednotiti drugače kot s subjektivno presojo.¹⁸ To je posledica različnega razumevanja ocenjevalnih meril in drugih napak ocenjevanja, med katere spada tudi pristranskost (prispevek o tem bomo strokovni javnosti predstavili ob drugi priložnosti).

Skladnost ocenjevanja in ujemanje med ocenjevalci se na osnovni ravni redno spremlja (s seminarji in z obiski pri izvajalcih), število ocenjevalcev omogoča tudi širšo, statistično podprto raziskavo. Na obeh višjih ravneh pa omejenost števila ocenjevalcev (ki v primeru izpitov z maloštevilnimi udeleženci pomeni prednost) ne omogoča uporabe statističnih metod, zato se v IC raziskovanja zanesljivosti ocenjevanja lotevamo kvalitativno.

2.3 Analiza rezultatov

V IC se sistematično zbirajo podatki o udeležencih in udeleženkah izpitov: poleg osebnih podatkov tudi podatki o prvem jeziku, stopnji izobrazbe, učenju slovenščine, namenu opravljanja izpita, ki jih uporabljamo za izdelavo letnih poročil IC (gl. op. 17), izpitne rezultate pa na Centru za psihološko merjenje svetovanje FF UL pri Oddelku za psihologijo¹⁹ statistično analizirajo, in sicer po klasični testni metodi. To velja za izpite na osnovni in višji ravni; za potrebe zunanje evalvacije je bila za izpite na osnovni ravni izvedena tudi analiza po Raschevem modelu.²⁰

Analizirajo se težavnost nalog in posameznih vprašanj v njih, njihova občutljivost in zanesljivost ter pristranskost glede na spol in udeležbo na tečaju. Ti podatki so izhodišče za redakcijo nalog oz. vprašanj v njih, naloge pa se nato s statističnimi podatki vred vključijo v banko nalog kot potencialna sidra za naslednje verzije testov.

Podatki za zdaj kažejo, da so izpiti na osnovni in višji ustrezno pripravljene in da naloge in vprašanja v njih dobro merijo zmožnosti in znanja, ki naj bi jih merili. Glede na ciljno publiko so razmeroma lahki: večina udeleženk in udeležencev izpitov je namreč govorcev katerega od slovenščini sorodnih jezikov, pozitivni prenos omogoča razmeroma visoko raven receptivnih zmožnosti. Nekoliko drugače je pri produktivnih zmožnostih (gl. letna poročila Centra za slovenščino kot drugi in tuji jezik).

Podatke bomo v IC zaradi vzdrževanja kakovosti statistično analizirali tudi v prihodnje; skušali pa bomo skrajšati čas med izvedbo izpita in analizo ter poenostaviti postopke, Raschev model pa bomo uporabili le občasno, kadar bo na voljo dovolj

¹⁸ Večjo objektivnost bo v prihodnje mogoče zagotoviti s strojnimi ocenjevanjem. Za angleščino kot drugi in tuji jezik so pri *Cambridge English* za pripravo na izpit že razvili orodje za samoevalvacijo, ki uporabnikovo pisno produkcijo uvrsti na ustrezno raven SEJO in označi dele besedila, ki zahtevajo popravke. Gl. <<https://writeandimprove.com/>>. Dostop 24. 7. 2019

¹⁹ <<http://cpms.si/>>. Dostop 24. 7. 2019.

²⁰ Ta je mogoča ob ustreznem številu udeležencev izpitov, ki je pri nas zagotovljeno le na osnovni ravni. Za ustrezen numerus v IC poskrbimo z uporabo istih testov na različnih izpitnih rokih.

podatkov. Zaželeno je, da bi statistična analiza imela bolj neposreden vpliv tudi na rezultate udeleženk in udeležencev izpitov (sploh v primerih, če se katero od vprašanj izkaže kot pretežno), da bi torej dosežke posameznikov prilagajali glede na rezultate po statistični analizi.²¹ A to v sedanjem sistemu, ko je mejna vrednost opredeljena v javno veljavnem programu, na katerem temeljijo izpiti, ni mogoče. Ob prihodnji nadgradnji sistema pa bi o takem vplivu vsekakor morali resno razmisliti.

Rezultatov izpitov na ravni odličnosti še ni mogoče analizirati kvantitativno, saj je število udeležencev na posameznem izpitnem roku premajhno, poleg tega pa tako analizo otežuje tudi zelo specifičen format izpita: gre za integrirani izpit, kjer se kombinirata branje in pisanje ter poslušanje in govorjenje, ocenjevanje pa poteka po opisnikih v analitičnih lestvicah (o oceni posamezne ocenjevane kategorije presoja ocenjevalec oz. ocenjevalka, kot velja tudi sicer za ocenjevanje govorjenja in pisanja). Dosežki na ravni odličnosti se za zdaj zato analizirajo kvalitativno.

3 Nadaljnji izzivi

Čas in prostor ne dopuščata natančnejše predstavitve vseh 17 standardov kakovosti. Tu smo izpostavili le tiste, ki so vsebinsko in organizacijsko najtežje dosegljivi, ker so zamudni in dragi – a so za vzdrževanje sistema in zagotavljanje njegove kakovosti nujni. Navsezadnje gre za izpite za uradne namene, pri katerih se odloča o pozitivnih in – kar je še veliko usodnejše – negativnih posledicah za posameznikovo življenje in delo v Sloveniji. Vsi postopki zahtevajo tudi transparentnost, biti morajo dokazljivi, to pa je povezano tudi z raziskovanjem. Zato se bodo naloge IC v prihodnje bolj usmerile prav na to področje: poleg razvoja gradiv, ki bo namenjeno udeleženkam in udeležencem za pripravo na izpite, in razvoja izpitov za posebne namene (med drugim na primer za zdravstvene delavce) bomo v IC bolj poglobljeno analizirali sporazumevalne prakse in potrebe naših uporabnikov, natančneje bomo raziskali, kje (pri katerih kategorijah) se ocenjevalke in ocenjevalci najbolj razhajajo in zakaj, ter kakšne so posledice izpitov iz znanja SDTJ oz. dosežkov na njih.

Dosežki na izpiti imajo za tiste, ki se jih udeležujejo, lahko daljnosežne posledice, še posebej, če so nizki. In dosežki v opisanem sistemu so nižji, kot bi jih pričakovali glede na dejstvo, da se izpitov v največji meri udeležujejo govorke in govorci slovenščini sorodnih jezikov (gl. letna poročila Centra za slovenščino kot drugi in tuji jezik). A vprašanje dosežkov in njihovih posledic presega okvire tega prispevka in se mu bomo podrobneje posvetili ob kaki drugi priložnosti.²²

Dejstvo ostaja, da izvajanje izpitov iz znanja SDTJ s sabo prinaša tudi odgovornosti, povezane s kakovostjo in pravičnostjo. Vsakršni testi so sicer v svojem bistvu

²¹ Ali pa bi posameznikove dosežke prilagajali glede na povprečni dosežek, kot to velja na primer za preizkuse pri splošni in poklicni maturi ter nacionalnem preverjanju znanja (<<https://www.ric.si/>>. Dostop 24. 7. 2019).

²² O dosežkih na izpiti iz znanja SDTJ pri branju gl. npr. Ferbežar 2018.

nepošteni in nepravični: ker dopuščajo, da na podlagi omejenega števila odgovorov sklepamo o določeni zmožnosti, ki je nikoli ni mogoče izmeriti v celoti, ker so vprašanja pogosto zastavljena tako, da zavajajo (na primer vprašanja izbirnega tipa s ponujanjem nepravilnih možnosti), in ker so morda postopki ob tem premalo standardizirani, ker se jih udeležujejo tako raznoliki posamezniki in – če drugega ne, ker testi vedno merijo tudi zmožnosti, ki niso neposredni predmet merjenja (npr. strategije reševanja nalog). V globalnem svetu pa delujejo nepošteno tudi zato, ker nekako diktirajo (nerealno) enojezičnost. Družbe so pač fluidne, kulturno in jezikovno pluralne, posameznike v njih pa zaznamuje raznolikost njihovih (razno)jezikovnih repertoarjev – že od nekdaj. V takšnem *sojezičju* (Stabej 2019) se govorke in govorci hote ali nehoti upiramo vsakršni enojezičnosti z uporabo hibridnih in fluidnih praks, na primer kodnega preklapljanja in medjezikovnega prepletanja (angl. *translanguaging*).²³ Takšne prakse pri reševanju (eno)jezikovnih testov seveda niso zaželene in se kaznujejo.

Zato vprašanje kakovosti jezikovnega testiranja nima dokončnega odgovora. Prav tako kot dinamična narava posameznikove jezikovne zmožnosti ne dopušča, da bi jo dokončno opisali – gre pač za kontinuum. »Dobrih« jezikovnih izpitov ni, lahko so le bolj ali manj uporabni za namen, za katerega so bili razviti, in bolj ali manj pravični do tistih, ki se jih udeležujejo.²⁴ Če naj bi bili pravični bolj, to zahteva nenehno prizadevanje oz. prevzemanje odgovornosti, da bodo izpiti pripravljene v skladu s strokovnimi zahtevami in etičnimi načeli, predvsem pa tudi odgovornosti za posledice njihovih rezultatov – tako za posameznike kot za širšo skupnost.

Literatura

Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors. Provisional Edition. Council of Europe, September 2017. <<https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>>. Dostop 24. 7. 2019.

Ferbežar, Ina, 2013: (Jezikovni) testi: moč argumenta ali argument moči?. V: Bjelčević, Aleksander (ur.). *Etika v slovenskem jeziku, literaturi in kulturi: zbornik predavanj*. Ljubljana: Znanstvena založba Filozofske fakultete. 99–105. <http://www.centerslo.net/files/file/ssjlk/49_SSJLK/ferbezar.pdf>. Dostop 24. 7. 2019.

Ferbežar, Ina, 2016: Kako realno je idealno? O konceptu avtentičnosti pri pouku slovenščine ob Toporišičevem Zakaj ne po slovensko. V: Kržišnik, Erika, in Hladnik, Miran (ur.). *Toporišičeva obdobja*. Ljubljana: Znanstvena založba Filozofske fakultete. 489–496. <<http://centerslo.si/wp-content/uploads/2016/11/Ferbezar.pdf>>. Dostop 23. 7. 2019.

²³ Koncepta, v angleščini imenovan *translanguaging* (García in Wei 2013), si ne predstavljamo zgolj tehnično, kot prenašanje prvin iz različnih jezikov med oblikovanjem (»tkanjem«) besedila oz. tvorenjem njegovega pomena; v te prvine so vpletene tudi izkušnje, ki si jih govorke in govorci pridobivamo v (jezikovni) interakciji s svetom – torej gre zares za nekakšno medjezikovno (in izkušnjsko) prepletanje. Angleški *languaging* naj bi v 70. letih prejšnjega stoletja prva uporabila čilska biologa in filozofa Humberto R. Maturana in Francisco J. Varela v svojem Drevesu spoznanja (García in Wei 2013: 7); Uroš Kalčič je izraz v slovenščino prevedel kot *jezikovanje* (Maturana in Varela 1998: 196).

²⁴ O etiki v jezikovnem testiranju gl. Ferbežar 2013.

Ferbežar, Ina, 2018. O besedilnih spretnostih tujih govorcev slovenščine. V: Javrh, Petra (ur.). *Spretnosti odraslih*. Ljubljana: Andragoški center Slovenije. 61–76.

Ferbežar, Ina, Piriš Svetina, Nataša, in Lutar, Mateja, 2014: The Common European Framework of Reference: a reference for Slovene. V: Lah, Meta (ur.). *Skupni evropski jezikovni okvir – navzkrižni pogledi. Cadre européen commun de référence pour les langues – regards croisés. Lingusitica* 54/1. Ljubljana: Znanstvena založba Filozofske fakultete. 277–291.

<<https://revije.ff.uni-lj.si/linguistica/article/view/2615/2738>>. Dostop 24. 7. 2019.

García, Ofelia, in Wei, Li, 2014: *Translanguaging*. Language, Bilingualism and Education. New York: Palgrave Macmillan.

Lah, Meta (ur.), 2014. *Skupni evropski jezikovni okvir – navzkrižni pogledi. Cadre européen commun de référence pour les langues – regards croisés. Lingusitica* 54/1. Ljubljana: Znanstvena založba Filozofske fakultete.

Manual = Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Svet Evrope 2009. <<https://rm.coe.int/1680667a2d>>. Dostop 24. 7. 2019.

Maturana, Humberto R., in Varela, Francisco J., 1998: *Drevo spoznanja*. Ljubljana: Studia humanitatis.

Skupni evropski jezikovni okvir: učenje poučevanje, ocenjevanje, 2011. Ljubljana: Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva. <http://www.mizks.gov.si/fileadmin/mizks.gov.si/pageuploads/podrocje/razvoj_solstva/Jeziki/Publikacija_SEJO_komplet.pdf>. Dostop 24. 7. 2019.

Skela, Janez 2011. Spremna beseda. *Skupni evropski jezikovni okvir: učenje poučevanje, ocenjevanje*. Ljubljana: Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva. 5–14.

Slovenščina kot drugi in tuji jezik. Izobraževalni program za odrasle. Ljubljana 2014. <http://arhiv.acs.si/programoteka/Slovenscina_kot_drugi_in_tuji_jezik.pdf>. Dostop 24. 7. 2019.

Slovenščina za tujce. Program. Ljubljana 2000. <http://programoteka.acs.si/PDF/slo_za_tujce.pdf>. Dostop 24. 7. 2019.

Stabej, Marko, 2019. Sojezičje: utopija ali resnična možnost? V: Požgaj Hadži, Vesna, Ljubešič, Marko, in Zihlerl, Jerica (ur.). *Ususret dijalogu: zbornik posvečen Mirjani Benjak*. Novigrad/Cittanova: Museo Lapidarium, Pula: Filozofski fakultet, Sveučilište Jurija Dobrile u Puli, Ljubljana: Znanstvena založba Filozofske fakultete. 281–290.