

# Stupanj sigurnosti u točnost odgovora kao dodatna informacija pri psihometrijskoj analizi testa kognitivnih sposobnosti

Damir Ljubotina

Filozofski fakultet Sveučilišta u Zagrebu, Odsjek za psihologiju, Ulica Ivana Lučića 3, 10000 Zagreb

**Sažetak:** Uradak u zadacima testova znanja i sposobnosti tradicionalno se vrednuje dihotomno na osnovi točnosti odgovora. U ovom radu razmatra se mogućnost uvođenja procjene stupnja sigurnosti (SS) u točnost odgovora kao dodatne informacije o uratku ispitanika u zadacima i testu. Analizirane su teorijske i konceptualne prednosti te metodološki problemi uključivanja ovih mjera u psihometrijsku analizu testnih rezultata i vrednovanje ukupnih rezultata. Teorijski okvir za uključivanje procjena stupnja sigurnosti u točnost odgovora povezan je s konceptima metakognicije te dispozicijskim mjerama samopoimanja i samoeфикаsnosti (Stankov i Lee, 2008).

Empirijski rezultati zasnovani su na rezultatima primjene računalne verzije neverbalnog testa kognitivnih sposobnosti (Test S) na uzorku od 168 studenata. Analiza rezultata provedena je na razini testnih zadataka, intraindividualnih i interindividualnih razlika zasnovanih na ukupnim testnim rezultatima. Na razini testnih zadataka prosječan SS povezan je s indeksom lakoće zadataka, a kod zadataka s većom diskriminativnom valjanošću registrirana je veća razlika u procijenjenoj sigurnosti između točnih i netočnih odgovora. Provedenom regresijskom analizom objašnjeno je 69,1% varijance prosječnog SS u točnost svih odgovora, pri čemu su najbolji prediktori metakognitivna procjena uratka u testu i stvarni broj točnih odgovora. Latentna struktura matrice zasnovane na procjenama SS upućuje na dominaciju jednog generalnog faktora te na visoku pouzdanost prosječnog SS od 0,96.

Diferencijalno ponderiranje točnosti stupnjevima sigurnosti povećava pouzdanost i homogenost cjelokupnog testa. Rezultati upućuju na zanimljiv potencijal uključivanja procjena SS u točnost odgovora u psihometrijsku analizu testova kognitivnih sposobnosti.

**Ključne riječi:** stupanj sigurnosti, točnost odgovora, analiza zadataka

## UVOD

U okviru klasične i moderne teorije testova vrednovanje uratka ispitanika u testnim zadacima tradicionalno je usmjereno na toč-

nost uratka. Najveći broj kvantitativnih matematičko-statističkih modela za analizu testnih rezultata polazi od binarne matrice u kojoj je uradak ispitanika dihotomiziran na točno i netočno riješene zadatke. Ovaj pristup rezultira

razmjerno niskom osjetljivošću te gubitkom informacija o mogućem razlikovanju ispitanika unutar ove dvije kategorije ispitanika. Predloženi su različiti postupci kojima se nastoji unaprijediti vrednovanje uratka u zadatku, pri čemu se uradak može ponderirati različitim numeričkim koeficijentima koji uzimaju u obzir težinu zadatka, diskriminativnu valjanost ili saturaciju generalnim faktorom koji test mjeri kao i prediktivnu valjanost zadatka. Pojedini teoretičari su skeptični prema učinku ponderiranja čestica, osobito kod kompozita duljih od 20 čestica (Nunnally i Bernstein, 1994). Pri bodovanju se mogu uključiti ispravnost korištenog procesa ili algoritma pri rješavanju, korekcija zbog mogućnosti slučajnog pogađanja, karakteristike odabranih netočnih odgovora te različite mjere brzine rada. Suvremena tehnologija omogućuje registraciju niza dodatnih parametara vezanih uz *ponašanje ispitanika tijekom rješavanja zadatka* kao što su neurofiziološke reakcije, pokreti očiju, neverbalna facijalna ekspresija, verbalni komentari i sl.

Stupanj sigurnosti (SS) ili uvjerenja u točnost odgovora (eng. CL - confidence level) podrazumijeva procjenu ispitanika o vjerojatnosti da su odgovor, rješenje ili sud koji su ponudili točni. Procjena SS u točnost odgovora ili iskaza korištena je u različitim područjima kao što su psihofizika, teorija detekcije signala, ispitivanje sposobnosti i znanja, pamćenje, svjedočenje, stavovi i sudovi, percepcija, metakognicija i sl. (Koriat, 2012; Stankov, 2000). Ranija istraživanja bila su primarno usmjerena na uporabu procjene SS u eksperimentalnim istraživanjima, a tek od devedesetih jača interes za ispitivanje individualnih razlika u ovom području (Stankov i Lee, 2008). U recentnom pregledu istraživanja u području procjena i donošenja odluka Fischhoff i Broomell (2020) razlikuju probleme vezane uz načine kako osobe predviđaju ishode pojedinih odluka, kako vrednuju različite moguće odluke te kako se

pojedinačne procjene i sudovi kombiniraju u konačnu odluku. U ovom tekstu nastojat ćemo kroz teorijsku elaboraciju i analizu empirijskih rezultata vrednovati potencijale informacije o stupnju sigurnosti u točnost odgovora u psihometrijskoj analizi testa i vrednovanju individualnih rezultata.

Hevner još 1932. koristi ovu paradigmu i nalazi da ponderiranje točnosti uratka procjenama SS može povećati procjene pouzdanosti i poboljšati postupke za korekciju rezultata zbog mogućnosti slučajnog pogađanja. Među ranim istraživanjima Soderquist (1936) boduje zadatke alternativnog tipa stupnjevima sigurnosti, pri čemu uputom i postupkom bodovanja ispitanike upozorava na to da visok stupanj sigurnosti u netočne odgovore dovodi do većeg gubitka pri bodovanju, a slično navodi i Ebel (1965). Coombs (1953) kod zadataka s višestrukim ponuđenim odgovorima traži od ispitanika da prekriže sve odgovore koje smatraju netočnim. Na osnovi ovakvog pristupa sugerira razlikovanje potpunog znanja (eliminacija svih netočnih odgovora) od djelomičnog znanja (eliminacija samo nekih odgovora) i pogrešnog znanja (visoka sigurnost u netočan odgovor). Ovaj pristup zahtijeva kvalitetnu pripremu netočnih odgovora, a neke studije pokazuju da i analiza odabranih netočnih odgovora nudi određenu informaciju o razvijivosti konstrukta. Svi navedeni autori navode da ponderiranje uratka stupnjevima sigurnosti povećava pouzdanost, pri čemu oduzimanje netočnih odgovora može imati i suprotan efekt. Unatoč razmjerno pozitivnim nalazima ranijih istraživanja, interes za uključivanje SS u psihometrijsku analizu testova raste tek devedesetih godina, pri čemu je najveći broj istraživanja potaknut skupinom autora vezanim uz L. Stankova. Sporadični primjeri uporabe ovakvog oblika ponderiranja mogu se pronaći u kontekstu razvoja i evaluacije računalnog adaptivnog testiranja (Yen i sur., 2010), razvoja i provjere formativnih i sumativnih postu-

paka vrednovanja znanja (Gardner-Medwin, 1998) te validacije računalnih sustava za ispitivanje učenja i razumijevanja (Brewer, 2004).

SS u točnost određene odluke ili suda može se razmatrati u različitim teorijskim okvirima. Ponderiranje SS ponekad se koristi u modelima koji se bave procesom donošenja odluke, tj. pri analizi sklonosti ispitanika da preuzme rizik u određenoj situaciji. Izbor konačnog odgovora u zadatku u osnovi ima elemente donošenja odluke. U kontekstu primjene SS u teoriji testova često se koristi koncept metakognicije. Metakognicija se, u širem smislu, odnosi na uvid koji pojedinac ima o vlastitoj kogniciji, znanju i procesima koji su u podlozi učenja i rješavanja problema. Metakognitivni procesi su usko povezani sa samoregulacijom te omogućuju osobi da kontrolira učenje i rješavanje problema, da odredi ciljeve i strategije koje koristi pri rješavanju problema. Metakognicija je važna jer uključuje motrenje vlastitih kognitivnih procesa i omogućuje upravljanje tim procesima s ciljem efikasnijeg rješavanja problema. Kognitivne sposobnosti omogućuju rješavanje problema ili zadatka, dok metakognicija omogućuje razumijevanje procesa ili načina na koji je zadatak riješen (Schraw i sur., 2006). Pored toga metakognicija posredno, uz kogniciju, omogućuje provjeru i evaluaciju ispravnosti rješenja zadatka. Slična uloga, osobito kod zadataka dosjećanja, može se pripisati ulozi metamemorije, koja podrazumijeva znanje osobe o vlastitom pamćenju i subjektivno zadovoljstvo vlastitim sposobnostima pamćenja (Vranić i Tonković, 2011). Stankov i sur. (2014) navode dva smjera istraživanja u području individualnih razlika u području SS. Dio istraživanja koja su predmet našeg interesa u ovom radu, odnosi se na neposredne procjene uvjerenja u točnost izvedbe konkretne kognitivne ili bihevioralne aktivnosti. Drugi smjer proučava širi spektar dispozicijskih mjera koje opisuju način na koji ispitanik doživljava sebe i svoju uspješnost poput samo-

efikasnosti, općeg samopoštovanja, samosviđanja, samopouzdanja ili samokompetencije. Samoefikasnost se odnosi na uvjerenje da će netko uključen u određenu aktivnost ostvariti pozitivan ishod vezan uz specifični zadatak ili domenu. Zbog šireg pozitivnog ili negativnog doživljaja sebe pojedinac može općenito biti manje ili više siguran u ispravnost svog konkretnog uratka, osobito u situacijama kada mu nedostaju metakognitivne informacije o realnoj točnosti uratka. Konceptualno je moguće od ispitanika tražiti i prospektivnu procjenu vjerojatnosti točnog rješavanja konkretnog zadatka te je usporediti s procjenom SS nakon rješavanja. Korelacija između ova dva konstrukta je srednje veličine ( $r = 0,53$ ) i istraživanja pokazuju da su uvjerenja u točnost uratka bolji prediktor stvarnog uratka od mjera samoefikasnosti (Morony i sur., 2013). Ponekad se ova prospektivna procjena naziva „osjećaj znanja“ (eng. FOK – feeling of knowing), a može se zatražiti i u slučaju kada ispitanik nije ponudio odgovor na zadatak s uputom da procijeni vjerojatnost da bi ga ipak mogao riješiti. Istraživanja upućuju na činjenicu da su metakognitivne evaluacije u području pamćenja podijeljene između semantičkih i epizodičkih zadataka pamćenja. Mazancieux i sur. (2020) nalaze povezanost metakognitivne efikasnosti pri procjeni SS kod zadataka epizodičkog i semantičkog pamćenja. Autori navode da takvu povezanost ne nalaze kod procjena FOK. Zaključuju da su procjene FOK drugačije povezane s epizodičkim i semantičkim pamćenjem u odnosu na opće procjene SS nakon uratka.

Stankov i Lee (2014) navode da su SS u različitim vrstama kognitivnih testova povezani, što upućuje na određeni stupanj generalnosti ovih procjena (Stankov i Kleitman, 2008). Stankov i sur. (2012) sugeriraju da je stupanj sigurnosti odvojena psihološka dimenzija koja se nalazi u prostoru između kognitivnih sposobnosti i osobina ličnosti, dijelom povezana s drugim

konstruktima kao što su različite mjere samo-poimanja, intelektualne angažiranosti te emocionalne inteligencije. Istraživanja pokazuju da je SS najbolji nekognitivni prediktor postignuća u testovima engleskog jezika i matematike, pored drugih mjera kao što su samoefikasnost, anksioznost i druge mjere samopoimanja (Stankov i sur., 2014). Autori naglašavaju da SS i srodne mjere predstavljaju različite faktore niže razine koji su u osnovi faktora drugog reda koji nazivaju vjerovanje u sebe. Neki autori navode da SS može predvidjeti fenomen kada ispitanici doživljavaju da nisu toliko dobri koliko to pokazuju njihovi rezultati, odnosno postignuće (Want i Kleitman, 2006) te mogu biti korisni u procesu formativnog vrednovanja pri učenju novih sadržaja (Kleitman i Costa, 2014). Stupanj sigurnosti nije povezan s karakteristikama ličnosti s izuzetkom otvorenosti iskustvu (oko 0,3), dok su korelacije s mjerama metakognicije također niske i kreću se oko 0,4 (Kleitman i Stankov, 2007). Šabić (2008) pri primjeni testa znanja iz psihologije na gimnazijalcima ne nalazi povezanost SS u točne odgovore sa samosviđanjem i samokompetencijom, i nalazi tek nisku korelaciju ove dvije mjere sa SS u netočne odgovore (oko 0,2). Za uporabu SS pri vrednovanju testnih rezultata načelno nije poželjna povezanost ovih procjena s nekim dispozicijskim osobinama ispitanika osim u onoj mjeri u kojoj su te osobine povezane s glavnim predmetom mjerenja nekim konkretnim testom. Temeljna logika uporabe SS pri psihometrijskoj analizi testa zasniva se na povezanosti stvarne razvijenosti mjerenog konstrukta i procijenjene sigurnosti odgovora u vlastite točne i netočne odgovore, odnosno očekivanju da će superiorniji ispitanici u prosjeku imati viši stupanj sigurnosti u svoje odgovore od inferiornih ispitanika. Lichtenstein i Fischhoff (1977) postavljaju jedno od ključnih pitanja u ovom području: „Imaju li ispitanici koji znaju više, ujedno i veće znanje o tome koliko znaju?“. Superiorni ispitanici imaju bolji uvid u procese koji vode

do točnog odgovora, češće su dobivali povratne informacije o svom znanju ili sposobnosti te evaluirali točnost svojih ranijih odgovora, imaju više iskustva u rješavanju određenih zadataka, a metakognitivno su više usmjereni na analizu svojih odgovora i provjeru njihove valjanosti. Dodatna pretpostavka, za koju nema potvrde u literaturi, jest da će sigurnost superiornih ispitanika u njihove netočne odgovore biti niža od inferiornih ispitanika jer oni u slučaju odabira netočnog odgovora realističnije postavljaju hipotezu o mogućoj pogrešci pri odgovaranju. Eksperti znaju točnije procijeniti svoje znanje pa time i jasnije razlikovati sadržaje koje znaju od sadržaja koje ne znaju. Boldt i suradnici (2019) naglašavaju važnost točnosti ranijih procjena sigurnosti na kasnije procjene te nalaze da se procjene sigurnosti zasnivaju na više izvora uključujući i očekivanja. Bujas i suradnici (1975) nalaze da povratne informacije o ranijim odgovorima dovode do većeg stupnja sigurnosti u točne odgovore, a manjeg u netočne.

Dio istraživanja usmjeren je na intraindividualnu razinu procjena i u tom kontekstu su česta istraživanja kalibracije i razlučivosti procjena. Kalibriranost procjena upućuje na stupanj u kojemu su stupnjevi sigurnosti usklađeni sa stvarnom točnošću odgovora. Kod savršeno kalibriranog procjenjivača sve procjene stupnja sigurnosti usklađene su sa stvarnom proporcijom točnih odgovora, a pristranost se očituje u mogućem precjenjivanju i podcjenjivanju stvarnog uratka (Lichtenstein i Fischhoff, 1977). Precjenjivanje sigurnosti u točnost odgovora u području kognicije jedan je od sustavnih nalaza u brojnim istraživanjima (Koriat, 2012; Lichtenstein i Fischhoff, 1977; McKenzie, 1997). Pallier i sur. (2002) navode da su procjene točnosti odgovora u klasičnim testovima inteligencije s neverbalnim sadržajem uglavnom realistične. Kalibracija nije povezana s razlikama u inteligenciji, stručnosti, korištenju ekstremnih odgovora te

nekim aspektima konteksta u kojima se zadaci zadaju (Lichtenstein i Fischhoff, 1977). Razlučivost ili diskriminacijska točnost korištena je u različitim istraživanjima (Koriat, 2012; Lichtenstein i Fischhoff, 1977), a odnosi se na sposobnost ispitanika da na osnovi procijenjene sigurnosti razlikuje svoje točne i netočne odgovore. Autori zaključuju da je razlučivost neovisna o kalibraciji.

Razmjerno je malo rezultata u literaturi koji bi upućivali na psihometrijske karakteristike SS kao dodatne mjere uratka u testu. Stankov i Lee (2008) navode da se korelacije između točnosti uratka i SS kreću u rasponu od 0,2 do 0,6. Šabić (2008) navodi korelaciju između broja točnih odgovora u testu znanja i prosječnog SS od 0,36, pri čemu su SS u točne odgovore korelirali 0,33 s brojem točnih odgovora, dok kod SS u netočne odgovore ove povezanosti nema. Rezultati studija upućuju na postojanje generalnog faktora SS, koji je odvojen (iako pozitivno povezan) od faktora inteligencije (Kleitman i Stankov, 2007; Stankov i Crawford, 1996; Stankov i Lee, 2008). Nema podataka o latentnoj strukturi procjena SS u jednom testu, ali većina autora upućuje na visoku pouzdanost ovih procjena uz zadatke istog testa koje se kreću od 0,90 do čak 0,94 (Stankov i Lee, 2008), dok je Cronbach alpha za mjere točnosti u istom testu niži i kreće se oko 0,80.

U literaturi su korišteni i opisani različiti metodološki pristupi ispitivanju i registraciji stupnja sigurnosti koji ponekad ovise o općem cilju istraživanju, vrsti zadatka, predmetu mjerenja te drugim karakteristikama cjelokupnog procesa mjerenja. Curran (2012) u nizu simulacijskih i empirijskih analiza potvrđuje korist od uključivanja procjena sigurnosti u psihometrijsku analizu testnih rezultata te zaključuje da se većina različitih postupaka ponderiranja može razmatrati kao specifičan oblik generalnijeg modela ponderiranja točnosti stupnjevim sigurnosti. Autor zaključuje da cijelo područje ponderiranja sigurnošću

zahtijeva određenu integraciju. Pojedini autori navode da format zadatka (verbalno-numerički ili grafičko-vizualni) nije povezan s procjenama stupnja sigurnosti, ali da je registriran značajan efekt interakcije formata zadatka i vremenskog pritiska tijekom rješavanja (Agus i sur., 2020). Paradigma procjene sigurnosti često se kombinira uz procjene točnosti. U području percepcije uz donošenje odluke o dva vizualna podražaja od ispitanika se zahtijeva i procjena koja od dvije odluke je vjerojatnije točna (Mamassian, 2020). Ovi postupci, pored dihotomiziranja ispitanika u dvije kategorije (točno-netočno), omogućuju dodatno diferenciranje s obzirom na sposobnost rangiranja različito vjerojatnih odluka. Ta paradigma se ponekad operacionalizira kroz zadatak ispitaniku da odabere odgovor/odgovore koji nisu sigurno točni ili da rangira moguće odgovore prema vjerojatnosti da su točni. Sposobnost eliminacije pojedinih rješenja nudi dodatnu diskriminacijsku informaciju o razvijenosti konstrukta. Među razmjerno rijetkim istraživanjima primjene SS na testovima kognitivnih sposobnosti Stankov i Crawford (2006) usmjereni su na problem pristranosti, a manje na psihometrijska svojstva procjena SS. Uz test RPM navode bolju diskriminaciju i sporije odgovaranje ispitanika i ne nalaze povezanost točnosti uratka s mjerama općeg samopoštovanja ili matematičkog samopoimanja. Jedno od rijetkih novijih istraživanja usmjerenih na psihometrijske aspekte uključivanja procjena SS u analizu testa navode Stankov i Lee (2008) na testovima verbalnog sadržaja. Autori nalaze značajnu, ali nisku do srednju povezanost procjena SS i točnosti te upućuju na to da su procjene SS jedan od najboljih nekognitivnih prediktora uratka u testovima znanja. Stoga je ovom radu naš interes usmjeren na analizu mogućnosti uključivanja stupnja sigurnosti u psihometrijsku analizu neverbalnog testa kognitivnih sposobnosti i to na tri razine. Prva razina odnosi se na uporabu informacija o SS na razini analize testnih zadataka i evaluacije njihovih psihometrijskih karakteristika. Druga razina odnosi se

na intraindividualne karakteristike procjena SS te pristranost i razlikovanje vlastitih točnih i netočnih odgovora. Treća razina usmjerena je na mogućnost ponderiranja mjera točnosti stupnjema sigurnosti te analizu individualnih razlika u stupnjema sigurnosti na razini ukupnih testnih rezultata. Širi cilj ovog rada jest ispitati informacijsku vrijednost uključivanja stupnja sigurnosti u psihometrijsku analizu testa i vrednovanje individualnih razlika. Na osnovi općeg cilja formulirani su sljedeći problemi:

1. ispitati povezanost prosječnih procjena stupnja sigurnosti u točnost odgovora na pojedine zadatke s drugim psihometrijskim karakteristikama zadataka (indeks lakoće, diskriminativna valjanost)
2. ispitati latentnu strukturu i unutarnju konzistenciju procjena stupnja sigurnosti na skupu testnih zadataka
3. ispitati neke prediktore prosječnog stupnja sigurnosti u točnost odgovora na testne zadatke
4. ispitati prosječni stupanj sigurnosti u točne i netočne zadatke u funkciji razvijenosti predmeta mjerenja
5. ispitati utjecaj diferencijalnog ponderiranja uratka u zadacima stupnjem sigurnosti u točnost odgovora na mjere unutarnje konzistencije testa.

## METODA

### Sudionici

Ciljana populacija u ovom istraživanju bili su studenti preddiplomskih i diplomskih studija.

Realizirani uzorak je prigodni uzorak od 168 studenata Sveučilišta u Zagrebu te u osnovi dobro reprezentira ciljanu populaciju s obzirom na spol i dob. Ukupno je sudjelovalo 107 žena (63,7%) te 61 muškarac (36,3%). S obzirom na vrstu studija, studente možemo po-

dijeliti na studente psihologije, kojih je bilo 70 (41,6%), te studente drugih, 46 različitih, studijskih grupa. Sto deset studenata (65,6%) studenti su preddiplomskih studija, a 58 (34,4%) diplomskih studija. Prosječna ocjena ostvarena na studiju iznosila je 3,8. Studenti psihologije dobrovoljno su se javili za sudjelovanje u istraživanju te je svaki od njih imao zadatak uključiti u istraživanje 1-2 redovita studenta Sveučilišta u Zagrebu koji studiraju neki drugi studij.

### Instrumenti

#### Test kognitivnih sposobnosti – Test S (računalna primjena)

Test S (Test slova) je test kognitivnih sposobnosti razvijen 2007. u okviru projekta *Razvoj, standardizacija i psihometrijska validacija testova kognitivnih sposobnosti*. Testom se ispituje sposobnost logičkog rezoniranja na apstraktnom sadržaju. Testni zadaci sastoje se od skupina slova koja imaju neko zajedničko obilježje. Zadatak ispitanika jest da to zajedničko obilježje utvrdi i od ponuđenih 6 odgovora odabere jedan odgovor koji također posjeduje isto obilježje. Autor testa je D. Ljubotina, a psihometrijske karakteristike provjeravane su u više navrata i upućuju na visoku pouzdanost koja iznosi oko 0,90 (Tonković, 2007). Teorijski raspon ukupnih rezultata kreće se od 0 do 40, a pouzdanost testa izražena Cronbachovim alpha koeficijentom u ovom istraživanju iznosila je 0,89. Prosječna interkorelacija među česticama testa iznosi  $r = 0,17$ .

U istraživanju je primijenjena računalna verzija testa od 40 zadataka, koja je omogućila registraciju točnosti uratka, stupnja sigurnosti te nekih mjera vremena rada u zadatku.

#### Mjere stupnja sigurnosti

Ponderiranje točnosti uratka može se zasnivati na različitim heuristikama. U ovom

radu korištena je paradigma prema kojoj se od ispitanika traži da neposredno nakon izbora odgovora u pojedinom zadatku procijeni stupanj sigurnosti u točnost odabranog odgovora na ljestvici od 0 (potpuna nesigurnost) do 100 (potpuna sigurnost). Ovaj postupak upotrebljavaju Stankov i Crawford (1996).

Nakon odabranog odgovora na ekranu računala pojavila bi se ljestvica na kojoj je ispitanik mogao uporabom miša odrediti stupanj sigurnosti na ljestvici od 0 do 100. Za to vrijeme bilo je zaustavljeno vrijeme predviđeno za rad na testnim zadacima. Kod zadataka u kojima ispitanik nije odabrao rješenje ne postoji procjena SS. Na ovaj način registrirano je maksimalno 40 procjena SS uz svaki zadatak iz čega su izvedene dodatne mjere: prosječni SS uz sve zadatke, prosječni SS uz točno riješene zadatke, prosječni SS uz netočno riješene zadatke. Kod ispitanika koji nisu odgovorili na sve zadatke prosječni rezultati zasnivaju se na manjem broju procjena SS.

### Procjene metakognicije za test kognitivnih sposobnosti

Metakognicija u testu kognitivnih sposobnosti procijenjena je dvama pitanjima koja upućuju na samoprocjenu ispitanika o vlastitom uratku na testu kognitivnih sposobnosti. *Prvo pitanje*, „Procijenite broj zadataka koje mislite da možete točno riješiti u Testu S (0-40)“ postavljeno je prije rješavanja testa nakon što je ispitanik vidio vrstu zadataka od kojih se test sastoji. Ova procjena je prospektivna metakognitivna procjena.

*Drugo pitanje* postavljeno je neposredno nakon što je ispitanik završio s radom u testu kognitivnih sposobnosti i glasilo je: „Procijenite broj zadataka za koje mislite da ste ih točno riješili u Testu S (0-40)“. Ova druga procjena je retrospektivna metakognitivna procjena nakon iskustva rješavanja testa.

### Mjera testne anksioznosti

Za mjerenje testne anksioznosti korištena je kratka skala od 6 pitanja konstruirana za potrebe ranijeg istraživanja (Ljubotina i sur., 2015). Pitanja su formulirana kako bi bila usmjerena na anksioznost vezanu upravo uz situaciju rješavanja testa. Zadatak ispitanika bio je da na ljestvici od 1 (nikada) do 5 (gotovo uvijek) označe u kojoj mjeri tvrdnja opisuje njihovo uobičajeno ponašanje za vrijeme rješavanja testa (primjer čestice „Za vrijeme rada na testu osjećam se uznemireno“). Ukupni rezultat izražava se kao prosječan odgovor na svih 6 pitanja, pri čemu viši rezultat upućuje na višu testnu anksioznost i može varirati u rasponu od 1 do 5. Visok rezultat upućuje na ispitanika koji ima općenitu odbojnost prema situaciji testiranja, nervozan je i uznemiren za vrijeme rješavanja testa, koncentracija mu je niska te je zabrinut oko rezultata na testu. Pouzdanost skale procijenjena Cronbachovim alpha koeficijentom iznosila je 0,83, uz aritmetičku sredinu 2,81. Eksploratorna analiza zajedničkih faktora upućuje na jedan generalni faktor, pri čemu sve čestice imaju saturacije s faktorom u rasponu od 0,49 do 0,80.

### Mjera samopouzdanja

Mjera općeg samopouzdanja odnosi se na uvjerenje osobe da je sposobna postići uspjeh pri rješavanju testa. Kratka skala konstruirana za potrebe istraživanja sastojala se od 7 pitanja usmjerenih na samopouzdanje tijekom rješavanja testa.

Zadatak ispitanika bio je da na ljestvici od 1 (nikada) do 5 (gotovo uvijek) označe u kojoj mjeri tvrdnja opisuje njihovo uobičajeno ponašanje za vrijeme rješavanja testa (primjer čestice „Pri rješavanju zadataka sam uporan i ne odustajem lako“). Ukupni rezultat izražava se kao prosječan odgovor na svih 7 pitanja, pri čemu viši rezultat upućuje na veće samopouzdanje i upornost pri rješavanju zadataka te

može varirati u rasponu od 1 do 5. Pouzdanost skale procijenjena Cronbachovim alpha koeficijentom iznosila je 0,70, uz aritmetičku sredinu 2,87. Eksploratorna analiza zajedničkih faktora upućuje na jedan generalni faktor, pri čemu sve čestice imaju saturacije s faktorom u rasponu od 0,35 do 0,64.

### Sociodemografski podaci

Kratkim upitnikom prikupljeni su i podaci o spolu, godini i vrsti studija te prosječnoj ocjeni tijekom studija.

### Postupak ispitivanja

Ispitivanje je bilo anonimno i provedeno je u računalnoj dvorani, a trajalo je približno 50-60 minuta. S obzirom na veličinu računalne dvorane ispitanici su test rješavali u skupinama od po 20 osoba uz nadzor dva ispitivača. Ispitanici su u prvom dijelu rješavali računalnu verziju testa kognitivnih sposobnosti. Prije same primjene testa, a nakon opće upute i uvida u zadatke za vježbu, ispitanici su procijenili broj zadataka u testu za koje očekuju da ih mogu točno riješiti. Nakon rješavanja testa ispitanici su procijenili broj zadataka za koje smatraju da su ih točno riješili. Nakon toga su ispunjavali dodatne upitnike i sociodemografske podatke metodom papir-olovka. Uputa u testove uključuje napomenu da je vrijeme za rad ograničeno, te stoga treba raditi brzo i točno.

## REZULTATI

---

### Karakteristike procjena stupnja sigurnosti za pojedine zadatke

Analize podataka provedene su u statističkom programu IBM SPSS Statistics ver-

zija 27. U prvom dijelu analize rezultata prikazana je deskriptivna statistika parametara pojedinih zadataka te Spearmanove rang korelacije između parametara. Pored klasičnih mjera težine/lakoće i diskriminativne valjanosti za svaki zadatak izračunat je prosječni SS za sve odgovore, prosječni SS uz netočne odgovore, prosječni SS uz točne odgovore te razlika između ove dvije mjere. Indeks pristranosti zadatka izračunat je kao razlika između indeksa lakoće (proporcija točnih odgovora transformirana je na ljestvicu od 0 do 100) i prosječnog SS. U tom slučaju pozitivna vrijednost indeksa upućuje na precjenjivanje stvarnog uratka, nula upućuje na nepristranu procjenu, a negativne vrijednosti upućuju na podcjenjivanje uratka u zadatku.

Rezultati prikazani u Tablici 1 predstavljaju deskriptivnu statistiku za svih sedam parametara te njihove interkorelacije. Prosječni SS (PSS) izračunat za svih 40 zadataka iznosi  $M = 68,66$ , pri čemu su ispitanici značajno višom ( $t = -16,35$ ;  $df = 39$ ;  $p < 0,001$ ) procjenjivali sigurnost u točne odgovore ( $M = 76,97$ ) od sigurnosti u točnost netočnih odgovora ( $M = 41,46$ ). Indeks pristranosti pokazuje da su procjene stupnjeva sigurnosti tek neznatno podcijenile stvarnu točnost rješavanja zadataka. Prosječna proporcija ispitanika koji su točno riješili zadatak (prosječni indeks lakoće) upućuje na to da su u prosjeku ispitanici ostvarili 71,1% točnih odgovora. Kako je broj jedinica analize (zadataka) u ovom slučaju razmjerno mali ( $k = 40$ ) kao mjera povezanosti korišten je Spearmanov koeficijent rang korelacije. U pojedinim zadacima prosječne vrijednosti parametara izračunate su samo na odgovorima ispitanika koji su procijenili SS tako da se prosjeci vezani uz SS zasnivaju na broju ispitanika koji varira od 150 do 168.



**Tablica 1.** Deskriptivna statistika za parametre zadataka korištene u istraživanju i Spearmanovi koeficijenti rang korelacije između parametara ( $k = 40$ )

	1.	2.	3.	4.	5.	6.	Minimum	Maksimum	M	SD
1. Indeks lakoće	1						0,21	0,98	0,71	0,19
2. Diskriminativna valjanost (nespuriozna)	-0,27	1					0,09	0,68	0,39	0,15
3. Prosječni stupanj sigurnosti (PSS)	0,80***	-0,18	1				33,07	94,52	68,66	14,35
4. PSS samo za netočne odgovore	0,18	-0,48**	0,46**	1			23,53	84,69	41,46	13,03
5. PSS samo za točne odgovore	0,56***	0,00	0,91***	0,40**	1		42,50	95,73	76,97	12,38
6. Razlika između PSS za točne i PSS za netočne odgovore	0,37*	0,51***	0,36*	-0,52***	0,45**	1	7,76	60,69	35,51	13,74
7. Indeks pristranosti (razlika PSS i indeksa lakoće)	-0,60***	0,01	-0,08	0,35*	0,19	-0,26	-20,69	26,48	-2,45	10,31

Napomena: \*\*\*  $p < 0,001$ ; \*\*  $p < 0,01$ ; \*  $p < 0,05$ ; Minimum – najmanji opaženi rezultat; Maksimum – najveći opaženi rezultat;

Rezultati u Tablici 1 pokazuju da postoji visoka povezanost između težine zadatka i prosječnog stupnja sigurnosti, što očekivano upućuje na veću sigurnost u točnost odgovora kod lakših zadataka ( $r = 0,80$ ;  $p < 0,001$ ). Zanimljivo je da indeks lakoće zadatka ima značajnu povezanost s PSS u točne odgovore ( $r = 0,56$ ;  $p < 0,001$ ), ali ne i s PSS u netočne odgovore ( $r = 0,18$ ;  $p > 0,05$ ). Pored toga, uz teže zadatke veže se i veći stupanj precjenjivanja točnosti uratka ( $r = -0,60$ ;  $p < 0,001$ ). Veća diskriminativna valjanost zadatka povezana je s manjim PSS uz netočne odgovore ( $r = -0,48$ ;  $p < 0,01$ ), ali ne i s PSS uz točne odgovore. Pored toga kod zadataka koji imaju veću diskriminativnu valjanost postoji tendencija veće razlike između PSS u točne i netočne odgovore. Prosječni stupnjevi sigurnosti uz točne i netočne zadatke umjereno su povezani ( $r = 0,40$ ;  $p < 0,01$ ).

### Latentna struktura i pouzdanost procjena stupnja sigurnosti točnosti odgovora uz testne zadatke

Drugi dio analize rezultata odnosi se na psihometrijske karakteristike procjena stupnjeva sigurnosti na 40 testnih zadataka. S ciljem provjere dimenzionalnosti latentnog prostora koji definira 40 procjena stupnja sigurnosti uz zadatke testa kognitivnih sposobnosti provedena je eksploratorna faktorska analiza uz primjenu modela zajedničkih faktora. U ovom slučaju potrebno je razriješiti metodološki problem vezan uz činjenicu da ispitanici koji nisu odabrali odgovore na sva pitanja u nekim zadacima nemaju procjenu stupnja sigurnosti. Kako je za provedbu faktorske analize i procjenu Cronbachova alpha koeficijenta nužno raspolagati s kompletnom matricom podataka, oba postupka provedena su na dva načina analizom matrica nazvanih A i B. Matrica A predstavljala je uzorak svih 168 ispitanika, pri čemu su po-

daci koji nedostaju zamijenjeni vrijednošću nule uz pretpostavku da ispitanik koji nije odabrao odgovor ima nultu sigurnost. Kod 53 ispitanika koji nemaju sve podatke o procjenama SS najvećem broju nedostaje jedan podatak ( $N = 18$ ), dok se kod ostalih broj zamjena kreće od 2 do 15. Ukupno je 276 praznih mjesta zamijenjeno nulom. Matrica B predstavlja uzorak od 115 ispitanika koji su odabrali sve odgovore i dali sve procjene SS. Rezultati analiza na oba uzorka upućuju na slične rezultate. Na uzorku A Bartlettov test ( $\chi^2 = 4334,93$ ;  $df = 780$ ;  $p < 0,001$ ) i Kaiser-Meyer-Olkinov indeks ( $KMO = 0,928$ ) pokazuju da je korelacijska matrica podobna za provedbu faktorske analize. Vrijednosti sedam karakterističnih korjenova sukcesivno ekstrahiranih glavnih komponenti iznose  $\lambda_1 = 16,46$ ,  $\lambda_2 = 2,41$ ,  $\lambda_3 = 1,87$ ,  $\lambda_4 = 1,56$ ,  $\lambda_5 = 1,17$ ,  $\lambda_6 = 1,12$ ,  $\lambda_7 = 1,04$  što ih čini značajnim na osnovi Kaiser-Guttmanovog kriterija (standardna opcija zadržavanja faktora u programu SPSS odnosi se na komponente s korjenovima većim od 1). Analiza Cattellova „scree-plot“ crteža te paralelne analize sugeriraju zadržavanje jednog generalnog faktora koji objašnjava 39,69% ukupne varijance. Saturacije svih čestica s generalnim zajedničkim faktorom kreću se u rasponu od 0,37 do 0,84, što zadovoljava uobičajeni kriterij od 10% zajedničke varijance između faktora i latentne varijable.

Na matrici B rezultati Bartlettova testa ( $\chi^2 = 3507,05$ ;  $df = 780$ ;  $p < 0,001$ ) i Kaiser-Meyer-Olkinov indeks ( $KMO = 0,918$ ) ponovno pokazuju da je korelacijska matrica podobna za provedbu faktorske analize. Vrijednosti osam karakterističnih korjenova sukcesivno ekstrahiranih glavnih komponenti iznose  $\lambda_1 = 18,17$ ,  $\lambda_2 = 2,38$ ,  $\lambda_3 = 2,01$ ,  $\lambda_4 = 1,59$ ,  $\lambda_5 = 1,35$ ,  $\lambda_6 = 1,08$ ,  $\lambda_7 = 1,03$ ,  $\lambda_8 = 1,01$ , što ih čini značajnim na osnovi Kaiser-Guttmanova kriterija. Analiza Cattellova „scree-plot“ crteža i paralelne analize sugeriraju

raju zadržavanje jednog generalnog faktora koji objašnjava 44,1% ukupne varijance. Saturacije svih čestica s generalnim zajedničkim faktorom kreću se u rasponu od 0,44 do 0,83. U oba slučaja analiza sadržaja pitanja koja imaju projekcije na drugom faktoru te visoka prosječna korelacija među svim pitanjima upućuju na jedan generalni faktor stupnja sigurnosti u ovom testu. U Tablici 2 prikazane su neke karakteristike kompozita sačinjenog od procjena stupnja sigurnosti izračunate na matricama A i B.

Kao i kod provedbe faktorske analize pouzdanost je procijenjena na uzorku A i uzorku B. Razlike su minimalne i u oba slučaja se pouzdanost unutarnje konzistencije kreće oko vrijednosti 0,96 uz prosječnu korelaciju među procjenama od 0,39 (matrica A) odnosno 0,43 (matrica B). Prosječan koeficijent diskriminativne valjanosti, tj. korelacija pojedinačnih procjena s prosječnom procjenom SS u cijelom testu iznosi 0,61 (matrica A) odnosno 0,65 (matrica B).

**Tablica 2.** Deskriptivna statistika i pouzdanost za kompozit sačinjen od procjena stupnjeva sigurnosti na 40 zadataka Testa S

	<i>M</i>	<i>SD</i>	<i>min</i>	<i>max</i>	$\alpha$	$r_{ij}$	$r_{iu}$
Matrica A (N = 168)	66,25	21,30	10,75	100	0,96	0,39	0,61
Matrica B (N = 115)	68,16	21,43	10,75	100	0,97	0,43	0,65

Matrica A = prazna mjesta uz procjenu stupnja sigurnosti zamijenjena su nulom; Matrica B = uzorak ispitanika koji su dali odgovore na sva pitanja;  $\alpha$  – Cronbachov alpha koeficijent pouzdanosti;  $r_{ij}$  – prosječna interkorelacija čestica u upitniku;  $r_{iu}$  – prosječna korelacija čestica s ukupnim rezultatom

### Analiza ukupnih individualnih rezultata i predikcija prosječnog stupnja sigurnosti

Treći dio analize odnosio se na karakteristike ukupnih individualnih rezultata u testu te mogućih korelata mjera točnosti i stupnja sigurnosti. Prosječni SS za pojedinog ispitanika izračunat je na raspoloživom broju valjanih procjena, što znači da je kod 115 ispitanika broj procjena 40, dok je kod 53 ispitanika prosjek izračunat na manjem broju procjena (u rasponu od 22 do 39). U Tablici 3 prikazana je deskriptivna statistika te interkorelacije između točnosti uratka, mjera stupnja sigurnosti, metakognicije, anksioznosti, samopouzdanja i spola. Ukupan broj ispitanika koji su imali valjane podatke u dodatnim

varijablama koje se odnose na metakogniciju, anksioznost i samopouzdanje iznosio je 151.

Iz rezultata u Tablici 3 razvidno je da retrospektivna metakognitivna procjena očekivanog broja bodova u testu predstavlja najbolji pojedinačni prediktor prosječnog SS uz sve odgovore ( $r = 0,79$ ;  $p < 0,001$ ), dok je korelacija sa stvarnim brojem točnih odgovora 0,61 ( $p < 0,001$ ). Prospektivna metakognitivna procjena uratka u testu povezana je s prosječnim SS 0,44 ( $p < 0,001$ ), ali nije povezana sa stvarnim brojem točnih odgovora. Viši prosječni SS u sve odgovore nisko, ali značajno, povezan je s manjom anksioznošću i većim samopouzdanjem. Muškarci također imaju nešto viši prosječni stupanj sigurnosti od žena. Važno je uočiti da je klasični broj točnih odgovora povezan samo s prosječnim SS

**Tablica 3.** Pearsonovi koeficijenti korelacije između točnosti uratka, mjera SS, metakognicije, anksioznosti, samopouzdanja i spola ( $N = 151$ )

	1.	2.	3.	4.	5.	6.	7.	8.	<i>M</i>	<i>SD</i>
1. Broj točnih odgovora	1								28,64	7,01
2. Prosječni stupanj sigurnosti	0,61***	1							69,62	20,04
3. Prosječni SS uz točne odgovore	0,37***	0,86***	1						80,60	14,91
4. Prosječni SS uz netočne odgovore	0,15	0,69***	0,70***	1					51,07	21,99
5. Metakognicija 1 - prospektivna procjena	0,07	0,44***	0,49***	0,47***	1				26,71	6,77
6. Metakognicija 2 - retrospektivna procjena	0,63***	0,79***	0,63***	0,51***	0,39***	1			24,60	9,26
7. Testna anksioznost	-0,12	-0,26***	-0,20*	-0,22**	-0,21**	-0,24**	1		2,81	0,74
8. Samopouzdanje	0,15	0,24**	0,18*	0,19*	0,22**	0,27***	-0,67***	1	2,87	0,55
9. Spol	0,07	-0,18*	-0,22**	-0,30**	-0,26***	-0,12	0,18*	0,14	-	-

Napomena: \*\*\*  $p < 0,001$ ; \*\*  $p < 0,01$ ; \*  $p < 0,05$

u točne odgovore, ali ne i s prosječnim SS u netočne odgovore. Pored toga broj točnih odgovora korelira s retrospektivnom procjenom broja bodova 0,63 ( $p < 0,001$ ). PSS u točne i PSS u netočne odgovore na sličan način su povezani s obje metakognitivne procjene kao i s mjerama anksioznosti i samopouzdanja.

Više metakognitivne procjene povezane su nisko, ali značajno s nižom anksioznošću te višim samopouzdanjem. Ako bismo na osnovi prosječnog stupnja sigurnosti uz sve zadatke pokušali prognozirati opaženi broj točnih odgovora u testu dobivamo vrijednost 27,8 (prosječni stupanj sigurnosti 69,6, tj. vjerojatnost točnog odgovora od 0,696 pomnožena s 40), što je prilično točna procjena aritmetičke sredine broja točnih odgovora u testu koja iznosi 28,6 ( $t = 1,47$ ;  $df = 167$ ;  $p > 0,05$ ). Analiza spolnih razlika pokazuje da na uzorku muškaraca postoji značajno precjenjivanje stvarnog uratka od 2,14 bodova

( $t = -2,54$ ;  $df = 60$ ;  $p < 0,05$ ), dok je na uzorku žena registrirano značajno podcjenjivanje od 2,51 boda ( $t = 3,65$ ;  $df = 106$ ;  $p < 0,001$ ).

### Predikcija individualnog prosječnog stupnja sigurnosti uz sve zadatke na testu

U Tablici 4 prikazani su rezultati hijerarhijske regresijske analize pri predikciji prosječnog stupnja sigurnosti u točnost odgovora na testu kognitivnih sposobnosti.

Statistički pokazatelji kolinearnosti (indeks tolerancije i *VIF*) za sve varijable uključene u analizu pokazuju da nisu prisutni problemi singulariteta i multikolinearnosti podataka.

Logika redoslijeda uvođenja prediktorskih varijabli zasnivala se na očekivanju da je najlogičniji teorijski prediktor stupnja sigurnosti razvijenost konstrukta odmjeren testom te na-

**Tablica 4.** Rezultati hijerarhijske regresijske analize za kriterij Prosječni stupanj sigurnosti ( $N = 151$ )

	$\beta_{M1}$	$\beta_{M2}$	$\beta_{M3}$
Broj točnih odgovora	0,61***	0,25***	0,27***
Metakognicija 1 - prospektivna procjena		0,21***	0,19***
Metakognicija 2 - retrospektivna procjena		0,55***	0,53***
Spol			-0,08
Testna anksioznost			0,05
Samopouzdanje			-0,08
$R^2$ ( $R^2_{adj}$ )	0,37 (0,37)	0,68 (,68)	0,69 (0,68)
$F$	88,57	104,66	53,76
$P$	<0,001	<0,001	<0,001
$\Delta R^2$		0,31	0,01
$F_{\Delta R^2}$		71,06	1,59
$p_{\Delta R^2}$		<0,001	>0,05

Napomena: \*\*\*  $p < 0,001$ ;  $\beta$  – vrijednost standardiziranog regresijskog koeficijenta; M1, M2, M3 – grupe prediktora u hijerarhijskoj regresijskoj analizi (modeli);  $R^2$  – ukupni doprinos objašnjenju varijanci;  $R^2_{adj}$  – korigirani ukupni doprinos objašnjenju varijanci;  $F$  – vrijednost ukupnog  $F$  omjera;  $\Delta R^2$  – doprinos dodatne grupe prediktora objašnjenju varijanci;  $F_{\Delta R^2}$  – vrijednost  $F$ -omjera za dodatnu grupu prediktora.

kon toga metakognitivne procjene razvijenosti konstrukta. Na osnovi te logike u prvom koraku kao prediktor uveden je broj točnih odgovora koji objašnjava 37,3% varijance kriterija. U drugom koraku uvedene su dvije metakognitivne procjene koje objašnjavaju dodatnih 30,8% varijance prosječnog stupnja sigurnosti. Ovdje treba upozoriti da su prospektivna i retrospektivna procjena stvarnog broja bodova u testu umjereno povezane ( $r = 0,39$ ;  $p < 0,001$ ), ali obje mjere imaju značajne beta pondere u regresijskoj jednadžbi, što sugerira da objašnjavaju dijelom različite dijelove varijance u kriteriju. U trećem koraku uvedene su varijable testne anksioznosti, samopouzdanja i spola čiji doprinos objašnjenju varijance kriterija nije statistički značajan. U slučaju provedbe zasebne regresijske analize uz uporabu anksioznosti, samopouzdanja i spola kao prediktora dobiva

se koeficijent multiple korelacije od  $R = 0,31$  ( $p < 0,01$ ), čime se objašnjava 9,3% varijance u kriteriju, međutim pojedinačni prediktori ne dostižu statističku značajnost.

### Analiza razlika u prosječnom stupnju sigurnosti uz točne i netočne odgovore

Uzorak ispitanika podijeljen je u tri skupine obzirom na broj točnih odgovora u testu kognitivnih sposobnosti te su za svaku skupinu izračunati prosječni SS uz točne i netočne odgovore (Tablica 5). S obzirom na asimetričnu distribuciju ispitanici su podijeljeni u tri skupine na osnovi normativnih podataka iz ranijeg istraživanja kako bi se osiguralo maksimalno razlikovanje skupina i dovoljan broj ispitanika u svakoj od skupina (Ljubotina i sur., 2015).

**Tablica 5.** Prosječni stupanj sigurnosti uz točne i netočne odgovore za 3 skupine ispitanika koji se razlikuju prema uratku na testu

	N	PSS točni odgovori	PSS netočni odgovori
Ispodprosječni ispitanici (0 do 22 boda)	36	74,07	46,68
Prosječni ispitanici (23 do 33 bodova)	65	79,49	50,53
Iznadprosječni ispitanici (34 do 40 bodova)	50	86,75	54,93
	151	80,60	51,07

Složena analiza varijance za djelomično zavisne varijable rezultirala je statistički značajnom razlikom između SS uz točne i netočne odgovore ( $F = 493,94$ ;  $df = 1/148$ ;  $p < 0,001$ ), pri čemu postoji generalni efekt višeg stupnja sigurnosti uz točne odgovore (Tablica 5). Nezavisni efekt razlike između 3 skupine ispitanika koji se razlikuju prema uratku u testu kognitivnih sposobnosti statistički je značajan ( $F = 2188,46$ ;  $df = 1/148$ ;  $p < 0,05$ ) te upućuje na viši stupanj sigurnosti kod ispitanika s razvijenijim konstruktom. Interakcijski efekt u ovom slučaju nije statistički značajan ( $F = 0,90$ ;  $df = 2/148$ ;  $p > 0,05$ ) što upućuje na zaključak da su veličine razlika podjednake na svim razinama razvijenosti konstrukta. Za kompletnu analizu Wilks' lambda iznosi 0,23 ( $df = 1/148$ ;  $p < 0,001$ ).

### Ponderiranje točnosti uratka stupnjem sigurnosti

U završnom dijelu analize provedeno je diferencijalno ponderiranje uratka u zadacima na način da je točnost odgovora u svakom zadatku ponderirana stupnjem sigurnosti koji je ispitanik procijenio uz taj zadatak. Ovaj oblik ponderiranja nije linearna transformacija rezultata u zadatku jer se za svakog ispitanika koristi njegov koeficijent važnosti stoga originalni i transformirani rezultati u svakom

zadatku nisu u potpunoj korelaciji. Nakon ovako provedenog ponderiranja izračunat je Cronbachov alpha koeficijent za cijeli test koji iznosi 0,94 i povećan je u odnosu na koeficijent izračunat na 40 binarnih mjera točnosti uratka koji je iznosio 0,90. Prosječna korelacija među zadacima povećana je s ranijih 0,17 na 0,29. Prosječni koeficijent diskriminativne valjanosti povećan je s 0,39 na 0,52. Broj različitih opaženih rezultata u testu povećan je s početnih 31 na 134 nakon ponderiranja.

## RASPRAVA

Cilj provedenog istraživanja bio je ispitati mogućnosti uključivanja SS u točnost odgovora u psihometrijsku analizu testa kognitivnih sposobnosti. Analiza je provedena na razini karakteristika testnih zadataka, na razini intraindividualnog razlikovanja SS u točne i netočne odgovore te na razini ukupnih intraindividualnih rezultata. Empirijski rezultati upućuju na višestruke mogućnosti uključivanja SS u analizu testa i vrednovanje ukupnih rezultata. Na razini analize zadataka SS pokazuje povezanost s indeksom lakoće (tj. općom težinom zadatka), ali tome doprinosi primarno procjena sigurnosti u točne odgovore, dok se za SS u netočne odgovore može pretpostaviti

da zahvaća druge pretežito dispozicijske osobine poput anksioznosti. Informacija o SS u točne i netočne odgovore kao i prosječni SS dostupni su za svaki pojedini zadatak i pored klasičnih parametara mogu upućivati na opću kvalitetu zadatka i procese koje izaziva kod ispitanika. Rezultati pokazuju da zadaci s višom diskriminativnom valjanošću imaju tendenciju veće razlike u SS između točnih i netočnih odgovora. Podjednaka (ne)sigurnost u točne i netočne odgovore može upućivati na probleme u konceptualizaciji zadatka te poslužiti kao dodatni kriterij pri izboru zadataka u konačnu verziju testa. U dostupnoj literaturi nema raspoloživih podataka o psihometrijskim svojstvima parametara zadataka pri procjenama SS izuzev istraživanja Šabića (2008). On nalazi da su prosječan SS i težina zadatka (uz različite vrste zadataka) povezani između 0,7 i 0,8, s izuzetkom zadataka tipa točno-netočno kod kojih nema povezanosti indeksa lakoće zadatka i sigurnosti. Navedena studija potvrđuje veće razlike između SS u točne i netočne odgovore kod zadataka višeg stupnja valjanosti. Čini se da uključivanje ove vrste informacije omogućuje autoru testa da pri konstrukciji i izboru zadataka raspolaže dodatnim informacijama o njihovoj kvaliteti. Pregled literature upućuje na činjenicu da se ova vrsta informacija vrlo rijetko koristi u procesu konstrukcije kompozitnih testova.

Druga razina analize odnosila se na karakteristike rezultata izvedenih iz procjena SS u točnost testnih zadataka. Analiza matrice dobivene na procjenama stupnja sigurnosti upućuje na zadovoljavajuće psihometrijske karakteristike. Latentna struktura pokazuje snažnu tendenciju generalnom faktoru, na što upućuje i razmjerno visoka homogenost kompozita od 40 procjena. Nalazi iz literature upućuju na postojanje generalnog faktora sigurnosti u točnost odgovora, koji je u određenoj mjeri stabilan kroz različite vrste zadataka (Kleitman i Stankov, 2007; Pallier i sur., 2002; Stankov i

Crawford, 1996). Pouzdanost tipa unutarnje konzistencije visoka je i iznosi 0,96, što odgovora rijetkim nalazima u literaturi (Costermans i sur., 1992; Stankov i Lee, 2008). Ovi nalazi upućuju na povezanost SS s mjerama točnosti uratka u zadacima, pri čemu je prosječna korelacija među procjenama nešto viša u odnosu na interkorelacije među mjerama točnosti uratka (Stankov i Crawford, 1996). Tretman zadataka uz koje ispitanici nisu procijenili SS ostaje otvoreno metodološko pitanje, ali rezultati ove studije sugeriraju da zamjena nedostajućih podataka nultim vrijednostima može biti jedna od mogućnosti u slučaju kada broj takvih vrijednosti nije velik.

Analize pokazuju da su dva najbolja prediktora prosječnog stupnja sigurnosti metakognitivna procjena uratka u testu te stvarni broj točnih odgovora. Uvođenje anksioznosti i samopouzdanja u regresijsku jednadžbu ne rezultira povećanjem prediktivne valjanosti, što se može objasniti na osnovi povezanosti ovih mjera s metakognicijom, ali ne i sa stvarnim brojem točnih odgovora. Ovaj nalaz na tragu je nalaza da se prosječni SS teorijski može pozicionirati između konstrukta mjerenog zadacima, metakognicije i dispozicijskih mjera ličnosti (Stankov i Lee, 2008). Curran (2012) navodi da nisu utvrđene korelacije između opće sklonosti riziku i testnih rezultata. Kroz daljnja istraživanja važno je utvrditi potencijalne faktore koji mogu utjecati na procjene SS, a povezane su s mogućim strategijama ispitanika i utjecajem upute. Rezultati ove studije naglašavaju važnost razlikovanja prosječnog SS u točne i netočne odgovore, što ostala istraživanja uglavnom zanemaruju.

Analize pokazuju da je na osnovi prosječnih SS u sve zadatke moguće razmjerno točno procijeniti stvarni broj točnih odgovora u testu, pri čemu su žene podcjenjivale, a muškarci precjenjivali stvarni uradak. To je na tragu nalaza Palliera i sur. (2002), koji navode da su procjene točnosti odgovora u klasičnim te-

stovima inteligencije s neverbalnim sadržajem uglavnom realistične, ali naglašavaju razliku prema kojoj žene pokazuju manju pristranost u kognitivnim zadacima. Međutim, i nalazi o prosječnoj nepristranosti impliciraju da se na individualnoj razini može analizirati tendencija precjenjivanja ili podcjenjivanja uratka koja može biti korisna pri individualnoj dijagnostici. Lundeborg i sur. (1994) nalaze da spolne razlike ovise o kontekstu (ovisno jesu li odgovori točni ili netočni) te o sadržaju koji se mjeri testom. Autori navode nalaze prema kojima žene iskazuju manje uvjerenje u vlastite sposobnosti u odnosu na muškarce u područjima matematike i rješavanja problema. Pored toga, žene su točnije u procjeni svojih netočnih odgovora od muškaraca koji više precjenjuju svoje netočne odgovore i zaključuju da je moguće da žene nemaju toliko nižu sigurnost koliko muškarci precjenjuju vlastitu točnost. Pored toga, nešto manje spolne razlike izražene su kod ispitanika s općenito višim stupnjem znanja/sposobnosti koji se mjere testom, ali se rezultati dijelom razlikuju ovisno o području mjerenja. Važno je naglasiti da u testu korištenom u ovom istraživanju nije bilo spolne razlike u stvarnom broju točnih odgovora. Dio autora naglašava da ispitanici često nisu svjesni vlastitih netočnih odgovora i često precjenjuju vlastito znanje (Lichtenstein i Fischhoff, 1977), stoga analiza pristranosti na individualnoj razini može upućivati na kvalitetu uvida u procese rješavanja zadataka koju ispitanici imaju. Rezultati ove studije upućuju na to da žene skromnije prospektivno procjenjuju svoj uradak u testu te da imaju nešto višu testnu anksioznost, što je konzistentno s ranijim nalazima i navodi na oprez pri analizi spolnih razlika u području procjena SS.

Složena analiza varijance, u skladu s očekivanjima, pokazala je da ispitanici s više razvijenim konstruktom ostvaruju viši SS u svoje odgovore, pri čemu su procjene SS u točne odgovore značajno više od procjena u netoč-

ne odgovore. Provjera interakcijskog efekta pokazala je da je razlika između SS u točne i netočne odgovore podjednaka kod ispitanika s različito razvijenim konstruktom. Moguće je da je ovaj nalaz dijelom povezan s činjenicom da je korišteni test bio razmjerno lagan uzorku studenata. U daljnjim istraživanjima i na drugim predmetima mjerenja bilo bi zanimljivo provjeriti pretpostavku o mogućem boljem razlikovanju točnih i netočnih odgovora između superiornih i inferiornih ispitanika. Iskustvo pokazuje da ispitanici s više razvijenim konstruktom, prilikom naknadnog uvida u testne rezultate, točnije predviđaju zadatke koje su netočno riješili u odnosu na lošije ispitanike.

Provjera učinkovitosti ponderiranja točnosti uratka u zadacima stupnjevima sigurnosti potvrdila je nalaze ranijih istraživanja o povećanju pouzdanosti unutarnje konzistencije tako dobivenih ukupnih rezultata (npr. Curran, 2012; Ebel, 1965; Hevner, 1932), pri čemu novija istraživanja rijetko navode primjenu ovakvog oblika ponderiranja (npr. Curtis i sur., 2013). Ponderiranje stupnjevima sigurnosti rezultiralo je povećanjem homogenosti kompozita, povećanjem osjetljivosti ukupnih rezultata i koeficijentata diskriminativne valjanosti zadataka u odnosu na klasično određenje ukupnog rezultata na osnovi zbroja točnih odgovora.

Ovi nalazi upućuju na potencijal opisanog postupka ponderiranja, ali je potrebno provesti dodatne analize vezane uz statističke karakteristike ponderiranih zadataka poput strukture varijance, asimetrija i promjena kriterijske valjanosti.

Prikazani rezultati upućuju na potencijale uključivanja SS u točnost odgovora na različitim razinama analize testa i vrednovanja individualnih rezultata. Jedan od dobitaka odnosi se na mogućnost dodatne informacije o ispitanikovo razini uvida u kvalitetu i ispravnost procesa u osnovi rješavanja zadatka te moguću kategorizaciju ispitanika na os-



novi kombinacije informacija o točnosti i SS osobito u situacijama nekonzistentnosti (npr. visoka točnost - niska sigurnost, niska točnost - visoka sigurnost). Ispitanici koji su visoko sigurni u netočne odgovore nekritični su spram korištenih procesa odgovaranja i u određenim poslovima mogu učiniti opasne pogreške. Dodatni problem jest da viši stupanj sigurnosti često vodi k bržem i odlučnijem odabiru odgovora odnosno rješenja (Jonsson i Allwood, 2003). Dio autora naglašava i važnost informacije o procesima koje ispitanici koriste pri rješavanju zadataka te razlikovanja netočnih odgovora koji proizlaze iz neznanja od onih koji proizlaze iz pogrešnog znanja (Curtis i sur., 2013). Koriat i Ackerman (2010) nalaze da su učenici u uvjetima u kojima im je odgovor bio važan ili kontroliran više bili usmjereni na odabir odgovora uz koje su imali visoki stupanj sigurnosti te zadatke kod kojih su imali kraće vrijeme latencije, a istraživači su suglasni da je stupanj SS viši kod ispitanika koji mogu argumentirati svoj odgovor. Pored toga jedno od praktičnih pitanja odnosi se na uputu u ovakvu vrstu zadataka, pri čemu je opća preporuka da uputa ispitanicima bude neutralna.

Ograničenja zaključaka i mogućnosti generalizacije rezultata mogu se odnositi na specifičnost testa korištenog u istraživanju i područje ispitivanja kognitivnih sposobnosti zadacima neverbalnog sadržaja. Uzorak ispitanika uključivao je studentsku populaciju, pretežito studente psihologije, za koju je korišten test bio razmjerno lagan, što je moglo dovesti do redukcije varijance i na mjerama točnosti i SS. Pored toga rezultat u testu nije utjecao na status ispitanika te bi stoga bilo korisno dobivene rezultate usporediti s rezultatima dobivenim u situaciji selekcije. Pored toga bilo bi iznimno korisno provjeriti promjene u konvergentnoj i diskriminativnoj valjanosti testnih rezultata nakon ponderiranja stupnjeva sigurnosti.

Nedostaci uključivanja procjena SS u proces primjene testa mogu se odnositi na činjenicu da u određenim uvjetima (npr. kod testova visokog uloga) ovaj dodatni zadatak procjene može biti dodatno opterećenje za pojedine ispitanike i kolidirati s primarnim procesom rješavanja zadataka ili provocirati različite strategije procjene SS. Pored toga ostaje metodološki problem tretmana zadataka uz koje ispitanici nisu procijenili SS. Ovaj zadatak oduzima jedan dio ukupnog vremena rješavanja testa te ga treba razmotriti u svim situacijama primjene testa uz izraženije vremensko ograničenje. S druge strane, još u pionirskom radu, Hevner (1932) navodi da su ovaj dodatni zadatak procjene ispitanici dobro primili, a taj zaključak navode i brojni drugi autori (npr. Stankov i Lee, 2008). Ispitanici pokazuju općenito pozitivne reakcije na mogućnost da različito vrednuju ponuđene odgovore i procjenjuju stupanj vlastite sigurnosti, čime mogu na određeni način upravljati rizikom pri odabiru odgovora u koje nisu posve sigurni. Ovaj pristup na određeni način povećava i pojavnu valjanost mjerenja. Na osnovi izloženih rezultata možemo zaključiti da uključivanje SS u proceduru primjene testa i psihometrijsku analizu nudi zanimljiv potencijal i opravdava daljnje provjere i validaciju ovako dobivenih rezultata.

---

## Literatura

- Agus, M., Peró-Cebollero, M., Guardia-Olmos, J., Portoghese, I., Mascia, M.L. i Pietronilla Penna, M., (2020). What's about the Calibration between Confidence and Accuracy? Findings in Probabilistic Problems from Italy and Spain. *EURASIA Journal of Mathematics, Science and Technology Education*. 16(2), 1-21.
- Boldt, A., Schiffer, A-M, Waszak, F. i Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation

- in Perceptual Decision Making. *Scientific Reports*, 9:4031, 1-17.
- Brewer, C. A. (2004). Near real-time assessment of student learning and understanding in biology courses. *BioScience*, 54(11), 1034-1039.
- Bujas, Z., Krizmanić, M., i Krković, A. (1975). Confidence rating as a complement of the cognitive test score. *Acta Instituti Psychologici Universitatis Zagabiensis*, 77, 31-37.
- Coombs, C. H. (1953). On the use of objective examinations. *Educational and Psychological Measurement*, 13(2), 308-310.
- Costermans, J., Lories, G. i Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1), 142-150.
- Curran, P. G. (2012). *Building a Better Test with Confidence (testing)*. (Neobjavljena doktorska disertacija). Michigan State University.
- Curtis, D. A., Lind, S. L., Boscardin, C. K. i Dellinger, M. (2013). Does student confidence on multiple-choice question assessments provide useful information? *Medical Education*, 47, 578-584.
- Ebel, R. L. (1965). Confidence weighting and test reliability. *Journal of Educational Measurement*, 2(1), 49-57.
- Fischhoff, B. i Broomell, S. B. (2020). Judgment and Decision Making. *Annual Review of Psychology*. 71:331-355.
- Gardner-Medwin, A. R. (1998). Updating with confidence: Do your students know what they don't know?. *Health Informatics*, 4, 45-46.
- Hevner, K. (1932) A Method of Correcting for Guessing in True-False Tests and Empirical Evidence in Support of IT. *The Journal of Social Psychology*, 3(3), 359-362.
- Jonsson, A. C. i Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, 34(4), 559-574.
- Kleitman, S. i Costa, D. S. (2014). The role of a novel formative assessment tool (Stats-mIQ) and individual differences in real-life academic performance. *Learning and Individual Differences*, 29, 150-161.
- Kleitman, S. i Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and individual differences*, 17(2), 161-173.
- Koriat, A. i Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science* 13(3), 441-453.
- Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological review*, 119(1), 80-113.
- Lichtenstein, S., i Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational behavior and human performance*, 20(2), 159-183.
- Lundeberg, M. A., Fox, P. W. i Punčohař, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of educational psychology*, 86(1), 114-121.
- Ljubotina, D., Pavlin-Bernardić, N. i Salkičević, S. (2015) *Metodološki aspekti računalne primjene testova - 22. ljetna psihologijska škola*. FF press.
- Mamassian, P. (2020). Confidence Forced-Choice and Other Metaperceptual Tasks. *Perception*. 49(6) 616-635.
- Mazancieux, A., Dinze, C., Souchay, C. i Moulin, C. J. A. (2020). Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. *Neuroscience of Consciousness*, 6(1), 1-11.
- McKenzie, C. R. (1997). Underweighting alternatives and overconfidence. *Organizational Behavior and Human Decision Processes*, 71(2), 141-160.
- Morony, S., Kleitman, S., Lee, Y. P. i Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, 58, 79-96.
- Nunnally, J. C. i Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. McGraw-Hill.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L. i Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of general psychology*, 129(3), 257-299.

- Schraw, G., Crippen, K. J. i Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in science education*, 36(1), 111-139.
- Soderquist, H. O. (1936). A new method of weighting scores in a true-false test. *The Journal of Educational Research*, 30(4), 290-292.
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 28(2), 121-143.
- Stankov, L. i Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, 21(6), 971-986.
- Stankov, L. i Crawford, J. D. (2006). Self-Confidence and Performance on Tests of Cognitive Abilities. *Intelligence*, 25(2), 93-109.
- Stankov, L. i Kleitman, S. (2008). *Processes on the borderline between cognitive abilities and personality: Confidence and its realism*. U G. J. Boyle, G. Matthews, i D. H. Saklofske (Ur.), The SAGE handbook of personality theory and assessment, Vol. 1. *Personality theories and models* (str. 545–559). Sage Publications, Inc.
- Stankov, L., Kleitman, S. i Jackson, S. A. (2014). Measures of the trait of confidence. U G. J. Boyle, D. H. Saklofske, i G. Matthews (Ur.), *Measures of personality and social psychological constructs* (str. 158–189). Elsevier Academic Press.
- Stankov, L. i Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100(4), 961.
- Stankov, L. i Lee, J. (2014). Overconfidence across world regions. *Journal of Cross-Cultural Psychology*, 45(5), 821-837.
- Stankov, L., Lee, J., Luo, W. i Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety?. *Learning and Individual Differences*, 22(6), 747-758.
- Stankov, L., Morony, S. i Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement?. *Educational psychology*, 34(1), 9-28.
- Tonković, M. (2007). *Psihometrijska validacija Testa S*. [Neobjavljeni diplomski rad]. Odsjek za psihologiju Filozofskog fakulteta u Zagrebu.
- Šabić, J. (2008). *Stupanj sigurnosti u točnost odgovora kao dodatna mjera pri psihometrijskoj analizi testa znanja*. [Neobjavljeni diplomski rad]. Odsjek za psihologiju Filozofskog fakulteta u Zagrebu.
- Vranić, A. i Tonković, M. (2011). Što ispitujemo testovima pamćenja? Odnos metamemorije i objektivnih mjera pamćenja. *Suvremena psihologija*, 14(2), 201-210.
- Want, J. i Kleitman, S. (2006). Imposter phenomenon and self-handicapping: Links with parenting styles and self-confidence. *Personality and individual differences*, 40(5), 961-971.
- Yen, Y. C., Ho, R. G., Chen, L. J., Chou, K. Y., & Chen, Y. L. (2010). Development and evaluation of a confidence-weighting computerized adaptive testing. *Journal of Educational Technology & Society*, 13(3), 163-176.

## Confidence level in response accuracy as additional information in psychometric analysis of cognitive abilities test

**Abstract:** Items in knowledge and ability tests are traditionally scored dichotomously based on their accuracy. This paper also considers using confidence level (CL) in response accuracy as additional information about responder's item and test scores. We analyse theoretical and conceptual advantages, as well as methodological problems of incorporating these measures in psychometric analysis of test results and evaluation/interpretation/formation of total scores. The theoretical frameworks on which integration of CL and response accuracy is based on are the concept of metacognition and the dispositional measures of self-concept and self-efficacy (Stankov & Lee, 2008).

Empirical results demonstrated here are based on data collected with the computer version of the non-verbal cognitive ability test (Test S) on 168 students. The analyses were performed on both item and test score intraindividual and interindividual level. On the item score level, the average CL is related to item difficulty, and for items with higher discriminative validity the difference in CL between correct and false answers is more pronounced. Regression analysis indicated that 69.1% of CL averaged across all responses could be explained, with metacognitive assessment of one's score and the actual number of correct responses being the best predictors. The latent structure of the matrix of CLs across items indicates the dominance of one general factor, as well as high reliability of the average CL of .96.

Differential weighting with confidence level results with higher reliability and homogeneity of the test. The results indicate the interesting potential of incorporating CL with response accuracy when analysing cognitive abilities tests.

**Keywords:** confidence level, response accuracy, item analysis

Korespondencija: dljubotina@ffzg.hr

Primljeno: 10. 6. 2021.  
Ispravljeno: 1. 9. 2021.  
Prihvaćeno: 10. 12. 2021.  
Online: 17. 12. 2021.