

Psihometrijska provjera adaptacije Verbalnog testa kognitivne refleksivnosti u Bosni i Hercegovini

Vesna Kovačić♦*, Nina Bobić, Siniša Subotić*****

* Udruga roditelja djece sa posebnim potrebama

„Put u život“ (PUŽ), Livno

** DDC Multilingual Solutions, Banja Luka

*** Katedra za psihologiju & Prirodno-matematički fakultet,
Univerzitet u Banjoj Luci

SAŽETAK: Kognitivna refleksivnost podrazumijeva sposobnost potiskivanja pogrešne početne intuicije. Ovaj konstrukt se tradicionalno mjerio numeričkim zadacima, zbog čega je razvijen verbalni CRT-V test, sa ciljem preciznijeg razdvajanja u odnosu na domen matematičkih sposobnosti. U ovoj studiji, CRT-V test je provjeren na 318 ispitanika iz BiH (45% ženskog pola; 65% srednjoškolaca, 35% studenata). Uporedo s CRT-V zadani su testovi verbalnog i matričnog rezonovanja. Jednofaktorski model CRT-V pokazao je dobar fit i zadovoljavajuću konvergentnu validnost. Prosječni skorovi su umjereni veći kod studenata u odnosu na srednjoškolce i blago veći kod žena u



vesnakovacic87@gmail.com

odnosu na muškarce. Uočena je nekolicina psihometrijskih problema. Većina zadataka je previše laka (70%+ uspješnosti na 6/10 zadataka), a informativnost im je uglavnom locirana ispod prosjeka latentne crte kognitivne refleksivnosti. Nekoliko zadataka pokazuje metrijske pristrasnosti (DIF), a prisutne su i terminološko-kulturne specifičnosti neoptimalne za multikulturalno bosanskohercegovačko okruženje. Iako trenutna verzija CRT-V može biti korisna, poželjna je temeljnija revizija ili zamjena određenih zadataka radi smanjenja pristrasnosti i povećanja težine i informativnosti, čime bi se poboljšala praktična vrijednost testa.

KLJUČNE RIJEČI: kognitivna refleksivnost, psihometrijska provjera testa, kulturalna primjerenost testa, klasična testna teorija (CTT), teorija odgovora na stavke (IRT).

Uvod

Kognitivna refleksivnost odnosi se na sposobnost pojedinca da potisne intuitivan, mahom pogrešan prvi odgovor na problem i umjesto toga uloži dodatni napor u „refleksiju“, tj. dublje razmišljanje o zadatku (Bubić & Erceg, 2015; Frederick, 2005; Otero et al., 2022; Sirota et al., 2021). Kognitivna refleksivnost se dovodi u vezu sa Modelom dualnih procesa, kojim se postuliraju dva tipa kognitivnih procesa – Tip 1 podrazumijeva brzo, heurističko i intuitivno procesuiranje, dok Tip 2 označava sporije, analitičko i promišljeno rezonovanje (Evans & Stanovich, 2013; Otero et al., 2022; Sinayev & Peters, 2015). Kroz inhibiciju početne, automatske intuicije, kognitivna refleksivnost omogućava preciznije prosuđivanje i donošenje racionalnijih odluka u raznovrsnim situacijama.

Istraživanja pokazuju da pojedinci s višim nivoom kognitivne refleksivnosti rjeđe podliježu tipičnim kognitivnim pristrasnostima, da bolje procjenjuju rizik i manje impulsivno reaguju (Frederick, 2005; Krašić, 2018; Toplak et al., 2011). Dodatno, veća kognitivna refleksivnost predviđa povoljnije ishode u akademskim i radnim kontekstima (Corgnet et al., 2015; Otero et al., 2021; Salgado et al., 2019;

Toplak et al., 2011). Kognitivna refleksivnost dijeli veliki dio varijanse s opštim faktorom inteligencije i numeričkim sposobnostima – prema procjenama oko 70% – što implicira da ne predstavlja u potpunosti poseban konstrukt, već radije aspekt intelektualnog funkcionisanja, koji može biti od koristi u specifičnim praktičnim situacijama, u kojima treba donijeti neke brze (npr. selekcijske) odluke (Otero et al., 2022).

Mjerenje kognitivne refleksivnosti

Najpoznatiji test kognitivne refleksivnosti – CRT (engl. Cognitive Reflection Test), razvio je Frederick (2005). U klasičnom obliku, CRT sadrži tri numeričke zagonetke, za koje je, neiznenađujuće, ustanovljeno da dijelom mjere i matematičko-numeričke sposobnosti (Campitelli & Gerrans, 2014; Otero et al., 2022), što predstavlja problem u kontekstu sadržinske i diskriminativne valjanosti ove mjere. Pored toga, samo tri zadatka mogu biti nedovoljna za zadovoljavajuću internu konzistentnost, diskriminativnost ili stabilnu faktorsku strukturu (npr., Sirota et al., 2021). Zbog toga su kasnije nastale duže forme ovog testa (Thomson & Oppenheimer, 2016; Toplak et al., 2013).

Uprkos proširivanju, numeričke verzije CRT-a i dalje pokazuju relativno visok nivo preklapanja sa matematičkim kompetencijama, a ustanovljene su i polne pristranosti testa u korist muškaraca (Avram, 2018; Brañas-Garza et al., 2019), koje izgleda ne proističu iz stvarnih razlika u samoj kognitivnoj refleksivnosti, već iz razlika u nivou matematičke sposobnosti i matematičke anksioznosti (Juanichich et al., 2020). Kako bi se prevazišli ovi nedostaci testa, Sirota et al. (2021) konstruisali su verbalni test kognitivne refleksivnosti – CRT-V. U CRT-V formi se koriste kratki verbalni problemski zadaci, koji ne zahtijevaju računanje, ali takođe potiču ispitanika da uoči netačnu „intuitivnu zamku“. Dosadašnje validacije ove verbalne verzije testa ukazuju na adekvatna psihometrijska svojstva, manju zavisnost od numeričkih vještina, uz očuvanu sposobnost predviđanja sličnih ishoda kao i klasični, numerički CRT (Sirota et al., 2021). Treba naglasiti da postoje i specifični oblici CRT-a namijenjeni za djecu (npr. CRT-D; Shtulman & Young, 2023), čije su stavke takođe date u verbalnom formatu i prilagođene su kognitivnom nivou djece. Sve ovo sugeriše da se kognitivna

refleksivnost potencijalno može mjeriti različitim zadacima, uključujući i one bez numeričke komponente.

Viši skorovi na mjerama kognitivne refleksivnosti ukazuju na bolje inhibiranje intuitivno pogrešnih odgovora i upuštanje u dublje promišljanje, čak i kada zadatak formalno ne zahtijeva složene proračune ili apstraktnu matematiku. Ipak, faktori poput verbalnih sposobnosti, poznavanja jezika ili kulturalnih osobnosti mogu da utiču na uspješnost rješavanja i verbalnog, nenumeričkog CRT-a (Sirota et al., 2021).

Cilj istraživanja: razlozi za ispitivanje CRT-V u Bosni i Hercegovini

Sve referentne postojeće mjere kognitivne refleksivnosti, uključujući i verbalni CRT-V, koji je predmet ispitivanja u ovom radu, nastale su i primarno su validirane na engleskom govornom području. S obzirom na samu formu i sadržaj CRT-V zadataka, koji su uobličeni kao kratke verbalne „mozgalice“ i koje, uz netačnu intuitivnu privlačnost, sadrže i terminološke i svjetonazorske specifičnosti izvorno engleskog podneblja, Sirota et al. (2021) naglašavaju da su kulturalni kontekst i nivo familijarnosti sa zadacima važni aspekti koje je neophodno uzeti u obzir prilikom provjere, adaptacije i primjene testa na drugim kulturama. Stoga adekvatan prevod i lokalizacija zadataka iziskuje prilagođavanje leksičkih, geografskih i kulturoloških elemenata (npr. nazivi država, uobičajeno kulturalno-religijsko znanje i sl.), kako bi se izbjegli pristrasnost i nerazumijevanje kod ispitanika. Ovo se čini naročito važnim u multikulturalnom i multietničkom kontekstu Bosne i Hercegovine (BiH), gdje koegzistiraju različite religije, jezičke varijante i identitetske odrednice jezika tri konstitutivna naroda. Ako se CRT-V test valjano adaptira ciljanim populacijama, to bi moglo donijeti višestruke koristi, s obzirom na to da su verbalne verzije zadataka kognitivne refleksivnosti manje zavisne od matematičkih sposobnosti i pogodnije za širi spektar ispitanika, uključujući srednjoškolce i studente različitih fakultetskih usmjerenja (Shtulman & Young, 2023; Sirota et al., 2021). Takođe, prilikom adaptacije i provjere verbalnog testa, trebalo bi uzeti u obzir i provjeriti postojanje polnih razlika i pristrasnosti svojstvenih za numeričke verzije

testova kognitivne refleksivnosti (Avram, 2018; Brañas-Garza et al., 2019; Juan-chich et al., 2020), kao i potencijalne razlike između obrazovno-uzrasnih skupina, poput npr. srednjoškolaca i studenata (Avram, 2018; Sirota et al., 2021), tj. trebalo bi provjeriti da li su zadaci jednako prikladni za različite kategorije ispitanika.

U skladu s navedenom argumentacijom, cilj ovog rada jeste vršenje prvog koraka u psihometrijskoj provjeri verbalnog testa kognitivne refleksivnosti (CRT-V), prilagođenog za bosanskohercegovačko govorno područje. Konkretno, ispituju se: (1) faktorska struktura i pouzdanost; (2) konvergentna valjanost, tj. odnos CRT-V postignuća sa drugim kognitivnim varijablama; (3) nivo težine stavki i testa i njihova informativnost na različitim intervalima latentne crte kognitivne refleksivnosti; (4) razlike po polu i obrazovnom statusu (srednjoškolci u odnosu na studente), prije svega u funkciji provjere moguće metrijske pristrasnosti testa za određene kategorije navedenih sociodemografskih grupa.

U vezi sa ranijim provjerama konvergentne valjanosti, prethodna istraživanja (Avram, 2018; Brañas-Garza et al., 2019; Sirota et al., 2021) sugerišu povezanost kognitivne refleksivnosti sa različitim mjerama inteligencije, uključujući matrično, numeričko i verbalno rezonovanje. U kontekstu konvergentne valjanosti, optimalno bi bilo dobiti pozitivne korelacije sa navedenim mjerama koje ne prelaze umjerene intenzitete (npr., Sirota et al. (2021) navode konvergentne korelacije sa drugim kognitivnim varijablama u načelnom rasponu od .2–.4), kako bi se očuvala diskriminativna valjanost testa.

Ispitivanje navedenih problemskih aspekata u vezi sa CRT-V testom može doprinijeti boljem razumijevanju konstrukta kognitivne refleksivnosti, validnosti mjerenja i praktičnoj primjeni u BiH kontekstu. Ako se pokaže prikladnim, CRT-V test bi mogao biti vrijedan dodatak mjerama kognitivnih sposobnosti, posebno za procjenu refleksivnosti nezavisno od matematičkih kompetencija. S druge strane, eventualno utvrđeni psihometrijski nedostaci ukazali bi na potrebu revizije instrumenta, otvarajući prostor za unapređenje i razvoj novih, prikladnijih zadataka za mjerenje kognitivne refleksivnosti u BiH.

Metod

Uzorak, postupak i mjere

Istraživanje je sprovedeno na uzorku od 318 ispitanika (s udjelom ženskog pola od 45%) iz Bosne i Hercegovine. Oko 65% uzorka sačinjavali su srednjoškolci (uglavnom iz Kantona 10 iz Federacije BiH, uz nešto manji broj učenika iz Republike Srpske), prosječnog uzrasta od 17.71 ($SD=1.20$) godinu starosti. Preostalih 35% uzorka činili su studenti (prosječnog uzrasta od $M=22.56$, $SD=1.71$ godina starosti), primarno sa Univerziteta u Banjoj Luci.

Podaci su prikupljeni putem onlajn ankete (platforma 1KA) u slučaju srednjoškolskog poduzorka i papir-olovka metodom, u slučaju studentskog poduzorka. Razlika u modalitetima testiranja je bila uslovljena logističko-administrativnim faktorima.

Ispitanicima je zadan verbalni test kognitivne refleksivnosti (CRT-V; Sirota et al., 2021), sačinjen od 10 zadataka, koji su adaptirani za lokalne jezičke i kulturološke specifičnosti. Kao osnova je korišten postojeći radni prevod testa sa teritorije Republike Srbije (Lazić & Žeželj, 2020). Ovaj prevod je prebačen sa ekavice na ijekavicu i izvršene su manje adaptacije terminologije, u zavisnosti od toga da li je testiranje vršeno u FBiH ili R. Srpskoj. Na primjer, u slučaju FBiH zadavanja, u jednom od zadataka (#3) korištene su geografske odrednice zagrebačkog aerodroma, uz letnu destinaciju Hrvatska – BiH, dok su u slučaju R. Srpska zadavanja ostavljene polaznim srbijanskim prevodom predložene geografske odrednice: beogradski aerodrom, letna destinacija Srbija – Mađarska.

Pored CRT-V testa, u svrhu provjere njegove konvergentne valjanosti, primijenjeni su i testovi verbalnog rezonovanja, sačinjenog od 15 zadataka (preuzet iz: Budiša et al., 2021) i matričnog rezonovanja, sačinjenog od 11 zadataka (preuzet iz ICAR repozitorijuma; Condon & Revelle, 2014). Dodatno, na poduzorku studenata, zadan je i test numeričkog rezonovanja, tj. nastavljanja numeričkih nizova, sačinjen od 13 zadataka (riječ je o podselekciji zadataka koji su takođe preuzeti iz ICAR repozitorijuma; Condon & Revelle, 2014).

Statistička obrada podataka

Psihometrijska evaluacija CRT-V testa je sprovedena korišćenjem kombinacije pristupa klasične testne teorije (CTT) i teorije odgovora na stavke (IRT). Faktorska struktura je ispitana konfirmativnom faktorskom analizom (CFA), koja je sprovedena u programu „lavaan“ (Rosseel, 2012) i zasnovana je na matrici tetrahoričnih korelacija i WLSMV/DWLS postupcima procjene parametara faktorskog modela. Kao indeksi fita odabrani su: CFI, TLI, RMSEA i SRMR (Hooper et al., 2008; Kline, 2023).

Iz ugla CTT, težina zadataka je operacionalizovana kao procenat tačnosti ($\%_T$) na svakom individualnom CRT-V zadatku, dok je diskriminativnost iskazana kao korigovana aitem-total korelacija individualnog zadatka i testa ($rcit$). Iz ugla IRT pristupa, korištena je 2PL analiza, sprovedena u programu „ltm“ (Rizopoulos, 2006), za procjenu težine (b), diskriminativnosti (a) i informativnosti (I).

Provjera metrijske pristrasnost stavki, tj. njihovog diferencijalnog funkcionisanja (DIF) je takođe testirana iz IRT okvira, oslanjajući se na Raju proceduru iz „difR“ programa (Magis, et al., 2010), implementiranog u okviru „snowIRT“ modula za „jamovi“ program (The jamovi project, 2024). DIF provjere su iz logističkih razloga izvršene samo za polne i obrazovne (srednjoškolci u odnosu na studente) kategorije. Premda bi bilo visoko poželjno, logistički nije bilo moguće obezbijediti ujednačenost ispitanika iz različitih dijelova BiH i za oba modaliteta zadavanja (onlajn i papir-olovka), kako bi se pristrasnost provjerila i po entitetsko-samoupravnim cjelinama BiH (idealno uz dodatno ukrštanje sa polnim i obrazovnim kategorijama, metodom zadavanja testova i drugim relevantnim sociodemografskim aspektima od interesa).

Sve konvencionalne analize su takođe izvršene u „jamovi“ platformi (The jamovi project, 2024). Ovo uključuje i konvergentne provjere, koje su operacionalizovane putem prostih bivarijacionih korelacija CRT-V sumativnog skora sa postignućem na testovima matričnog, verbalnog i numeričkog rezonovanja (uz opasku da su korelacije sa numeričkim testom dostupne samo za studentski poduzorak, jer ovaj test iz logističkih razloga nije mogao biti uvršten u testovnu bateriju za srednjoškolce).

Rezultati

Psihometrijske karakteristike CRT-V testa

Testirana je faktorska struktura CRT-V testa (Sirota et al., 2021) na objedinjenom srednjoškolsko-studentskom uzorku, pri čemu je potvrđena jednodimenzionalnost, sa dobrim fitom modela prema uobičajenim ad hoc konvencijama (Hooper et al., 2008; Kline, 2023): CFI=.994; TLI=.992; SRMR=.060; RMSEA=.025, 90% CI [.006, .046] i solidnom pouzdanosti interne konzistencije (ω =.80). Faktorska zasićenja (Λ), težina (CTT: %T; IRT: *b*) i diskriminativnost (CTT: *rcit*; IRT: *a*) prikazani su u Tabeli 1. Većina zadataka pokazuje faktorska zasićenja blizu .70, što se konvencionalno smatra poželjnim (Hair et al., 2019; Kline, 2023). Ipak, zadatak #9 („Da li bi bilo etički ispravno da muškarac oženi sestru svoje udovice?“) pokazuje nisko faktorsko zasićenje prema CFA standardima. Kline (2023, p. 280) navodi da bi minimalna Λ vrijednost za jednofaktorske modele trebalo bi da bude $\geq .55$ (iako preporuke u širem kontekstu faktorske analize dozvoljavaju i nešto niže donje pragove, npr. $\Lambda \geq .32$; Tabachnick & Fidell, 2019). Problem je izraženiji na srednjoškolskom poduzorku, gdje zadatak #9 ima vrlo nisko faktorsko zasićenje (Λ =.16), dok je na studentskom poduzorku ono značajno više (Λ =.82), što ukazuje na varijabilnost njegove metrijske prikladnosti u različitim obrazovnim grupama.

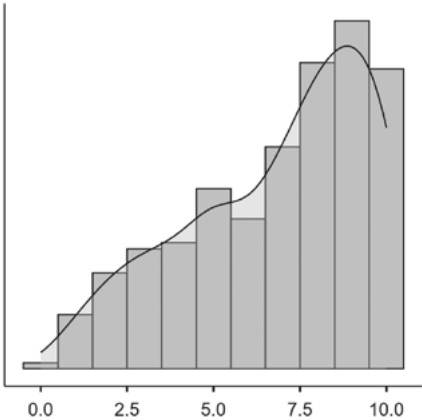
Tabela 1.

Psihometrijske karakteristike CRT-V testa

CRT-V stavke (skraćeni opisi)	Λ	%T	<i>rcit</i>	<i>b</i>	<i>a</i>
CRTV1: Majin otac ima pet ćerki	.70	74%	.48	-0.94	1.61
CRTV2: Trka – preticanje osobe na 2. mjestu	.58	85%	.35	-1.79	1.24
CRTV3: Avionska katastrofa – sahrana preživjelih	.61	75%	.41	-1.11	1.33
CRTV4: Trka životinja do vrha drveta	.85	75%	.57	-0.82	2.69
CRTV5: Prizemna kuća – boja stepenica	.84	68%	.61	-0.56	2.56
CRTV6: Mojsijeva barka – broj životinja	.69	39%	.44	0.41	1.63
CRTV7: Električni voz – smjer dima	.83	57%	.58	-0.22	2.56
CRTV8: Mračna soba – šta upaliti prvo	.67	55%	.46	-0.20	1.39

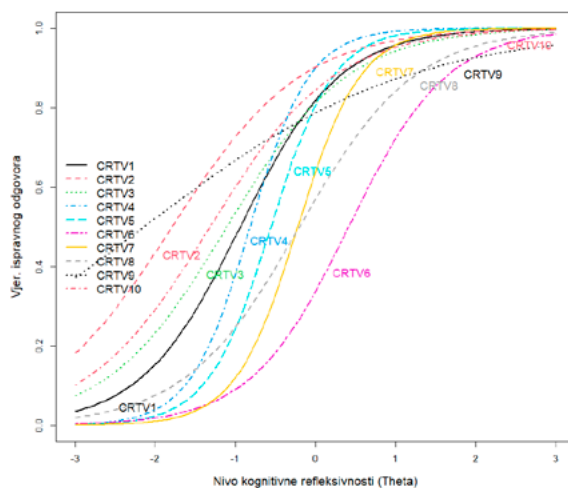
CRT-V stavke (skraćeni opisi)	Λ	%T	rcit	b	a
CRTV9: Etika braka muškarca sa sestrom udovice	.34	77%	.22	-2.14	0.61
CRTV10: Tačnost rečenice o žumancu jajeta	.63	79%	.40	-1.32	1.29

Pitanje diskriminativnosti stavki usko je povezano sa faktorskim zasićenjem. Prema Nunnally & Bernstein (1994, pp. 304-305), $rcit \geq .30$ predstavlja arbitrarni prag prihvatljive CTT diskriminativnosti, dok Baker (2001, p. 35) klasifikuje IRT parametar diskriminativnosti u pet kategorija: vrlo niska ($a < 0.34$), niska ($0.35 \leq a \leq 0.64$), umjerena ($0.65 \leq a \leq 1.34$), visoka ($1.35 \leq a \leq 1.69$) i vrlo visoka ($a \geq 1.70$). Držeći se ovih kriterijuma, svi zadaci osim #9 imaju prihvatljivu diskriminativnost prema CTT, dok većina zadataka prema IRT modelu pokazuje diskriminativnosti oko granice između umjerene i visoke, pri čemu zadaci #4, #5 i #7 dostižu vrlo visoke vrijednosti. Zadatak #9 ponovo se izdvaja kao problematičan, sa niskim parametrima diskriminativnosti, naročito na poduzorku srednjoškolaca ($rcit = .10$, $a = 0.27$), dok su vrijednosti kod studenata zadovoljavajuće ($rcit = .53$, $a = 2.33$). Dodatno, na studentskom poduzorku uočava se niže faktorsko zasićenje i diskriminativnost zadatka #10 („Koja rečenica je tačna: a) „Žumance od jajeta je bijelo.“ ili b) „Žumance jajeta je bijelo.“?): $\Lambda = .38$, $rcit = .17$, $a = 0.98$, dok su njegove vrijednosti na srednjoškolskom poduzorku adekvatne ($\Lambda = .59$, $rcit = .41$, $a = 1.19$).



Slika 1.
Distribucija sumativnih skorova CRT-V testa

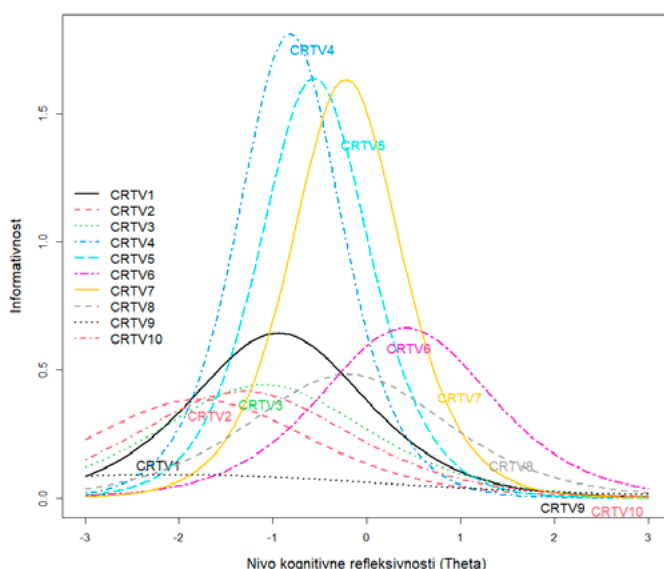
Test se u cjelini pokazao kao lagan, sa prosjekom: $M=6.84$, $SD=2.61$ i ukupnom uspješnošću od 68% (prikaz distribucije skorova na Slici 1). Najčešći rezultat (*Mod*) bio je 9 od maksimalnih 10 poena (18.2% ispitanika). Ovo postignuće je za oko jednu standardnu devijaciju više u odnosu na referentne vrijednosti iz istraživanja Sirota et al. (2021; $M=4.21$, $SD=2.81$), pri čemu je razlika statistički značajna: $t(317)=17.98$, $d=1.01$, $p<.001$. Na šest od deset zadataka (#1, #2, #3, #4, #9 i #10) prosječna uspješnost (%T) prelazi 75%, dok je najteži zadatak #6 („Po koliko od svake životinje je Mojsije ukrcao na barku?“) – sa 39% tačnih odgovora. Iz IRT ugla (pored b vrijednosti iz Tabele 1, vidjeti i karakteristične krivulje na Slici 2), svi zadaci osim jednog zahtijevaju ispodprosječan nivo refleksivnosti za 50% šanse za ispravan odgovor. Najlakši su zadaci #9 i #10, za koje su već konstatovani problemi sa diskriminativnošću, a zahtijevaju nešto preko dva, odnosno nešto više od jednog logita ispod prosjeka latentne crte za 50%-nu vjerovatnoću tačnog odgovora. Pored njih, izdvaja se i zadatak #2 („Zamislite da učestvujete u trci i preteknete osobu na 2. mjestu. Na kojem biste mjestu sada bili?“), koji za 50%-nu vjerovatnoću tačnog odgovora traži nešto manje od dva logita ispod prosjeka. Jedini donekle teži zadatak je #6, ali čak i za njega je 50%-na vjerovatnoća tačnog odgovora ostvariva već uz manje od pola logita iznad prosjeka latentne crte.



Slika 2.

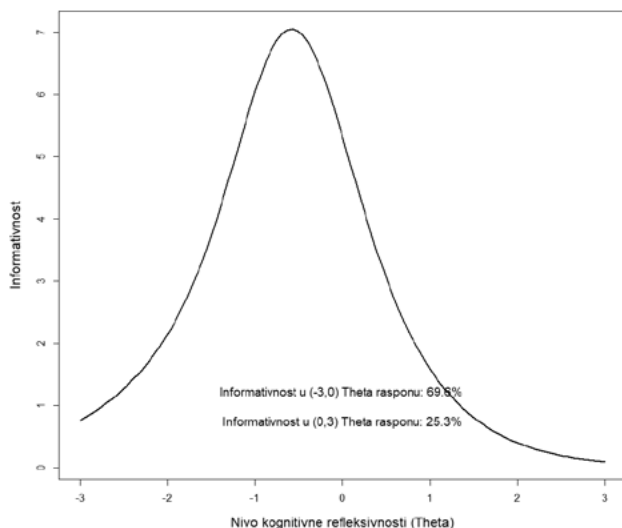
Karakteristične krivulje (ICC) za CRT-V test, prema 2PL IRT modelu

Kada je riječ o informativnosti (Slike 3 i 4), primarno je koncentrisana oko prosjeka i nešto ispod prosjeka latentne crte kognitivne reflektivnosti. Većina informativne vrijednosti dolazi od zadataka #4 („Majmun, vjeverica i ptica se trkaju do vrha kokosovog drveta. Ko će prije stići do banane – majmun, vjeverica ili ptica?“), #5 („U prizemnoj ružičastoj kući, nalazila se ružičasta osoba, ružičasta mačka, ružičasta ribica, ružičast kompjuter, ružičasta stolica, ružičast sto, ružičast telefon, ružičast tuš – sve je bilo ružičasto! Koje su vjerovatno boje bile stepenice?“) i #7 („Vjetar duva u zapadnom smjeru. Električni voz se kreće u istočnom smjeru. U smjeru koje od četiri strane svijeta ide dim iz lokomotive?“), dok je informativnost testa iznad prosjeka crte minimalna i potiče uglavnom od „najtežeg“ zadatka #6. Dakle, test je najprikladniji za ispitanike sa nešto ispodprosječnim do prosječnim nivoom kognitivne reflektivnosti, pri čemu najveći dio informativne vrijednosti dolazi iz manje od polovine zadataka.



Slika 3.

Informativnost pojedinačnih zadataka CRT-V testa, prema 2PL IRT modelu



Slika 4.

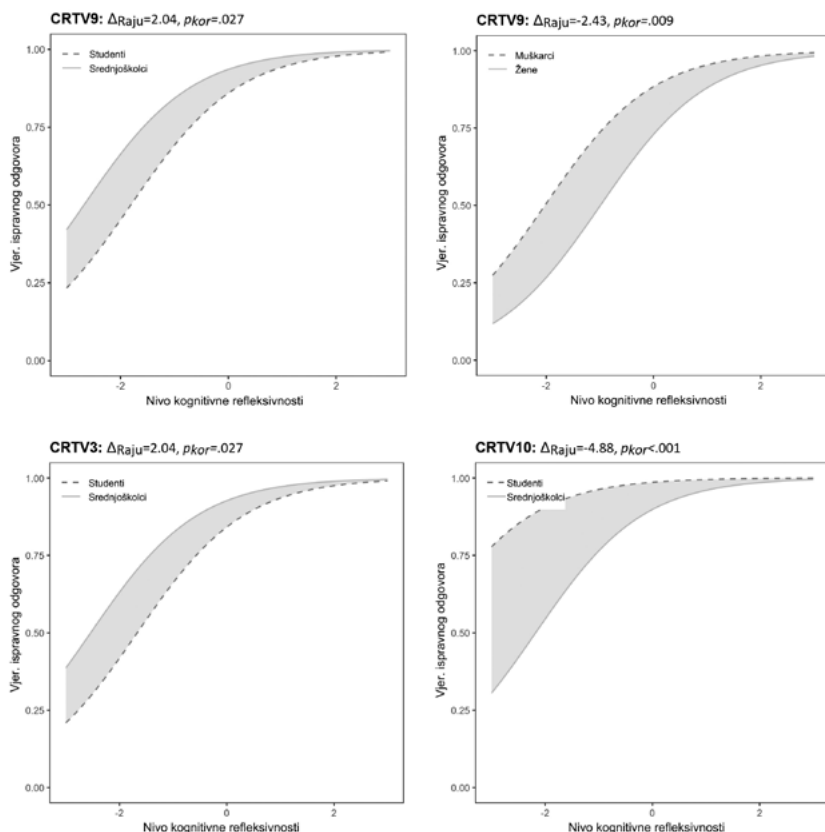
Informativnost ukupnog CRT-V testa, prema 2PL IRT modelu.

Metrijska pristrasnost (DIF) CRT-V testa

Metrijska pristrasnost, tj. diferencijalno funkcionisanje (DIF) CRT-V zadataka ispitana je u odnosu na pol i obrazovne skupine (srednjoškolci vs. studenti), kako bi se utvrdilo da li ispitanici sa istim nivoom kognitivne refleksivnosti imaju nejednake šanse za tačan odgovor u zavisnosti od grupe pripadnosti. Rezultati DIF testova prikazani su na Slici 5, pri čemu su navedeni samo zadaci sa statistički značajnim efektima (nakon Holm–Bonferroni korekcije). Na zadatku #9 uočena je pristrasnost i po polu i po obrazovnoj skupini – zadatak favorizuje muškarce i srednjoškolce, koji imaju veću vjerovatnoću tačnog odgovora u poređenju sa ženama i studentima iste refleksivnosti. Na zadacima #3 i #10, pristrasnost se odnosi samo na obrazovne skupine, pri čemu #3 favorizuje srednjoškolce, a #10 studente. Takođe, ovi efekti su izraženiji na nižim do umjerenim intervalima latentne crte, dok na višim intervalima razlike nestaju, a šanse za tačne odgovore postaju funkcionalno jednake između grupa.

Ne postoji konsenzus o intenzitetu DIF efekata. Prema ad hoc kriterijumu iz „snowIRT“ modula za „jamovi“ (The jamovi project, 2024): $\Delta R_{aju} < |1.0|$ označava zanemarljiv

DIF, $\Delta\text{Raju} > |1.0|$ umjeren, a $\Delta\text{Raju} > |1.5|$ visok DIF. Sve četiri utvrđene metrijske pristrasnosti spadaju u viši intenzitet. Djelimičnu provjeru impakta DIF-a omogućava poređenje skorova sa i bez pristrasnih zadataka. Razlika između srednjoškolaca i studenata na sumativnom skor CRT-V testa pokazuje prednost studenata: $t(263.98)=4.53$, $d=0.52$, $p<.001$. Isključivanjem pristrasnih zadataka (#3, #9, #10), razlika se tek neznatno smanjuje: $t(270.81)=4.30$, $d=0.49$, $p<.001$. U pogledu polnih razlika, žene imaju blagu, ali statistički neznačajnu prednost: $t(307.76)=-1.81$, $d=-0.20$, $p=.072$. Nakon isključenja pristrasnog zadatka #9, prednost žena blago raste i dostiže statističku značajnost: $t(314.55)=-2.24$, $d=-0.25$, $p=.026$.



Slika 5.

Provjere pristrasnosti (DIF) CRT-V zadataka po polu i obrazovnim skupinama.

Konvergentna valjanost CRT-V testa

Konvergentna valjanost provjerena je kroz korelacije skora kognitivne refleksivnosti sa postignućima na testu matričnog i verbalnog rezonovanja, kao i numeričkog rezonovanja (ali dostupno samo za studentski poduzorak). Dobijene su sljedeće korelacije:

- matrično rezonovanje: $r=.35, p<.001$,
- verbalno rezonovanje: $r=.44, p<.001$,
- numeričko rezonovanje (samo studenti): $r=.36, p<.001$.

Diskusija

Rezultati istraživanja upućuju na više važnih implikacija o psihometrijskim karakteristikama i primjeni CRT-V verbalnog testa kognitivne refleksivnosti (Sirota et al., 2021) u BiH kontekstu. Test se pokazao vrlo lakim – prosječna uspješnost (68%) znatno nadmašuje referentne vrijednosti Sirota et al. (2021), pri čemu više od polovine zadataka rješava preko 75% ispitanika. Mogući uzrok je što su mnogi zadaci CRT-V testa prepoznati kao „dječje mozgalice“, široko dostupne putem interneta.¹ Iako su Sirota et al. (2021) upozorili na porast familijarnosti kroz ponovljene upotrebe i adaptacije testa, vjerovatnije je da su autori potcijenili popularnost ovih zadataka (ili porast popularnosti i dostupnosti u funkciji vremena), naročito u edukativno-zabavnom kontekstu za djecu. Ova široka dostupnost umanjuje efekat „intuitivne zamke“, smanjuje težinu, snižava diskriminativnost i sužava informativni opseg testa, naročito iznad prosječnog nivoa kognitivne refleksivnosti, otežavajući diferencijaciju ispitanika sa višim stepenom ove osobine.

Analiza diferencijalnog funkcionisanja stavki (DIF) pokazala je da neki zadaci metrijski favorizuju određene podskupine ispitanika, čak i pri istom nivou kognitivne

1 Npr. u trenutku pisanja ovog rada, na stranici: <https://www.pjesmicezadjecu.com/mozgalice/mozgalice.html> bilo je moguće pronaći više varijacija na temu CRT-V zadataka. „Sudjelujete u utrci. Prestigli ste osobu koja je bila na drugom mjestu. Na kojem ste vi mjestu sada? (Ako ste odgovorili da ste prvi, niste u pravu! Ako prestignete drugu osobu i zauzmete njezino mjesto, onda ste drugi!)“ „Marijin otac ima 5 kćeri: 1. Nana, 2. Nene, 3. Nini, 4. Nono. Kako se zove peta kćer? (Nunu? NE! Njezino ime je Marija.)“ I drugo.

refleksivnosti. Zadatak #9 („Da li bi bilo etički ispravno da muškarac oženi sestru svoje udovice?“), koji ima nisku diskriminativnost i faktorsko zasićenje, pokazao je metrijsku pristrasnost u korist muškaraca i srednjoškolaca. Zadatak #3 metrijski favorizuje srednjoškolce, dok zadatak #10 favorizuje studente. Već pretpostavljena familijarnost sa zadacima i učestalost izlaganja „mozgalicama“ mogu djelimično objasniti metrijske pristrasnosti između srednjoškolaca i studenata. U slučaju zadatka #3, prisutne su specifične geografske odrednice koje nisu dosljedno lokalizovane za sve dijelove BiH², što može uticati na razumljivost i obradu zadatka. Konkretno, u FBiH (uglavnom srednjoškolski poduzorak) lokalizacija je uključivala „zagrebački aerodrom“ i let na relaciji Hrvatska – BiH, dok je u R. Srpskoj (uglavnom studentski poduzorak) korištena verzija preuzeta iz srbijanskog prevoda testa, gdje se pominju: beogradski aerodrom i let na relaciji Srbija – Mađarska. Ova druga verzija možda je zahtijevala viši nivo geografskog znanja i refleksivnog razmišljanja (npr. razmatranje teritorijalnih pozicija i graničnih linija zemalja), što može objašnjavati dio uočenog DIF-a. Stoga bi bilo korisno isprobati i uporediti više verzija lokalizacije ovog zadatka kako bi se osigurala njegova konzistentna metrijska svojstva u različitim dijelovima BiH.

DIF efekt na zadatku #10 („Koja rečenica je tačna: a) ‘Žumance od jajeta je bijelo.’ ili b) ‘Žumance jajeta je bijelo.’?“) može proizaći iz razlike u akademskoj pripremi. Studenti, sa većim iskustvom u analitičkom razmišljanju, lakše prepoznaju da je riječ o logičkoj, a ne gramatičkoj dilemi, dok su srednjoškolci skloniji fokusiranju na sintaksičke i stilske aspekte rečenica (što, u prosjeku, više odlikuje srednjoškolski, u odnosu na fakultetski nastavni fokus), što ih onda navodi na intuitivan odgovor (prepoznajući da formulacija „žumance od jajeta“ zvuči stilski neprirodno). Ova pristrasnost sugerise da zadatak ne mjeri isključivo kognitivnu refleksivnost, već može biti i pod uticajem kognitivne fleksibilnosti i metalingvističkih sposobnosti.

2 Za ispitanike iz R. Srpske zadržana je lokalizacija iz postojećeg srbijanskog prevoda: „Tokom jedne olujne noći, avion polijeće sa beogradskog aerodroma. Oluja se pogoršava i avion se sruši – jedna polovina sleti u Srbiji, druga polovina u Mađarskoj. U kojoj državi će se sahraniti preživjeli?“ Za ispitanike iz FBiH, korištena je adaptirana lokalizacija: „3. Tokom jedne olujne noći, avion polijeće sa zagrebačkog aerodroma. Oluja se pogoršava i avion se sruši – jedna polovina sleti u Hrvatskoj, druga polovina u Bosni i Hercegovini. U kojoj državi će se sahraniti preživjeli?“

Polna pristrasnost na zadatku #9 može biti posljedica različite percepcije rodni uloga u samoj formulaciji zadatka. Kako je „glavni akter“ muškarac, moguće je da muškarci lakše uočavaju logičku grešku jer se neposrednije identifikuju sa situacijom, što može doprinijeti bržem prevazilaženju intuitivne zamke. Ovaj zadatak, kao i zadatak #10, potencijalno mjeri i aspekte kognitivne fleksibilnosti, što bi trebalo uzeti u obzir prilikom revizije testa.

U vezi sa ustanovljenim diferencijalnim funkcionisanjem stavki #3, #9 i #10 – treba iznijeti i određene ograde.³ Naime, premda prisustvo DIF-a po definiciji predstavlja problem, trenutno ne postoje dobro definisani kriterijumi po kojima se utemeljeno može zaključiti – u kojoj mjeri to zaista jeste praktični problem, ali može se iznijeti teza da to, u ovom konkretnom slučaju i za sadašnju formu testa, nije nužno od suviše velike konsekventnosti. Prije svega, korišteni Δ_{Raju} kriterijumi (The jamovi project, 2024), iako jesu ukazali da je detektovani DIF relativno velikog intenziteta, relativno su arbitrarni, a nakon uklanjanja DIF stavki razlike u prosjecima između poredbenih grupa mijenjaju se vrlo malo. Dalje, utvrđeni DIF ovih stavki je većinski lociran ispod prosjeka latentne crte kognitivne refleksivnosti i uglavnom se umanjuje i iščezava na intervalima koji prelaze prosjek. Uz to, DIF nije utvrđen na stavkama koje nose najveću informativnu vrijednost (#4, #5 i #7). Uzevši sve ovo u obzir, nekoliko ustanovljenih značajnih DIF efekata ne mora nužno biti impaktna prepreka za upotrebu testa za mjerenja i odlučivanja niskog rizika.

Zadatak #6, koji pominje Mojsija umjesto Noe u kontekstu ukrcavanja životinja na barku, pokazao se kao najteži na testu. Iako za ispitivane sociodemografske grupe (pol i obrazovanje) nije utvrđena metrijska pristrasnost ovog zadatka, logističke prepreke u sprovođenju istraživanja onemogućile su provjeru DIF-a u odnosu na relevantnije mjere za sadržaj ovog zadatka – etničku pripadnost i religijsku orijentaciju. Anegdotski, veći broj ispitanika izrazio je nesigurnost ili pritužbe zbog religijskog sadržaja ovog zadatka, što su kao potencijalni problem već naglasili i Sirota et al. (2021, p. 340). Moguće je da povećana težina ovog zadatka proizlazi iz nefamilijarnosti ispitanika sa sadržajem, a ne iz kognitivne reflek-

3 Alternativna interpretacija DIF efekata predložena je na osnovu komentara anonimnog recenzenta, kojem ovim putem izražavamo zahvalnost.

sivnosti. Stoga bi u procesu dalje adaptacije CRT-V testa trebalo razmotriti zamjenu i ovog zadatka.

Konvergentna valjanost CRT-V testa potvrđena je umjerenim korelacijama sa testovima matričnog, verbalnog i numeričkog rezonovanja, što ukazuje da test dijeli dio varijanse s opštim kognitivnim sposobnostima (Sirota et al., 2021), ali istovremeno mjeri i specifične aspekte refleksivnog razmišljanja. Intenziteti povezanosti slični su ranijim nalazima, pri čemu je veza sa verbalnim rezonovanjem nešto izraženija nego kod Sirota et al. (2021). Ovo je očekivano, s obzirom na verbalnu prirodu CRT-V testa i veću sličnost verbalnog testa inteligencije korištenog u ovom istraživanju u odnosu na onaj iz originalne validacije.

U pogledu praktične primjenjivosti testa, CRT-V bi mogao poslužiti kao brza mjera kognitivne refleksivnosti za širu populaciju, naročito kada je cilj razlikovati ispitanike s ispodprosječnim, u odnosu na makar prosječni nivo kognitivne refleksivnosti, primjera radi, prilikom selekcije za zanimanja poput policije ili administrativno-šalterskih radnika. Ipak, detekcija više DIF efekata, prevelika lakoća zadataka i sužena informativnost dovode u pitanje univerzalnu upotrebljivost instrumenta u multikulturalnom kontekstu BiH, gdje različite populacije (srednjoškolci, studenti, pripadnici različitih entiteta, religija i polova) mogu drugačije doživljavati iste stavke.

Prethodno navedeno upućuje na potrebu temeljne revizije CRT-V testa u kontekstu BiH, sa fokusom na: (a) modifikaciju ili zamjenu psihometrijski i sadržinski problematičnih stavki (#9, #3 i #10, potencijalno i #6); (b) uvođenje težih zadataka ili zadataka koji nisu već poznate „mozgalice“, radi proširenja opsega informativnosti na višim nivoima latentne crte; (c) konzistentnu lokalizaciju geografsko-kulturoloških elemenata (npr. naziva država, religijskih i jezičkih odrednica) za različite dijelove i etnose BiH.

Uzimajući u obzir nedostatke ove studije, potrebno je testirati CRT-V na raznovrsnijem i balansiranim uzorku, koji bi ravnomjerno obuhvatio i direktno komparirao ispitanike različitih polova, uzrasta, obrazovnih nivoa, etničke i religijske pripadnosti, uz bolju kontrolu metodoloških varijabli, poput načina zadavanja testa, koji je u ovom istraživanju zbog logističkih ograničenja bio različit za različite obrazovne skupine – isključivo onlajn za srednjoškolce i isključivo

papir-olovka za studente. Takođe, potrebno je sprovesti i provjeru kriterijumske valjanosti u različitim relevantnim domenima (npr. obrazovnim ili radnim). Budući da informativnost sadašnje verzije testa primarno nose tri zadatka, dok bi tri-četiri trebalo revidirati ili izbaciti, te da je potrebno proširiti težinski i informativni raspon, umjesto revizije, možda bi bilo svrsishodnije konstruisati novi test. To bi možda mogla da bude i prilika da se ponovo razmotri ideja o numeričkim mjerama kognitivne refleksivnosti, tj. revidirani test bi možda mogao kombinovati verbalne (Sirota et al., 2021) i numeričke zadatke (Frederick, 2005), čime bi se potencijalno obuhvatio širi spektar refleksivnog mišljenja.

Literatura

- Avram, L.-A. (2018). Gender differences and other findings on the cognitive reflection test. *Studia UBB Oeconomica*, 63(3), 59–69. <https://doi.org/10.2478/sub-boec-2018-0014>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation
- Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, 82, Article 101455. <https://doi.org/10.1016/j.socec.2019.101455>
- Bubić, A., & Erceg, N. (2015). Uloga kognitivnih stilova u razumijevanju kognitivnog funkcioniranja pojedinaca. *Suvremena psihologija*, 18(2), 159–173.
- Budiša, D., Petković, T., Pušić, Z., Karić, M., & Subotić, S. (2021). Provjera inkrementalnog doprinosa samoprocijenjene emocionalne inteligencije u objašnjenju akademskog postignuća srednjoškolaca i studenata [Examination of the incremental contribution of self-rated emotional intelligence in explaining the academic achievement of secondary school and university students]. In S. Lakić (Ed.), *Proceedings of the Banja Luka November encounters 2021. scientific conference* (Vol. 22) (pp. 379–400). University of Banja Luka, Faculty of Philosophy.
- Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42(3), 434–447. <https://doi.org/10.3758/s13421-013-0367-9>
- Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- Corgnet, B., Hernán Gonzalez, R., & Mateo, R. (2015). Cognitive reflection and the diligent worker: An experimental study of millennials. *PloS One*, 10(11), Article e0141243. <https://doi.org/10.1371/journal.pone.0141243>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>

- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Juanchich, M., Sirota, M., & Bonnefon, J.-F. (2020). Anxiety-induced miscalculations, more than differential inhibition of intuition, explain the gender gap in cognitive reflection. *Journal of Behavioral Decision Making*, 33, 427–443. <https://doi.org/10.1002/bdm.2165>
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.
- Krašić, S. (2018). Test kognitivne refleksije kao prediktor podložnosti pristranostima pri likom donošenja odluka. *Psychē*, 1(1), 73–79.
- Lazić, A., & Žeželj, I. (2020). *Serbian translation of the VerbalCognitiveReflectionTest (CRT-V)* [Questionnaire translation]. Open Science Framework. <https://osf.io/ejdqx>
- Magis, D., Beland, S., Tuerlincks, F., & De Boeck, P. (2010). difR: A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847–62. <https://doi.org/10.3758/BRM.42.3.847>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.
- Otero, I., Salgado, J. F., & Moscoso, S. (2021). Criterion validity of cognitive reflection for predicting job performance and training proficiency: A meta-analysis. *Frontiers in Psychology*, 12, Article 668592. <https://doi.org/10.3389/fpsyg.2021.668592>
- Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence*, 90, Article 101614. <https://doi.org/10.1016/j.intell.2021.101614>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analysis. *Journal of Statistical Software*, 17(5), 1–25. <https://doi.org/10.18637/jss.v017.i05>

- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Salgado, J. F., Otero, I., & Moscoso, S. (2019). Cognitive reflection and general mental ability as predictors of job performance. *Sustainability*, 11(22), 6498. <https://doi.org/10.3390/su11226498>
- Shtulman, A., & Young, A. G. (2023). The development of cognitive reflection. *Child Development Perspectives*, 17(1), 59–66. <https://doi.org/10.1111/cdep.12476>
- Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6, Article 532. <https://doi.org/10.3389/fpsyg.2015.00532>
- Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., & Marshall, A. C. (2021). Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making*, 34(3), 322–343. <https://doi.org/10.1002/bdm.2213>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- The jamovi project (2024). jamovi. (Version 2.6) [Computer Software]. <https://www.jamovi.org>
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99–113. <https://doi.org/10.1017/S1930297500007622>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275–1289. <https://doi.org/10.3758/s13421-011-0104-1>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147–168. <https://doi.org/10.1080/13546783.2013.844729>

IZJAVA O UPOTREBI GENERATIVNIH AI TEHNOLOGIJA

Prilikom izrade teksta članka korišten je ChatGPT-4o i o1, u svrhu gramatičke i stilske provjere teksta i referenci i konciznijeg formulisanja i prevođenja određenih navoda, kao i dopunske provjere kvaliteta napisane argumentacije. Ovaj alat je korišten isključivo kao pomoćno sredstvo, sve predložene izmjene su pažljivo verifikovane, gdje je potrebno modifikovane i odobrene od strane autora.

Psychometric Evaluation of the Adaptation of the Verbal Cognitive Reflection Test in Bosnia and Herzegovina

Vesna Kovačić*, Nina Bobić, Siniša Subotić*****

* Volunteer of the Association of parents of children with special needs Path to Life, Livno

** DDC Multilingual Solutions, Banja Luka

*** Department of Psychology & Faculty of Natural Sciences and Mathematics, University of Banja Luka

SUMMARY: Cognitive reflection refers to the ability to suppress an incorrect initial intuition. This construct has traditionally been measured using numerical tasks, which led to the development of the verbal CRT-V test, aimed at more clearly distinguishing it from the domain of mathematical ability. In this study, the CRT-V test was examined on a sample of 318 participants from Bosnia and Herzegovina (45% female; 65% high school students, 35% university students). Along with the CRT-V, tests of verbal and matrix reasoning were administered. The one-factor CRT-V model exhibited excellent fit indices (CFI=.994; TLI=.992; SRMR=.060; RMSEA=.025, 90% CI [.006, .046]) and adequate convergent validity. However, several issues emerged. Specifically, item #9 had low factor loading and poor discrimination. Most items were too

easy (70%+ success on 6/10 tasks), and test informativeness was mainly located below the average latent cognitive reflection trait level. Item #9 displayed bias (DIF) favoring males and high school students, while item #3 favored high school students and item #10 favored university students. Some items also contain terminological and cultural nuances not optimal for the multicultural BiH context. The one-factor CRT-V model showed good fit and satisfactory convergent validity. Average scores were moderately higher among university students compared to high school students, and slightly higher among women compared to men. Several psychometric issues were identified. Most tasks were too easy (70%+ success rate on 6 out of 10 items), with informativeness primarily located below the average level of the latent cognitive reflection trait. Some items showed metric bias (DIF), and several contained terminological or cultural specificities that may be suboptimal for the multicultural context of Bosnia and Herzegovina. Although the current version of the CRT-V may be useful, a more thorough revision or replacement of certain items is recommended to reduce bias and increase difficulty and informativeness, thereby enhancing the practical utility of the test.

KEYWORDS: cognitive reflection, psychometric test evaluation, test cultural appropriateness, classical test theory (CTT), item response theory (IRT).