

**Bojan Janičić**

Odsek za psihologiju,  
Filozofski fakultet,  
Novi Sad

## Jednakost merenja sposobnosti računarski podržanom i klasičnom formom testa\*

U radu je ispitana ekvivalencija računarski podržane i klasične forme dva kognitivna testa: AL-4 (verbalni) i M-9 (neverbalni). Na uzorku od 238 ispitanika je ispitan i uticaj pola, starosti, nivoa obrazovanja, računarskog iskustva, računarske anksioznosti i anksioznosti kao stanja, na skorove ispitanika na dva modaliteta ispitivanih testova.

Analiza strukture i rangova ispitanika na dva modaliteta nije dala rezultate na osnovu kojih bi se moglo zaključiti da li je promenom modaliteta testiranja došlo do promene predmeta merenja testova.

Analizom skorova ispitanika nisu utvrđene razlike u modalitetima. Nije utvrđeno ni postojanje značajnog efekta interakcije posmatranih prediktora i modaliteta testiranja. Na osnovu toga zaključeno je da su dva modaliteta testiranja psihometrijski i iskustveno ekvivalentni.

**Ključne reči:** računarski podržano testiranje, psihometrijska ekvivalencija, iskustvena ekvivalencija, jednakost merenja, kognitivni testovi

### Uvod

Početak *računarski podržanog testiranja* (RPT) možemo vremenski locirati u 60-te godine XX v. Povećana dostupnost računara bila je dovoljna da se jave ideje za njihovu primenu u psihologiji. Prvobitno, računari su u oblasti psihološkog testiranja korišćeni za skorovanje testova papir-olovka, a nedugo za tim i za interpretaciju testova. Potom je počela njihova primena za administraciju psiholoških testova. Očigledne prednosti računara, te razvoj računarske tehnologije dali su podstrek automatizaciji postojećih i razvoju novih testova

\* Rad u okviru projekta broj 149008b "Psihološke karakteristike društva u tranziciji" kojeg finansira Ministarstvo nauke i zaštite životne sredine Republike Srbije.  
e-mail adresa autora: janicic@uns.ac.rs

(Bartram & Bayliss, 1984). 70-ih godina XX v. pojavile su se računarski podržane verzije mnogih klasičnih testova. Približno deceniju kasnije počinju da se javljaju i testovi konstruisani isključivo za računarsku administraciju.

Kod nas su retki primeri programa za računarsku administraciju testova, a ako i postoje o njima se malo zna. Kostić i Anđelković (1996) su konstruisali programski paket psiholoških testova „Psiho“ koji sadrži programe za računarski podržanu administraciju Plutchikovog testa Profil indeks emocija, te za skrovanje i interpretaciju testova MMPI-201 i EPQ. Milin je (1999) konstruisao program „Kolokvijum“ za računarski podržano testiranje znanja iz predmeta Statistika na Odseku za psihologiju Filozofskog fakulteta u Novom Sadu, gde se ovaj program neko vreme uspešno i primenjivao. Autor ovog istraživanja je konstruisao nekoliko računarski podržanih testova koji se upotrebljavaju u nastavi predmeta Psihometrija 1 i 2 na Odseku za psihologiju FF u Novom Sadu, ali nijedan od njih nije u komercijalnoj upotrebi. To su programi za testove: MMPI - različite forme (Janičić, 2004), PIE, Harrower-Erickson, TMTns – nova verzije neuropsihološkog Trail Making testa (Todorovski, Salević–Obradović i Janičić, 2006; Todorovski, Salević-Obradović, Janičić i Kopitović, 2006a).

Pored dostupnosti računara, na izuzetan napredak računarski podržanog testiranja (RPT) uticale su i brojne prednosti koje ono ima nad klasičnim testovima (Bugbee, 1996; British Psychological Society<sup>1</sup>, 1999; Wijndaele et al., 2007; Bartram & Bayliss, 1984; Fajgelj, 2005; Haladyna & Roid, 1983). Naravno, RPT nije bez mana ali su one zanemarljive u odnosu na prednosti (Bugbee & Bernt, 1990; Bartram & Bayliss, 1984).

Zbog pomenutih prednosti RPT je dobilo na popularnosti. Još 1987, u SAD je čak 267.000 ispitanika bilo ispitano pomoću RPT (Bugbee & Bernt, 1990).

Međutim, konstrukcija računarski podržane verzije nekog psihološkog mernog instrumenta ne znači samo izradu softvera. Naime, mnogi psihometrijski modeli pred testove postavljaju zahtev za jednodimenzionalnošću – test mora da meri jednu stvar. S obzirom da u svakom ponašanju osobe, pa i u odgovaranju na stavke testa, učestvuje celokupna ličnost, zahtev da test meri samo jednu stvar nije ostvariv. Psihološko testiranje, kao proces merenja, podrazumeva interakciju sa ispitanicima (Fajgelj, 2005). Uvođenje računara u taj proces dodaje nove elemente i menja prirodu te interakcije. Ono može promeniti njegov predmet i samim tim narušiti *validnost testa*. Npr., ako na nekom računarski podržanom, kognitivnom testu, ispitanici sa istim nivoom sposobnosti dobiju različite skorove samo zato što imaju različito iskustvo u radu sa računarima,

.....

<sup>1</sup> U daljem tekstu - BPS

više ne možemo reći da test meri samo sposobnost za čije merenje je namenjen. Zbog toga, konstrukcija računarski podržane forme postojećeg testa, ne može se završiti konstrukcijom odgovarajućeg softvera. Njegova izrada prvi je korak nakon kojeg je nužno ispitati da li su dve forme testa (klasična i računarska) ekvivalentne (Schulenberg & Yutrzenka, 2001).

### **Ekvivalencija računarski podržane i klasične forme testa**

APA (1986) je u *Smernicama za računarski zasnovano testiranje i interpretaciju*, upozorila na potrebu utvrđivanja ekvivalencije između klasične i računarske forme testa. Ekvivalenciju definiše ovako: (a) rangovi skorova individualnih ispitanika, testiranih na alternativne načine, približni su ili jednaki, i (b) aritmetičke sredine, raspršenja i oblici distribucije skorova približno su jednaki ili se mogu učiniti približno jednakim, reskaliranjem rezultata dobijenih računarskim testiranjem

I BPS, u *Smernicama za razvoj i upotrebu računarski podržanog ocenjivanja* iz 1999, kaže se da je prilikom prevođenja nekog postojećeg testa u računarski podržan oblik potrebno jasno dokumentovati da su ocene dobijene pomoću dve forme ocenjivanja *ekvivalentne*. Isto zahteva i Međunarodna komisija za testove u *Međunarodnim smernicama za računarski podržano i Internet testiranje* (International Test Commission – ITC, 2005).

Ovi zahtevi u suštini predstavljaju potrebu da se dokaže *jednakost* (engl. *equity*) merenja različitim modalitetima. Drugim rečima, mere ispitanika ne bi trebalo da zavise od testa (ili modaliteta testa) kojim je merenje izvršeno. Na taj način se pitanje ekvivalencije svodi na objektivnost merenja (Fajgelj, 2005).

Lord (1977) *jednakost* testova definiše tako da je ispitaniku na bilo kojem nivou osobine svejedno koji test rešava (u ovom slučaju to se odnosi na modalitet administracije testa). Ako želimo da poredimo skorove ispitanika na dva testa koji mere istu stvar, ili skorove ispitanika u dve vremenske tačke koristi se jednačenje. Jednačenje je procedura kojom se skorovi jednog testa transformišu i izražavaju u terminima skorova drugog testa i na taj način dobijaju ekvivalentni skorovi.

Iako su stavke u različitim modalitetima testova iste (i trebalo bi da imaju iste metrijske karakteristike), može se desiti da ispitanici ne postižu iste skorove. Razlog za to bi mogla biti promena predmeta merenja. Na primer, kognitivne operacije potrebne za rešavanje zadataka nekog računarski administriranog testa sposobnosti ne razlikuju se od onih za rešavanje klasične forme istog testa. Međutim, spretnost u unosu odgovora u računarskoj formi može uticati na

skorove ispitanika<sup>2</sup>. Na taj način kao drugi predmet merenja testa može se pojaviti spretnost u radu sa računarom.

Kako se mogu jednačiti samo testovi koji imaju iste predmete merenja i približno iste pouzdanosti i težine (Lord, 1977), rešenje bi trebalo tražiti u posebnim normama za svaki modalitet. Ako modalitet testiranja jednako utiče na skorove svih ispitanika, norme za modalitete mogu biti opšte. Međutim, ako grupe ispitanika jednake po merenoj osobini, dobijaju različite skorove na različitim modalitetima testova onda je najverovatnije u pitanju diferencijalno funkcionisanje testova (engl. *differential test functioning – DTF*) ili bolje reći modaliteta. U tom slučaju, rešenje bi trebalo tražiti u posebnim grupnim normama za modalitete. Kada je u pitanju računarski podržano testiranje, grupe o kojima pričamo mogu biti formirane na osnovu individualnih razlika koje će biti razmatrane kasnije.

U skladu sa preovlađujućom terminologijom ispitivanja *jednakosti merenja pomoću dva modaliteta testova* zvaćemo ispitivanjem *ekvivalencije modaliteta testova*.

Honaker govori o dva vida ekvivalencije (Honaker, 1988):

- 1) *psihometrijska ekvivalencija* i
- 2) *iskustvena (doživljajna) ekvivalencija*.

Pod *psihometrijskom ekvivalencijom* podrazumeva se da između grupnih aritmetičkih sredina i distribucija skorova ispitanika na dve forme testa ne postoje statistički značajne razlike. Modaliteti administracije testova ne smeju uticati na grupne skorove. Ni varijanse skorova na dve forme testa ne smeju se značajno razlikovati, jer bi u suprotnom to značilo da se opseg i standardna devijacija dvaju formi testova ne mogu porediti. Takođe, rang korelacije skorova dve forme nam govore o njihovoj pouzdanosti i sličnosti (Williams & McCord, 2006).

*Iskustvena ekvivalencija* znači da ne postoje značajne razlike u stavu prema testu i doživljavanju testa od strane ispitanika. Smatra se važnom za dokazivanje konstruktivne validnosti. McDonald (2002) kaže da rangovi ispitanika na dve forme testa, pouzdanost formi i faktorska struktura daju odgovor na ovo pitanje. Razlike u ovim pokazateljima ukazuju da sve forme mere različite konstrukte ili da je jedna od formi kontaminirana dodatnim, neželjenim predmetom merenja.

.....  
<sup>2</sup> Posebno kada su u pitanju testovi kod kojih je vreme izrade značajniji izvor varijanse.

### **Istraživanja ekvivalencije računarski podržanih i klasičnih modaliteta testova**

Istraživanja ekvivalencije računarski podržanih i klasičnih formi testova su brojna, a mi ćemo ovde prikazati samo neka.

Prema Williamsu i McCordu (2006) većina dosadašnjih ispitivanja računarski podržanih formi verbalnih testova, pokazuju da oni poseduju *psihometrijsku ekvivalenciju*. Za neverbalne dobijeni su nejasni rezultati. Oni dalje kažu da nijedno istraživanje ekvivalencije neverbalnih testova nije uzelo u obzir iskustvenu ekvivalenciju, odnosno efekte individualne percepcije računarski podržane forme testa.

Mead i Drasgow (1993) obavili su meta-analizu istraživanja ekvivalencije kognitivnih testova. Ispitano je 159 koeficijenata korelacije između dva modaliteta testova, od kojih su se 123 odnosila na *testove snage* i 36 na *testove brzine*. Meta-analizom, dobijena je prosečna korelacija između modaliteta za svih 159 testova koja je iznosila ,91 (,95 za testove snage, a ,72 za testove brzine). Utvrdili su značajan efekat ubrzanja testa i veličine uzorka i pripadnosti baterije testova. Modalitet testiranja (računarski ili klasični) nije imao značajni efekat. Takođe, zaključuju da su računarski podržane varijante testova neznatno teže, sa malim variranjem od studije do studije. Zaključuju da se testovi snage mogu, a testovi brzine ne mogu smatrati ekvivalentnim. Razlog vide u različitim motornim veštinama koje su potrebne za unošenje odgovora. Takođe, zaključuju da se računarski podržane varijante testova moraju posebno validirati, odnosno ne sme se automatski pretpostaviti da mere istu stvar kao njihovi klasični parnjaci.

Istraživanja koja se bave *iskustvenom ekvivalencijom* su retka i najčešće uključuju ispitivanje *računarske anksioznosti*<sup>3</sup> (engl. *computer anxiety*).

Kubinger, Formann i Farkas (1991) su se bavili uticajem računarske forme prezentacije na rezultate ispitanika na Ravenovim standardnim progresivnim matricama (RSPM). Prema njima, računarski podržana forma davala je niže skorove ispitanika od standardne za IQ poena. Niže rezultate ispitanika objašnjavaju uticajem *računarske anksioznosti*. Novije istraživanje (Arce-Ferrer & Martínez Guzmán, 2009) RSPM, sprovedeno na uzorku od 300 srednjoškolaca dalo je slične rezultate.

Ford, Vitelli i Stuckless (1996.) kažu da je ekvivalenciju dva modaliteta testa potrebno utvrditi posebno za svaku subpopulaciju (kliničku, nekliničku...). King i Miles (1995) te Wright, Aquilino i Supple (1998) kao moderatore efekta modaliteta administracije koje treba uzeti u obzir prilikom utvrđivanja ekvivalencije navode: pol, starost, stepen obrazovanja i iskustvo u radu sa računarima.

<sup>3</sup> O računarskoj anksioznosti će biti više reči kasnije.

### ***Mogući uzroci nejednakosti različitih modaliteta testova***

Faktore koji mogu uticati na nejednakost merenja različitim modalitetima testa možemo podeliti na one koji se tiču fizičkih osobina testa i na individualne razlike među ispitanicima.

Što se tiče *fizičkih aspekata testa* – razlike su očigledne. U klasičnoj formi testa obično imamo list papira sa stavkama, dok u računarskoj formi (ne nužno) pitanja bivaju prezentirana na ekranu jedno po jedno. Neki autori smatraju grupno prezentiranje stavki u testovima sposobnosti može olakšavati rešavanje testa, a drugi da takav način prezentiranja dovodi do ishitrenog odgovaranja i nepažnje (Dimock & Cormier, 1991).

Kada su u pitanju faktori koji mogu uticati na ekvivalenciju modaliteta, McDonald (2002) stavlja akcenat na *individualne razlike* a posebno ističe: *računarsko iskustvo, računarsku anksioznost i stavove prema računarima*. Colley, Gale i Harris (1994.) te Brosnan (1998) pominju i ulogu *pola*.

Brojna istraživanja (Brosnan 1998; Rosen, Scars & Weil, 1987; Teasdale & Lupart, 2001; Cooper, 2006; Durndell, Haag & Laithwaite, 2000) u vezi sa polom (nekad i polnim identitetom) ukazuju da muškarci poseduju više *računarskog samopouzdanja*, više vole i više rade sa računarima, te imaju niže nivoe računarske anksioznosti.

*Računarsko samopouzdanje* koje Brosnan (1998), te Henderson, Deane i Ward (1995) definišu kao *samoprocenu pojedinca o sposobnosti korišćenja računara*. Oni nalaze da osobe sa višim računarskim samopouzdanjem imaju niži nivo računarske anksioznosti.

*Računarsko iskustvo* se operacionalizuje: posedovanje ili dostupnost računara, učestalost upotrebe, stavovi prema računarima, pohađanje kurseva o ili na računarima...

Iako se može činiti kao najbitniji faktor koji može uticati na uspeh ispitanika na računarskim kognitivnim testovima, samo u 2 od 11 studija koje navodi McDonald (2002), pronađen je njegov značajan efekat (pozitivna povezanost). Međutim, autori tih studija efekat procenjuju kao mali i kažu da bi se uz minimalno računarsko iskustvo mogao korigovati.

Kay (1993) pozitivne *stavove prema računarima* smatra preduslovom za razvoj računarskih veština. Stavovi utiču na to koliko će osoba biti spremna da radi za računarom, a računarsko iskustvo može uticati na formiranje i menjanje stavova prema računarima.

Najčešće pominjani moderator efekta modaliteta administracije je *računarska*

*anksioznost*. Definiše se kao: stanje povišene anksioznosti prilikom rada sa računarima koje doživljavaju osobe sa malim računarskim iskustvom (Schulenberg i Yutrzenka, 2001) ili kao: strah od računara (prilikom njegove upotrebe) i kao strah od mogućnosti da će osoba koristiti računar (Chua, Chen, & Wong, 1999).

Smith i Capputi (1997) mehanizam delovanja računarske anksioznosti objašnjavaju *modelom kognitivne interferencije računarske anksioznosti*. Ovaj model na računarsku anksioznost gleda kao na crtu ličnosti i zasniva se na *samoregulacionom modelu anksioznosti*<sup>4</sup> Carvera i Scheiera (1986) i *teoriji efikasnosti procesiranja*<sup>5</sup> Eysencka i Calvoa (1992). Po ovom modelu motivaciona funkcija anksioznosti vodi ka usmeravanju dodatnih resursa na zadatak ili upotrebi strategija koje omogućavaju povećanje kapaciteta radne memorije. Na ovaj način biva oštećena efikasnost procesiranja (brzina), ali ne i rezultati (tačnost). Ovo važi za testove snage. Kod testova brzine, s obzirom na vremensko ograničenje, model predviđa da računarski anksiozne osobe neće imati dovoljno vremena da upotrebe strategije prevladavanja, pa će u ovom slučaju biti oštećena i brzina i tačnost.

Schulenberg i Yutrzenka (2001) smatraju da je računarska anksioznost povezana sa polom i starošću osobe, a da je činilac koji ih povezuje – računarsko iskustvo (žene i stariji su manje vični radu sa računarom). U skladu sa time je i mišljenje Whitakera (2007) koji kaže da veći broj studija ukazuje da žene pokazuju viši nivo računarske anksioznosti.

Chua i saradnici (1999) su izvršili meta-analizu novijih istraživanja računarske anksioznosti i njenih korelata (pola, starosti i računarskog iskustva). Ova analiza obuhvatila je 28 studija objavljenih u periodu 1990–1996. Rezultati nisu jednoznačni. U zavisnosti od primenjenog instrumenta za merenje računarske anksioznosti, razlike među polovima nisu bile značajne ili su bile niske. Najkonzistentniji nalaz ove meta-analize je da je računarska anksioznost u negativnoj korelaciji sa računarskim iskustvom, ali jačina ove veze varirala je od studije do studije pa se ne može doneti definitivan zaključak.

### ***Istraživanja uticaja računarske anksioznosti na testiranje različitim modalitetima testova***

Brojna su istraživanja uticaja računarske anksioznosti na postignuće na računarski podržanim kognitivnim testovima. Neka nalaze negativnu povezanost

<sup>4</sup> engl. *self-regulation model of anxiety*

<sup>5</sup> engl. *processing efficiency theory - PET*

računarske anksioznosti i postignuća (Kanfer i Heggstad, 1997, Bloom i Hautaluoma, 1990), neka nalaze takvu povezanost samo sa određenim vrstama testova (Shermis i Lombard, 1998), dok neka nalaze negativnu povezanost računarske anksioznosti sa postignućem na kognitivnim testovima bez obzira na modalitet (Powers, 1999, Tseng, Tiplady, Macleod, & Wright, 1998). Zanimljiv je rad Vogelove (1994, prema: McDonald, 2002) koja je pronašla odnos u obliku obrnutog slova „U“ između računarske anksioznosti i postignuća na računarski podržanoj formi *Graduate Record Examination – GRE*. Na ovom testu najbolje postignuće imali su ispitanici sa umerenim nivoom računarske anksioznosti.

Cilj ovog rada je ispitati mogu li se različite forme administracije kognitivnih psiholoških testova smatrati ekvivalentnim. To će biti ispitano na primeru dva kognitivna testa – verbalnom i neverbalnom. Biće ispitano i efekat sledećih prediktora: *računarske anksioznosti, anksioznosti kao stanja, pola, starosti, nivo obrazovanja te računarskog iskustva* na ponašanje ispitanika u različitim modalitetima administracije testova.

## Metod

### *Uzorak i postupak*

Uzorak je bio prigodan,  $N=238$ , Mahom, su ga činili studenti druge godine psihologije Odseka za psihologiju, Filozofskog fakulteta u Novom Sadu (3 generacije). 88,66% ispitanika je ženskog pola. Prosečna starost ispitanika je 20,77 godina ( $s=7,19$ ,  $Me=20$ ). Najmlađi ispitanik imao je 18, a najstariji 67 godina. Prosek godina školovanja ispitanika je 13,83 ( $s=1,503$ ,  $Me=14$ ). 61,4% ispitanika je iz grada, 96,2% poseduje računar, a svakodnevno ili gotovo svakodnevno ga upotrebljava njih oko 80%. 89,41% ispitanika računar rado koristi i to najčešće kod kuće i na poslu/fakultetu. Većina svoje računarsko iskustvo procenjuje kao osrednje (55,04%).

Svi ispitanici testirani su istim testovima. Redosled zadavanja bio je sledeći: 1. supskala stanja anksioznosti iz testa POMS, 2. AL-4 verbalni test sposobnosti, 3. numerički test sposobnosti M-9 i 4. Skala procene računarske anksioznosti - CARS. Ispitanici su testirani u dva navrata, u razmaku od jedne do dve nedelje. Posle drugog testiranja ispitanicima je dat i upitnik koji se tiče računarskog iskustva, interesovanja, upotrebe računara, te utiska o dva primenjena modaliteta testiranja.

AL-4 i M-9 su u dva navrata zadavani u različitim modalitetima: u računarski podržanoj i klasičnoj formi. Ostali instrumenti bili su zadavani isključivo u klasičnoj formi.



Ispitanici su slučajnim izborom podeljeni u četiri grupe koje su se razlikovale po modalitetu u kom su radile testove AL-4 i M-9, te po redosledu modaliteta. I i II grupa su radile testove u oba modaliteta s tim da se razlikovao redosled modaliteta, dok kod III i IV grupe modalitet nije menjan. I grupa je testove radila prvo na računaru, a II na papiru. III grupa je testove oba puta radila na računaru, a IV na papiru. Ovakav nacrt odabran je da bi se kontrolisalo efekat redosleda, vežbe, pamćenja i zamora. Testiranje je bilo anonimno.

Ispitivanje je u većini slučajeva obavljeno grupno u računarskoj učionici Filozofskog fakulteta u Novom Sadu u kojoj se nalazilo 20 računara istih performansi (Intel® Celeron® procesori na 1,60 GHz, sa 1Gb RAM, TFT 19"). Testiranje je obavljano u popodnevnim i večernjim časovima.

### **Instrumenti**

U istraživanju su korišćeni: supskala stanja anksioznosti iz testa Profile of Mood States (POMS – forma 1, prevod B. Popova, 2007), supstest AL-4 iz baterije KOG-3 (Wolf, Momirović i Džamonja, 1992), supstest M-9 iz M-serije Bujasa i Petza (Bujas, 1967), Skala procene računarske anksioznosti CARS (Heinssen, Glass, & Knight, 1987) i ranije pomenuti upitnik.

Konstruisane su računarske verzije testova AL-4 i M-9. U oba modaliteta ovih testova davano je isto uputstvo, a u računarski podržanoj formi ono dopunjeno uputstvom za odgovaranje putem miša i tastature. I u računarskom modalitetu ovih testova ispitanici su mogli na ekranima da vide sva pitanja odjednom kao u papirnoj formi.

Skala CARS je za potrebe ovog istraživanja prevedena sa engleskog jezika.

### **Statistička obrada**

Razlike u skorovima ispitanika na dva modaliteta analizirane su procedurom koja se može svrstati u familiju linearnih mešovityh modela. Korišćena je procedura reuzorkovanja (*resampling* i *bootstrapping*) kako bismo kompenzovali nedostatke uzorka. Analize su obavljene u statističkim paketima SPSS for Windows 16.0 i R 2.9.2 za Windows. Takođe, pre svake od ovih analiza proveravano je da li klasična i računarska forma testova mere iste konstrukte, odnosno, da li je prevodjenjem testa u računarski podržanu formu došlo do promene predmeta merenja. To je učinjeno poređenjem faktorskih rešenja dvaju formi.

Za sve instrumente kao pokazatelj pouzdanosti, za svako testiranje u svakom od modaliteta izračunat je intraklasni koeficijent korelacije za relativno odlučivanje i prosečnu meru. On je jednak Cronbachovoj alfi i ovde je tretiran tako.

Prikazane su i test-retest korelacije. S obzirom da je razmak između testa i retesta bio relativno mali, upitno je koliko ove koeficijente korelacije možemo tretirati kao koeficijente pouzdanosti. Osim toga korelacije u grupama koje su menjale modalitet pre bi se mogle posmatrati kao pouzdanost alternativnih formi.

## Rezultati

### *Metrijske karakteristike korišćenih instrumenata*

Pouzdanost skala CARS i POMS-anx je visoka u oba testiranja. Što se tiče kognitivnih testova, ona je zadovoljavajuća za AL-4 i dosta bolja u računarskoj formi. Što se tiče testa M-9 pouzdanost je generalno niska i slična za oba testiranja u papirnoj formi i 2. testiranju računarskim modalitetom. Odskače prvo testiranje na računaru koje ima najveću pouzdanost.

Ako uporedimo modalitete testiranja, mogli bismo reći da računarski podržana forma ima tendenciju da daje pouzdanije skorove.

**Tabela 1.** Pouzdanost upotrebljenih instrumenata

	m	PAPIR		RAČUNAR	
		1. testiranje	2. testiranje	1. testiranje	2. testiranje
POMS-ANX	10	,89	<b>.90</b>		
CARS	19	<b>.85</b>	,84		
AL-4	40	,76	,70	<b>.91</b>	,88
M-9	15	,56	,59	<b>.70</b>	,57

Napomena: za testove POMS i CARS N=238, za AL-4 i M-9 za 1. testiranje na papiru i 2. na računaru N=121 i za 2. testiranje na papiru, te prvo na računaru N=117.

**Tabela 2.** Koeficijenti korelacije između testa i retesta

GRUPA	svi	I g. (r-p)	II g. (p-r)	III g. (r-r)	IV g. (p-p)
POMS – ANX	,52	,50	<b>.67</b>	,43	,55
CARS	,87	,88	,81	<b>.89</b>	<b>.89</b>
AL-4	,61	,61	,58	<b>.74</b>	,15
M-9	,67	,74	,55	,67	<b>.76</b>

Napomena: za sve korelacije  $p < ,05$  osim za AL-4 za IV grupu (papir-papir)  $p = ,23$  (ukupni N=238, 1. grupa N=59, 2. grupa N=63, 3. grupa N=58, 4. grupa N=58).

Iz tabele 2 vidimo da su korelacije između testa i retesta najveće (podvučene) u grupama koje nisu menjale modalitet testiranja (IV i III), pa bismo mogli pretpostaviti da modalitet testiranja utiče na skorove. Test-retest korelaciju za test AL-4 u IV grupi ( $r=,15$ ,  $p>,05$ ) možemo smatrati artefaktom i objasniti je smanjenim varijabilitetom skorova ispitanika IV grupe na ovom testu (tabela 3)

**Tabela 3.** Aritmetičke sredine i standardne devijacije ukupnih skorova na testu AL-4

	1. testiranje		2. testiranje	
	AS	SD	AS	SD
I grupa	37,15	4,08	38,07	2,48
II grupa	37,22	3,16	37,84	3,47
III grupa	36,45	5,82	37,62	4,01
IV grupa	37,98	2,41	38,48	1,81

Međutim, test-retest korelacije grupa koje nisu menjale modalitet, ne razlikuju se drastično od onih u grupama koje jesu. Kasnije će biti utvrđeno mogu li se te razlike pripisati modalitetu testiranja ili su rezultat drugih činilaca. U tom kontekstu skrenućemo pažnju na sledeću činjenicu. Testovi POMS i CARS rađeni su samo u klasičnoj formi, a u tabeli 2 su i oni prikazani po grupama. Uočavamo da CARS ima najveću test-retest korelaciju u III i IV grupi, a kada uporedimo razlike među koeficijentima korelacije po grupama (kako za CARS tako i za POMS), one su slične kao one na instrumentima kod kojih se menjao modalitet.

### Struktura skale CARS

Ispitana je struktura skale CARS (koja ovde zbog obima rada neće biti prikazana). S obzirom na sličnost strukture CARS na testu i retestu (u oba slučaja izolovan jedan značajan faktor), te na visoku korelaciju skorova sa prvog i drugog testiranja (tabela 2) možemo reći da ovaj instrument ima jedan glavni, relativno stabilan, predmet merjenja - računarsku anksioznost.

Test se na ovom uzorku pokazao „težim“ ( $\bar{X}_{1.test}=34,39$ ,  $\bar{X}_{2.test}=33,00$ , min. 0 – max. 76). Reprezentativnost testa je dobra i po Kaiserovoj klasifikaciji i spada u nivo „zaslužnih“ ( $KMO_{1.test}=,845$  i  $KMO_{2.test}=,845$ ).

### Analiza grupa po prediktorima

Jednosmernom analizom varijanse (uz korišćenje bootstrapping procedure na 1.000 slučajnih uzoraka) ustanovljeno je da se grupe ispitanika ne razlikuju po prediktorima koji će biti uvršteni u analize (tabela 4).

**Tabela 4.** ANOVA – *F* testovi

Prediktor	F	p
starost	,48	,70
broj godina školovanja	1,01	,39
samoprocena računarskog iskustva	1,23	,30
POMS – anx	1,81	,15
POMS – anx (retest)	,12	,95
CARS	1,86	,14
CARS (retest)	1,57	,20

Svoje računarsko znanje na skali od 1 do 5 ispitanici su procenili ocenama od 3,22 do 3,35, pa možemo reći da ga procenjuju kao nešto više od prosečnog.

Anksioznost merena supskalom testa POMS se u prvom testiranju kretala između 5,71 i 7,82, a na drugom testiranju između 5,00 i 5,58. S obzirom da se mogući skorovi na ovom testu kreću od 0 do 36, možemo reći da su ispitanici u sve 4 grupe ispoljili nizak nivo anksioznosti.

Na skali računarske anksioznosti CARS, prosečni skorovi ispitanika su se u prvom merenju kretali od 37,90 do 40,17, a u drugom između 36,77 i 39,82. Mogući skorovi na ovom testu se kreću između 19 i 95, pa možemo reći da su ispitanici iz ovog uzorka ispoljili niži nivo računarske anksioznosti.

Polna struktura grupa ne razlikuje se značajno ( $\chi^2[5, N=238]=399$ , MC  $p=,137$ )

### ***Razlike u distribucijama skorova na ispitivanim testovima***

Korišćen je Wilcoxonov test rangova i bootstrapping procedura sa 10.000 uzoraka.

**Tabela 5.** Poređenje distribucija skorova na testu i retestu u zavisnosti od modaliteta testiranja

	1. računar/papir-olovka		2. računar-računar		3. papir-olovka/papir-olovka	
	Z	p (2-s)	Z	p (2-s)	Z	p (2-s)
AL4	-0,84	,40	-1,84	,07	-1,82	,07
M-9 <sub>1,2,3</sub>	<b>-2,29</b>	<b>,02</b>	<b>-4,68</b>	<b>,00</b>	<b>-5,19</b>	<b>,00</b>
POMS <sub>3</sub>	-1,54	,12	-1,08	,28	<b>-3,11</b>	<b>,00</b>
CARS <sub>2,3</sub>	-0,34	,73	<b>-2,35</b>	<b>,02</b>	<b>-2,21</b>	<b>,03</b>

Legenda: Indeksima su označene pozicije na kojima postoje značajne razlike u distribucijama. Indeksi se odnose na redne brojeve iz prvog reda tabele.

U zavisnosti od modaliteta testiranja postoje značajne razlike u distribucijama skorova na testu M-9, ali su one čak češće u grupama kod kojih nije bilo promena modaliteta (najčešće u grupi koja je u oba navrata radila klasičnu formu testova). Ako pogledamo rezultate za CARS i POMS koje su ispitanici radili samo u klasičnoj formi, vidimo da u grupi ispitanika koja je na ostalim testovima menjala modalitet testiranja (spojene I i II grupa) nema razlike u distribucijama. U grupi koja je ostale testove oba puta radila na računarima, razlikuje se distribucija samo za CARS. Na kraju, u grupi koja je ostale testove radila samo na papiru, razlikuju se distribucije i na testu POMS i na testu CARS. To može ukazivati da su ispitanici grupa koje su menjale modalitet testiranja, generalno doslednije odgovarali.

### ***Testovi sposobnosti***

U analizama oba testa sposobnosti zavisna varijabla je skor na testu (AL-4 i M-9 –respektivno), a inicijalni skup fiksnih prediktora je isti i njega čine: modalitet testiranja, navrat, pol (faktori), starost, broj godina školovanja, skorovi na POMS – anksioznost i CARS (kovarijable). U svakoj analizi kontrolisan je uticaj slučajnog efekta subjekta.

S obzirom da za svaku od skala postoje dva merenja, nacrt je u suštini ANCOVA sa ponovljenim merenjima.

#### ***AL-4***

AL-4 se na ovom uzorku pokazao kao lagan. U klasičnoj formi prosečan skor ispitanika je 37,58 za prvo testiranje i 38,27 za retest. U računarskoj formi prosečan skor za prvo testiranje je 36,80, a za retest 37,73. Maksimalan mogući skor na ovom testu je 40.

Stavke testa AL-4 su ovom uzorku bile izuzetno lake u oba modaliteta. Za veliku većinu stavki indeks težine je veći od ,90 (u oba modaliteta). Težim su se pokazale stavke pred kraj testa što nam govori da je test u izvesnoj meri ubrzan.

#### ***Poređenje strukture dva modaliteta AL-4***

Za poređenje strukture dva modaliteta ovog testa korišćena je analiza glavnih komponenti sa promax rotacijom. Iz analize su izbačene stavke koje bar u jednom modalitetu imaju nultu varijansu: 1, 2, 5, 12, 18. Sve takve stavke su sa početka testa, što govori u prilog tome da ovaj test ima komponente testa brzine.

Analiza nije dala rezultate iz kojih bi se mogao izvući definitivan zaključak.

Iako je u oba modaliteta dobijeno isto rešenje (4 komponente), izdvojene komponente se razlikuju, ali bi se u oba slučaja mogle protumačiti kao *komponente brzine*. To nam ništa ne govori o predmetu merenja, niti da li je modalitet testiranja doveo do njegove promene.

### **Poređenje skorova ispitanika na dva modaliteta**

Nakon izbacivanja prediktora koji se nisu pokazali značajni, u modelu su zadržani *navrat*, *pol* i *starost*.

Prema pokazateljima fita, redukovani model bolje opisuje podatke.

**Tabela 6.** Pokazatelji fitovanja modela

	AIC	BIC	log Likelihood
inicijalni model	2460	2514	-1217
redukovani model	2035	2060	-1011

Standardna devijacija odsečka kao rezultat slučajnog efekta subjekta iznosi 1,86, a reziduala 1,48.

**Tabela 7.** MCMC procena parametara i njihove značajnosti

	$\beta$	95% i.p.		p
		donji	gornji	
odsečak	41,73	39,68	42,07	,0001
NAVRAT	,51	,02	,86	,0356
STAROST	-,43	-,46	-,23	,0001
STAROST: god. škol.	,02	,01	,02	,0001

Procenjeni  $R^2 = ,88$  za celokupan model, a samo za fiksne efekte  $R^2_F = ,42$ .

Modalitet testiranja nema značajan efekat na skorove ispitanika, pa skorove dobijene na papirnoj ili računarskoj formi ovog testa možemo smatrati jednakim.

Kao značajni pokazali su se glavni efekat starosti i efekat interakcije godina školovanja i starosti. S obzirom da nisu u interakciji sa modalitetom, nisu od interesa za ovo istraživanje pa ih nećemo interpretirati.

Izračunat je i Gulliksenov koeficijent ubrzanja testa (Fajgelj, 2005) za računarsku i klasičnu formu, koji se zasniva na odnosu varijansi broja pitanja na koje ispitanik nije dao odgovor u odnosu na ukupnu varijansu broja pogrešnih odgovora i pitanja na koje ispitanik nije dao odgovor. Za računarski podržanu formu iznosi ,79, a za klasičnu ,60. Na osnovu toga možemo zaključiti da je ta forma ubrzanija.

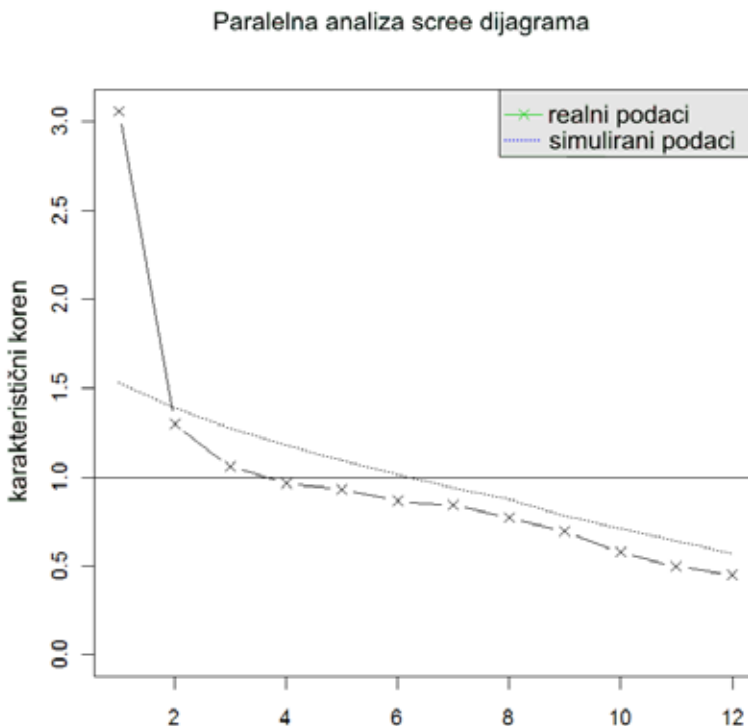
**M-9**

I M-9 se pokazao relativno lagan. Prosečan skor u klasičnom modalitetu na testu bio je 10,85, a retestu 11,85. U računarski podržanoj formi prosečan skor na testu bio je 10,15, a na retestu 11,74. Maksimalan skor na ovom testu je 15 bodova.

**Poređenje strukture dva modaliteta M-9**

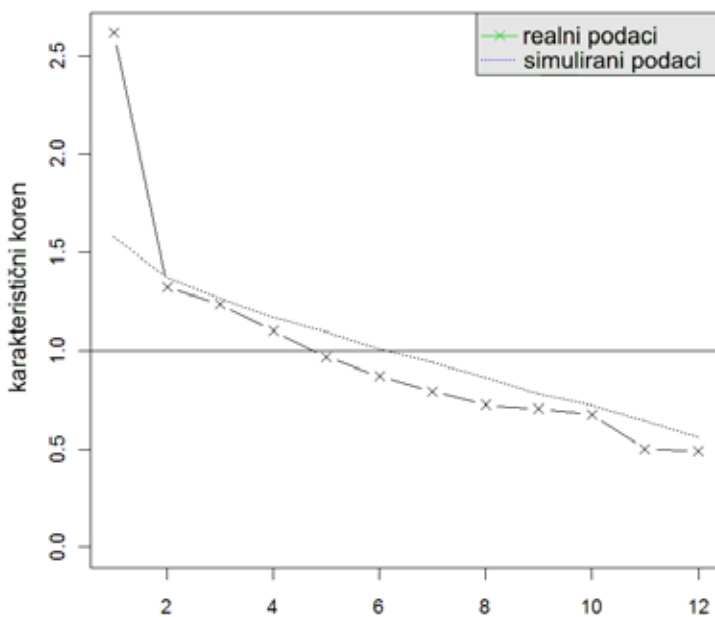
Struktura dva modaliteta testa M-9 ispitivana je komponentnom analizom sa promax rotacijom. U analizu nisu uključene stavke pod rednim brojevima 1-3 koje imaju nultu varijansu u bar jednom od modaliteta.

Na osnovu kriterijuma paralelne analize, u oba modaliteta izdvaja se po jedna značajna glavna komponenta. Ovakvo rešenje sugerise i Velicerov MAP kriterijum koji za oba modaliteta, minimum od ,14 dostiže sa jednom glavnom komponentom.



**Grafikon 1.** Paralelna analiza scree dijagrama za računarski podržani modalitet M-9

## Paralelna analiza scree dijagrama



**Grafikon 2.** Paralelna analiza scree dijagrama za klasični modalitet M-9

Koeficijent korelacije između dobijenih komponenti iznosi  $r=,66$ , a koeficijent kongruencije dve komponente je ,99. Možemo zaključiti da promenom modaliteta nije došlo do promene predmeta merenja testa M-9.

### Poređenje skorova ispitanika na dva modaliteta

Kada su eliminisani prediktori koji nisu značajni, u modelu su ostali: modalitet testiranja, navrat, pol i starost. Međutim, nakon bootstrapping procedure ispostavilo se da modalitet testiranja nije značajan pa je i on izbačen iz modela.

**Tabela 8.** Pokazatelji fitovanja modela

	AIC	BIC	log Likelihood
inicijalni model	1964	2014	-970,2
redukovani model (uključen modalitet)	1851	1880	-918,4
finalni model (bez modaliteta)	1843	1868	-915,7

Ovako redukovani model još bolje opisuje podatke.



Standardna devijacija odsečka kao rezultat slučajnog efekta subjekta iznosi 1,60, a reziduala 1,13.

**Tabela 9.** MCMC procena parametara i njihove značajnosti

	$\beta$	95% i.p.		p
		donji	gornji	
odsečak	11,83	11,08	12,68	,0001
NAVRAT	1,20	,84	1,52	,0001
POL (m):STAROST	-,06	-,09	-,03	,0001
POL (ž):STAROST	-,12	-,14	-,09	,0001

Procenjeni  $R^2=,83$  za celokupan model, a samo za fiksne efekte  $R^2_F=,20$ .

Značajni su efekti navrata i interakcija pola i starosti (koji nisu od interesa za ovo istraživanje). Modalitet testiranja ne utiče na promenu skorova ispitanika na ovom testu.

Gulliksenov koeficijent ubrzanja za računarski podržanu formu iznosi ,37, a za klasičnu ,42. Klasična forma je neznatno ubrzanija.

## Diskusija

S obzirom da analiza rangova i strukture dvaju modaliteta testa ispitivanih testova nije dala jasne zaključke ne možemo tvrditi da je promena modaliteta testiranja dovela do promene predmeta merjenja primenjenih instrumenata. Ako tome dodamo i nalaze koji govore u prilog činjenici da nema razlike u skorovima ispitanika na dva modaliteta mogli bismo reći da su oni *psihometrijski i iskustveno ekvivalentni*.

Osvrnućemo se na raspodele ukupnih skorova testova. One se razlikuju u zavisnosti od modaliteta. S druge strane, razlikuju se i distribucije skorova testa i retesta u grupama kod kojih nije dolazilo do promena modaliteta. Uzroke ovome možemo tražiti u efektu pamćenja i vežbe, te kratkom vremenskom periodu između dva testiranja. Tome u prilog govori činjenica da je efekat navrata za oba testa značajan. Zanimljivo je da je razlika u distribucijama najmanje značajna upravo u situacijama kada je modalitet testiranja menjan. Skloni smo da to tumačimo smanjenim efektom vežbe i pamćenja upravo zbog menjanja modaliteta.

Dalja potvrda teze da su modaliteti testova i iskustveno ekvivalentni mogli bi biti i nalazi da se skorovi na dve forme ne razlikuju. Nije pronađen značajan efekat modaliteta testiranja niti efekat interakcije modaliteta i nekog drugog prediktorra. Drugim rečima, visina skorovi ispitanika ne zavise od modaliteta testa.

Međutim, razlozi za nedostatak efekata pomenutih prediktora mogu biti i druge prirode.

Prvo, ispitivani testovi bili su laki za ovaj uzorak pa uticaj, npr. računarske anksioznosti ili iskustva, ako je i postojao nije bio dovoljan da utiče na skorove ispitanika. Težina testova ima veze sa strukturom uzorka koji su uglavnom sačinjavali obrazovani ispitanici. Veći deo njih su studenti psihologije, koji su kroz prijemni ispit selekcionisani i po nivou sposobnosti.

Drugo, uzorak je bio homogen po polu, starosti, školovanju, računarskom iskustvu ali i po nivoima anksioznosti i računarske anksioznosti. Ispitanici su generalno pokazali niske nivoe anksioznosti (kako stanja, tako i računarske) sa malim variranjem oko proseka. Prema *modelu kognitivne interferencije računarske anksioznosti* (Smith, Capputi, 1997), anksiozni ispitanici bi prilikom suočavanja sa ugrožavajućim stimulusima (a kod računarski anksioznih – to je sve što se tiče računara) morali da odvoje izvesne kognitivne resurse za njihovo prevladavanje. Očito da nivo anksioznosti u ovom uzorku ne zahteva prevelike kognitivne resurse, pa su preostali bili sasvim dovoljni za neometano rešavanje zadataka iz testova koji su ispitivani.

Slično je i kada posmatramo uticaj računarskog iskustva. Iako dosadašnja istraživanja daju oprečne rezultate (Burke & Normand, 1987; De Beer & Visser, 1998; Russel, 1999; Taylor, Kirsch, Jamieson, & Eignor, 2002; McDonald, 2002), bilo je očekivano da postoji uticaj interakcije nivoa računarskog iskustva i modaliteta testiranja na postignuće ispitanika. Međutim, ni taj efekat se nije pokazao značajnim. Ispitanici u preko 96% slučajeva poseduju računar koji upotrebljavaju svakodnevno ili gotovo svakodnevno, a svoje računarsko iskustvo procenjuju kao osrednje (sa vrlo malim variranjem). Johnson i White (1980) tvrde da je samo malo računarsko iskustvo (npr. instrukcija od sat vremena) dovoljno da eliminiše njegov uticaj na postignuće na računarski podržanim kognitivnim testovima. Ovaj uzorak ispitanika ima mnogo više od toga, pa pretpostavljamo da je to bilo dovoljno da eliminiše uticaj ovog prediktora.

Da je nekog uticaja promene modaliteta možda i bilo, govori nam činjenica da je test AL-4 ubrzaniji u računarskoj formi. Moguće je da prilikom rešavanja testa u ovom modalitetu bitnu ulogu igra i vičnost radu sa mišem (pošto se na računarski podržanoj verziji ovog testa odgovara klikom na miša). S druge strane, ne mora se nužno raditi o umešnosti ispitanika, već se može raditi i o tehničkoj (ne)ispravnosti stanju miševa u računarskoj učionici koja je mogla uticati na brzinu rada.

Kao zaključak u vezi sa ispitivanim kognitivnim testovima, mogli bismo reći da računarske verzije zadovoljavaju zahteve koje u svojim smernicama za raču-

narski podržano testiranje postavljaju APA (1986) i Međunarodna komisija za testove (2005), da bi se mogli smatrati ekvivalentnim. Drugim rečima, merenje sposobnosti obavljeno različitim modalitetima je jednako, a nije ustanovljeno ni postojanje njihovog diferencijalnog funkcionisanja niti pristrasnosti. Ovo važi, kako za verbalni tako i za neverbalni test. Naravno ostaju dileme oko jednakosti strukture modaliteta testa AL-4 koju nismo sa sigurnošću dokazali (ali ni opovrgli).

U vezi sa prethodnim, McDonald (2002) kaže, da prilikom procene ekvivalencije modaliteta testiranja treba uzeti u obzir individualne razlike u iskustvu, računarskoj anksioznosti i sl. Međutim, on kaže i da na njih treba gledati kao na proizvod svog vremena, te da one za deset godina možda neće predstavljati tako bitan faktor. Na osnovu dobijenih nalaza mogli bismo reći da je McDonal-dova procena bila suviše pesimistična.

Na kraju, trebalo bi se osvrnuti na neke nedostatke ovog istraživanja. Kao što je već pominjano u diskusiji najveći nedostatak je uzorak koji je bio prigodan, a sačinjavali su ga uglavnom mladi, obrazovani ispitanici, pretežno ženskog pola, studenti psihologije. Međutim, ako pogledamo pregled dosadašnjih istraživanja u svetu, vidimo da su i tamo u velikoj većini uzorci bili prigodni, i neretko su ih sačinjavali studenti (psihologije). Teško da bi se ijedan od njih mogao smatrati reprezentativnim. Nedostatke uzorka u ovom istraživanju pokušali smo kompenzovati primenom bootstrapping procedure. Međutim, i ova procedura svoj pravi smisao ima na tek na reprezentativnom uzorku. Stoga, generalizacije iz ovog istraživanja bismo morali ograničiti na populaciju mladih, obrazovanih žena. Naime, uzorak je takav da nemamo pravo da *tvrdimo* drugačije. Što se tiče pola, na osnovu dosadašnjih istraživanja mogli bismo *pretpostaviti* da situacija ne bi trebalo da bude bitno drugačija kod muških ispitanika slične starosti i obrazovnog statusa. Takođe, možemo pretpostaviti da se situacija ne bi bitnije promenila ni kod sredovečnih ispitanika. Kada su u pitanju stariji i neobrazovani situacija bi mogla biti sasvim drugačija. Da bismo zaključke mogli generalizovati na celu populaciju, uzorak bi svakako trebalo dopuniti. Ovakav uzorak, je verovatan razlog restrikcije opsega skorova na kognitivnim testovima u oba modaliteta. Prosečni skorovi na oba testa su visoki i smanjenog varijabiliteta. Postoji mogućnost da bi na drugačijem uzorku, na kome bi skorovi kognitivnih testova imali veću varijansu, razlike među modalitetima dostigle nivo statističke značajnosti.

Mišljenja smo da je proveravanje ekvivalencije modaliteta, odnosno jednakosti mera dobijenih njima, posao koji bi trebalo da rade institucije sa većim resursima, koje za relativno kratko vreme mogu da obezbede reprezentativan uzorak i ispitaju ga na računarima. Zbog brzog napredovanja računarske tehnologije

i softvera, esencijalno je da ovakva istraživanja budu obavljena u što kraćem roku. To prevazilazi mogućnosti pojedinca.

Završićemo još jednom McDonaldovom izjavom (2002). On kaže da se danas računarski podržane verzije porede sa papirnim standardom, ali da se u budućnosti ta situacija može okrenuti. RPT u bliskoj budućnosti mogu postati standardi sa kojim će biti poređeni i na osnovu kojih će biti procenjivani drugi modaliteti. U bliskoj budućnosti biće sve veći broj testova koji će imati samo (ili na prvom mestu) računarski podržanu formu.

## Literatura

American Psychological Association (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: APA

Arce-Ferrer, A.J., & Martínez Guzmán, E. (2009). Studying the equivalence of computer-delivered and paper-based administrations of the Raven Standard Progressive Matrices. *Test Educational and Psychological Measurement*, 69, 855-867.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology*, 57, 221-237.

Bloom, A.J., & Hautaluoma, J.E. (1990) Anxiety management training as a strategy for enhancing computer user performance, *Computers in Human Behavior*, 6, 337-349.

British Psychological Society (1999). *Guidelines for the development and use of computer-based assessments*. Leicester: British Psychological Society

Brosnan, M.J. (1998). The impact of psychological gender, gender-related perceptions, significant others, and the introducer of technology upon computer anxiety in students. *Journal of Educational Computing Research*, 18, 63-78.

Bugbee, A.C.Jr. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28, 282-290.

Bugbee, A.C.Jr., & Bernt, F.M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, 23, 87-101.

Bujas, Z. (1967). *M-Serija, Priručnik*. Zagreb: Republički zavod za zapošljavanje Zagreb.

Burke, M.J., & Normand, J. (1987). Computerized psychological testing: Overview and critique. *Professional Psychology Research and Practice*, 1, 42-51.

Carver, C.S., & Scheier, M.F. (1986). Functional and dysfunctional responses to anxiety: the interaction between expectancies and self-focused attention. U: R. Schwarzer (ur.), *Self related cognitions in anxiety and motivation* (pp. 111-141). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chua, S.L., Chen, D., & Wong, A.F.L. (1999). Computer anxiety and its correlates: a meta-analysis. *Computers in Human Behavior*, 15, 609-623.

Colley, A., Gale, M., & Harris, T. (1994). Effects of gender role identity and experience on computer attitude components. *Journal of Educational Computing Research*, 10, 129-137.

Cooper, J. (2006). The digital divide: the special case of gender. *Journal of Computer Assisted Learning*, 22, 320-334.

de Beer, M., & Visser, D. (1998). Comparability of the paper-and-pencil and computerized adaptive versions of the General Scholastic Aptitude Test. *South African Journal of Psychology*, 28, 21-27.

Dimock, P.H., & Cormier, P. (1991). The Effects of format differences and computer experience on performance and anxiety on a computer-administered test. *Measurement and Evaluation in Counseling and Development*, 24, 119-126.

Durndell, A., Haag, Z., & Laithwaite, H. (2000). Computer self efficacy and gender: a cross cultural study of Scotland and Romania. *Personality and Individual Differences*, 28, 1037-1044.

Eysenck, M.W., & Calvo, M.G. (1992). Anxiety and performance: Processing efficiency theory. *Cognition and Emotion*, 6, 409-434.

Fajgelj, S. (2005). *Psihometrija - Metod i teorija psihološkog merjenja*. Beograd: CPP.

Ford, B.D., Vitelli, R., & Stuckless, N. (1996). The effects of computer versus paper-and pencil administration on measures of anger and revenge with an inmate population. *Computers in Human Behavior*, 12, 159-166.

Haladyna, T.M., & Roid, G.H. (1983). A comparison of two approaches to criterion referenced test construction. *Journal of Educational Measurement*, 20, 271-282.

Heinssen, R.K. Jr., Glass, C.R., & Knight L.A. (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior*, 3, 49-59.

Henderson, R.D., Deane, F.P., & Ward, M.J. (1995) Occupational differences in computer-related anxiety: Implications for the implementation of a com-

puterized patient management Information System. *Behavior and Information Technology*, 14, 23-31.

Honaker, L. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8, 561-577.

International Test Commission (2005). *International guidelines on computer-based and internet delivered testing*. Preuzeto sa <http://www.intestcom.org/Downloads/ITC%20-Guidelines%20on%20Computer%20%20version%202005%20approved.pdf>. 23.09.2008.

Janičić, B. (2004). *Konstrukcija i evaluacija ekspertskog sistema za interpretaciju MMPI* (neobjavljena magistarska teza). Novi Sad: Filozofski fakultet.

Johnson, D., & White, C. (1980). Effects of training on computerized test performance in the elderly. *Journal of Applied Psychology*, 65, 357-358.

Kanfer, R., & Heggstad, E.D. (1997). Motivational traits and skills: a person-centered approach to work motivation. *Research in Organizational Behavior*, 19, 1-56.

Kay, R.H. (1993). An exploration of theoretical and practical foundations for assessing attitudes towards computers: The computer attitude measure (CAM). *Computers in Human Behavior*, 9, 371-386.

King, W.C., & Miles, E.W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology*, 80, 643-651.

Kostić, P., i Anđelković, Z. (1996). Primena računara u profesionalnom radu psihologa. *Merenje u psihologiji*, 1, 169-173.

Kubinger, K.D., Formann, A.K., & Farkas, M.G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices, in particular for computerized testing. *European Review of Applied Psychology*, 41, 295-300.

Lord, F.M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 117-138.

McDonald, A.S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers and Education*, 39, 299-312.

Mead, A.D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.

Milin, P. (1999). Računarski podržano testiranje znanja: Prikaz jednog računarskog programa. *Info Science*, 7, 49-52.

Popov, B. (2007). *Uticaoaj kognitivnih činilaca na emocionalno reagovanje u situaciji negativne povratne informacije* (neobjavljena magistarska teza). Novi Sad: Filozofski fakultet.

Powers, D.E. (1999). *Test anxiety and test performance: Comparing paper-based and computer-adaptive versions of the GRE General Test (RR-99-15)*. Princeton: Educational Testing Service.

Rosen, L.D., Scars, D.C., & Weil, M.M. (1993). Treating technophobia: a longitudinal evaluation of the computerphobia reduction program. *Computers in Human Behavior*, 9, 27-50.

Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7 (20), on-line. Preuzeto sa <http://epaa.asu.edu/epaa/v7n20/>, 17.08.2009.

Schulenberg, S.E., & Yutrzenka, B.A. (2001). Equivalence of computerized and conventional versions of the Beck Depression Inventory-II (BDI-II). *Current Psychology*, 20, 216-230.

Shermis, M.D., & Lombard, D. (1998). Effects of computer-based test administrations on test anxiety and performance. *Computers in Human Behavior*, 1, 111-123.

Smith, B., & Caputi, P. (2007). Cognitive interference model of computer anxiety: Implications for computer-based assessment. *Computers in Human Behavior*, 23, 1481-1849.

Taylor, C., Kirsch, I., Jamieson, J., & Eignor, D. (2002). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, 49, 219-274.

Teasdale, S., & Lupart, J. (2001). *Gender differences in computer attitudes, skills, and perceived ability*. Canadian society for studies in education, Quebec, Maj, 2001.

Todorovski, Z., Salević–Obradović, J., i Janičić, B. (2006). *TMTns – Valjanost prema kriterijumu patoloških promena u endokranijumu definisanih kompjuterskom tomografijom*. 54. Naučno stručni skup psihologa Srbije – Sabor, Izažovi za primenjenu psihologiju u Srbiji. Zlatibor 24-27. maj 2006. Knjiga rezimea str. 45

Todorovski, Z., Salević-Obradović, J., Janičić, B., i Kopitović, A. (2006a). *Trail*

*making test – valjanost prema kriterijumu prisustva „mekih“ neuroloških znakova.* Drugi kongres neurologa Bosne i Hercegovine s međunarodnim sudjelovanjem, Mostar 9–12. novembar, Zbornik sažetaka, str. 155

Tseng, H.-M., Tiplady, B., Macleod, H.A., & Wright, P. (1998). Computer anxiety: A comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology*, 89, 599-610.

Vogel, L.A. (1994). Explaining performance on p&p versus computer mode of administration for the verbal section of the Graduate Record Exam. *Educational Computing Research*, 11, 369–383.

Whitaker, B.G. (2007). Internet-based attitude assessment: Does gender affect measurement equivalence?. *Computers in Human Behavior*, 23, 1183-1194.

Wijndaele, K., Matton, L., Duvigneaud, N., Lefevre, J., Duquet, W., Thomis, M., De Bourdeaudhuij, I., & Philippaerts, R. (2007). Reliability, equivalence and respondent preference of computerized versus paper-and-pencil mental health questionnaires. *Computers in Human Behavior*, 23, 1958-1970.

Williams, J.E., & McCord, D.M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Computers in Human Behavior*, 22, 791-800.

Wolf, B., Momirović, K., i Džamonja, Z. (1992). *KOG 3 (Baterija testova inteligencije)*. Beograd: Savez društava psihologa Srbije - Centar za primenjenu psihologiju.

Wright, D.L., Aquilino, W.S., & Supple, A.J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*, 62, 331-353.



Bojan Janičić

## Equity of computer-assisted and classic form measures of cognitive abilities

In this paper the equivalence of computer-assisted and traditional forms of two cognitive tests have been examined. In a sample of 238 subjects, the effects of gender, age, education level, computer experience, computer anxiety and state anxiety on the scores on different modalities of tests have been investigated.

The analysis of the test structure and analysis of the ranks of the test scores, in two modalities, yielded no information on which to conclude that change of modality affected constructive validity of tests.

There have been no significant differences in scores in two different modalities of test. Significant effects of interaction between the modality of tests and other employed predictors have not been detected. On this basis it has been concluded that the two modalities of testing have psychometric and experiential equivalence.

**Key words:** computer-assisted testing, psychometric equivalence, experiential equivalence, equity of measurement, cognitive tests