Univerzitet u Novom Sadu
Filozofski fakultet

Bojana Dinić

PRINCIPI PSIHOLOŠKOG TESTIRANJA

Novi Sad, 2019.

UNIVERZITET U NOVOM SADU
FILOZOFSKI FAKULTET
Dr Zorana Đinđića 2.
21000 Novi Sad
Tel: +38121485390
www.ff.uns.ac.rs

Za izdavača
Prof. dr Ivana Živančević Sekeruš, dekan

Bojana Dinić
PRINCIPI PSIHOLOŠKOG TESTIRANJA

Recenzenti
Prof. dr Goran Opačić, Filozofski fakultet, Univerzitet u Beogradu
Prof. dr Vladimir Hedrih, Filozofski fakultet, Univerzitet u Nišu

Lektura
mr Nataša Belić

Tehnička priprema
Igor Lekić

ISBN
978-86-6065-540-2

URL
http://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-540-2

Novi Sad, 2019.

Reprints Courtesy of International Business Machines Corporation,
© International Business Machines Corporation, pages 381–385, 387, 391–395, 397–398, 402, 405, 407–408, 410–416.
Nijedan deo ovog udžbenika ne sme se reprodukovati ni prenositi u bilo kom obliku, niti bilo kojim sedstvima – elektronskim, mehaničkim, fotokopiranjem, snimanjem ili na drugi način, u bilo koji informatički sistem za skladištenje i korišćenje bez prethodne dozvole vlasnika prava. Sva prava zadržava izdavač i autor.

Citirati kao:
Dinić, B. (2019). *Principi psihološkog testiranja*. Novi Sad, RS: Filozofski fakultet. Preuzeto sa http://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-540-2.

Sadržaj

Umesto predgovora ... 15

PSIHOLOŠKI TEST I NJEGOVA PRIMENA ... 17
Šta je psihološki test? ... 17
Test kao alat ... 17
Standardizovanost testa ... 19
Objektivnost testa ... 20
Test kao proizvod ... 20
Test i slični pojmovi ... 22
Odnos psihološkog testiranja i procene ... 31
Primena psiholoških testova ... 35
Klasifikacija ... 36
Selekcija ... 36
Trijaža ... 36
Sertifikat ... 37
Raspoređivanje ... 37
Dijagnostika ... 38
Evaluacija programa ... 39
Samospoznaja i lični razvoj ... 41
Naučna istraživanja ... 44
Uloge u psihološkom testiranju ... 46
Indirektne uloge ... 47
Autor ili konstruktor testa ... 47
Autor softvera ili osoba koja skoruje test ... 52
Naručilac testa ... 52
Recenzent ... 53
Direktne uloge ... 55
Korisnik testa ... 55
Ispitivač ili administrator testa ... 57
Ispitanik ili rešavač testa ... 57
Informacije o testovima ... 58
Pitanja za razmišljanje ... 61
Reference ... 61

INTERPRETACIJA SKORA NA TESTU ... 66
Razlike između normativne i kriterijumske interpretacije ... 67

NORMATIVNA INTERPRETACIJA 71
Uslovi za adekvatne norme 74
Reprezentativnost uzorka 74
Šta je referentna grupa? 77
Kolika treba biti veličina referentne grupe? 82
Aktuelnost normi 85
Vrste normi 88
Razvojne norme 89
Razvojne norme zasnovane na bihejvioralnim pokazateljima 90
Razvojne norme zasnovane na teorijskim pretpostavkama 92
Ostale vrste razvojnih normi 92
Prednosti *94*
Nedostaci *94*
Percentili 96
Prednosti 102
Nedostaci 103
Ukupni z-skor 109
Prednosti 110
Nedostaci 111
Izvedeni standardni skorovi 112
T-skor 112
Prednosti 115
Nedostaci 115
IQ-skor 116
Prednosti i nedostaci 118
Normalizovani standardni skorovi 118
Prednosti i nedostaci 121
Stenajn skorovi 121
Prednosti 125
Nedostaci 125
Sten skorovi 126
Prednosti 127
Nedostaci 128
C-skorovi 128
Činioci koji utiču na izbor normi 129
Oblik distribucije skorova 129
Preciznost procene 130

Standardna greška 130
Standardna greška razlike 139
Pitanja za razmišljanje 142
Reference 143

KRITERIJUMSKA INTERPRETACIJA 149
Prednosti 152
Nedostaci 153
Kombinovanje kriterijumske i normativne interpretacije 153
Vrste kriterijumske interpretacije 156
Granični skor 156
Kako odrediti granični skor? 158
Kvotni metod 160
Metod baziran na merama centralne tendencije 160
Metod baziran na evaluaciji ispitanika (eng. method based on evaluation of examineers) 161
Metod baziran na evaluaciji testa (eng. method based on evaluation of the test) 166
Metod s kriterijumskom grupom 167
Metod baziran na regresionoj analizi 176
Jedan ili više graničnih skorova? 178
Nedostaci graničnih skorova 179
Tabele i grafikoni očekivanih vrednosti 180
Kako napraviti tabelu očekivanih vrednosti? 184
Interpretacija u odnosu na sadržaj ispitivanog područja 186
Vrste testova u odnosu na kriterijum 187
Testovi postignuća 187
Kako se određuje kriterijum? 190
Prikaz rezultata 190
Testovi učinka 190
Razlike testova postignuća i testova učinka 192
Kako se određuje kriterijum? 192
Prikaz rezultata 193
Testovi podobnosti 193
Problemi razlikovanja testova postignuća i podobnosti 198
Kako se određuje kriterijum? 200
Prikaz rezultata 201
Pitanja za razmišljanje 201

Reference ... 202

INTERPRETACIJA ORIJENTISANA NA OSOBU ... 207
Ipsativni skorovi ... 207
Vrste ipsativnih skorova ... 209
Prednosti i nedostaci ... 216
Pitanja za razmišljanje ... 219
Reference ... 219

ODGOVARAČKE PRISTRASNOSTI ... 222
Vrste odgovaračkih pristrasnosti ... 225
Socijalna poželjnost/disimulacija i simulacija ... 226
Definicija ... 226
Dimenzije ... 229
Problemi ... 240
Detekcija ... 240
Rešenja ... 257
Povlađivanje/slaganje i neslaganje ... 268
Definicija ... 268
Dimenzije ... 271
Problemi ... 272
Detekcija i rešenja ... 272
Problemi negativno formlusanih tvrdnji ... 275
Preferiranje ekstremnih ili srednjih odgovora ... 279
Definicija ... 279
Dimenzije ... 280
Problemi ... 280
Detekcija i rešenja ... 282
Nepažljivi odgovori ... 283
Dimenzije ... 284
Problemi ... 284
Detekcija ... 285
Rešenja ... 288
Izvori odgovaračkih pristrasnosti ... 289
Izvori koji potiču od testa i uslova testiranja ... 290
Format ajtema ... 290
Koji broj podeoka je optimalan? ... 291
Nejasni i složeni ajtemi ... 292

Medijum testa....293
Kognitivno opterećenje....293
Dužina testa....294
Vremensko organičenje....295
Ispitivač....295
Jezik testa....295
Predmet merenja....296
Izvori koji potiču od ispitanika....296
Pol....297
Starost....297
Ekonomski status....298
Obrazovanje....298
Inteligencija....298
Osobine ličnosti....299
Motivacija....301
Rasa....302
Kultura....302
Opšte tehnike prevazilaženja....303
Pitanja za razmišljanje....305
Reference....305

IZAZOVI PRIMENE TESTA....321
Da li je psihološki test potreban?....324
Osnovni razlozi za primenu testa....325
Nepristrasnost testiranja....327
Vrste pristrasnosti testiranja....331
Konstruktna pristrasnost....331
Detekcija konstruktne pristrasnosti.... 333
Prediktivna pristrasnost....336
Detekcija prediktivne pristrasnosti....337
Pravedna upotreba testa....341
Izvori pristrasnosti....343
Sadržajni izvori pristrasnosti testa....343
Izvori pristrasnosti povezani s odgovorima....343
Problemi u procesu testiranja....344
Problemi koji potiču od korisnika testa....344
Sistem kvalifikacije korisnika testa....348
Odgovornosti korisnika testa....351

Problemi prilikom selekcije testova ... 354
Problemi prilikom administracije testova ... 357
Priprema uslova testiranja ... 357
Priprema ispitanika ... 359
Priprema ispitivača ... 364
Problemi prilikom skorovanja testova ... 368
Problemi prilikom tumačenja rezultata testova ... 370
Prava i odgovornosti ispitanika ... 374
Pitanja za razmišljanje ... 375
Reference ... 375

PRIRUČNIK ... 380
Ajtem-analiza ... 380
Preduslov ... 380
Matrica za vežbu ... 380
Rekodiranje ... 381
Koje opcije treba odabrati za ajtem-analizu? ... 383
Tumačenje autputa iz SPSS-a ... 385
Pouzdanost interne konzistencije - Kronbahov alfa koeficijent (α) ... 385
Pouzdanost interne konzistencije ... 385
Ajtem statistici - osnovni deksriptivni pokazatelji ... 387
Težina ajtema ... 388
Standardna devijacija ajtema ... 390
Oblik distribucije u kontekstu određenja težine ajtema ... 391
Sumirane vrednosti za ajtem statistike ... 394
Težina testa ... 394
Homogenost ... 394
Ajtem-total statistici ... 397
Diskriminativnost ajtema ... 399
Koeficijent determinacije ... 401
Promena alfe ukoliko se ajtem ukloni ... 404
Statistici za skalu – deskriptivni podaci za skalu ... 405
Težina testa ... 406
Oblik distribucije u kontekstu određenja težine testa ... 406
Primer izveštaja ... 409
Računanje normi ... 409
Percentili ... 409

Standardni skorovi 412
Izvedeni standardni skorovi 413
Normalizovani standardni skorovi 414
Sten i stenajn 417
Reference 418

Umesto predgovora

U ovom udžbeniku se obrađuju savremeni zahtevi prilikom praktične primene testa, te svaki korak u psihološkom testiranju, od postavljanja cilja testiranja i razlikovanja uloga u procesu testiranja, standarda za administraciju, do referentnih okvira za interpretaciju rezultata i izveštavanje o njima. Takođe, obrađuju se problemi prilikom testovne prakse u svakoj fazi testiranja i praktične smernice za rešavanje tih problema.

Ovaj udžebnik je nastao iz potrebe da budućim psiholozima pruži osnovne informacije o procesu testiranja, te je, pre svega, namenjen studentima psihologije, ali može koristiti i svim drugim praktičarima i istraživačima koji u svom radu primenjuju testove za različite ciljeve. Užbenik je prvenstveno namenjen studentima psihologije viših godina osnovnih studija koji pohađaju istoimeni kurs na Odseku za psihologiju Filozofskog fakulteta u Novom Sadu, Principi psihološkog testiranja, te razumevanje teksta zahteva poznavanje osnovnih statističkih i psihometrijskih pojmova i analiza koji se obrađuju na nižim godinama osnovnih studija. Ipak, na kraju udžbenika može se naći priručnik za sprovođenje ajtem-analize koji ujedno predstavlja i podsetnik o osnovnim psihometrijskim karakteristikama. Udžbenik je nastao s idejom da može pružiti pomoć u samostalnom sprovođenju testiranja i evaluaciji testovnih rezultata, te u razvoju kritičkog pristupa prilikom selekcije i ocenjivanja testova.

Ovom prilikom želim da se zahvalim studentima generacije upisa 2015. i 2014. koji su pohađali ovaj kurs, kao i doktorantima Milanu Oljači, Bojanu Branovačkom, Selki Sadiković i Iliji Milovanoviću za korisne sugestije i komentare za poboljšanje razmljivosti i opšte čitljivosti teksta. Takođe, veliku zahvalost dugujem i recenzentima, prof.

dr Goranu Opačiću i prof. dr Vladimiru Hedrihu, za upućene komentare i sugestije. Želela bih da se zahvalim i kolegi Zoranu Svilaru koji je uradio originalne određene grafikone i slike, kao i kolegi doc. dr Dejanu Pajiću koji je napravio onlajn vizualizaciju preseka dve distribucije.

Mojim roditeljima i Sveti

BD

Šta je psihološki test?

Posao psihologa je teško zamisliti bez upotrebe testa. Test je najkorišćeniji alat na osnovu kojeg psiholog donosi odluke u praktičnom i istraživačkom radu. Mnogi bi se složili sa iskazom da je test neizbežan ili obavezan alat psihologa pri donošenju odluka, iako svakako nije jedini. Stoga, na početku objašnjenja pojma test, važno je istaći dve osnovne činjenice o testu. Prva je da je test alat psihologa, a druga, podjednako važna, da je test proizvod. Druga činjenica se često zanemaruje, ali svakako dopunjuje određenje testa. U narednom odeljku će ove dve osnovne činjenice o testu biti detaljnije obrazložene.

Test kao alat

Test je alat ili instrument kojim psiholog obavlja svoju delatnost. U samim definicijama testa se ističe ova činjenica o testu. Na primer, najosnovnija definicija testa glasi da je test psihološki merni instrument (Fajgelj, 2013). Ova definicija je pogodna jer se njome određuju bitne odrednice testa:

1. naučna disciplina u kojoj se test primenjuje, tj. *psihologija*;
2. suština postojanja i primene testa koja je u *merenju*;
3. test se smatra *instrumentom*, tj. sredstvom ili alatom koji pomaže u donošenju odluka o osobi, grupi ljudi, programu ili psihološkom konstruktu i teoriji.

Treba napomenuti da, iako se test definiše kao *psihološki* merni instrument, njegova primena nije ograničena samo na psihološku

praksu, već se koristi i u ostalim naučnim disciplinama. Tako, testove često srećemo u praksi pedagoga, sociologa, doktora medicine i na svim mestima gde postoji potreba za testovnim ispitivanjem različitih aspekata funkcionisanja ljudi.

U Standardima za pedagoško i psihološko testiranje (u daljem tekstu Standardi), test se definiše kao „evaluativni instrument ili postupak u kojem se uzorak ponašanja ispitanika u određenom području registruje i posle skoruje i evaluira u standardizovanom procesu" (AERA, APA, & NCME, 2006, str. 24). Svaki element navedene definicije je bitna karakteristika i uslov za određenje testa:

1. test je *instrument ili postupak*, dakle profesionalni alat na osnovu čijih rezultata se donose odluke o osobama, grupama ljudi ili programu;
2. test je *uzorak ponašanja*, tj. treba da predstavlja reprezentativan set širokog spektra ponašanja koji obuhvata mereni konstrukt, a na osnovu kojeg se mogu izvesti pretpostavke koje doprinose donošenju odluke;
3. uzorak ponašanja odnosi se na *određeno područje*, tj. relevatan je za kognitivno, afektivno i/ili socijalno funkcionisanje osobe, što je utvrđeno na osnovu empirijskih ili praktičnih, odnosno iskustvenih podataka;
4. uzorak ponašanja se testom *registruje*, što govori o testu kao i o načinu prikupljanja podataka;
5. uzorak ponašanja se *skoruje i evaluira* – nad rezultatima testa se primenjuje utvrđen numerički ili kategorijalni sistem, te se rezultati interpretiraju shodno utvrđenim pravilima;
6. za adekvatnu evaluaciju testovnog rezultata, potrebno je da test bude *standardizovan*.

Standardizovanost testa

Standardizovanost testa je njegova veoma važna odlika i odnosi se na tri aspekta: standardizaciju procedure testiranja, interpretacije i skorova (Murphy & Davidshofer, 1994). Standardizacija procedure odnosi se na uniformnost davanja instrukcija za rad, odnosa prema ispitaniku, izgleda testa (na primer, redosled i format ajtema), načina zadavanja testa (na primer, grupno ili individualno, uz list za odgovore ili uz direktno davanje odgovora u samom testu), uslova zadavanja testa i slično. Standardizacija interpretacije odnosi se na adekvatnost dobijenih normi kako po pitanju reprezentativnosti uzorka na kojem su dobijene tako i po pitanju aktuelnosti normi. Konačno, standardizacija skorova odnosi se na izbor adekvatnih normi u odnosu na distribuciju skorova i druge relevantne karakteristike. U opštem smislu, standardizacija se odnosi na uniformnost samog testa i testovne procedure, bez obzira na to ko test zadaje, skoruje i interpretira rezultate. Svrha standardizovanosti procedure testiranja jeste da se sve varijable koje potiču od ispitivača ujednače što je više moguće, držeći se tako pod kontrolom. Na taj način se obezbeđuje da skor svakog ispitanika zavisi samo od izraženosti njegove osobine ili sposobnosti, a ne (i) od procedure testiranja. Standardizacija doprinosi osiguravanju istih mogućnosti svim ispitanicima da pokažu svoje kompetencije i osobine. Time se omogućuje preciznost i uporedivost testovnih skorova. Treba napomenuti da pored standardizacije, i održavanje sigurnosti testa jedan je od načina obezbeđivanja istog tretmana za sve ispitanike. Pod sigurnošću testa podrazumeva se da se test ne umnožava bez dozvole i ne deli potencijalnim ispitanicima pre samog testiranja, da se ne pokazuju rešenje testa i tumačenje, niti da se objašnjava način rešavanja testa i slično.

Objektivnost testa

Standardizovanost je važna zbog postizanja objektivnosti u procesu testiranja. Objektivnost je jedna od metrijskih karakteristika testa i podrazumeva da skor na testu, tj. njegovo tumačenje ne zavisi od subjektivne procene procenjivača, već da se bazira na objektivnim kriterijumima. Da bi tumačenje testovnog skora bilo objektivno, a to znači adekvatno i u skladu sa utvrđenim pravilima, mora se poštovati standardizovanost procedure testiranja. U skladu sa prepoznavanjem važnosti za objektivnošću, Anastasi (Anastasi, 1988) definiše test kao objektivnu i standardizovanu meru uzorka ponašanja.

Test kao proizvod

Pored toga što je test alat ili instrument koji pomaže u donošenju odluka, postoji još jedna važna činjenica o testu koja se često zanemaruje, a to je da je test proizvod. Test je proizvod kao i bilo koji drugi, kao što je to, na primer, čokolada. Kao takav, za njegovo plasiranje potrebno je znati kom tržištu je namenjen, kojoj ciljnoj grupi, kakve dobiti može da pruži i slično, što je povezano i sa marketinškim veštinama predstavljanja testa.

Test kao proizvod proizvođaču treba da donese neku vrstu dobiti. Primena testa autoru testa pruža određene dobiti kao što su doprinos akademskoj zajednici, sticanje naučne reputacije, uticajnosti, citiranosti i slično. Ipak, ne treba zanemariti i finansijske dobiti koje test može da ostvari. Važno pomenuti podelu testova na komercijalne i one koji su u javnom domenu. **Komercijalni testovi** nisu javno dostupni, već je za njihovo korišćenje neophodno otkupiti dozvolu od autora ili češće od izdavačke kuće koja raspolaže autorskim pravima i/ili pravima na distribuciju testa. U ovom slučaju, autor i izdavačka kuća imaju dobit

od distribucije testa, koja se odnosi na sticanje reputacije, a takođe imaju i materijalnu dobit od prodaje dozvole za korišćenje testa i samog testovnog materijala. Komercijalni testovi nisu uopšte jeftini, te na domaćem tržištu za neki komercijalni test (priručnik i testovni materijal za određeni broj ispitanika) treba odvojiti od 10.000 do 130.000 RSD. Pored otkupljivanja dozvole za korišćenje i testovnog materijala, vlasnici autorskih prava imaju i posebne uslove za objavljivanje sadržaja testova u različitim publikacijama, o čemu takođe treba voditi računa. Naime, sama kupovina dozvole za korišćenje testa ne podrazumeva da se sadržaj testa može objavljivati u publikacijama kao što su različiti izveštaji, naučni članci itd. Najčešće, ukoliko je u pitanju test koji se komercijalno distribuira, u objavljenim publikacijama ne bi smeo da se dâ sadržaj stavki ili zadataka testa. To je eventualno moguće, ali uz posebnu pisanu dozvolu vlasnika autorskih prava, i to najčešće podrazumeva prikaz samo nekoliko stavki, nikako celog testa. Neke izdavačke kuće imaju pripremljene primere stavki ili zadataka koje se mogu objavljivati u publikacijama, ali to svakako treba proveriti. Dakle, sadržaj stavki komercijalnog testa sme da se navede u celosti samo u okviru priručnika za test i testovnog materijala, a ne i u okviru ostalih publikacija, osim ako to dozvolom za korišćenje nije drugačije regulisano. Na ovaj način se štite autorska prava i sprečava zloupotreba testova koji se koriste za donošenje različitih odluka.

S druge strane, **testovi u javnom domenu** mogu se slobodno koristiti, bez otkupljivanja dozvole i prava na korišćenje. Jedino što treba proveriti pre primene testa koji je u javnom domenu, su uslovi za njegovo korišćenje, a to se proverava direktno od autora testa. Na primer, pojedini autori zauzvrat traže prikupljene podatke, dozvoljavaju modifikacije testa, a drugi ne, pojedini traže dokaz o rigoroznoj

proceduri prevoda i adaptacije testa, dok su drugi fleksibilniji po tom pitanju itd. Svakako, kontaktiranje autora testa je lep akademski gest, a na taj način se može proveriti i da li već postoji preveden test na jezik koji vam je potreban, pa ne radite dupli posao.

Test i slični pojmovi

Pojam testa često se meša sa srodnim pojmovima kao što su inventar, upitnik, skala, anketa i slično. Prema užem određenju, pod pojmom testa podrazumevaju se samo oni postupci u kojima se odgovori ispitanika evaluiraju na osnovu tačnosti i kvaliteta. Takvi postupci uključuju procenu kognitivnog funkcionisanja, znanja, veština i sposobnosti i pripadaju tzv. prvoj paradigmi psihološkog merenja. U okviru ove paradigme, zadatak ispitanika je jasan i ispitanik ima jasan cilj – uspešno rešenje zadatka ili obavljanja neke aktivnosti, pri čemu se nedvosmisleno zna šta je tačno rešenje ili uspešno obavljena aktivnost. Uspešnost se procenjuje na osnovu spoljašnjeg kriterijuma – skorovanje i evaluaciju vrši ispitivač/ocenjivač ili posmatrač. Stoga se ova paradigma često naziva i paradigmom spoljašnjeg ocenjivanja. Još jedan naziv koji je u upotrebi je paradigma maksimalne aktivnosti, zato što je njen cilj utvrđivanje maksimuma onoga što ispitanik može da postigne i pruži, tj. utvrđivanje dometa ili tzv. plafona. Testovi koji se primenjuju u okviru ove paradigme zajednički se nazivaju testovima sposobnosti.

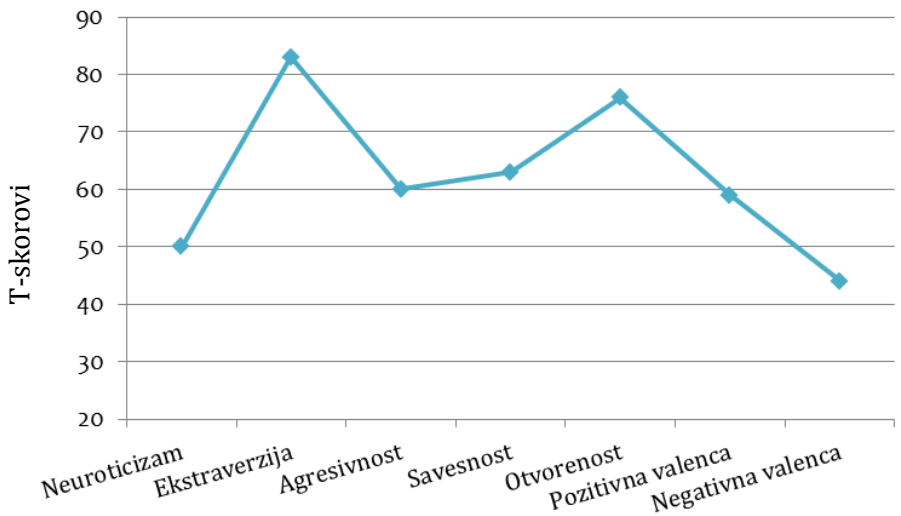
S druge strane, prema blažem stanovištu, i postupci koji ne uključuju opisan način evaluacije, mogu se takođe nazivati testovima. Ti postupci uključuju procenu ličnosti, stavova i uverenja, motivacije, interesovanja, i nazivaju se inventarima, upitnicima, skalama, anketama, ček-listama ili projektivnim tehnikama. Ovi testovi pripadaju drugoj paradigmi psihološkog merenja, gde nije nužno da postavljena pitanja

budu jasna, a ispitanik sam procenjuje svoje ponašanje, misli i osećanja. Pri tome, ispitivač ne mora nužno znati ispitanikov cilj prilikom davanja odgovora. S obzirom na to da ispitanik vrši samoprocenu, ova paradigma se još i naziva paradigmom unutrašnjeg ocenjivanja. Kako je cilj utvrđivanje tipičnog ponašanja ispitanika, tj. njegovog uobičajenog ponašanja, ova paradigma se naziva i paradigmom tipične aktivnosti. Testovi koji se primenjuju u okviru ove paradigme se zajedničkim nazivom zovu testovima ličnosti, iako, kao što je navedeno, njihova upotreba nije ograničena na procenu osobina ličnosti, već i drugih konstrukata kao što su stavovi, motivacija, emocije i slično. U stvari, ovi testovi su svi testovi kojima se *ne* procenjuju sposobnosti, već drugi konstrukti (osobine ličnosti, stavovi itd.). U ovoj knjizi će se pod testom podrazumevati i testovi sposobnosti i testovi ličnosti (usvojiće se fleksibilnije određenje testa).

Zabuna oko terminologije odnosi se, u stvari, na testove u okviru paradigme tipične aktivnosti. U ovoj paradigmi se često pojam test meša sa pojmovima kao što su inventar, upitnik, skala i slično. Odmah treba napomenuti da razlike između navedenih pojmova nisu jasno definisane u literaturi, ali i istaći činjenicu da se svi ovi pojmovi odnose na testove ličnosti, a ne sposobnosti. Drugim rečima, ne bi trebalo da naiđete na inventar kognitivnih sposobnosti, niti skalu kognitivnih sposobnosti.

Inventar (eng. *inventory*) bi trebalo da predstavlja procenu nekoliko karakteristika, tj. da podrazumeva multidimenzionalnost. Drugim rečima, u okviru inventara bi trebalo da se dobije nekoliko skorova, tačnije onoliko skorova koliko se karakteristika meri inventarom. Pri tome, svaka karakteristika se meri jednom zasebnom skalom. Suština inventara bi trebalo da bude u dobijanju različitih profila – ličnosti, interesovanja, motivacije, vrednosti i slično. Važno je

napomenuti da ukupni ili total skor u nekim inventarima ima smisla, a u nekima nema. Tako, inventar ličnosti Velikih pet plus dva – VP+2 (Smederevac, Mitrović i Čolović, 2010) meri sedam bazičnih osobina ličnosti, pa njime dobijamo sedam skorova, ali ne i jedan ukupni skor, jer on nema interpretativnog smisla. Primer profila baziranog na skorovima na VP+2 dat je na Slici 1. Na osnovu datog profila možemo zaključiti da ispitanik ima izrazito povišene skorove na ekstraverziji i otvorenosti, viši skor na savesnosti, a niži skor na negativnoj valenci, dok su skorovi na ostalim skalama prosečni.



Slika 1. Primer profila ličnosti na inventaru Velikih pet plus dva.

S druge strane, postoje inventari koji nude i ukupni skor. Takvi inventari su, u stvari, jednodimenzionalni – mere jedan konstrukt, ali

zbog složenosti konstrukta, sastoje se od nekoliko **supskala ili faceta**[1] (u ovom slučaju opravdanije je koristiti termin "supskala" ili "faceta" nego "skala" koji je primereniji inventarima u kojima nije predviđen ukupan skor). Takav je, na primer, Maslak inventar sindroma izgaranja (Maslach Burnout Inventory – MBI: Maslach & Jackson, 1981), koji je doživeo čak četvrto izdanje. Ovaj instrument, u svojoj verziji za stručnjake pomagačkih zanimanja, meri jedan konstrukt – sindrom izgaranja na poslu, ali meri različite aspekte ovog konstrukta i sastoji se od tri supskale: emocionalne iscrpljenosti, depersonalizacije i smanjenog ličnog postignuća. Ovim inventarom se, pored skorova na tri supskale, može dobiti i ukupni skor. U ovom kontekstu, ukupni skor ima smisla jer se sve supskale odnose na isti predmet merenja – sindrom izgaranja. Isto tako, u okviru pomenutog inventara VP+2, svaka od skala ima svoje supskale. Na primer, skala neuroticizma sastoji se od supskala anksioznosti, depresivnosti i negativnog afekta, pri čemu se mogu dobiti skorovi na navedenim supskalama, ali i ukupan skor na skali neuroticizma. Pritom, supskale i skale mogu se zadavati individualno u istraživačke svrhe, što zavisi od samog cilja istraživanja, ali nema puno smisla zadavati ih individualo u praktične svrhe, kada treba da se donese odluka o pojedincu.

Treba pomenuti da postoje jednodimenzionalni instrumenti koji se nazivaju inventarima, iako nemaju skale ili supskale. Takav je, na primer, Bekov inventar depresivnosti (aktuelna je verzija II, tj. Beck Depression Inventry II – BDI-II: Beck, Steer, & Brown, 1996). Navedeni instrument meri samo jedan konstrukt – depresivnost i daje samo

[1] U ovom kontekstu "faceta" ima isto značenje kao i "supskala" i ne odnosi se na facete kao izvore varijabilnosti u teoriji generalizabilnosti ili G-teoriji.

ukupni skor (ne sastoji se od skala ili supskala). Ipak, ono što se čini važnim prilikom određenja nekog testa kao inventara je mogućnost uvida u profil. Može se zaključiti da, bez obzira na to da li su multidimenzionalni ili jednodimenzionalni, inventari bi trebalo da pruže informaciju o profilu ispitanika, tj. da se sastoje od nekoliko skala ili supskala. S obzirom na to, čini se neopravdanim Bekov inventar depresivnosti nazivati inventarom.

Ako se usvoji fleksibilnije određenje testa, tj. ako se pod pojmom testa podrazumevaju i testovi koji pripadaju paradigmi tipične aktivnosti, inventar se može smatrati testom. U odnosu na ostale vrste instrumenata, pre svega na upitnik i skalu, može se smatrati da je inventar najbliži pojmu testa. Tome doprinosi i praksa u kojoj su inventari često standardizovani.

Upitnik (eng. *questionnaire*) i **skala** (eng. *scale*) češće su jednodimenzionalni instrumenti, a mogu se sastojati od nekoliko supskala. Pojam skala se više vezuje za procenu stavova (skala stavova), mada ne nužno. Takođe, moguće je da autori termin "skala" pripisuju onim instrumentima u kojima se kao format prezentovanja primenjuju skale sa uređenim kategorijama, poput skale Likertovog tipa. Zabunu više unose nazivi nekih multidimenzionalnih instrumenata koji se pre mogu okarakterisati kao inventari, a nazivaju se upitnicima. Na primer, postoji Upitnik šesnaest faktora ličnosti (16 Personality Factors – 16PF: Cattell, Cattell, & Cattell, 1993) kojim se dobijaju procene na osobinama ličnosti i uvid u profil ličnosti, što više odgovara inventaru. Svi navedeni primeri ukazuju na terminološku neusaglašenost kada je reč o instrumentima u okviru paradigme tipične aktivnosti, te je savet da se prilikom konstrukcije novog instrumenta vodi računa o primerenosti određenog termina nazivu instrumenta – da li je u pitanju inventar,

upitnik ili skala. Prilikom navođenja naziva postojećeg instrumenta treba koristi termin koji je autor instrumenta dao, čak i ako možda termin nije adekvatan. Drugim rečima, ako je izvorni naziv nekog instrumenta, npr. Aggressiveness Questionnaire, ne treba ga prevoditi kao Skala agresivnosti, već kao Upitnik agresivnosti. Kada je u pitanju status upitnika i skala, ukoliko se usvoji fleksibilnije stanovište shvatanja testa, onda se i oni svakako mogu smatrati testovima. U prilog tome ide i činjenica da se upitnici i skale mogu standardizovati.

Na ovom mestu skrenula bih pažnju na čestu grešku koju srećem kod studenata, a to je da se niz pitanja o socio-demografskim i sličnim karakteristikama naziva upitnikom (ili socio-demografskim upitnikom). U ovom slučaju svakako nije reč o upitniku nego o setu pitanja, budući da se na osnovu odgovora o socio-demografskim karakteristikama ne može dobiti neki ukupni skor. Stoga je za pitanja ovakvog tipa pravilno reći da je reč o setu pitanja, a nikako da je reč o upitniku.

Anketa, ili kako se još naziva i anketni upitnik (eng. *survey*) ima jasniju distinkciju u odnosu na ostale vrste instrumenata u okviru paradigme tipične aktivnosti. Naime, anketom se ne meri neki latentni konstrukt. Tako, u anketi možemo pitati za pol, bračni status, političko opredeljenje, učestalost konzumacije alkohola i sl., ali ova pitanja nemaju zajedničko jezgro u nekom latentnom konstruktu. Drugim rečima, odgovori na pitanja u anketi nisu manifestacija jednog uzroka, tj. latentnog konstrukta. Anketa ne podrazumeva nužno dobijanje nekog skora, već se najčešće pitanja u anketi analiziraju zasebno, pa se na osnovu uvida u pojedinačne odgovore ili grupu kritičnih pitanja izvode zaključci o predmetu merenja. Ipak, treba pomenuti da pitanja u anketi mogu nekad služiti za dobijanje određenog indeksa (Edwards & Bagozzi, 2000). Na primer, može se dobiti indeks stresa na osnovu odgovora

ispitanika o prisustvu negativnih događaja (npr. smrt bliske osobe, bolest, razvod, gubitak posla...). Pretpostavka koja leži u osnovi ovog indeksa je da iskustvo negativnih događaja vodi ka stresu, ali ne i da latentni konstrukt, kao što je stres, uzrokuje ove događaje. Druga karakteristika ankete je da ona uobičajeno sadrži pitanja različitog formata prezentovanja (npr. pitanja višestrukog izbora, sa uređenim kategorijama ili otvorenog tipa). Treća karakteristika ankete je da se ona uobičajeno konstruiše *ad hoc*, dakle za jednu namenu. Najčešće se anketom mere stavovi i preferencije, zadovoljstvo konkretnim uslugama i ispituje javno mnjenje. S obzirom na navedene karakteristike, anketa svakako ne predstavlja standardizovani instrument i ne može se smatrati testom.

Na kraju treba pomenuti i na šta se odnosi baterija testova. **Baterija testova** predstavlja grupu testova (ili suptestova) koji se zadaju odjednom, jednoj osobi (Urbina, 2004). Baterija se može odnositi na dva načina kompletiranja testova. Prvi način obuhvata baterije koje su standardizovane i kontruisane s namerom da mere neki složeni konstrukt. Čim su standardizovane, znači da su im psihometrijske karakteristike poznate. U Standardima (AERA, APA, & NCME, 2006, str. 258) pod baterijom se upravo na to misli, i ona se definiše kao „skup testova standardizovanih na istoj populaciji tako da se njihovi normirani rezultati mogu upoređivati ili u kombinaciji upotrebljavati za donošenje odluke". One obično već i sadrže termin "baterija" u svom nazivu. Takva je baterija testova inteligencije – KOG-3 (Wolf, Momirović i Džamonja, 1992) za procenu intelektualnih sposobnosti, koja se sastoji od tri testa, od kojih svaki ima poseban predmet merenja: Test upoređivanja slika (IT-1) služi za procenu perceptivne sposobnosti, Test sinonima-antonima (AL-4) služi za procenu sposobnosti verbalnog razumevanja i

Test spacijalizacije (S-1) služi za procenu sposobnosti vizualizacije prostornih odnosa. Svi testovi su fizički odvojeni jedan od drugoga, i mogu se zadavati pojedinačno za različite potrebe, a može se zadavati i cela baterija, kada je cilj procena generalne intelektualne sposobnosti. Pri tome se dobija IQ mera zbrajanjem rezultata na pojedinačnim testovima i prevođenjem u norme. Dakle, ovom baterijom predviđeni su i skorovi na pojedinačnim testovima, i globalni skor na bateriji.

Kada je reč o baterijama za procenu fenomena ličnosti, čini se da autori često mešaju bateriju sa inventarima, upitnicima i skalama. Nazivajući baterijom neki set skala ili testova, verovatno se želi istaći multidimenzionalnost ili složenost jednodimenzionalnih konstrukata koji su predmet merenja. Kibernetička baterija konativnih testova – KON-6 (Momirović, Wolf i Džamonja, 1992) sastoji se od šest testova za procenu različitih regulativnih sistema ličnosti. Ovim testovima dobijaju se mere različitih regulatornih funkcija, ali ne i njihova ukupna mera. Dakle, reč je o multidimenzionalnom konstruktu. S druge strane, postoji Baterija prosocijalne ličnosti (Penner, Fritzsche, Craiger, & Freifeld, 1995) koja meri dva aspekta prosocijalnosti – empatiju orijentisanu ka drugima i spremnost na pomaganje, pri čemu svaki aspekt sadrži facete ili supskale, ali je baterijom predviđen i ukupni skor prosocijalnosti. U ovom slučaju je reč o složenom jednodimenzionalnom konstruktu prosocijalnosti. Iako je postojanje ukupnog skora uobičajeno kod baterije testova, on nije obavezan. Međutim, postoje razlike između baterije testova s jedne strane, i inventara, upitnika i skala, s druge strane. Name, u bateriji su testovi fizički odvojeni, dok u inventarima, upitnicima i skalama stavke različitih skala i supskala obično ne idu redom, jedna za drugom, već su slučajno ili sekvencijalno raspoređene u testu (mada ovo nije pravilo). Pored toga, testovi u okviru baterije se

mogu zadavati samostalno – informacija o skoru na samo jednom testu može biti od praktičnog interesa za neke specifične potrebe procene. S druge strane, informacija o skoru na jednoj skali inventara ili upitnika obično nije od praktičnog interesa, već je njihova samostalna primena ograničena na istraživačke svrhe. Može se zaključiti da testovi u okviru baterije imaju širu praktičnu primenu, za razliku od skala u inventarima i upitnicima. U tom smislu, skala u inventaru i upitniku se ne može izjednačiti sa testom u okviru baterije.

Drugi način kompletiranja testova u bateriju odnosi se na *ad hoc* sastavljanje baterije, za specifične potrebe i pružanje pomoći u donošenju specifičnih odluka. Ovakve baterije su mnogo češće u praksi i njih sastavlja sam korisnik. Fajgelj (2013) u svojoj definiciji baterije ističe upravo ovaj način sastavljanja baterije, te definiše bateriju kao skup nezavisnih testova koje korisnik sastavlja za neku određenu potrebu. Na primer, za procenu kandidata za posao obično se zadaje jedan test opšte intelektualne sposobnosti, jedan test sposobnosti koja je uže povezana sa radnim mestom (npr. test znanja iz oblasti ili rešavanja konkretnih problema koji su u vezi sa vrstom posla) i jedan test ličnosti, koji, po proceni korisnika, može sadržati informacije o stilu rukovođenja kandidata, motivaciji i relevantnim osobinama ličnosti za konkretan posao. Obično takve baterije nemaju unapred poznate metrijske karakteristike, jer baterija nije standardizovana, već se karakteristike izračunavaju tek nakon primene. Moguće je da su poznate metrijske karakteristike pojedinačnih testova, kada se oni zadaju samostalno, ali ne i kada se zadaju u okviru baterije, što nas opet dovodi do toga da karakteristike baterije nisu unapred poznate. Podsetiću, svaka promena u testu, uključujući redosled zadavanja testova, broj

testova koji se zadaje odjednom i sl., sa stanovišta psihometrije predstavlja nov test čije se karakteristike trebaju posebno izračunati.

Odnos psihološkog testiranja i procene

Psihološko testiranje predstavlja postupak primene i skorovanja psiholoških testova, kao i interpretacije testovnih rezultata. Kao što je istaknuto prilikom određenja testa, test predstavlja alat/instrument koji pomaže u donošenju odluka. Test je samo jedan od alata koji pomažu u donošenju odluke. Pored testa, u donošenju odluke mogu se primeniti i drugi izvori podataka, dobijeni putem biografskih podataka, anamneze, intervjua, posmatranja, procene od strane drugih i slično. Kada na osnovu svih podataka donosimo odluku o osobi, grupi ljudi ili programu, mi, u stvari, vršimo psihološku procenu (eng. *psychological assessment*). Psihološka procena je širi pojam od psihološkog testiranja. **Psihološka procena** predstavlja proces donošenja suda, u kojem se sud donosi na osnovu sinteze širokog spektra informacija među kojima su najčešće i rezultati psihološkog testa (Domino & Domino, 2006). U Standardima se navodi da je „procena širi pojam koji se obično odnosi na proces u kojem se integrišu informacije prikupljene primenom testova s onima iz drugih izvora (npr. podaci iz društvene, obrazovne, radne ili psihološke istorije pojedinca)" (AERA, APA, & NCME, 2006, str. 24). Koen i Sverdlik (Cohen & Swerdlik, 1999) psihološko testiranje definišu kao administraciju i skorovanje testova i tumačenje rezultata psiholoških testova, dok psihološku procenu definišu kao objedinjavanje rezultata psiholoških testova i rezultata dobijenih drugim metodama. Prema nekim autorima (Walsh & Betz, 2001) na psihološku procenu se gleda kao na proces rešavanja problema, koji uključuje sve etape ovog procesa – definiciju problema, prikupljanje podataka (informacija), njihovo razumevanje i,

na kraju, rešavanje problema. Za razliku od procesa psihološkog testiranja, proces psihološke procene je fleksibilniji i nije standardizovan (Maloney & Ward, 1976). Naime, psihološka procena nije standardizovana jer poslednji korak uključuje sintezu i donošenje odluke na osnovu ekspertize, stručnosti i iskustva onoga ko vrši psihološku procenu, te se tako može desiti da dva stručnjaka, na osnovu istih podataka, donesu različite odluke. S druge strane, testiranje može, a i ne mora biti standardizovano, ali se obično u psihološkoj proceni koriste standardizovani testovi. Može se zaključiti da su testovi koji se koriste u okviru psihološke procene najčešće standardizovani, ali sam proces psihološke procene sa svim svojim koracima nije standardizovan.

Psihološka procena obuhvata nekoliko koraka:

1. identifikaciju cilja procene, koji mora biti jasan, realističan i u skladu sa mogućnostima;
2. selekciju metoda prikupljanja podataka;
3. administraciju i prikupljanje podataka;
4. skorovanje testova i ocenjivanje drugih primenjenih metoda prikupljanja podataka;
5. interpretaciju rezultata testova i rezultata prikupljenih drugim metodama;
6. prosuđivanje, tj. sintezu/integraciju rezultata koji su dobijeni različitim metodama i donošenje argumentovane odluke u skladu sa ciljem procene.

Najvažniji korak psihološke procene je identifikacija cilja. Bez jasno postavljenog cilja, i realističnog koji je u skladu sa mogućnostima, donošenje odluke na osnovu procene neće biti validno. Svi koraci psihološke procene podrazumevaju poznavanje barem osnovnih psihometrijskih načela, osim prvog i poslednjeg koraka. Ova dva koraka

prevazilaze psihometrijsku ekspertizu i zahtevaju ekspertizu iz oblasti u kojoj je postavljen cilj procene, npr. u oblasti zdravstvene nege, psihopatologije, psihologije organizacije, kriminologije i slično.

Koraci koji su zajednički psihološkom testiranju i proceni su administracija i prikupljanje podataka (3. korak), skorovanje testa (4. korak) i interpretacija testovnih rezultata (5. korak). Prvi korak, koji se odnosi na identifikaciju cilja, postoji takođe i u psihološkom testiranju, ali je on uže definisan nego u psihološkoj proceni. Tako, cilj psihološke procene u okviru socijalnog rada može biti odluka o dodeljivanju starateljstva, a cilj testiranja utvrđivanje afektivnog stila vezanosti, mentalnih sposobnosti, profila ličnosti itd. Psihološko testiranje, takođe, obuhvata i korak koji se odnosi na selekciju testova, za koji bi se moglo reći da odgovara koraku selekcije metoda prikupljanja podataka (2. korak) u psihološkoj proceni. Međutim, selekcija u okviru psihološkog testiranja je ograničena na izbor među dostupnim testovima, dok su u psihološkoj proceni na raspolaganju ne samo različiti testovi, već i različite metode prikupljanja podataka. Ključna razlika između psihološkog testiranja i procene je u poslednjem koraku, tj. prosuđivanju. Ovaj korak nije sadržan u procesu psihološkog testiranja i predstavlja glavnu karakteristiku psihološke procene. Sumiran prikaz razlika između psihološkog testiranja i psihološke procene dat je u Tabeli 1.

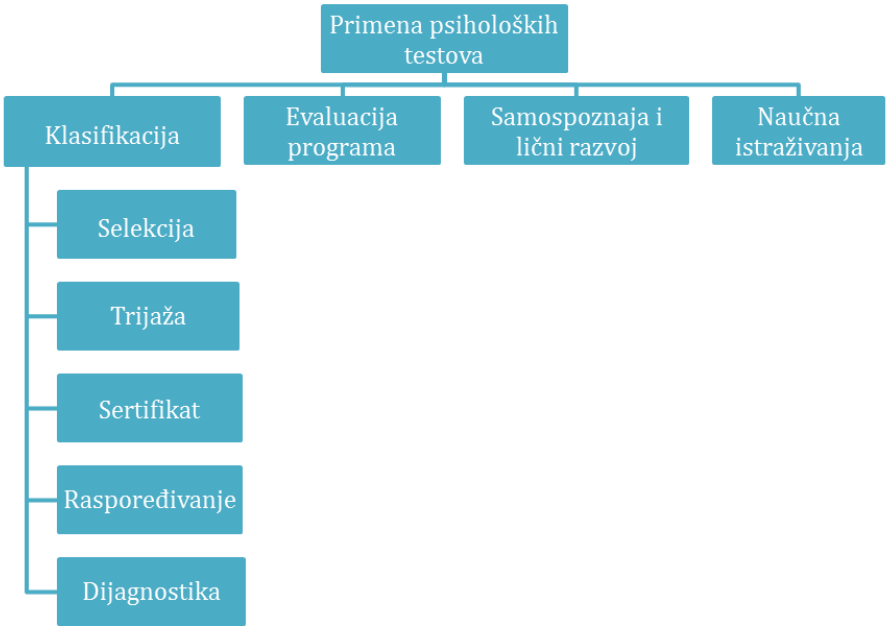
Tabela 1

Razlike psihološkog testiranja i psihološke procene

	Testiranje	Procena
Složenost	jednostavnije; uključuje jednu uniformnu proceduru i jedan aspekt procene (npr. samo ličnost, samo sposobnost…)	složenije; uključuje različite metode prikupljanja podataka (intervju, posmatranje, testiranje…) kojima se procenjuju različiti aspekti funkcionisanja
Trajanje	kraće (od nekoliko minuta do nekoliko sati)	duže (od nekoliko sati do nekoliko dana i više)
Izvor podataka	jedna osoba, osoba koja radi test	više osoba (porodica, učitelji, prijatelji…)
Fokus	poređenje osobe/grupe sa drugima (nomotetski pristup); rezultat na testu	jedinstvenost osobe/grupe/događaja (idiografski pristup); kako osoba funkcioniše
Kvalifikacije za upotrebu	znanje o testovima i proceduri testiranja	znanje o testovima i testiranju, ali i drugim metodama procene, kao i oblasti u kojoj se vrši procena (psihijatrijski poremećaji, zahtevi posla…)
Procedura	veća objektivnost, kvantifikacija	veća subjektivnost, kvantifikacija je retko moguća
Troškovi i resursi	minimalni, posebno kad se testiranje obavlja grupno	veći, zahteva se angažovanje visokokvalifikovanih profesionalaca; uloga procenjivača je veća
Svrha	prikupljanje podataka u cilju merenja nečega, pomoć u donošenju odluke	donošenje odluke
Strukturisanost	viša	uključuje i strukturisane i nestrukturisane aspekte
Evaluacija rezultata	jednostavnija, zahteva ispitivanje psihometrijskih karakteristika	složenija usled varijabilnosti metoda prikupljanja podataka, šireg cilja…

Napomena: Adaptirano prema Urbina (2004).

Psihološki testovi mogu biti ključni alat u donošenju odluke u psihološkoj proceni, ali nijedna odluka psihološke procene ne sme biti doneta samo na osnovu rezultata testova. Postoji nekoliko razloga za to. Prvi razlog je taj da su rezultati na testovima promenljivi i zavise od konteksta testiranja. Drugi je taj što se testovima ne mogu obuhvatiti svi aspekti funkcionisanja osobe koje je potrebno proceniti zarad donošenja odluke pri psihološkoj proceni. Važno je istaći i da su testovi samo uzorak ponašanja. Ova ograničenja testova treba uvek imati na umu prilikom psihološke procene i donošenja odluke uopšte.



Slika 2. Vrste primene psiholoških testova.

Primena psiholoških testova

Svrha primene psiholoških testova je, kao što je rečeno, pomoć pri donošenju odluka. U literaturi se najčešće sreću četiri osnovne

primene/upotrebe testova, koje je ponudio Kronbah (Cronbach, 1990) – klasifikacija, evaluacija programa, samospoznaja i naučna istraživanja (Slika 2). Ove vrste primena testova doprinose donošenju različitih odluka – o pojedincu, programu ili grupnim tendencijama, sebi samom i o naučnim pitanjima.

Klasifikacija

Klasifikacija se odnosi na donošenje odluke o pojedincu, a u vezi sa tim da li ispunjava neki uslov, da li treba da bude podvrgnut nekom tretmanu ili obuci, da li pripada određenoj dijagnostičkoj kategoriji i slično. Klasifikacija obuhvata pet podvrsta različitih primena.

Selekcija

U okviru selekcije donosi se odluka tipa da-ne ili prošao-pao. Primer ove vrste klasifikacije je donošenje odluke o tome da li će se neko zaposliti, upisati na studije, biti unapređen ili ne i slično. Nekada se selekcija može obaviti u jednom krugu (npr. selekcija kandidata za upis na fakultet), pod uslovom da upis podrazumeva jednokratni kvalifikacioni ispit za kandidate. S druge strane, nekada selekcija podrazumeva višeetapni proces, te se finalna odluka donosi nakon uvida u sve etape selekcije. Adekvatan primer bi bilo postojanje nekoliko krugova selekcije za dobijanje posla (npr. selekcija na osnovu dostavljenog CV-a, na osnovu rezultata testova i, u finalnoj fazi, na osnovu intervjua).

Trijaža

Trijaža predstavlja donošenje brze i grube procene, obično početne odluke o tome koga treba dalje ispitivati. Primer za to su dostavljeni CV-jevi u okviru prijave za posao, na osnovu kojih se vrši

odabir kandidata koji se pozivaju u naredni krug postupka selekcije. Kada je reč o primeni testova, primer je davanje kontrolnih zadataka deci pri upisu u školu, kako bi se utvrdilo koju decu treba poslati na dodatno ispitivanje za utvrđivanje mogućeg zaostajanja u razvoju. Ishod ove vrste klasifikacije, kao i selekcije, može biti donošenje odluke tipa da-ne, ali u zavisnosti od situacije, dalje ispitivanje se može vršiti nad onima koji su "prošli" ili onima koji su "pali". Ishod može biti i, na primer, postavljanje preliminarne dijagnoze, mada se i ova odluka može shvatiti kao da-ne tip odluke (ima datu dijagnozu ili nema). S obzirom na to da trijaža predstavlja grubu i prvu procenu, u njoj se mogu koristiti kraći testovi, o čemu će kasnije biti više reči. Testovi koji se koriste u ove svrhe obično se nazivaju trijažnim ili skrining testovima (eng. *screening test*).

Sertifikat

Davanje serfitikata podrazumeva klasifikaciju u odnosu na to da li neko zadovoljava postavljeni standard za taj serfitikat. Primeri sertifikata su vozačka dozvola, polaganje ispita, dobijanje licence za obavljanje prakse u određenom području rada i slično. Za razliku od selekcije, gde se po pravilu biraju samo neke od prijavljenih osoba, broj onih koji mogu dobiti sertifikat nije ograničen i zavisi samo od zadovoljenja postavljenih standarda, a ne i od drugih aspekata kao što je broj prijavljenih kandidata, broj slobodnih mesta za posao i slično.

Raspoređivanje

U okviru ove vrste klasifikacije, za svakog pojedinca odlučuje se o tome koji postupak treba primeniti nad njim. Primer za raspoređivanje je donošenje odluke o tome na koji nivo kursa poznavanja engleskog jezika treba rasporediti svakog studenta. Isto tako, imajući u vidu

rezultate testova, neke prijavljene kandidate možemo postaviti na jednu, a druge na drugu radnu poziciju. Za razliku od drugih vrsta klasifikacije, kod raspoređivanja se sa svakom osobom nešto dalje radi.

Dijagnostika

Ova vrsta klasifikacije podrazumeva šire tumačenje kategorije kojoj neko pripada. Prilikom donošenja odluke u cilju dijagnostike, nije dovoljno samo dati naziv dijagnoze, već se odluka mora obrazložiti i mora se dati predlog tretmana. Ova vrsta klasifikacije je posebno česta u praksi kliničkih i školskih psihologa.

Navedene vrste klasifikacije se međusobno ne isključuju, već se mogu koristiti u okviru istog procesa donošenja odluke. Na primer, prilikom odabira kandidata za posao, najpre se može uraditi trijaža na osnovu CV-ja, pa potom selekcija na osnovu rezultata na testovima i intervjua, i na kraju se može vršiti raspoređivanje na određeno radno mesto.

Kada se test koristi za donošenje odluke o pojedincu, trebalo bi uzeti u obzir kontekst u kojem se odluka donosi, ograničenja testa i druge izvore informacija o ispitaniku, a koji nisu prikupljeni testom. Drugim rečima, kada se donosi odluka o pojedincu, preporučljivo je vršiti psihološku procenu. Nažalost, u praksi je čest primer da se odluka o pojedincu donosi samo na osnovu rezultata testa. Moglo bi se zaključiti da je u takvim situacijama testovima data velika odgovornost. U odnosu na ozbiljnost ishoda i posledica, testovi se dele na one s "**niskim ulogom**" (eng. *low-stakes test*) i "**visokim ulogom**" (eng. *high-stakes test*, više u npr. Harlen, 1994). Razlika između njih je u funkciji koju imaju, odnosno u tome kako će se testovni rezultati koristiti. Ukoliko se oni koriste u cilju donošenja važne odluke o pojedincu, programu,

ustanovi ili zajednici, onda pripadaju grupi testova s visokim ulogom. Primeri takvih testova su maturski ispit za upis u srednju školu i kvalifikacioni ispit za upis na studije, testovi za dobijanje stipendije, licence za rad, posla i sl. Rezultat testa s visokim ulogom može biti i nešto što se tiče ne samo onog ko rešava test, već i drugih ljudi, kao što su npr. nastavnici i profesori, ili širih krugova, kao što su npr. reputacija škole ili fakulteta i slično. Tako, ukoliko puno učenika jedne škole padne kvalifikacioni ispit, onda reputacija te škole može biti ugrožena. Rezultati ovih testova se koriste za određenje sankcija (novčane kazne, negativni publicitet i sl.), priznanja (nagrade, pozitivni publicitet i sl.) ili stimulisanje napredovanja (povišica, bonus, sponzorisanje i sl.).

Evaluacija programa

Evaluacija programa podrazumeva primenu testova u cilju utvrđivanja efektivnosti određenog programa, tretmana, kursa, akcije i tome slično. Na primer, moguće je primeniti test u cilju utvrđivanja učinka određenog programa prevencije vršnjačkog nasilja, psihoterapeutskog tretmana, novog medikamenta, zadovoljstva arhitektonskim rešenjem studentskog kampusa, dizajnom sajta i dr. Rezultati testiranja u okviru ove kategorije primene testova mogu ukazati na to koji aspekti nekog programa se mogu poboljšati i korigovati kako bi ispunili predviđeni cilj. Ovakav način primene testova zahteva longitudinalno ispitivanje u minimalno dve vremenske tačke – pre i nakon primene određenog programa. Tek upoređivanjem rezultata pre i posle primenjenog programa, može se zasigurno znati da li je programom postignut očekivani ishod ili nije, da li je postignut samo u nekim aspektima ili u celosti.

Krajnji korisnici rezultata testiranja najčešće su organizatori programa, ali evaluaciju programa može zatražiti i neko spoljno javno telo. Na primer, rezultate evaluacije nekog fakultetskog nastavnog kursa može tražiti nastavnik koji drži kurs, ali i Komisija za obezbeđivanje kvaliteta na fakultetu ili Komisija za akreditaciju i proveru kvaliteta na nivou države.

Karakteristika primene testova u cilju evaluacije programa je što se na ovaj način ocenjuje grupna tendencija, tj. donosi se odluka na osnovu rezultata testiranja grupe ljudi. Drugim rečima, efektivnost nekog programa (npr. psihoterapeutskog tretmana) ne može se utvrditi na osnovu jednog ili nekoliko klijenata. Međutim, postoje izvesne razlike u odnosu na primenu testova u naučne svrhe, koju takođe karakteriše donošenje odluke o grupnim tendencijama. Kada je reč o evaluaciji programa, u fokusu je sam program i efekti koje program postiže. Kada je reč o primeni testova u naučnim istraživanjima, fokus je na problemu i ciljevima istraživanja koji mogu biti raznovrsni. Desiće se da je cilj istraživanja upravo usmeren na evaluaciju nekog programa, te je razliku između ove dve vrste primena teže odrediti. Neki logičan sled bi bio da se istraživanjima najpre utvrdi efektivnost nekog programa, pa da se program ponudi praktičarima i tržištu generalno. Pri tome, istraživanje s ciljem evaluacije programa bi trebalo da zadovolji sve metodološke uslove za evaluaciju programa, kao što su npr. obezbeđivanje kontrolne grupe koja neće prolaziti ni kroz kakav program, i obezbeđivanje alternativne grupe koja će prolaziti kroz neki drugi, pažljivo odabrani program. Ovi metodološki zahtevi se postavljaju prilikom primene testova u naučnim istraživanjima, ali se ne postavljaju prilikom primene testova u cilju evaluacije programa. To bi mogla biti jedna od razlika između ove dve primene, kada one imaju isti cilj. Druga razlika može biti

u vrsti publikacije u kojoj se objavljuju rezultati u vezi sa evaluacijom programa. Kada se testovi primenjuju u naučnim istraživanjima, onda se publikuju u naučnim publikacijama – naučnim časopisima, zbornicima, monografijama i slično. Kada se testovi primenjuju za evaluaciju programa, rezultati ne moraju biti objavljeni u naučnim publikacijama. Najčešće su oni deo priručnika za primenu datog programa, pri čemu priručnik ne mora imati naučni karakter. Često rezultati evaluacije ostanu na nivou interne evaluacije neke ustanove. Autori i organizatori programa bi, pored uputstva za primenu programa, trebalo da ponude i način evalucije programa i komunikacije sa organizatorima u vezi s evaluacijom programa. Tako bi oni koji primenjuju program mogli podatke o evaluaciji da šalju organizatorima progama, u cilju unapređenja programa.

U situacijama kada se testovi koriste za ispitivanje grupnih tendencija, postoji nekoliko saveta u vezi sa izborom testa. Ovde ću samo naglasiti da je u slučaju ispitivanja grupnih tendencija manji fokus na grešci merenja, pa se mogu odabrati kraći i manje pouzdani testovi, ukoliko nema uslova za primenu dužih testova.

Samospoznaja i lični razvoj

Informacije koje se dobijaju u okviru ove vrste primene testova služe za donošenje odluke o sebi. Dakle, krajnji korisnik rezultata testa je sam ispitanik/klijent. Takve informacije mogu već biti poznate ispitaniku, ali za koje ne postoji i formalni dokaz. Na primer, kada student želi da aplicira za neki master program, potrebno je da dostavi uverenje o poznavanju, na primer, engleskog jezika na određenom nivou. Iako student zna na kom nivou poznaje jezik, za prijavu mu je potrebna potvrda o tome, te se podvrgava testiranju znanja jezika kako

bi dobio sertifikat o tome. Međutim, češće su situacije kada informacije o samom sebi ispitaniku nisu sasvim poznate. Najjednostavniji primer primene testova u svrhu samospoznaje je profesionalna orijentacija. Savetnik za profesionalnu orijentaciju sa klijentom/ispitanikom diskutuje o njegovim rezultatima na testovima, svestan da konačnu odluku o budućem zanimanju treba da donese sam ispitanik. Na isti način, savetnik za razvoj karijere može sa zaposlenima da analizira rezultate na primenjenim testovima, u cilju biranja najoptimalnijeg načina za ostvarivanje željene profesionalne uloge. U svrhe samospoznaje koriste se i testovi kojima se proverava stepen ovladavanja nekim sadržajem ili veštinama. Ovi testovi mogu imati i motivacioni karakter. Naime, pružanje informacija o razvoju određenih sposobnosti i veština ili o određenim potencijalima ličnosti, mogu ispitanika motivisati da neki aspekt unapredi ili neki potencijal koji ima više iskoristi u datim okolnostima. Naravno, posebno motivišuće je to što ispitanik/klijent, u saradnji sa savetnikom ili terapeutom, dolazi do saznanja o sebi. Na taj način se pretpostavlja da će i ispitanik motivisanije pristupiti testovnom materijalu.

Sedamdesetih godina 20. veka zaživela je upotreba testova, u skladu sa tada razvijenom humanističkom i egzistencijalističkom strujom. Praksa primene testova u cilju samospoznaje i ličnog razvoja bila je osnova za kreiranje terapeutskog modela procene. Počeci ideje ovog pravca mogu se naći u radovima Konstanse Fišer koja je definisala tzv. saradničku procenu (eng. *collaborative assessment*) kao procenu u kojoj klijent i terapeut zajedno rade na uzajamno postavljenim ciljevima, dele informacije i razmenjuju utiske (Fischer, 1985/1994). Najzaslužniji za razvoj ovog pravca je Fin, koji je 1996. uobličio metod saradničke procene, tj. terapeutske procene (eng. *therapeutic assessment – TA*).

Prvu empirijsku podršku ovom tretmanu dali su Fin i Tonsager (Finn & Tonsager, 1992) kada su objavili rad u kojem su pokazali prednost saradničke procene. Naime, oni su sproveli studiju u kojoj su imali random raspoređene studente u dve grupe, pri čemu su svi studenti bili oni koji su čekali na tretman u okviru univerzitetskog centra za savetovanje. Prva grupa studenata (n = 32) je prošla kroz dve seanse i radila je Minesota višefazni inventar ličnosti – 2 (Minnesota Multiphasic Personality Inventory – MMPI-2), na osnovu kojeg im je dat fidbek u skladu sa metodama saradničke procene. Druga grupa studenata (n = 29) je dobila podjednaku kliničku pažnju, ali je fokus u seansama bio na aktuelnim brigama i problemima – nije rađen test, niti procena. Podaci o prisustvu simptoma i nivou samopouzdanja prikupljeni su pre i nakon tretmana. Rezultati su pokazali da je grupa u kojoj je primenjena saradnička procena ostvarila značajno niže rezultate na simptomima, a više na samopoštovanju u odnosu na početni nivo. S druge strane, studenti koji su prošli kroz "tradicionalni" tretman nisu pokazali značajne promene u odnosu na početni nivo. Ova studija bila je očetak ukazivanja na to da se testiranje i procena mogu koristiti u terapeutske svrhe. Naredne studije ponovile su ove rezultate, uz bolju kontrolu uslova i kontrolnog tretmana, kao i uz varijacije kontrolnog tretmana (npr. Newman & Greenway, 1997), ali je svakako najbolji dokaz učinka ovog tretmana meta-analitička studija Postona i Hensona (Poston & Hanson, 2010). Oni su identifikovali 17 studija (na ukupno 1.496 ispitnika) u kojima je psihološka procena upotrebljena kao terapeutska intervencija, pri čemu je u većini studija ispitivan terapeutski efekat fidbeka na osnovu psihološke procene. Rezultati su

pokazali da je veličina efekta umerena (0,42[2]). Iako je veličina efekta umerena, zaključak je da oni koji vrše procenu u svom poslu na uobičajeni način mogu propustiti priliku da ostvare pozitivne efekte po klijenta i njegovu promenu ukoliko propuste da na adekvatan način obrade rezultate testiranja sa klijentom.

Prva prednost ovog tretmana je u tome što ispitanik menja sliku o sebi, što otvara nove mogućnosti u njihovim životima. Na primer, neko može misliti za sebe da je glup, ali kroz saradničku procenu može uvideti da to nije tačno, već možda ima problema u organizaciji učenja. Druga prednost je u tome što se zajedničkim postavljanjem cilja procene može redukovati napetost i distres koji proizilazi iz disbalansa verifikacije slike o sebi i uvida o sebi. Treća prednost je u tome što samo trestiranje i procena stvaraju pogodno tlo za ispoljavanje empatije prema klijentu. Fin (Finn, 2007) testove slikovito opisuje kao "magnete za empatiju". Kada se u procenu uključe i druge osobe bliske klijentu, tu se svakako stvaraju dodatne mogućnosti za rast i promene.

Naučna istraživanja

Informacije koje se dobijaju u okviru ove vrste primene testova služe za donošenje odluke o konstruktima, teorijskim modelima i naučnim pretpostavkama. Odluka se donosi na osnovu grupne tendencije. Izuzetak su istraživački nacrti s jednim ispitanikom, kao što je studija slučaja, ali i ovi nacrti služe razumevanju teorija i izvođenju predikcija o budućem ponašanju.

[2] Veličina efekta je izražena kao Koenov *d* čije vrednosti 0,2 ukazuju na malu, 0,5 na umerenu ili srednju i 0,8 na veliku veličinu efekta (Cohen, 1988), a prema novijem stanovištu vrednosti 0,41 i više se mogu smatrati "praktično" značajnim vrednostima u društvenim naukama (Ferguson, 2009).

U istraživanjima se koriste kako standardizovani tako i nestandardizovani testovi. Među nestandardizovanim testovima nalaze se oni koji su konstruisani za određene istraživačke ciljeve ili za primenu samo u istraživačke svrhe. Kako je u istraživanjima najčešće fokus na grupnim tendencijama, manji je fokus na grešci merenja, pa se mogu koristiti kraći i manje pouzdani testovi. Na primer, postoje kraće verzije inventara ličnosti čija je primena ograničena na grubu procenu osobina ličnosti. Jedan takav instrument je kraća verzija Inventara Velikih pet koja sadrži 10 ajtema (Big Five Inventory-10 – BFI-10: Rammstedt & John, 2007) u kojem je svaka skala, tj. dimenzija ličnosti operacionalizovana samo sa dva ajtema. Zbog malog broja ajtema, skale kratkih verzija testova imaju nižu pouzdanost, ali je za njihovu upotrebu važno da poseduju zadovoljavajuću validnost. Oni se primenjuju kada procena ličnosti (ili koji je već predmet merenja u pitanju) nije glavni predmet istraživanja, kada je ispitivanje vremenski ograničeno, kada set zadatih instrumenata sadrži veliki broj instrumenata (kada je set zadatih instrumenata obiman). Takođe, prednost kratkih verzija instrumenata je u tome što olakšavaju procenu u istraživanjima s ponovljenim merenjima i olakšavaju situaciju kada ispitanik treba da proceni veći broj osoba, npr. drugove iz odeljenja. Nestandardizovanih testova je mnogo više nego standardizovanih, i prilikom njihovog prvog objavljivanja, oni bi svakako morali proći isti postupak psihometrijske validacije kao i standardizovani testovi. S obzirom na činjenicu da primena testova u istraživanjima ne uključuje donošenje odluke o pojedincima i programima, tj. praktične posledice su retke, njihova primena u ovom kontekstu je manje sporna. Drugim rečima, testovi koji se primenjuju u istraživanjima su testovi niskog uloga. Ipak, treba imati na umu da primena loših testova u istraživanjima može voditi ka

pogrešnim zaključcima, što se onda odražava i na teorijske postavke, pa posredno i na praksu.

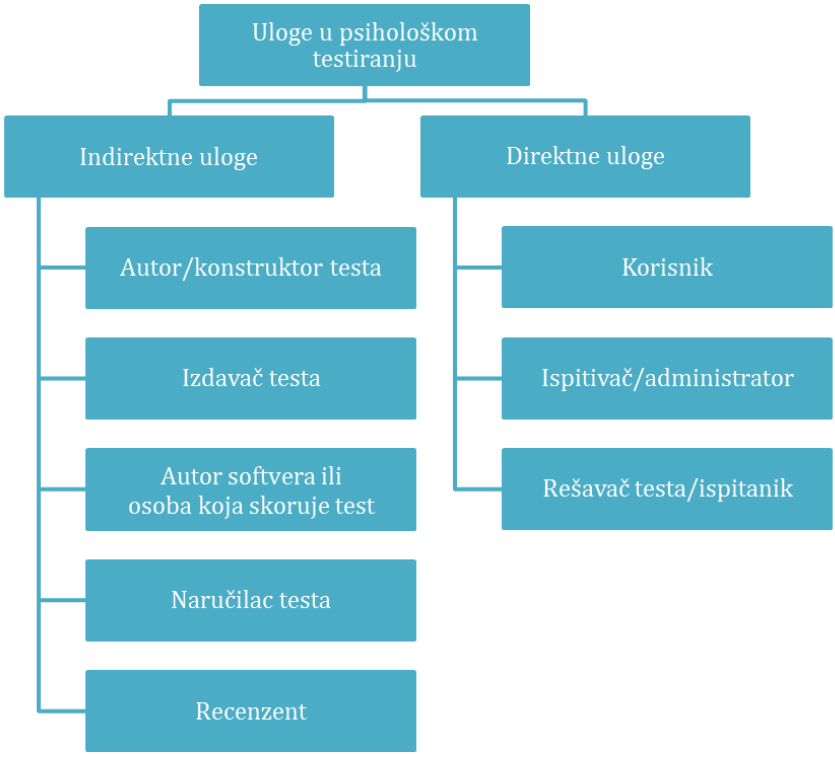
Veoma je važno da razvoj testova prati razvoj nauke. Izbor testova se ne bazira samo na njegovim psihometrijskim kvalitetima, već i na konceptualnoj osnovi na kojoj počiva i na njegovoj moći da predvidi buduće ponašanje. Tako, na primer, nije svejedno koji ćete test ličnosti primeniti i koju vrstu sposobnosti ćete procenjivati u cilju donošenja odluke o zapošljavanju na menadžerskoj poziciji ili na poziciji savetnika. U skladu s teorijskom i empirijskom utemeljenošću znanja o činiocima koji su povezani s uspešnošću na ove dve pozicije, za procenu uspešnosti na menadžerskoj poziciji će vam više informacija pružati upitnik stilova rukovođenja, a za procenu uspešnosti savetnika upitnik empatije. Takođe, klinički test kao što je MMPI vam neće pružati potrebne informacije prilikom selekcije kandidata za posao menadžera, ali hoće prilikom odluke o upućivanju na psihoterapeutski tretman.

Na kraju, treba pomenuti da navedene četiri primene testova nisu međusobno isključive. Već je pomenuto kako primena testova u istraživanjima može biti povezana sa primenom u cilju evaluacije programa. Isto tako, prilikom razvoja nekog testa u cilju klasifikacije, autori najpre sprovedu istraživanje u cilju psihometrijske evaluacije testa, pa onda test ponude praktičarima.

Uloge u psihološkom testiranju

Navedena je i obrazložena činjenica da test predstavlja proizvod koji može biti komercijalan, te da interesi učesnika u procesu testiranja mogu biti različiti. S obzirom na to, neophodno je definisati uloge u procesu testiranja. Osnovna podela uloga vrši se na osnovu toga da li učesnik u procesu testiranja direktno učestvuje u primeni testa ili ne. S

obzirom na to, mogu se razlikovati indirektne i direktne uloge. Obe vrste uloga sadrže po nekoliko specifičnih uloga (Slika 3).



Slika 3. Uloge u psihološkom testiranju.

Indirektne uloge

Indirektne uloge u psihološkom testiranju odnose se na pet specifičnih uloga.

Autor ili konstruktor testa

Ova uloga podrazumeva jednu ili više osoba koje na osnovu sprovedenih istraživanja konstruišu i predlažu test. Test može biti konstruisan na osnovu teorijskog modela ili može biti ateorijski, pri

čemu ateorijski može biti na osnovu postavljenog kriterijuma ili empirijski generisan. Spektor (Spector, 1992) prvi pristup, na osnovu teorijskog modela, naziva induktivnim ili konfirmatornim, a drugi pristup, koji je ateorijski, naziva deduktivnim ili eksplorativnim. Primer testa koji je nastao na osnovu teorijskog modela je Ajzenkov upitnik ličnosti EPQ (Eysenck Personality Questionnaire – EPQ: Eysenck & Eysenck, 1975). EPQ je nastao na osnovu Ajzenkovog PEN modela o tri bazične osobine ličnosti – psihoticizam, ekstraverzija i neuroticizam. Imajući u vidu teorijska očekivanja u vezi s tim koji skup ponašanja je karakterističan za svaku od tri osobine, sačinjene su stavke koje operacionalizuju ta ponašanja. Primer ateorijskog testa nastalog na osnovu kriterijuma je Minesota višefazni inventar ličnosti (Minnesota Multiphasic Personality Inventory – MMPI: Schiele, Baker, & Hathaway, 1943), čija je aktuelna verzija MMPI-2. Početni skup stavki ovog testa odabran je iz do tada postojećih inventara i akumuliranog psihijatrijskog znanja, tj. repertoara iskaza pacijenata. Ajtemi koji su pokazali diskriminativnost između pacijenata s određenim poremećajima i pripadnika opšte populacije, zadržani su u okviru inventara. Dakle, u ovom slučaju nisu unapred postojala teorijska očekivanja u vezi sa sadržajem ponašanja i simptoma koja razlikuju kliničku i opštu populaciju. Primer ateorijskog, empirijski generisanog testa je i Inventar ličnosti Velikih pet plus dva (Smederevac i sar., 2010), koji je nastao na osnovu psiholeksičkih istraživanja na srpskom jeziku. Prema psiholeksičkoj paradigmi pretpostavlja se da su osobine po kojima se ljudi razlikuju, a koje su se kroz svakodnevno iskustvo pokazale kao značajne, zapravo kodirane u jeziku. S obzirom na to, termini koji se koriste za opis ličnosti u nekom jeziku mogu poslužiti kao polazna osnova za ispitivanje bazične strukture ličnosti. Na osnovu tih termina,

mogu se sačiniti ajtemi testa koji se potom podvrgavaju faktorskoj analizi kako bi se dobio broj i ispitala struktura bazičnih dimenzija ličnosti. Dakle, faktori, odnosno izdvojene bazične dimenzije ličnosti nastali su na osnovu empirijskih podataka o strukturi testa, bez unapred definisanog broja i sadržaja tih dimenzija. S obzirom na to da ne postoje pretpostavke o broju faktora i faktorskoj strukturi, za ovaj inventar se kaže da je empirijski generisan. Ispitivanje strukture ovako konstruisanog testa na srpskom jeziku rezultiralo je sedmofaktorskim rešenjem (Smederevac i sar., 2010), dok je u nekim drugim jezicima/kulturama dobijeno pet (npr. Goldberg, 1990) ili šest faktora (npr. Ashton et al., 2004) u zavisnosti od metode selekcije termina.

Iako Spektor (Spector, 1992) daje prednost induktivnom metodu konstrukcije testa, važno je napomenuti da metod konstrukcije testa ne mora zavisiti nužno od toga da li postoji ili ne postoji teorijski okvir. Na primer, ukoliko želimo da konstruišemo test sklonosti ka agresivnom ponašanju, možemo poći od: 1. neke teorije agresivnog ponašanja (npr. frustracione teorije agresije) ili teorijske podele agresivnog ponašanja (npr. po funkciji na reaktivnu i proaktivnu agresiju), 2. uvida u tipične karakteristike koje ispoljavaju npr. osuđeni za neko krivično delo koje uključuje agresiju i nasilje, ili 3. kolekcije ajtema agresivnosti iz postojećih inventara za procenu ličnosti koje bismo podvrgli psihometrijskoj analizi. U slučaju poslednjeg navedenog pristupa, rezultati mogu ići u prilog nekom postojećem teorijskom stanovištvu, ili se na osnovu njih može formirati novo teorijsko stanovište o dispozicijama za agresivno ponašanje. Na osnovu jednog takvog istraživanja u kojem je analiziran skup stavki skala agresivnosti i prijatnosti (kao opozita agresivnosti) iz različitih inventara za procenu ličnosti, izolovana su četiri faktora agresivnosti – bes, osvetoljubivost,

dominacija i hostilnost (Dinić, Mitrović i Smederevac, 2014). Ekstrahovani faktori ne idu u prilog strukturi agresivnosti prema nekom određenom teorijskom okviru, već predstavljaju zajednički bazični sadržaj agresivnosti različitih teorijskih pristupa. Dobijeni faktori su potom poslužili kao polazište za razvoj novog upitnika agresivnosti kojim se operacionalizuju navedena četiri faktora, te mogu dalje poslužiti za razvoj novog modela agresivnosti.

Obično je glavni interes autora testa vezan za davanje doprinosa nauci i unapređenju psihološke prakse. Autor testa donosi odluku o načinu distribuiranju testa – da li će test imati komericijalnu distribuciju ili će se distribuirati slobodno, na druge načine, npr. preko profesionalnih publikacija. Bez obzira na to za koji način distribucije se odluči, autor testa može doprineti upoznavanju šire akademske zajednice sa testom preko objavljenih publikacija o testu.

U Standardima (AERA, APA, & NCME, 2006, str. 258) navodi se da je autor testa „osoba (ili osobe) ili agencija odgovorna za izradu testa, te za dokumentaciju njegovih tehničkih kvaliteta za planiranu svrhu". Dakle, autor testa je u obavezi da transparentno navede sve tehničke, tj. psihometrijske karakteristike testa, te je samim tim i odgovoran za iste. Cilj svakog autora je da test bude kvalitetan. Pored toga, autor testa može doprineti i daljem razvoju testa, čineći revizije, i unapređivanjem psihometrijskih i praktičnih aspekata testa. Iako autor testa preuzima najveću odgovornost za konstrukciju testa i njegove tehničke kvalitete, on bi trebalo da preuzima odgovornost i za dalji razvoj testa. Mada, ukoliko je reč o testu namenjenom komercijalnoj upotrebi, deo odgovornosti i za konstrukciju i za razvoj testa preuzima i izdavač testa.

Izdavač testa. Izdavač testa je uobičajeno udruženje ili izdavačka kuća koja test objavljuje, daje na tržište i obavlja i kontroliše

njegovu distribuciju i prodaju. Primer takvog udruženja kod nas je Društvo psihologa Srbije, koje je 1992. godine osnovalo Centar za primenjenu psihologiju koji se, između ostalog, bavi izradom i distribucijom psiholoških mernih instrumenata[3]. Primeri najpoznatijih izdavačkih kuća koje publikuju psihološke testove su PEARSON[4] i Guilford Press[5], a kod nas postoji samo jedna takva izdavačka kuća – Sinapsa edicije[6].

Izdavač testa bavi se njegovom promocijom i uglavnom mu je glavni interes profit. Stoga je u interesu izdavača da ponudi kvalitetan proizvod. Naime, izdavač testa bi trebalo da se pobrine da svi preduslovi za dobijanje što tačnijih psihometrijskih karakteristika budu obezbeđeni – da se obezbedi adekvatan i dovoljno velik uzorak za izračunavanje normi, da se test zadaje u odgovarajućim uslovima od strane stručnih osoba i slično. Ukoliko se ispostavi da test nema zadovoljavajuće psihometrijske karakteristike, izdavač ne bi trebalo da publikuje i promoviše takav test, ali može uložiti resurse za njegovu reviziju koja bi doprinela poboljšanju njegovog kvaliteta. Izdavači obezbeđuju procenu kvaliteta testa pre samog objavljivanja tako što daju test na recenziranje. Pored uloge u obezbeđivanju kvaliteta testa, izdavač testa bi trebalo da se bavi i održavanjem kvaliteta testa (izdavač ima odgovornost i za razvoj datog testa). To se postiže unapređenjem i revizijama testa, kao i restandardizacijom i izradom novih normi, za koje se preporučuje da ne budu starije od 10 godina. Izdavač bi trebalo da obezbedi uslove za

[3] http://www.dps.org.rs/merni-instrumenti
[4] http://www.pearsonclinical.com/
[5] http://www.guilford.com/browse/assessment-scales
[6] http://www.sinapsaedicije.rs/psihodijagnosticka-sredstva

sprovođenje restandardizacije i revizije testa ukoliko se pokaže da je potrebno.

Na ovom mestu treba samo napomenuti da i autor može profitirati od izdavanja testa, što se definiše ugovorom između izdavača i autora. Ugovorom se može definisati da autor dobije jednokratni honorar ili honorar na godišnjem nivou, ili procenat od prodaje testa (uobičajeno do 15%, ali to zavisi od pregovora).

Autor softvera ili osoba koja skoruje test

Ova uloga podrazumeva autora softvera za primenu i skorovanje testa, a nekad i za generisanje interpretacije rezultata testa na osnovu informacija koje je dobio od autora testa. Ukoliko primena i/ili skorovanje nisu kompjuterski podržani, ova uloga bi bila ekvivalentna ulozi osobe koja skoruje test i testovni rezultat prevodi u predložene norme. Pri tome, ova uloga ne podrazumeva poznavanje predmeta testa i referentnog okvira na osnovu kojeg je nastao test, niti način interpretacije, već samo tehničko poznavanje skorovanja i prevođenja skorova u norme.

Naručilac testa

Naručilac testa može biti neka organizacija koja ugovorom naručuje konstrukciju testa ili uslugu testiranja. Takva organizacija može biti odbor koji predstavlja određenu ustanovu ili državnu agenciju, kompanija/firma, izdavačka kuća, neko udruženje ili drugo javno telo. Razlog naručivanja testa može biti prepoznata potreba za procenom nekog fenomena za koji ne postoji standardizovani test, ili ne postoji u datoj državi. Mogu se, takođe, naručivati standardizacije postojećih testova, s ciljem dobijanja lokalnih ili nacionalnih normi, kao i restandardizacije postojećih normi. Pored toga, mogu se naručivati i

usluge testiranja. Primer naručivanja usluge testiranja je kada neka kompanija želi da sprovede profesionalnu selekciju kandidata za odabir najboljih kandidata za zapošljavanje na određenu radnu poziciju. U te svrhe, kompanija angažuje određenu HR agenciju koja pruža usluge profesionalne selekcije i procene kandidata. Odabrana HR agencija sprovodi testiranje i na osnovu rezultata testiranja vrši selekciju kandidata i predlaže naručiocu finalni izbor. Naručilac testiranja može biti i stručnjak koji zaključak o osobi ili grupi ljudi želi da donese (i) na osnovu psihološke procene. Primer ovakve prakse je kada se počinilac krivičnog dela od strane pravnih lica uputi na psihološku procenu zarad utvrđivanja stepena uračunljivosti pri izvršenju krivičnog dela.

Recenzent

Ova uloga podrazumeva kompetetnog stručnjaka koji, na osnovu svog znanja i ekspertize, procenjuje kvalitet testa. Ne postoje posebno propisani uslovi da bi neko bio recenzent, već se on bira na osnovu svoje reputacije (na osnovu kvaliteta objavljenih radova iz određene oblasti, citiranosti, učešća u pisanju pravilnika i zakona o psihološkoj delatnosti, iskustvu u praksi i zalaganju za unapređenje prakse itd.). Recenzija mora biti nepristrasna, a recenzent ne sme biti u sukobu interesa. Nepristrasnost recenzije odnosi se na to da recenzent, bez obzira na to kakvo teorijsko stanovište zastupa i da li poznaje autora testa, može objektivno da sagleda kvalitet testa i ukaže na sve prednosti i mane. Često recenzije testova nisu slepe, te recenzenti dobiju informaciju o tome ko je autor testa. Kako bi se nepristrasnost očuvala, trebalo bi da recenzent i autor nisu u saradničkom odnosu, tj. da nemaju zajedničke publikacije, ali treba voditi računa i o drugim aspektima sukoba

interesa, npr. da recenzent nije autor konkurentskog instrumenta ili član Komisije koja odobrava instrumente i slično.

Procena kvaliteta testa vrši se u odnosu na dva aspekta: tehnički aspekt, koji se odnosi na procenu metrijskih karakteristika, načina zadavanja i skorovanja testa i sl., i praktični aspekt koji podrazumeva ispunjenost svrhe i namene testa. S obzirom na to da se dva aspekta testa ocenjuju, praksa je obično da se za svaki test dobiju dve nezavisne recenzije – jednu daje recenzent iz oblasti primene testa, a drugu recenzent iz oblasti psihometrije ili metodologije. Kao što je pomenuto, izdavačka kuća može obezbediti recenzenta, zarad interne provere. No, bez obzira na to da li izdavačka kuća obezbeđuje recenzenta ili ne, kod nas, svaki novi test pre puštanja u komercijalnu prodaju, mora da prođe dve nezavisne recenzije koje obezbeđuje Institut za psihologiju u Beogradu, koji daje licencu za testove[7].

Pored institucije koja daje licencu za testove, recenziju može zahtevati i neko udruženje ili organizacija, mada kod nas takva praksa još nije ustanovljena. Recenzije mogu biti i javno dostupne. Najveću bazu recenzija testova ima Buros centar. To je nezavisna neprofitna organizacija u okviru Univerziteta u Nebraskoj – Linkoln, tačnije, Departmana za psihologiju obrazovanja na Koledžu za obrazovanje i humanističke nauke (više na str. 60). Ovakve publikacije su veoma korisne jer pomažu korisnicima da procene kvalitet testa i odluče da li vredi uložiti resurse u nabavljanje testa i edukaciju u vezi s njegovom primenom.

[7] Više o zakonima i pravilnicima može se videti na: http://dps.org.rs/o-drustvu-psihologa/zakoni-i-pravilnici

Direktne uloge

Direktne uloge u psihološkom testiranju obuhvataju tri specifične uloge – korisnika testa, ispitivača i ispitanika.

Korisnik testa

Korisnici testa su zasigurno najbrojniji među ulogama u procesu testiranja. Korisnik testa procenjuje adekvatnost i korisnost testa i donosi odluku o tome koji test će se primeniti za određenu svrhu testiranja. Pored toga, korisnik testa interpretira testovne rezultate. Samim tim, on mora da zna kako da protumači testovne rezultate, i kako da saopšti i predstavi rezultate drugima, npr. ispitaniku, naručiocu testiranja i slično. Onaj ko potražuje testovne rezultate, može biti sam korisnik, ali i naručilac testa ili ispitanik. Najčešće je sam korisnik ujedno i naručilac rezultata testiranja. Tada korisnik testa upotrebljava testovne rezultate za donošenje odluka i zaključaka o osobi, grupi ljudi ili programu. Primer za to je kada psiholog u okviru kliničke prakse obavi testiranje zarad donošenja odluke o tretmanu pacijenta.

Međutim, neretko će se dešavati da korisnik testa treba nekoj drugoj osobi ili organizaciji (naručiocima testa) da dostavi rezultate testiranja i njihovo tumačenje. U toj situaciji, korisnik testa može da sugeriše kakav bi ishod, odnosno odluka bila najbolja, ali odluku ne donosi korisnik, već naručilac testa. Već je spomenuta situaciju kada naručilac testa (npr. neka kompanija), od korisnika testa (npr. zaposlenih u HR agenciji) naruči testiranje u cilju profesionalne selekcije i donošenja odluka o tome koje kandidate treba zaposliti na određenu radnu poziciju. Takođe, psiholog u školi može na osnovu rezultata testiranja učenika predložiti direktoru koje učenike treba rasporediti u

specijalno odeljenje s talentovanim učenicima npr. za matematiku, strani jezik i slično.

Naručilac rezultata testiranja može biti i sam ispitanik. Ovo je najčešće slučaj u individualnoj proceni u okviru savetovanja i psihoterapije, kao i u okviru profesionalne orijentacije. Primer za to je i kada HR menadžer u svojoj kompaniji prati razvoj karijere zaposlenih, i na osnovu procene, koja uključuje i testiranje, zajedno sa zaposlenim razmatra u kom pravcu je najbolje da zaposleni razvija svoju karijeru.

Pre upotrebe testa, test se može evaluirati u odnosu na njegove tehničke kvalitete kao što su metrijske karakteristike, opremljenost i slično. Kada se test razmatra za potencijalnu primenu, primarni interes je za njegove tehničke kvalitete. Za ovaj aspekt su najodgovorniji autori testa, ali i sam izdavač. Čim se test počne primenjivati, primarni interes se pomera na veštine i kompetencije korisnika testa, kao i na način i svrhu primene testa. Drugim rečima, kada se stavi u upotrebu, test se ne može evaluirati odvojeno od načina korišćenja. U tom smislu, korisnik testa ima najveću odgovornost za upotrebu i primenu testova. U Standardima (AERA, APA, & NCME, 2006, str. 264) se navodi da je korisnik testa „osoba (ili osobe), ustanova ili agencija odgovorna za odabir i primenu testa, interpretaciju testovnih rezultata dobijenih u datom kontekstu, te za bilo kakve odluke ili akcije koje se delom baziraju na testovnim rezultatima". Treba napomenuti da osoba koja samo zadaje ili skoruje test nema ovakve odgovornosti kao korisnik testa i ne smatra se korisnikom testa (DeMers et al., 2000), mada se često dešava da korisnik testa ujedno ima i ulogu onoga ko zadaje i skoruje test.

Ispitivač ili administrator testa

Ova uloga podrazumeva osobu koja zadaje test i obuhvata obezbeđivanje adekvatnih uslova za testiranje, pripremu i podelu materijala za testiranje, davanja uputstva, primera i vežbe za rešavanje i popunjavanje testa, nadziranje rešavanja i popunjavanja testa, poštovanje vremenskog ograničenja ukoliko je to testom predviđeno, uredno prikupljanje urađenih testova. Administrator testa ima odgovornost za sprovođenje testiranja, stoga on mora poštovati standardizovanost procedure testiranja, kako bi se obezbedila adekvatna interpretacija testovnih skorova. Dakle, iako administrator testa ne mora nužno da zna teorijsku osnovu testa i sam predmet merenja, mora da zna šta znači standardizacija i kako se ona ostvaruje. Drugim rečima, neophodno je da poseduje bar neka bazična znanja iz psihometrije. Pored navedenog, administrator testa ima još jedan zadatak, a to je obezbeđivanje adekvatne atmosfere testiranja.

Ispitanik ili rešavač testa

Ispitanik je osoba koja se podvrgava testiranju na osnovu sopstvenog izbora, uputa ili zahteva. U slučaju upotrebe psiholoških testova za samospoznaju i lični razvoj, ispitanik se najčešće sam odlučuje da se podvrgne testiranju. U ovom slučaju ispitanik je krajnji korisnik kojem su potrebni testovni rezultati i interpretacija istih. Kao što je već navedeno, primer ovakve situacije je savetovanje za profesionalnu orijentaciju. U drugoj situaciji, ispitanik može dobiti uput za testiranje u cilju neke klasifikacije. Primer za to je kada doktor opšte medicine izda uput pacijentu za psihologa u cilju testiranja na organicitet.

Navedene uloge su često isprepletene i dešava se da jedna osoba može imati više uloga. Već je spomenuto da ispitanik može biti i naručilac

testa u slučaju primene testova u okviru samospoznaje. Takođe, izdavač isto tako može biti naručilac testa ukoliko traži potencijalne autore za konstruisanje testa ili reviziju postojećeg testa. Često je autor testa ujedno i administrator, osoba koja skoruje test i korisnik, kada sprovodi sopstveno istraživanje, a najčešće u praksi je korisnik testa ujedno i administrator i i osoba koja skoruje test. Iako zvuči paradoksalno, autor testa takođe može biti korisnik u slučaju kada svoja autorska prava na test prepiše izdavačkoj kući. Tada autor čak mora da otkupi svoje testove ukoliko želi da ih zadaje, što zavisi od uslova korišćenja testa od strane autora koji su definisani ugovorom između autora i izdavačke kuće.

Informacije o testovima

Psihološka udruženja su svakako prva adresa za dobijanje validnih informacija o testovima. Na primer, APA ne izdaje testove, ali izdaje publikacije o testovima, npr. „Directory of Unpublished Experimental Mental Measures" (Goldman & Mitchell, 2008). APA, takođe, ima poseban odeljak na internet-stranici „FAQ: Finding Information About Psychological Tests" gde se javnost može informisati o dostupnim testovima, komercijalnim, ali i o testovima u javnom domenu. Takođe, od velikih psiholoških udruženja može se spomenuti i Britanska psihološka asocijacija (British Psychological Society – BPS) koja, takođe, publikuje informacije o testovima. Kod nas, u okviru Društva psihologa Srbije, izdavačkom delatnošću se bavi Centar za primenjenu psihologiju koji objavljuje i testove.

Drugi izvor informisanja o testovima su izdavačke kuće koje objavljuju psihološke testove, o kojima je već bilo reči (npr. Pearson, a kod nas Sinapsa edicije). Postoji Asocijacija izdavača testova (Association of

Tests Publishers – ATP) a na njihovoj internet-stranici može se naći lista izdavača[8].

Treći izvor su posebne publikacije posvećene informacijama o testu, kao što je publikacija „Tests in Print", koju izdaje Buros centar. Ova publikacija obuhvata bibliografske informacije o testovima i predstavlja najobuhvatniji popis testova, ali je ograničena na testove na engleskom i španskom jeziku. Slična publikacija je i „Test Reviews" izdavačke kuće Proed koja datira od 1984. godine.

Postoje i onlajn baze testova koje sadrže i nekomercijalne testove u kojima se nude osnovne informacije o testu, kao što su baza Službe za testiranje u obrazovanju (Test Collection at Education Testing Service – ETC[9]), PsycTESTS[10] u kojoj se mogu naći i ajtemi različitih instrumenata, a postoje i različite bibliotečke baze instrumenata[11]. Izvori testova u ovim bazama su uglavnom iz objavljenih članaka, disertacija i drugih publikacija, ili su informacije dobijene direktno od autora. Naravno, sve bibliografske baze naučnih publikacija mogu služiti za pribavljanje informacija o testovima, ukoliko znamo koji nam je test potreban, ili za koji konstrukt (npr. ERIC). Kod nas ne postoji objedinjena baza instrumenata na srpskom jeziku, ali postoje neki pokušaji u tome. Tako je, u okviru časopisa „Primenjena psihologija" objavljen je popis instrumenata koji su korišćeni u radovima ovog časopisa u periodu od nastanka časopisa 2002. do 2013. godine (Jovin, 2018).

[8] https://www.testpublishers.org/
[9] https://www.ets.org/test_link/about
[10] https://www.apa.org/pubs/databases/psyctests
[11] http://tam.rc.usf.edu/

S popularnošću otvorenog pristupa nauci, mnogi instrumenti postali su dostupni. Možda najveći repozitorijum ajtema (preko 3.000 ajtema) koji predstavljaju simulaciju komercijalnih, ali i instrumenata u javnom domenu, predstavlja International Personality Item Pool – IPIP[12] (Goldberg et al., 2006). Istraživači s Odseka za psihologiju Filozofskog fakulteta u Novom Sadu, pod rukovodstvom prof. dr Snežane Smederevac, su veći deo ovih ajtema preveli na srpski jezik, te postoji i srpska verzija IPIP repozitorijuma.[13]

Međutim, vrlo retko se na navedenim adresama mogu naći recenzije testova. Broj testova koji se nude je sve veći, pa je vrlo teško odlučiti se koji test primeniti u određene svrhe. Stoga su potrebni kritički osvrti i evaluacije ponuđenih testova. Jedan od pokušaja informisanja javnosti o kvalitetu testova koji su u upotrebi predstavlja „Mental Measurement Yearbook" pomenutog Buros centra.

Buros centar

Ideja za osnivanje ovog centra nastala je publikacijom „Godišnjak mentalnih instrumenata" („Mental Measurement Yearbook") Oskara Krisena Burosa 1938. godine. Do tada je ova publikacija sadržala samo bibliografske podatke o testovima, ali od 1938. godine sadrži i recenzije. Naime, Buros je osnovao forum za vodeće naučnike gde bi mogli da objavljuju recenzije postojećih komercijalnih testova, a sa ciljem da se pomogne korisnicima u selekciji testova za njihove potrebe i ciljeve procene. Iako je Buros imao ideju da osnuje organizaciju za pružanje psihometrijskih usluga, nažalost, nije uspeo da nađe potrebna sredstva. Umro je 1978. godine, a njegova udovica, Luela Gubrud Buros, nastavila je da podržava njegovu ideju. Uz njenu pomoć, Buros centar je osnovan 1994. godine. Buros centar je osnovan s ciljem unapređenja testiranja i procene kako u naučne tako i u praktične svrhe, a preko pružanja usluga evaluacije, tj. recenziranja psihometrijskih konsultacija, edukacija u vezi sa testiranjem i procenom, kao i sprovođenja istraživanja s ciljem evaluacije testova. Buros centar je proširio svoje polje delovanja tako da se preko njega ne rade samo evaluacije komercijalnih, već i svih ostalih testova, a pružaju se,

[12] https://ipip.ori.org/
[13] https://www.ipiptesting.ml/

takođe, različite psihometrijske usluge. Buros centar je vodeći izvor informacija o testovima i ima reputaciju stručnih i kvalitetnih recenzija testova. Ovaj centar objavljuje dve ključne edicije. Prva je „Godišnjak mentalnih instrumenata" čije je poslednje, 20. izdanje, objavljeno 2017. i obuhvata preko 190 recenzija novih ili skoro revidiranih testova. Druga je „Štampani testovi" („Tests in Print") čije je poslednje, 9. izdanje, objavljeno 2016. i obuhvata bibliografske informacije o preko 3.000 testova, za sada samo na engleskom i španskom jeziku (više na https://buros.org/).

Pitanja za razmišljanje

1. Čemu služi standardizacija testa?
2. Koje su sličnosti i razlike između psihološke procene i psihološkog testiranja?
3. Koja je razlika između komercijalnih testova i testova u javnom domenu?
4. Navedi primere iz prakse u kojoj se mogu identifikovati različite primene testa u cilju klasifikacije.
5. Navedi situaciju u kojoj jedna osoba ima različite uloge u procesu testiranja.

Reference

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2006). *Standardi za pedagoško i psihološko testiranje*. Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York, NY: Macmillan Publishing Co, Inc.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of

personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356–366.

Beck, A. T., Steer, R. A., Brown, G. K. (1996). *Manual for the Beck Depression Inventory – II.* San Antonio, TX: Psychological Corporation.

Cattell, R. B., Cattell, A. K. S., & Cattell, H. E. P. (1993). *Soxteen Personality Factor Quiesstionnaire* (5. ed). Champaign, IL: Institute for Personality and Ability Testing, Inc.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to test and measurement* (4th ed.). Mountain View, CA: Mayfield.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York, NY: Harper Collins Publisher.

DeMers, S. T., Turner, S. M., Andberg, M., Foote, W., Hough, L., Rey-Casserly, C. M. (2000). *Report of the Task Force on test user qualifications.* Preuzeto sa https://www.apa.org/science/programs/testing/qualifications.pdf

Dinić, B., Mitrović, D. i Smederevac, S. (2014). Upitnik BODH (bes, osvetoljubivost, dominacija, hostilnost): novi upitnik za procenu agresivnosti. *Primenjena psihologija, 7*(3–1), 297–324.

Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction.* New York, NY: Cambridge University Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155–174.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Kent, UK: Hodder & Stoughton.

Fajgelj, S. (2013). *Psihometrija. Metod i teorija psihološkog merenja* (IV dopunjeno izdanje). Beograd, Srbija: Centar za primenjenu psihologiju.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers**.** *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4*(3), 278–287.

Finn, S. E. (2007). *In our clients' shoes: Theory and techniques of therapeutic assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fischer, C. T. (1985/1994). *Individualizing psychological assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216–1229.

Goldman, B. A., & Mitchell, D. F. (2008). *Directory of unpublidhed experimental mental measures (Vol. 9)*. Washington, DC: American Psychological Association.

Harlen, W. (Ed.) (1994). *Enhancing quality in assessment*. London, UK: Paul Chapman Publising.

Jovin, J. (2018). Bibliografija naučnih radova u časopisu *Primenjena psihologija* (2008–2013). *Primenjena psihologija, 11*(1-dodatak prvom broju), 3–47.

Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. Oxford, England: Oxford University Press.

Maslach, C., & Jackson, S. (1981). *MBI: Maslach Burnout Inventory*. Palo Alto, CA: Consulting Psychological Press.

Momirović, K., Wolf, B. i Džamonja, Z. (1992). *KON 6 – Kibernetička baterija konativnih testova*. Beograd, Srbija: Centar za primenjenu psihologiju.

Murphy, K. R., & Davidshofer, C. O. (1994). *Psychological testing: Principles and applications.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Newman, M. L., & Greenway, P. (1997). Therapeutic effects of providing MMPI-2 test feedback to clients at a university counseling service: A collaborative approach. *Psychological Assessment, 9*(2), 122–131.

Penner, L. A., Fritzsche, B. A., Craiger, J. P., & Freifeld, T. S. (1995). Measuring the prosocial personality. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment, Vol. 10* (pp. 147–163). Hillsdale, NJ: Erlbaum.

Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment, 22*(2), 203–212.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory

in English and German. *Journal of Research in Personality, 41*(1), 203–212.

Schiele, B. C., Baker, A. B., & Hathaway, S. R. (1943). The MMPI. *Lancet, 63*, 292–297.

Smederevac, S., Mitrović, D. i Čolović, P. (2010). *Velikih pet plus dva, primena i interpretacija.* Beograd, Srbija: Centar za primenjenu psihologiju.

Spector, P. E. (1992). *Sage university papers series: Quantitative applications in the social sciences, No. 82. Summated rating scale construction: An introduction.* Thousand Oaks, CA: Sage Publications, Inc.

Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons Inc.

Walsh, B. W., & Betz, N. E. (2001). *Test and assessment* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Wolf, B., Momirović, K. i Džamonja, Z. (1992). *KOG3 – Baterija testova inteligencije*. Beograd, Srbija: Centar za primenjenu psihologiju.

INTERPRETACIJA SKORA NA TESTU

Sirovi ili bruto skor (eng. *row score*) je broj koji se odnosi na neki aspekt funkcionisanja osobe u slučaju selekcionisanog uzorka ponašanja obuhvaćenog testom. Sirovi skor je uobičajeno prost sumacioni skor (npr. broj tačnih odgovora na testu sposobnosti), a šire posmatrano podrazumeva kombinaciju pojedinačnih ajtemskih skorova. Sirovi skor, sam po sebi, nije informativan i ne nosi nikakvo značenje ukoliko se ne interpretira u odnosu na neki kontekst. Tako je, na nekim testovima, poželjno ostvariti visok skor, dok je na drugima poželjno ostvariti nizak skor. Na primer, visok skor na testu inteligencije ukazuje na razvijene mentalne sposobnosti koje su pozitivno povezane sa adaptabilnošću osobe, dok visok skor na testu depresivnosti ukazuje na prisutnost depresivnog poremaćaja koji se negativno odražava na adaptabilnost i funkcionisanje osobe. Sirovi skor nam ne pokazuje koliko je merena karakteristika izražena kod ispitanika, tj. ne možemo odrediti šta je zaista "visoko", a šta "nisko" ukoliko nemamo neki referentni okvir za ove vrednosti.

Da bismo mogli da interpretiramo skor, potrebno je da ga uporedimo s odgovarajućim referentnim okvirom. Postoje dva osnovna referentna okvira za interpretaciju skorova – normativna interpretacija (eng. *norm referenced interpretation*) i kriterijumska interpretacija (eng. *criterion or standard referenced interpretation*). Kod **normativne intepretacije**, skor ispitanika se interpretira kroz njegov položaj u distribuciji skorova određene grupe drugih ispitanika. Primer takve situacije je kada se skor na testu sposobnosti jednog učenika poredi sa skorovima njegovih vršnjaka na istom testu. Ovaj tip interpretacije

primenjuje se kada postoji primarna potreba da se ispitanici porede međusobno ili da se porede sa referentnom grupom. Najčešća primena normativne interpretacije je u cilju klasifikacije, jer je poređenje ispitanika u fokusu, ali se primenjuje i u cilju samospoznaje i ličnog razvoja.

Kod **kriterijumske interpretacije** skor ispitanika se interpretira preko poređenja sa spoljašnjim kriterijumom ili utvrđenim standardom. Primer za to je kada se skor jednog učenika na testu znanja poredi sa zadatim standardom za ocenjivanje nivoa znanja. Tako, kao prolazni standard može se odrediti uspešno rešavanje minimum 55% pitanja na testu znanja, te učenik koji tačno uradi preko 55% dobija prelaznu ocenu, bez obzira na to kakvo je postignuće ostalih učenika. Ovaj tip interpretacije skorova na testu primenjuje se kada postoji primarna potreba da se utvrdi da li ispitanik zadovoljava određeni kriterijum, odnosno standard. Kada god dobijete povratnu informaciju: "Čestitamo, položili ste!", znajte da je reč o kriterijumskoj interpretaciji. Najčešća primena ove interpretacije je ista kao i u slučaju normativne interpretacije, dakle u cilju klasifikacije, ali se često koristi i kod samospoznaje i ličnog razvoja kada je u fokusu zadovoljavanje standarda.

Razlike između normativne i kriterijumske interpretacije

Osnovna razlika između normativne i kriterijumske interpretacije je u tome što u normativnoj interpretaciji skor ispitanika potpuno zavisi od skorova ostalih ispitanika iz referentne grupe, tj. pozicija ispitanika potpuno zavisi od rezultata ostalih ispitanika. Kod kriterijumske interpretacije nema poređenja sa drugima, već se ispitanik poredi sa utvrđenim standardom. Iako standard može biti postavljen na osnovu rezultata drugih ljudi, kod kriterijumske interpretacije nema direktnog

poređenja sa drugima, kao kod normativne. Samim tim, kod kriterijumske interpretacije je eliminisan efekat takmičenja među ispitanicima. U vezi s tim, Karver (Carver, 1974) navodi da je kod normativne interpretacije fokus na individualnim razlikama, dok je kod kriterijumske fokus na razvoju i napredovanju osobe. Koen i Sverdlik (Cohen & Swerdlik, 2009) još dodatno objašnjavaju razliku u fokusu, ističući da je kod kriterijumske interpretacije fokus na učinku, tj. tome šta osoba može, a šta ne može; šta je naučila, a šta nije; da li je osoba ispunila uslov da bude u nekoj grupi, dobije određeni tretman itd. Dakle, dok normativnom intereptacijom dobijamo odgovor na pitanje *koliko* je nešto izraženo kod osobe, kriterijumskom interpretacijom dobijamo odgovor na pitanje *šta* neko može da postigne, ispolji i slično (Kodžopeljić i Pekić, 2017). U tom smislu, u okviru kriterijumske interpretacije skor se direktno interpretira u terminina kriterijuma. Razlike između normativne i kriterijumske interpretacije sumirane su u Tabeli 2.

Tabela 2

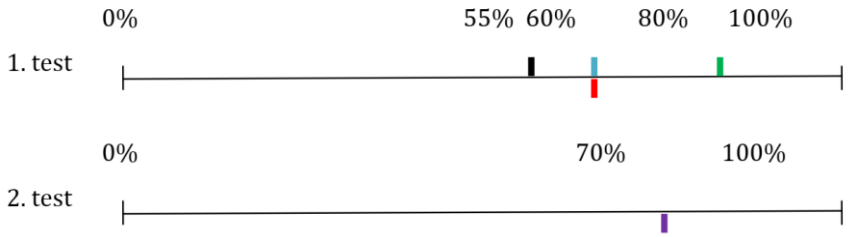
Razlike između normativne i kriterijumske interpretacije

	Normativna interpretacija	Kriterijumska interpretacija
Cilj	međusobno poređenje ispitanika	poređenje sa standardom
Fokus	individualne razlike; koliko je nešto izraženo?	učinak; napredovanje osobe...; šta osoba može?
Određivanje skora	unutrašnje, na osnovu skorova na testu	spoljašnje, preko veze testa sa kriterijumom
Interpretacija	položaj ispitanika u odnosu na ostale	da li je zadovoljen standard
Od čega zavisi skor?	od drugih skorova	od standard
Efekat takmičenja	postoji	ne postoji

Pored ovih osnovnih referentnih okvira za tumačenje skorova na testu, postoji i **interpretacija orijentisana na osobu** (eng. *self-referenced interpretation*). Flanagan (Flanagan 1939, prema De Vries, 2006) razlikuje dve vrste ove interpretacije, iako originalno o njima govori kao o trećoj i četvrtoj vrsti interpretacije skorova na testu. Prva od njih je interpretacija orijentisana na test (eng. *test-referenced interpretation*) u kojoj se skor jednog ispitanika na jednom testu poredi sa skorom istog ispitanika na nekom drugom testu. Primer za to je kada se porede skorovi jednog učenika na testu numeričkih i verbalnih sposobnosti. Da bi to bilo moguće potrebno je skorove na oba testa izraziti preko iste skale. Na osnovu ovih podataka može se utvrditi, na primer, da li je neko veštiji u baratanju brojevima ili u razumevanju teksta. Druga od njih je interpretacija orijentisana na okolnost (eng. *occasion-references tests*) u kojoj se skor jednog ispitanika na jednom testu poredi sa njegovim skorom na istom testu, ali u različitim okolnostima. Primer za to je kada se poredi skor jednog učenika na skali agresivnog ponašanja pre i nakon treninga nenasilne komunikcije. Najčešća primena interpretacije orijentisane na ispitanika je u cilju samospoznaje i ličnog razvoja, ali i u cilju provere efekta tretmana.

U skladu sa različitim referentnim okvirima postoji i podela testova na testove orijentisane na norme, testove orijentisane na kriterijum i testove orijentisane na osobu/ispitanika. Važno je napomenuti da se podela referentnih okvira za interpretaciju skorova u prvom planu odnosi upravo na to – referentne okvire za interpretaciju, a ne na vrste testova. Tako se, isti sirovi skor, može se interpretirati i u odnosu na normu, i u odnosu na kriterijum. Na primer, ukoliko neki učenik ostvaruje sirovi skor 30 od maksimalnih 50 poena, to znači da je

uspešno rešio 60%[14] zadataka (vidi Sliku 4). Ukoliko postavimo da je kriterijum za prolaz 55%, zaključićemo da je taj učenik dobio prelaznu ocenu. Međutim, pored toga što smo zaključili da učenik ima prelaznu ocenu, možemo želeti da dobijemo finiji uvid u njegovo postignuće. Da bismo to dobili, njegov skor možemo porediti sa skorovima ostalih učenika iz odeljenja. Ukoliko je prosečan skor svih učenika na testu 45 od maksimalno mogućih 50 poena (80%), onda bismo, uvidom u norme, zaključili da je postignuće tog učenika nisko u odnosu na postignuće ostalih učenika. Ukoliko je, na primer, prosečno postignuće na testu kao i učenikovo tj. 30 (60%), onda bismo zaključili da je postignuće tog učenika prosečno u odnosu na postignuće ostalih učenika. Ako pored toga znamo i da isti ispitanik na drugom testu ostvaruje 70% tačnih odgovora, možemo da izvršimo i interpretaciju orijentisanu na ispitanika i da uporedimo postignuća istog ispitanika na dva testa. Na osnovu toga bismo zaključili da je učenik bolje uradio drugi test u odnosu na prvi.



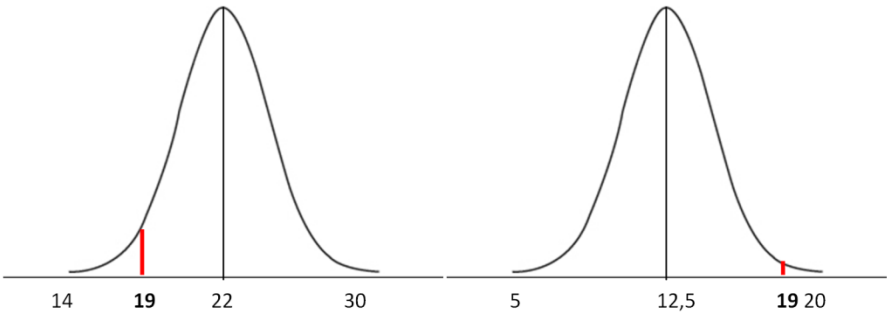
Slika 4. Primer kombinovanja normativne i kriterijumske interpretacije.

[14] 100 : X = 50 : 30, X = 100 * 30 / 50 = 60

Kao što je rečeno, sam sirovi skor nam nije informativan prilikom njegove interpretacije. Na primer, ukoliko je sirovi skor nekog ispitanika 19 na testu čiji teorijski raspon sirovih skorova iznosi od 0 do 30, mi ne možemo sa sigurnošću interpretirati ovaj skor kao iznadprosečan, prosečan ili ispodprosečan. Da bismo to mogli, potrebno je da znamo oblik distribucije. Ukoliko se skorovi distribuiraju normalno i potiču sa intervalne skale, poređenje datog skora sa teorijskom aritmetičkom sredinom može da zavara, te se skor može okarakterisati kao prosečan ili iznadprosečan. Na primer, ako sirovi skor 19 poredimo sa teorijskom aritmetičkom sredinom na testu, koja je u ovom slučaju 15[15], čini nam se da bi dati skor mogli okarakterisati kao prosečni ili iznadprosečni zato što je blizu ili iznad teorijskog proseka. Međutim, to ne bi bila ispravna interpretacija. Moguće je da je test veoma lak, pa da većina ispitanika ostvaruje skor u rasponu od 14 do 30 sirovih skorova (vidi Sliku 5). Ukoliko poznajemo i empirijsku aritmetičku sredinu, npr. 22 u ovom primeru, pod pretpostavkom da se svi postignuti skorovi normalno distribuiraju, skor 19 bi se pre okarakterisao kao ispodprosečan, kada bismo ga samo grubo poredili sa empirijskim prosekom. S druge strane, moguće je da je test veoma težak, pa da većina ispitanika ostvaruje skor u rasponu od 5 do 20 sirovih skorova, a da je empirijska aritmetička sredina testa 12,5. U ovom slučaju bismo skor 19 pre interpretirali kao iznadprosečan. Međutim, navedena

[15] Sledeći formulu za teorijsku aritmetičku sredinu, (Min + Max) / 2, dobija se teorijska aritmetička sredina (0 + 30) / 2 = 15.

interpretacija je, opet, proizvoljna. Čak i kada znamo empirijsku aritmetičku sredinu, ni tada ne bismo sa sigurnošću mogli da interpretiramo neki skor kao "visok" ili "nizak". Pored toga što je proizvoljna, ovakva interpretacija je i ograničena, u ovom slučaju, na dve kategorije ispitanika – ispodprosečnih i iznadprosečnih.



Slika 5. Primer različite interpretacije istog postignuća na testu (19) uz isti teorijski raspon skorova (0–30).

Da bismo pravilno interpretirali sirovi skor, potrebno je da ga lociramo u kontekst skorova koji potiču od referentne grupe za šta je potrebno da poznajemo distribuciju skorova. Da bismo znali distribuciju skorova, pored artmetičke sredine, potrebno je da znamo i standardnu devijaciju[16], a poželjno i skjunis[17] (zakošenost) i kurtozis[18] (spljoštenost).

[16] Mera odstupanja ili varijacije pojedinačnih skorova od aritmetičke sedine.
[17] Horizontalno (pozitivno ili negativno) odstupanje dobijene distribucije podataka od normalne distribucije, tj. koliko je distribucija asimetrična.
[18] Vertikalno (pozitivno ili negativno) odstupanje dobijene distribucije od normalne distribucije, tj. da li postoji grupisanje ili raspršenje odgovora oko aritmetičke sredine.

Na taj način možemo dobiti kategorijalne norme[19]:

1. ispodprosečni (sve vrednosti ispod *AS*–1 *SD*),
2. prosečni (sve vrednosti u opsegu *AS*–1 *SD* i *AS* + 1 *SD*) i
3. iznadprosečni (sve vrednosti iznad *AS* + 1 *SD*).

U našem primeru, ukoliko bismo znali da je, na primer *AS* = 22 i *SD* = 7, za ispitanika koji ostvaruje sirovi skor 19 mogli bismo da zaključimo da pripada kategoriji prosečnih ispitanika, budući da je opseg proseka između 15 (22 – 7) i 29 (22 + 7). Iako je ovaj način interpretacije precizniji i ispravniji u odnosu na grubo poređenje sa empirijskom aritmetičkom sredinom, i u nekim situacijama dovoljan, svakako nije adekvatan ukoliko želimo precizniju informaciju o rezultatu ispitanika. Dakle, i dalje je interpretacija skorova gruba i svodi se na tri ili više kategorija (ukoliko uzmemo u obzir i odstupanje od 2 standardne devijacije).

Ukoliko želimo preciznu informaciju o rezultatu ispitanika, onda skor ispitanika treba da poredimo sa **normom**. Norme su, inače, najčešće korišćeni referentni okvir za interpretaciju skorova na testu i svi komercijalni testovi najčešće sadrže normativnu interpretaciju. Norme su najčešći okvir za interpretaciju skorova jer se najčešće procenom nečijeg rezultata želi utvrditi njegov položaj u odnosu na druge ljude, opštu populaciju itd. Norma predstavlja prosečno postignuće ili tipično ponašanje jedne ili više referentnih grupa ispitanika. O tome govori i samo poreklo reči "norma" čiji je semantički koren u reči "normalno" ili "prosečno". Dakle, norma je prosek ili tipični

[19] Ovo je primer statističkog normiranja na osnovu uvida u distribuciju skorova i ne mora se poklapati sa dijagnostičkim normiranjem koje se bazira na uvidu u relacije sa nekom spoljašnjom, kriterijumskom varijablom.

skor (izražena kao aritmetička sredina ili medijana[20]) na konkretnom testu koji ostvaruje specifična populacija (Chadha, 2009). U Standardima (AERA, APA, & NCME, 2006, str. 266) norme se definišu kao „statističke vrednosti ili tabelarni podaci u kojima je prikazan učinak na testu za određene grupe, kao što su ispitanici različite starosne dobi ili razreda."

Da bi poređenje sa normom moglo da se izvede, potrebno je sirovi skor transformisati u neki drugi skor, koji bi imao značenje za interpretaciju. Transformacijom dolazimo, u stvari, do normi. Dobijanje normi je ključno za standardizaciju testova orijentisanih na norme. S obzirom na to, ovi testovi se nazivaju još i normativnim.

Uslovi za adekvatne norme

Postoje dva osnovna uslova koja treba da se ispune da bi norme bile upotrebljive i da bi se njima dobila verna i validna interpretacija skorova. Prvi je da su norme dobijene na reprezentativnom uzorku ispitanika za populaciju kojoj je test namenjen, a drugi je aktuelnost normi.

Reprezentativnost uzorka

Reprezentativan uzorak poseduje karakteristike slične onima u populaciji, tj. odražava populaciju u malom. Većina normativnih testova sadrži nacionalne norme (Airasian, 2005; Nitko, 2004), tj. norme dobijene na reprezentativnom uzorku populacije neke države. I u samoj definiciji normi u Standardima (AERA, APA, & NCME, 2006, str. 266)

[20] Medijana je centralni rezultat u distribuciji koji deli distribuciju na dva jednaka dela.

navodi se da se „norme često izrađuju tako da reprezentuju neku veliku populaciju kao što su npr. ispitanici iz cele zemlje". U vezi sa tim, treba napomenuti da je situacija lakša kada svi članovi jedne nacije govore istim jezikom jer se jezik obično uzima kao odrednica neke kulture (Hedrih, 2018). U tom smislu, moguće je dobiti norme na objedinjenim uzorcima stanovnika više zemalja, sve dokle oni govore istim jezikom. U slučaju kada u istoj zemlji postoji upotreba različitih jezika (npr. kao što u Belgiji postoji nemačko i francusko govorno područje), onda je potrebno dokazati da su verzije testa na tim jezicima ekvivalentne i da se mogu međusobno porediti. Drugim rečima, ne može se unapred pretpostaviti da su test na originalnom jeziku na kojem je nastao i preveden test isti, već to treba empirijski dokazati. Za utvrđivanje ekvivalentnosti, tzv. invarijatnosti merenja (eng. *measurement invariance – MI*), koristi se poseban statistički postupak baziran na konfirmatornoj faktorskoj analizi (više o tome može se naći na str. 335).

Prilikom određivanja reprezentativnosti uzorka treba voditi računa o svim *relevantnim* karakteristikama populacije, a to su one karakteristike od kojih zavisi rezultat na testu. Na primer, ukoliko postoje sistematske razlike između muškaraca i žena u postignuću na testu motornih sposobnosti, onda pol predstavlja relevantnu karakteristiku populacije za dati predmet merenja i potrebno je napraviti posebne norme za muškarce i za žene. Najčešće karakteristike uzorka o kojima se vodi računa su polna i starosna sturktura, eventualno i obrazovna struktura u slučaju kognitivnih testova. Ove karakteristike ostvaruju dosledne efekte na skoro sve konstrukte koji su predmet merenja testova (više u Mitrović i Trogrlić, 2014; Schaie i Willis, 2001), te se zato smatraju relevantim karakteristikama populacije. U zavisnosti od namene testa, mogu se uključiti i nacionalne,

etničke, socioekonomske, profesionalne i druge karakteristike za definisanje reprezentativnog uzorka. Svrha reprezentativnog uzorka je u smanjivanju potencijalnih pristrasnosti prilikom poređenja rezultata. Tako, na primer, ukoliko bi norme za neki test znanja dobili na uzorku učenika samo iz razvijenih mesta, onda bi tumačenje skorova učenika iz nerazvijenih mesta bilo pristrasno i neadekvatno. Da bi se to izbeglo, uzorak na kojem će se računati norme mora biti reprezentativan za populaciju kojoj je namenjen i obuhvatiti varijabilitet svih karakteristika ispitanika koje su važne za predmet merenja. O osnovnim načinima odabira reprezentativnog uzorka može se više saznati u Fajgelj (2007).

Nisu svi uzorci na kojima se računaju norme reprezentativni za datu populaciju. Nekada je to rezultat nedostatka resursa da se pribavi reprezentativni uzorak, nekada stvarna nemogućnost itd. Kod nas izdavačke kuće nekada pruže skromnu podršku autorima da prikupe podatke, ali se češće očekuje da autori sami obave terensko istraživanje i prikupe podatke, pomoću sopstvenih sredstava i resursa. S obzirom na to, uzorci često nisu reprezentativni, ali se autori trude da ne budu ni potpuno prigodni. Bez obzira na to kakve su karakteristike uzorka u odnosu na ciljnu populaciju kojoj je test namenjen, autori u priručnicima i drugim publikacijama treba jasno da okarakterišu vrstu uzorka i opišu način selekcije ispitanika, a na korisniku testa je da odluči kakav je kvalitet uzorka i da li se relevantne karakteristike uzorka poklapaju s onima koje su korisniku važne pri proceni.

Kao što je rečeno, kod normativne interpretacije skorova, interpretacija se vrši na osnovu poređenja skora jednog ispitanika sa skorovima drugih ispitanika koji pripadaju istoj populaciji ili subpopulaciji, tj. odgovarajućoj referentnoj grupi. Nekada je ciljna populacija mala, pa se ispitivanjem mogu obuhvatiti svi njeni članovi

(učenici jednog odeljenja). S druge strane, ciljna populacija može biti veća (stanovnici Republike Srbije). Osnovno pitanje ovde je šta je cilj interpretacije rezultata i za koje se svrhe ona koristi. Ukoliko samo želimo da znamo kojem učeniku iz jedne škole treba dati nagradu za najbolji uspeh, svakako da nam je poređenje sa svim učenicima te iste škole ili razreda dovoljno. Međutim, ukoliko želimo da znamo koje učenike iz cele zemlje treba nagraditi, onda svakako treba da znamo kakvo je postignuće učenika ne samo iz jedne škole, nego iz svih škola u zemlji.

Šta je referentna grupa?

S obzirom na to da u praksi najčešće nije moguće ispitati sve pripadnike jedne populacije, pribegava se uzorkovanju. U tom kontekstu, skor jednog ispitanika poredi se sa skorovima drugih ispitanika u okviru ispitanog uzorka, koji predstavlja referentnu grupu. Na ovaj način dolazimo do položaja jednog ispitanika na distribuciji skorova ostalih ispitanika iz uzorka. Ispitanik i ispitani uzorak za dobijanje referentnog okvira za interpretaciju, po pravilu, moraju biti iz iste populacije. Na primer, ukoliko želimo da saznamo kakvo je postignuće na numeričkom testu nekog učenika uzrasta 12 godina, referentni uzorak mora da se sastoji od učenika približnih godina kao i naš ispitanik, npr. od 11 do 13 godina.

Ponekad postoji potreba i da se poredi skor jednog ispitanika sa skorovima drugih ispitanika iz više specifičnih referentnih grupa, kojima pripada i sam ispitanik. Da bi to bilo moguće, potrebne su nam norme dobijene na različitim referentnim grupama. Na primer, možemo izvršiti poređenje skora na testu numeričkih sposobnosti jednog učenika sa skorovima njegovih vršnjaka iz škole, ali i sa skorovima njegovih

vršnjaka iz cele zemlje, iz drugih zemalja ili sa članovima Mense. U ovom primeru specifične referentne grupe su vršnjaci iz konkretne škole, vršnjaci iz cele zemlje, vršnjaci iz drugih zemalja i vršnjaci koji su članovi Mense. Moguće je, na primer, da je neki učenik najbolji u svom odeljenju po pitanju numeričkih sposobnosti, ali prosečan u odnosu na postignuće njegovih vršnjaka u celoj zemlji. Ovu informaciju možemo dobiti samo ukoliko koristimo druge referente grupe za interpretaciju skora učenika.

Međutim, postoji izuzetak kada ispitanik i referentna grupa ne pripadaju istim populacijama. Primer za to je kada se na osnovu rezultata na testu predviđa populacija kojoj ispitanik pripada, npr. kada se na osnovu skora ispitanika na bateriji testova sposobnosti određuje njegov mentalni uzrast (više o tome će biti reči na str. 91). Tako, ispitanik uzrasta 10 godina u poređenju sa svojim vršnjacima može ostvariti iznadprosečni rezultat, a u poređenju sa 12-godišnjacima prosečan, te se zaključuje da je njegov mentalni uzrast 12 godina, iako je kalendarski 10.

Često se termin referentna grupa koristi u istom značenju kao i termini standardizacioni uzorak, normativni uzorak ili validacioni uzorak. U praksi se često dešava da je zaista reč o istom uzorku ispitanika, ali ipak treba napraviti razliku u značenjima. Osnovna razlika je u stepenu opštosti, pri čemu najmanju opštost ima termin standardizacioni uzorak, pa normativni, pa referentna grupa i na kraju, validacioni uzorak.

Standardizacioni uzorak je uzorak na kojem je test originalno standardizovan. Dakle, to je uzorak koji je opisan u priručniku nad kojim su izračunate norme. Valja se podsetiti da je dobijanje normi samo jedan od aspekata standardizacije, kako bi se ostvarila ista interpretacija

testovih rezultata za sve. Pored toga, standardizacija još uključuje i isti format testa za sve, uslove primene, odnos prema ispitaniku i slično.

Normativni uzorak se često koristi kao sinonim standardizacionom, ali on ipak ima nešto šire značenje. Normativni uzorak se odnosi na bilo koji uzorak na kojem su sve moguće norme izračunate. On obuhvata standardizacioni uzorak, ali i druge uzorke na kojima su izračunate norme. Na primer, moguće je da se nakon standardizacije nekog testa ličnosti za odraslu populaciju ukaže potreba za primenom testa u okviru neke specifične starosne grupe, kakva je populacija starih ili populacija adolescenata. U tom slučaju bi se pristupilo prikupljanju podataka na ovim populacijama kako bi se izračunale njima primerene norme. Ove dodatne norme mogu se objaviti u posebnim publikacijama ili kao dodatak postojećem priručniku. U ovom primeru standardizacioni uzorak bi činila populacija odraslih, a normativni uzorak i populacija odraslih i druga populacija, npr. adolescenata. Sa stanovišta primene testa, normativni uzorak je najvažniji i on zapravo određuje upotrebnu vrednost testa.

Referentna grupa je širi pojam i obuhvata bilo koju grupu ljudi u odnosu na koju se poredi dati skor nekog ispitanika. Ona može obuhvatati standardizacioni i normativni uzorak, ali i sve druge uzorke koji služe za poređenje skora. Takvi uzorci mogu imati više lokalni karakter, npr. kada nas intersuje postignuće nekog učenika u odnosu na učenike njegovog odeljenja, razreda škole, razreda svih škola u opštini i slično. Isto tako, može nas interesovati kakvo je zadovoljstvo poslom novozaposlenog službenika u odnosu na sve ostale službenike u njegovoj filijali, ili u odnosu na sve ostale zaposlene u njegovoj filijali, u centralnoj filijali, u drugim bankama i slično. Dakle, referentna grupa može da varira na kontinuumu od speficične populacije (npr. učenici

jednog razreda jedne opštine) do opšte populacije (npr. stanovnici neke države, pripadnici neke nacije i slično).

Još jedan primer referentne grupe mogu biti ispitanici jedne studije, kada se sprovodi neko naučno istraživanje. To je grupa ispitanika koja nije služila za standardizaciju testa i dobijanje normi za opštu primenu, već su se samo za potrebe istraživanja ispitanici podelili u grupe u odnosu na izraženost skorova. Pritom, na ovaj način se ne dobijaju norme za ceo kontinuum skorova, već kategorijalne norme, tj. norme za svrstavanje u grupe. Često se u istraživanjima koristi podela po medijani (eng. *median split*) na dve grupe ispitanika, s niskim i visokim skorovima, ali moguće je i podela na osnovu *AS* i *SD* na tri grupe, tj. ispodprosečne, prosečne i iznadprosečne ispitanike. Na primer, moguće je u istraživanjima na osnovu neke skale agresivnog ponašanja ispitanike podeliti na one koji ne ispoljavaju nasilje, koji ga ispoljavaju u manjoj meri i one koji ga ispoljavaju u velikoj meri, pa se onda ovako dobijene grupe porede u odnosu na neke karakteristike kao što su osobine ličnosti, porodična kohezija, kognitivni stil i slično. U ovoj situaciji može se desiti da neko ko je svrstan, na primer, u kategoriju prosečnih, ne mora biti u istoj kategoriji kada bi se njegov skor poredio sa standardizacionim uzorkom na istom testu. Moguće je da uzorak konkretnog istraživanja potiče iz neke rizične grupe (npr. delinkvenata), pa da prosečni skor u toj grupi odgovara, zapravo, iznadprosečnom skoru u opštoj populaciji. Moguća je i obrnuta situacija. Naime, u istraživanjima na opštoj populaciji često se zaključuje o nekim kliničkim fenomenima, pa se tako na osnovu ponašanja ispitanika koji su svrstani u nekom istraživanju u grupu visokoanksioznih zaključuje o funkcionisanju osoba s anksiznim poremećajima. Međutim, moguće je da ispitanici visokoanksiozne grupe u opštoj populaciji ne ispunjavaju

uslove za dijagnostikovanje anksioznog poremećaja. Iako ovakva istraživanja mogu pružiti uvid u funkcionisanje osoba s anksioznim poremećajem, ona svakako imaju svoja ograničenja.

U istraživačkim radovima mogu se prikazati norme za neki test za ceo kontinuum, ali je to vrlo retko. To ima smisla kada su u pitanju testovi koji nisu prethodno korišćeni u datoj zemlji, na datom jeziku, ili kada se predstavlja neki novokonstruisani test, a postoji potreba da se na osnovu datog testa vrši psihološka procena. Ukoliko ne postoji potreba da se test koristi zarad psihološke procene, onda nema potrebe ni računati norme. Iako ovakvi podaci mogu biti korisni za razumevanje skorova na testu, njihova primena je ipak ograničena, budući da su istraživački uzorci uglavnom nereprezentativni. Ukoliko je istraživački uzorak dovoljno reprezentativan, onda su naravno i norme korisnije, ali u tim situacijama se pre preporučuje da se napiše priručnik umesto da se norme prikažu u okviru istraživačkog izveštaja. Priručnik, pored ostalog, treba da sadrži i primer tumačenja skorova na testu, tipične profile i slično, što nije uobičajeno navoditi u naučnim člancima i drugim, sličnim izveštajima.

Validacioni uzorak je najširi pojam i odnosi se na uzorak na kojem je proverena validnost nekog testa. To može biti bilo koji od prethodno navedena tri uzorka, ali i uzorak koji uopšte nije vezan za dobijanje bilo kakvih normi, već za utvrđivanje validnosti testa. Takav uzorak može biti uzorak za istraživačke svrhe ispitivanja validnosti konstrukta ili skorova na testu. Validacioni uzorak je jedini od pomenutih vrsta uzoraka koji ne mora rezultatirati normama.

Ukoliko imamo norme dobijene na različitim vrstama uzoraka, svakako se treba osloniti na norme dobijene normativnim uzorkom. Međutim, ukoliko imamo neku specifičnu populaciju koja nije

obuhvaćena normativnim uzorkom (npr. socijalno deprivirane osobe), a posedujemo norme na uzorku te populacije koji je okarakterisan kako referentna grupa, precizniji uvid u rezultate će nam dati poređenje s normama dobijenim na referetnom, a ne normativnom uzorku. Ipak, uvek treba voditi računa o cilju testiranja i relevantnim karakteristikama uzorka za dati predmet merenja. Ukoliko neka karakteristika nema velikog efekta na skor na testu (npr. nivo obzovanja na skor na testu ličnosti), onda nema potrebe da skor nekog ispitanika s nižim nivoom obrazovanja poredimo s referentnom grupom koju čine ispitanici nižeg nivoa obrazovanja, već skor treba da poredimo s normativnom grupom iako ne postoje posebne norme za one s nižim nivoom obrazovanja.

Na ovom mestu samo treba skrenuti pažnju na to da opštost pojmova nema veze sa veličinom uzorka. Na primer, iako je validacioni uzorak najopštiji pojam, moguće je da je on samo deo standardizacionog ili normativnog uzorka. Moguće je da su autori testa sproveli istraživanje na npr. 300 ispitanika s ciljem da utvrde konstrukt validnost testa, pa su uz test dali i druge mere za validaciju. Pored ovog uzorka, autori su prikupili podatke od još 500 ispitanika, ali su tim ispitanicima zadali samo test. U ovom primeru validacioni uzorak čini 300 ispitanika, a normativni uzorak čini 300 + 500, ukupno 800 ispitanika.

Kolika treba biti veličina referentne grupe?

Iako je opšta preporuka da je bolje imati što veći uzorak, od same veličine važnija je reprezentativnost uzorka za definisanu populaciju. Ukoliko je reprezentativnost adekvatna, postoje smernice u vezi s veličinom uzorka, kada je reč o nacionalnim normama. U recenzentskom modelu za opis i evaluaciju psiholoških i edukativnih

testova, koji je propisala Evropska federacija psiholoških udruženja (European Federation of Psychologists' Associations – EFPA, 2013) v. 4.2.6, date su smernice o veličini uzorka, imajući u vidu da li je reč o testovima niskog ili visokog uloga (Tabela 3). Ukoliko je u pitanju test visokog uloga, uzorak mora biti veći, ali je ova razlika u veličini uzorka evidentna samo u okviru preporučenog minimuma da bi se uzorak mogao smatrati adekvatnim. Ukoliko se norme računaju na poduzorcima, npr. ukoliko treba da postoje muške i ženske norme, uzorak bi trebalo onda povećati, ali ne nužno udvostručiti.

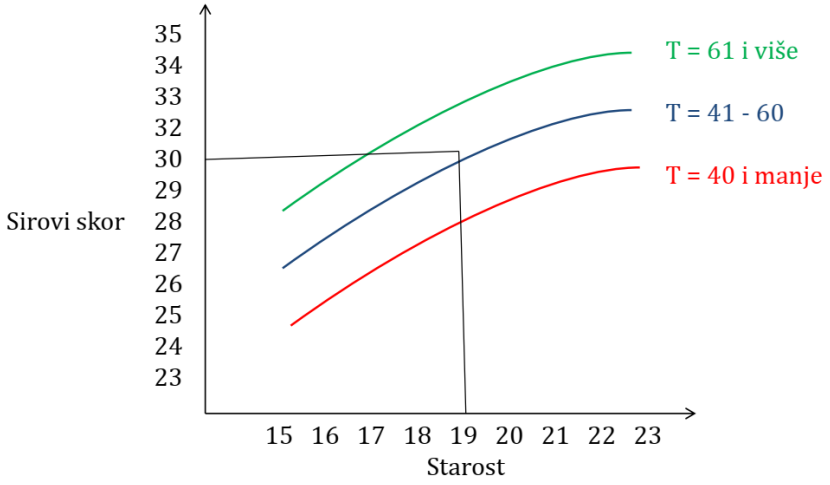
Tabela 3

Preporučena veličina uzorka prema EFPA modelu

Veličina uzorka	Testovi niskog uloga	Testovi visokog uloga
Adekvatna	200–299	300–399
Dobra	300–999	400–999
Odlična	preko 1000	preko 1000

Ukoliko želimo da dobijemo norme po nekoj karakteristici koja je originalno kontinuirana, autori najčešće pribegavaju veštačkoj dihotomizaciji ili podeli varijable na nekoliko kategorija ili rangova. Primer za to je kada želimo da dobijemo starosne norme, pa uzorak podelimo na mlađe i starije ispitanike, ili na grupe čiji je raspon starosti 5 ili 10 godina. Međutim, ovim načinom gubimo na informativnosti jer se ne uzima u obzir ceo varijabilitet karakteristike. Poseban je problem ako grupe nisu ujednačene po veličini ili ako neke grupe nisu dovoljno velike. S obzirom na ovaj problem, neki autori su ponudili procedure za korekciju skorova, tzv. glačanje skorova (engl. *smoothing*), ali najsofisticiranija procedura je tzv. kontinuirano normiranje (eng. *continuous norming,* Gorsuch, 1983). U **kontinuiranom normiranju** i

dalje imamo grupe, ali se norme dobijaju u funkciji kontinuirane varijable, tj. za svaku vrednost te varijable. Tako, na primer, možemo imati norme za 5-godišnjake, 6-godišnjake, 7-godišnjake i slično. Norme se u ovom slučaju prikazuju grafički (vidi Sliku 6), pri čemu se na X-osi nalaze vrednosti za datu kontinuiranu varijablu (npr. starost), a na Y-osi sirovi skorovi, i različitim krivama su označeni npr. ispodprosečni, prosečni ili iznadprosečni skorovi (može biti dat i detaljniji prikaz, npr. za decile). Ukoliko želimo da lociramo postignuće nekog ispitanika čiji je sirovi skor 30 na testu, a znamo da ima 19 godina, naći ćemo tačku preseka na grafikonu i očitati norme. U primeru sa Slike 6, ovaj ispitanik ima postognuće oko proseka, nešto iznad T-skora 41 (u ovom primeru se prosečnim skorovima smatraju T-vrednosti od 41 do 60, videti više u Tabeli 8).



Slika 6. Primer prikaza kontinuiranog normiranja.

Važna posledica ovakvog načina računanja normi je ta što kontinuirano normiranje zahteva manji uzorak u odnosu na tradicionalno. U EFPA modelu navode se smernice koje su dali Behger i

saradnici (Bechger, Hemker, & Maris, 2009), u vezi sa brojem ispitanika po poduzorku. Naime, ako imamo 8 poduzoraka, 70 ispitanika u svakom uzorku daje istu preciznost kao i da imamo po 200 ispitanika u svakom poduzorku, potom 100 ispitanika daje istu preciznost kao i da imamo 300, a 150 kao da imamo 400 ispitanika. Kada je broj poduzoraka veći, broj ispitanika u okviru tih poduzoraka je manji, i obrnuto. U slučaju testova visokog uloga, broj ispitanika po poduzorcima, naravno, treba biti veći. U praksi je nekada teško ostvariti preporučeni standard, ali ono što je sigurno je da se ne preporučuje ukupni uzorak manji od 200 ispitanika.

Kontinuirano normiranje ima smisla primenti kod procene razvojnih fenomena, kada se svojstvo kvalitativno ili kvantitativno menja s godinama. S obzirom na to, nije iznenađujuće što je kontunuirano normiranje najčešće našlo primenu u domenu procene kognitivnih i obrazovnih kapaciteta. U drugim situacijama, kada postoje relativno stabilne tendencija ispitanika u većem razvojnom periodu, kao što su npr. osobine ličnosti, nema potrebe za kontinuiranim normiranjem, već tu dovoljno informativne mogu biti podele u starosne grupe (npr. do 25 i od 25, ili da se izračunaju posebne norme za adolescente, odrasle i stare osobe).

Aktuelnost normi

U vezi s drugim uslovom, tj. aktuelnošću normi, preporuka iz Standarda (AERA, APA, & NCME, 2006) je da se ponovno normiranje vrši periodično. Postoji saglasnost da renormiranje treba raditi na 10 godina (Urbina, 2004), ali prema EFPA modelu adekvatnim normama smatraju se one koje nisu starije od 15 do 19 godina, dobre one koje nisu starije od 10 do 14 godina, i odlične one koje nisu starije od 10 godina.

Norme su često kontekstualno specifične i nestabilne tokom vremena, te je zbog toga potrebno njihovo renormiranje. Kao najbolja ilustracija ovog problema može poslužiti tzv. **Flinov efekat** (eng. *Flynn effect*), koji je dobio naziv po Džejmsu Flinu koji ga je opisao. Ovaj efekat se odnosi na fenomen da IQ skorovi ljudi stalno rastu tokom dekada. Postoji nekoliko teorija o uzroku napredovanja ljudske inteligencije. Sam Flin je kao razlog isticao generalno stimulativnije okruženje savremnog čoveka, koje se ogleda najviše u upotrebi tehnologija, vizuelnih medija, zahteva za intelektualnim radom i sl., i koja ima efekat na hipotetičko rešavanje problema. Drugi autori kao razlog ističu obrazovanje i veće iskustvo sa testovnim materijalom, kvalitetniju ishranu, smanjenje infektivnih bolesti itd. (više u Brouwers, van de Vijver, & van Hemert, 2009). Ipak, pojedini autori izražavaju sumnju u ovaj fenomen, čak tvrde kako IQ stagnira ili opada u novom dobu (npr. Dutton, van der Linden, & Lynn, 2016). Novije meta-analize ne daju jednoznačno rešenje, i dok je u jednoj Flinov efekat potvrđen, ali je manji nego što se to mislilo (Trahan, Stuebing, Hiscock, & Fletcher, 2014), u drugoj je pokazano da IQ skorovi variraju u zavisnosti od toga o kojoj sposobnosti je reč, ali da generalno opadaju u skorijim dekadama (Pietschnig & Voracek, 2015).

Kako se manifestuje Flinov efekat? Uspostavljanjem standardne skale IQ skorova, prosek na testu predstavlja 100 IQ jedinica. Kada se norme za isti test ponovo računaju kroz neko vreme, ponovo se uspostavljaju uz prosek od 100 IQ, ali ako bi se prosečni skor isptanika koji učestvuju u restandardizaciji uporedio sa starim normama, on bi gotovo uvek prelazio prosek od 100 IQ jedinica. Ovaj fenomen se često sreće u praksi kada se koriste testovi sa zastarelim normama. U ekstremnom slučaju dešava se da nekad za nekog ispitanika koji je

visoko inteligentan, u priručniku sa zastarelim normama, nema normi koje odgovaraju skoru koji taj isitanik postiže, tj. taj ispitanik bi se svrstao u onih 1% ispitanika s izrazito visokim skorovima. Ovo ne znači da taj ispitanik pokazuje genijalnost, nego prosto da su norme zastarele. Stoga je potrebno stalno obnavljati norme kako bi se dobila realna slika nečijeg postignuća.

Flinov efekat je primetan i u slučaju skorova na testovima ličnosti (Jokela, Pekkarinen, Sarvimäki, Terviö, & Uusitalo, 2017). Na primer, na velikom uzorku muškaraca iz Finske koji su rođeni između 1962. i 1976. pokazano je da postoji trend rasta skorova na skalama samopouzdanja, socijabilnosti i motivaciji za vođstvom, potom na skalama aktiviteta i težnje za postignućem, pa na skalama promišljenosti i poslušnosti, dok na skali maskulinosti nema jasnog trenda. Pokazano je, takođe, da je ovaj trend rasta povezan sa visinom zarade.

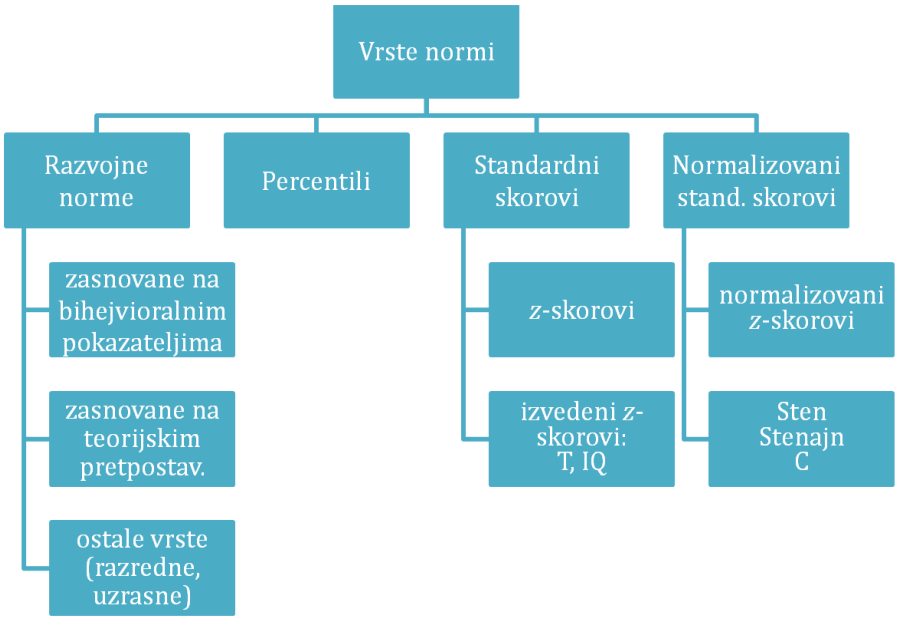
Valja napomenuti da se u Standardima (AERA, APA, & NCME, 2006) navodi da je dužnost izdavača testa da osigura dovoljno često ponovno normiranje sve dok je test u distribuciji. Za prvo izdanje testa izdavač može, a i ne mora da se uključi u proces pribavljanja uzorka i slično, ali kada se test objavi i pusti u promet, izdavač ima odgovornost za dostavljanje aktuelnih normi, a ne autor testa. Kada norme zastare, izdavač u dogovoru s autorima inicira restandardizaciju. Iako je uobičajeno da autor učestvuje u restandardizaciji, ovo nije nužno, a nekada nije ni izvodljivo (npr. usled smrti autora), te restandardizaciju može realizovati i neki drugi istraživački tim. Iako je izdavač odgovoran za aktuelnost normi, korisnik je odgovoran za donošenje odluke o korišćenju testa, pa bi trebao da izbegne upotrebu testova sa zastarelim normama.

Nažalost, u domaćoj praksi je čest slučaj da se ne radi restandardizacija i da psiholozi koriste testove sa zastarelim normama, iz 90-ih, pa čak i 80-ih godina. Takođe se, iz potrebe za aktuelnim normama, pribegava svežim normama koje su dobijene na nekoj drugoj populaciji, što je možda i veća greška nego koristiti zastarele norme. Ukoliko su norme iz druge populacije bazirane na populaciji koja je nama kulturološki bliska, ili bliska po obrazovnom sistemu kada je reč o kognitivnim testovima, onda se možemo koliko-toliko osloniti na te norme. Suočeni s onim što imaju, psiholozi treba pažljivo da tumače rezultate i da imaju u vidu zastarelost normi, populaciju na kojoj su nastale norme i druge nedostatke primenjenog testa, te da prilikom izveštaja obavezno napomenu ograničenost pri tumačenju rezultata.

Vrste normi

Postoji nekoliko podela normi. Jedna od njih odnosi se na sveobuhvatnost uzorka, tj. referentne grupe. S obzirom na ovu podelu, mogu se razlikovati lokalne norme (koje se odnose na ograničenu referentnu grupu od posebnog interesa za korisnika testa, kao što je npr. neka opština, škola itd.), podgrupne norme (npr. za muškarce i žene, adolescente i odrasle), nacionalne norme i sl. Druga vrsta podele odnosi se na primenu određenih matematičkih transformacija sirovog skora. U zavisnosti od vrste transformacije koja je primenjena nad sirovim skorovima, mogu se razlikovati transformacije koje ne menjaju oblik distribucije sirovih skorova (npr. linearne transformacije kao što je standardizacija i dobijanje *z*-skorova) i transformacije koje ga menjaju (npr. nelinearne transformacije, a uobičajeno se koristi normalizacija kojom dobijamo npr. normalizovane *z*-skorove). Treća vrsta podele odnosi se na primenjenu skalu izražavanja skorova, pa se tako mogu

razlikovati percentilni skorovi, standardni skorovi, sirovi skorovi i slično. U ovom udžbeniku usvojiće se podela u kojoj se kombinuje vrsta primenjene transformacije i vrsta skale izražavanja skorova (Slika 7). Kako se razvojne norme izdvajaju u odnosu na sve ostale opisane, one će biti izdvojene kao posebna vrsta normi.



Slika 7. Vrste normi.

Razvojne norme

Ljudski razvoj se odvija u sekvencijalnim fazama, pri čemu svaku fazu karakterišu razvojne promene koje se manifestuju u različitim ponašanjima (npr. Santrock, 2013). Razvojne norme bazirane su na pretpostavci da se neka osobina, sposobnost, veština ili druga karakteristika razvija, pogoršava ili na drugi način menja pod uticajem hronološke dobi, školskog razreda ili životnog stadijuma (Cohen &

Swerdlik, 2009). One daju podatak o tipičnom postignuću različitih uzrasnih grupa u trenutku ispitivanja i služe tome da individualni rezultat poredimo s tipičnim postignućem odgovarajuće uzrasne grupe. Na osnovu poređenja zaključujemo da li je individualni razvoj u skladu sa normativnim (Jerković i Zotović, 2017).

Postoji nekoliko vrsta razvojnih normi u odnosu na to na koji način su uspostavljene. Osnovna podela je na ordinalne skale zasnovane na bihejvioralnim pokazateljima i na teorijskim očekivanjima (Urbina, 2004).

Razvojne norme zasnovane na bihejvioralnim pokazateljima

Ove vrste razvojnih normi zasnovane su na opservacionim metodama tipičnog sleda faza i perioda javljanja različitih ponašanja tokom razvoja. Najčešće se kao primer ovih razvojnih normi prikazuje motorni razvoj. Tako, od 3. meseca života, dete čini prvi, ali neuspešni pokušaj hvatanja predmeta, u 6. mesecu uspeva da uhvati predmete, ali bez opozicije palca, u 7. mesecu se javlja i opozicija palca, dok krajem prve godine dete uspeva da planira i izvodi pokrete dosezanja i hvatanja i da sistematski ispituje svojstva predmeta (stavljanje u usta, opipavanje, stezanje, bacanje i sl.). U okviru ove vrste normi poredi se ono što dete aktuelno može da uradi sa onim što se dobija kao tipična aktivnost na referentnoj grupi dece iste hronološke dobi. Razlika u ovim aktivnostima može biti indikator prolongiranog razvoja kod dece, te signal za praćenje razvoja deteta ili primenu određenog tretmana.

Najpoznatiji primer ove vrste razvojnih normi predstavljaju norme koje je ponudio Arnold Gezel 1925. godine, koje su nastale na osnovu longitudinalnog istraživanja ranog razvoja dece (od 4. nedelje do

6. godine), a kasnije i adolescenata. Gezel je pratio 10 različitih aspekata razvoja, uključujući kognitivni, motorni, socijalno-emocionalni razvoj i dr. Kao rezultat istraživanja 1940. godine nastala je Gezelova razvojna skala koja je doživela nekoliko revizija.

Prilikom pominjanja testova koji su bazirani na bihejvioralnim pokazateljima, nezaobilazno je spomenuti Bine-Simonovu skalu za procenu inteligencije. Nju su 1905. konstruisali Alfred Bine i Teodor Simon na zahtev francuskog ministarstva u cilju procene spremnosti dece za polazak u školu. Skala obuhvata ispitivanje nekoliko aspekata inteligencije pod pretpostavkom da svi oni konvergiraju u jedan opšti (G) faktor inteligencije. Skala je kasnije revidirana tako da uključuje procenu inteligencije dece od 3 do 15 godina. Kod nas postoje dve revizije ove skale: Beogradska revizija Bine-Simonove skale (Stevanović, 1934) i Nova beogradska revizija Bine-Simonove skale (Ivić, Milinković, Rosandić i Smiljanić, 1985), ali su obe standardizovane prilično davno.

Na osnovu postignuća na testu određuje se mentalni uzrast (MU) koji predstavlja broj tačno rešenih zadataka koji se smatra prosečnim za dati uzrast deteta. Taj broj se pretvara u mentalne mesece, npr. svaki rešeni zadatak vredi 2 meseca mentalnog uzrasta. Kasnije je Vilijam Štern predložio računanje umnog količnika preko formule: IQ = MU / KU * 100, gde KU predstavlja kalendarski uzrast. Ukoliko se mentalni uzrast poklapa sa kalendarskim, reč je o detetu prosečnih intelektualnih sposobnosti, i ono će imati IQ = 100. Uopšteno rečeno, ukoliko je mentalni uzrast manji od kalendarskog, mentalne sposobnosti deteta su ispodprosečne, a ukoliko je mentalni uzrast veći od kalendarskog, onda su iznadprosečne. Međutim, opisani način određivanja nivoa intelektualnih sposobnosti je kritikovan (više u Kodžopeljić i Pekić, 2017; Urbina, 2004). Jedna od kritika koja mu se upućuje je da se ne

može primeniti na odraslima. Druga kritika se odnosi na to što je umni količnik nestalan na raznim uzrastima, te je stoga preporučen devijacioni umni količnik. tj. IQ-skor koji je izveden na osnovu standardnog skora (vidi str. 116).

Razvojne norme zasnovane na teorijskim pretpostavkama

Ova vrsta razvojnih normi, kao što joj i samo ime kaže, zasnovana je na teorijskim pretpostavkama u vezi s tim koja ponašanja deteta treba da se jave na određenom uzrastu, a koja govore o razvoju mentalnih sposobnosti, socio-emocionalnom razvoju, moralnom razvoju i slično. Najčešći primeri su teorija kognitivnog razvoja Žana Pijažea i teorija moralnog razvoja Lorenca Kolberga. Ova vrsta normi češće se koristi u istraživačke svrhe, nego za individualnu procenu. Upotreba ovih normi zavisi od toga koliko je teorija validna i koliko je primenljiva na datoj populaciji, odnosno referentnoj grupi.

Ostale vrste razvojnih normi

Postoje i druge vrste razvojnih normi, kao što su razredne norme ili razredni ekvivalenti (eng. *grade-equivalent*) ili uzrasne norme ili uzrasni ekvivalenti (eng. *age-equivalent*). Obe vrste normi su korisne za merenje i opis razvoja u mentalnim sposobnostima i kompetencijama kao što su čitanje, matematika i druge sposobnosti koje pokazuju prilično konzistentan obrazac rasta u okviru formalnog i institucionalizovanog obrazovanja (Mertler, 2007). Ove vrste normi bi se mogle podvesti pod razvojne norme zasnovane na bihejvioralnim pokazateljima u širem smislu, budući da su dobijene na osnovu tipičnog nivoa razvoja sposobnosti, kompetencija i veština na određenim uzrastima ili školskim razredima. Ipak, kako one imaju poseban način izražavanja skorova i češće se odnose na kognitivne pokazatelje, biće

označene kao posebna vrsta normi. **Razredne norme** ukazuju na prosečno postignuće u određenom razredu kao referentnoj grupi (npr. Oosterhof, 2001). Izražavaju se preko dva numerička parametra, od kojih prvi predstavlja razred, a drugi mesec tokom školske godine koji može da varira od 0 (početak škoske godine – septembar) do 9 (kraj školske godine – jun). Na primer, ako neki učenik trećeg razreda ostvaruje sirovi skor 15, koji odgovara razrednoj normi 3;8[21], to znači da njegovo postignuće odgovara tipičnom postignuću učenika koji rešava test u maju (8. mesec školske godine) u trećem razredu. Ujedno, to znači da skor 15 na testu za učenike trećeg razreda ima razredni ekvivalent 3;8 za tu referentnu grupu. Ukoliko bi učenik koji je 5. razred ostvario sirovi skor 15, to bi značilo da je njegov razredni ekvivalent drugi razred u 8. mesecu školske godine (2;8, vidi Tabelu 4). Dakle, na osnovu ovih normi se sirovi skor učenika izjednačava sa određenim razrednim ekvivalentom za referentnu grupu na osnovu čega se zaključuje o postignuću učenika. Razredni ekvivalent se može računati kao aritmetička sredina ili kao medijana.

Tabela 4

Primer prikaza razrednih normi

	Razred učenika				
	3;0	3;6	4;0	4;6	5;0
Sirovi skor	Razredni ekvivalent				
14	3;7	3;4	3;2	3;0	2;7
15	3;8	3;5	3;3	3;1	2;8
16	3;9	3;6	3;4	3;2	2;9
17	3;10	3;7	3;5	3;3	2;10

[21] Namerno se razred i mesec odvajaju tačkom i zarezom da se ne bi norma pomešala sa decimalnim zapisom, mada se u literaturi može naći i da se odvajaju tačkom.

Slične razrednim su i **uzrasne norme** koje pokazuju kalendarski uzrast u određenoj populaciji za koji dati sirovi skor predstavlja prosečnu vrednost ili medijanu. Drugim rečima, na osnovu uzrasnih normi utvrđuje se s kojom uzrasnom grupom se sirovi skor ispitanika može izjednačiti. Uzrasne norme se izražavaju preko godine i meseca uzrasta, pa samim tim raspon meseci može da varira od 0 do 12. Tako, ukoliko dete od 7 godina ostvaruje sirovi skor 10, a 10 u okviru normi za 7-godišnjake predstavlja medijanu postignuća za uzrasnu grupu od 7 godina i 6 meseci, možemo reći da je za ispitanikovo postiguće uzrasni ekvivalent 7;6. Ovakva vrsta normi primenjuje se, na primer, u okviru Bine-Simonove skale inteligencije.

Prednosti

Prednost razvojnih normi je u tome što su jednostavne za tumačenje kako stručnjacima tako i laicima. Takođe, pored poređenja sa vršnjacima, razvojne norme pružaju uvid u progres ispitanika iz godine u godine, te se u tom smislu mogu koristiti i za interpretaciju orijentisanu na ispitanika.

Nedostaci

Nedostatak razvojnih normi je to što one pripadaju ordinalnim skalama merenja. Čak iako su bazirane na aritmetičkim sredinama, a ne na medijanama, i dalje je skala izražavanja skorova ordinalna (npr. 6;1, 6;2, 6;3...), tj. intervali između njih nisu jednaki. Tako, učenik 7. razreda koji na setu zadataka čitanja ostvaruje razredni ekvivalent 6. razreda ne pokazuje isti kvantitet ni kvalitet zaostajanja u čitalačkoj kompetenciji kao učenik koji je 2. razred, a čiji je razredni ekvivalent 1. razred. S obzirom na to da se čitalačka kompetencija treba razviti u 1. razredu, zaostajanje u ovoj kompetenciji učenika koji je 2. razred je zapravo mnogo veće nego

kod učenika koji je 7. razred. Procentualno izraženo, to bi značilo da je zaostajanje učenika 2. razreda 50% (ova vrednost se dobija tako što se od 1 oduzme količnik razrednog ekvivalenta sa aktuelnim razredom učenika, što je u ovom primeru 1 – (1 / 2) = 1 – 0,50 = 0,50, odnosno 50%), a zaostajanje učenika 7. razreda 14% (1 – (6 / 7) = 1 – 0,86 = 0,14, više u Cermak, 1989). Ordinalna skala ima svoja ograničenja koja se odnose na nemogućnost primene parametrijske statistike. To znači da se za njih ne može računati aritmetička sredina, ni standardna devijacija. Ipak, postoje razvijeni statistički metodi za obradu podataka koji su nižeg nivoa od intervalnog, ali nisu uvek svi dostupni u komercijalnim softverima.

Pored toga, dešava se da se razvojne norme pogrešno shvataju kao standard koji neki učenik treba da postigne u određenom periodu školovanja, tj. zanemaruje se činjenica da se interpretacija vrši s obzirom na normu, a ne na kriterijum (Oostefhof, 2001). Na primer, testovi inteligencije za mlađi uzrast uobičajeno sadrže za svaki razred poseban set zadataka. Ukoliko neki učenik koji je testiran u 3. razredu ostvaruje postignuće ekvivalentno razrednoj normi 5;1, pogrešno bismo mogli zaključiti da je njegov kognitivni nivo na nivou nekoga ko je 5. razred, te da je taj učenik spreman za 5. razred. Budući da se set zadataka razlikuje od razreda do razreda, to samo znači da to dete ostvaruje visok rezultat u svojoj referentnoj grupi, tj. u 3. razredu. Takođe, kako za svaki razred ili uzrast postoji različit set pitanja, skorovi se ne mogu porediti između različitih razreda ili uzrasta, već samo *unutar* svoje referentne grupe (npr. grupe učenika 3. razreda). Na primer, ukoliko neko ko je 5. razred ostvari razredni ekvivalent 5;4, i ukoliko neko ko je 7. razred ostvari razredni ekvivalent takođe 5;4, to ne znači da je njihovo postignuće u istoj kategoriji jer se zadaci za 5. i za 7. razred međusobno razlikuju.

U praksi se javlja još jedan problem koji je vezan za organizaciju samog prikupljanja podataka. Zbog različitih razloga često se dešava da uzorak na kojem se radi standardizacija ne obuhvata sve uzraste ili razrede za svaki mesec (npr. nedostaju deca stara 6 godina i 3 meseca). Kako se pretpostavlja da razvoj pokazuje konzistentni rast, posebnim postupcima vrši se predikcija nedostajućih starosnih grupa. Iako su ovi statistički postupci prihvaćeni, mora se voditi računa o ispunjenju uslova za njihovu primenu kako bi tumačenje skorova bilo adekvatno, a pored toga, pretpostavka o sekvencijalnom rastu je upitna (više u Cermak, 1989; Nitko, 2004).

Na kraju treba napomenuti da se razvojni fenomeni ne izražavaju samo razvojnim normama, već se mogu koristiti i druge vrste normi. Na str. 84 opisan je kontinuirani način normiranja koji se takođe koristi kod merenja razvojnih fenomena, a mogu se koristiti i druge norme sve dok postoji dovoljna varijansa skorova unutar referentne grupe.

Percentili

Percentili predstavljaju najčešće korišćenu vrstu normi. Njihova popularnost je u lakoći interpretacije koja je bliska i laicima, što se mogu izračunati za bilo koju vrstu testa, i što su nezavisni u odnosu na to kakvu distribuciju tvore sirovi skorovi. Percentili su zasnovani na ordinalnoj skali, tj. na rangiranju. Kako rangiranje predstavlja najjednostavniji način poređenja skorova na testu (Jackson, 2003), ne čudi što su percentili dominantni način računanja normi.

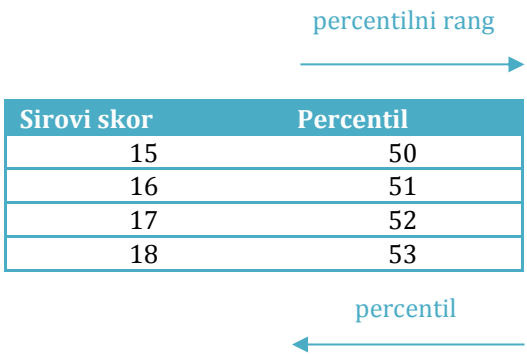
Prilikom određenja percentila, najpre treba napraviti razliku između percentilne skale, percentila i percentilnog ranga. **Percentilna skala** predstavlja skalu koja se dobija transformacijom frekvencije distribucije testovnih skorova u skorove koji sadrže po 1% ispitanika iz

referentnog uzorka. Transformacija koja se primenjuje menja distribuciju odgovora tako da ona bude uniformna – jedan percentil ili podeok na skali sadrži 1% ispitanika iz referentne grupe.

Percentilna skala služi, u stvari, za očitavanje percentila. Razlika između percentila (eng. *percentile*) i percentilnog ranga (eng. *percentile rank*) u literaturi je prilično nejasna i neretko se oba pojma koriste u istom značenju. Ipak, postoje razlike, a one se očitavaju u različitim načinima iskazivanja iste informacije – informacije o relaciji između sirovog skora i procenta ispitanika koji imaju isti ili skor ispod datog sirovog skora. Jednostavnije rečeno, razlika je u načinu i smeru interpretacije skora. **Percentil ili percentilni skor** (označava se kao *P*) je *skor* na testu ispod ili na kojem se nalazi određeni procenat ispitanika. Preciznije, percentil je vrednost na percentilnoj skali koja odgovara datom sirovom skoru na testu, a kojim se označava procenat skorova koji je isti ili niži od tog sirovog skora. Dakle, percentil je vezan za izvornu skalu skorova, tj. vezuje se za sirove skorove. Na primer, ako neko ostvari sirovi skor 15 na testu, možemo izračunati da taj skor odgovara 25. percentilu (što bi se označilo kao P_{25}), tj. možemo da kažemo da 25. percentil odgovara sirovom skoru 15. Ujedno, to znači da 25% ispitanih ima isti ili niži skor od 15. **Percentilni rang** (označava se kao *Pr*) je *procenat* ispitanika u distribuciji skorova koji imaju isti ili niži skor od datog. U navedenom primeru, percentilni rang ispitanika je 25, tj. sirovi skor 15 odgovara 25. percentilnom rangu (25% ispitanika ima isti ili niži skor od 15). Valja primetiti da je reč o inverznim operacijama. Dok se pri tumačenju percentila polazi od percentilne skale kako bi se dobio sirovi skor koji se vezuje za dati percentil, kod percentilnog ranga se kreće od sirovog skora kako bi se dobio procenat ispitanika sa istim ili nižim skorom (Slika 8). U svakom slučaju, percentil uvek ima istu vrednost kao i

percentilni rang, a razlika je samo u smeru tumačenja skora – da li se ona odnosi na poziciju na skali skorova ili na poziciju ispitanika.

Ne postoji slaganje u vezi s tim da li se percentil odnosi na skor na testu *ispod* kojeg se nalazi određeni procenat ispitanika, ili *na kojem ili ispod kojeg* se nalazi određeni procenat ispitanika (za detalje videti Barrett, 2011). Ukoliko želimo da odredimo tačan percentilni rang (npr. 79,14%), onda je primerenije percenil računati kao skor *ispod* kojeg pada određeni procenat ispitanika. Prema ovoj koncepciji, pretpostavlja se da su skorovi na testu kontinuirani, ali da su mereni kao diskretne, cele vrednosti. Ukoliko nam takva preciznost nije važna i ukoliko smo zadovoljni uvidom u to koliko ispitanika ima sa *istim ili skorom ispod* datog, onda i percentil tako računamo i definišemo. U ovom slučaju se pretpostavlja da su skorovi takvi kako su i zabeleženi, kao diskretne vrednosti. U praksi najčešće precizno određenje percentilnog ranga nije potrebno, te je u ovom udžbeniku usvojena definicija percentila po kojoj on predstavlja skor **na kojem ili ispod kojeg** se nalazi određeni procenat ispitanika.



Slika 8. Prikaz direkcije tumačenja percentila i percentilnog ranga.

Uzmimo, na primer, da Miljana ostvaruje sirovi skor 26 na testu numeričkih sposobnosti. Na osnovu pregleda kumulativnih frekvenci[22], možemo odrediti broj osoba koji ima isti ili niži skor od Miljaninog. Iz Tabele 5 možemo videti da je taj broj 17 (2 + 5 + 13). Da bismo našli percentil, moramo konvertovati kumulativne frekvence u procente koji se nazivaju kumulativnim procentima. Kumulativni procenti pokazuju ukupni procenat osoba sa istim i nižim skorovima, kako se krećete po skali. Na primer, na uzorku od 10 ispitanika ako dve osobe imaju skor 1 na nekom testu, a tri osobe imaju skor 2, kumulativna frekvenca za skor 2 je 2 + 3 = 5, a kumulativni procenat je 50% (10 : 5 = 100 : X; X = (5 x 100) / 10 = 50). U našem primeru, kumulativni procenat za Miljanin skor 26 je 85% i odgovara 85. percentilnom rangu (Tabela 5). U ovom slučaju, rekli bismo da 85. percentil odgovara sirovom skoru 26 i da sirovi skor 26 ima percentilni rang 85, odnosno da 85% ispitanika ima isti ili niži skor od Miljane. Na osnovu ovog zaključujemo da je Miljana prilično dobro uradila test, tačnije isto ili bolje od 85% drugih isptanika koji su radili isti test.

Tabela 5

Primer uporednog prikaza sirovih skorova, frekvenci i procenata

Sirovi skor	Frekvenca skorova	Proporcija skorova	Procenat skorova	Kumulativna frekvenca skorova	Kumulativi procenat
27	3	,15	15	20	100
26	**4**	**,20**	**20**	**17**	**85**
25	8	,40	40	13	65
24	3	,15	15	5	25
23	2	,10	10	2	10
...					

[22] Kumulativna frekvenca je frekvenca dobijena postupnim sabiranjem frekvenci prethodnih vrednosti. To je ukupna frekvenca pojavljivanja neke vrednosti i svih prethodnih (nižih) vrednosti.

Percentilni rang se koristi da se opiše svojstvo ispitanika, a ne grupna tendencija. Za interpretaciju percentilnog ranga potrebno je znati karakteristike referentne grupe s kojom se skor poredi. Na primer, nije isto ako neko na testu znanja ima percentilni rang 25 u grupi svojih vršnjaka iz razreda i u grupi učenika istog razreda iz cele zemlje. S druge strane, za interpretaciju percentila, poželjno je znati mogući opseg sirovih skorova. Na primer, ako 25. percentil na testu odgovara skoru 15, nije svejedno da li je maksimalni ostvareni skor na skali 20 ili 100. Iako je u pitanju isti percentil (25.), u prvom slučaju mnogo veći raspon sirovih skorova se nalazi do 25. percentila, nego u drugom slučaju, što nam govori o obliku distribucije skorova (u ovom primeru bi distribucija bila negativno zakošena).

Percentil i percentilni rang mogu se odrediti za svakog pojedinačnog ispitanika. U praksi, najmanji mogući percentil je 1, jer niko ne može dobiti skor kome nije dodeljen percentilni rang, a najveći je 99, jer niko ne može dobiti skor koji bi bio iznad svih skorova tj. iznad 100% ispitanika. Ako se neko nalazi na 20. percentilu, to znači da 20% ispitanih ima isti ili manji skor na istom testu, i posledično bi značilo da 80% ispitanika ima veći skor, međutim u praksi je to 79% (99 – 20).

Percentilni rang za konkretnog ispitanika može se dobiti na različite načine, a jedna od najkorišćenijih formula koja se sreće u udžbenicima je sledeća:

$$\text{Percentilni rang} = \frac{\text{Broj vrednosti ispod } X}{\text{Veličina uzorka}} \text{ x } 100$$

gde X predstavlja ostvareni sirovi skor ispitanika. Tako, ako imamo grupu od 75 ispitanika i ako neki ispitanik ostvaruje sirovi skor 40,

ispod kojeg se nalazi 50 ispitanika, njegov percentilni rang će biti (50 / 75) x 100 = 66,67.

Ovakav račun je moguć kada ne postoje ispitanici sa istim skorom. U slučaju kada je to dobijeno, tj. kada postoje tzv. združeni skorovi (eng. *tied scores*), postoji nekoliko načina njihovog tretiranja, a najjednostavnija je sledeća formula:

$$\text{Percentilni rang} = \frac{\text{Broj vrednosti ispod } X + \ 0{,}5 \text{ x broj istih skorova } X}{\text{Veličina uzorka}} \text{ x } 100.$$

U našem primeru, ukoliko imamo dva ispitanika sa istim skorom 40, njihov percentilni rang će biti (50 + 0,5 x 2) / 75 x 100 = 68, a ukoliko imamo tri ispitanika sa istim skorom 40, njihov percentilni rang će biti 68,67.

Percentili mogu da služe za podelu distribucije na dva ili više delova. Na primer, ukoliko želimo da podelimo skorove na testu na one koji su položili i pali neki test, onda ćemo ispitanike podeliti na one koji imaju percentil 50 ili ispod 50 i one koji imaju percentil 51 i više. Najjednostavnija podela je na četiri dela, tzv. kvartila (vidi Tabelu 6). U skladu sa ovim određenjem, možemo neki skor grubo protumačiti kao ispodprosečni, prosečni ili iznadprosečni.

Tabela 6

Podela na kvartile

Raspon percentila	Procenat skorova	Kumulativni procenat	Kvartil	Tumačenje
1–25	25%	25%	Q_1	ispodprosečno
26–50	25%	50%	Q_2	prosečno
51–75	25%	75%	Q_3	prosečno
76–99	25%	100%	Q_4	iznadprosečno

Pored podele na kvartile, postoji i podela na decile pri čemu opseg jednog decila obuhvata 10 rangova (do 10, od 11 do 20, od 21 do 30...). Decili se obično označavaju kao D1, D2 ... D10. Zbog česte potrebe za identifikacijom darovitih osoba i osoba s potencijalnim problemima, informativno je detaljnije znati skorove koji označavaju esktreme, te se u priručnicima često prvi i poslednji decil razlože (videti Tabelu 7). Ovo je najčešći način prikazivanja percentilnih normi u priručnicima.

Tabela 7

Primer prikaza percentila u priručnicima i njhov raspon

Prikaz u priručniku	Raspon percentila
1	1–2
3	3–4
5	5
10	6–10
20	11–20
...	...
80	80–89
90	90–94
95	95–96
97	97–98
99	99

Prednosti

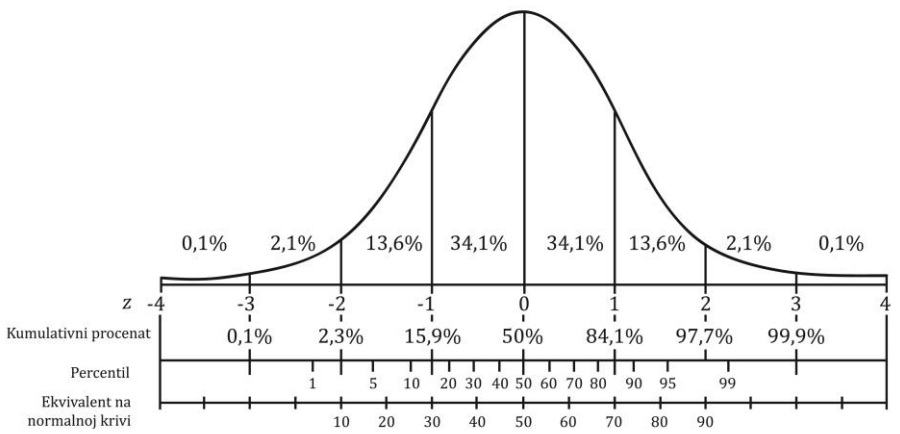
Postoje dve osnovne prednosti percentilnih normi. Prva je što su ove norme jednostavne za tumačenje i stručnjacima i laicima. Druga je što se ove norme mogu primeniti na svim testovnim skorovima, bez obzira na to kako su distribuirani. Naime, percentilne norme ne zahtevaju normalnu distibuciju kao što je to slučaj s nekim drugim vrstama normi kakvi su npr. standardni skorovi.

Nedostaci

Percentilne norme, kao i razvojne norme, potiču sa ordinalne skale merenja koja ima svoja ograničenja u pogledu interpretacije i primene metoda statističke obrade podataka. Ipak, važniji nedostatak ogleda se u samoj prirodi percentilne skale. Naime, razmak između percentila nije jednak duž čitave percentilne skale. Kada se skorovi distribuiraju normalno, postoji tendencija grupisanja skorova oko centralne vrednosti, dok se na ekstremima skale skorovi više rasipaju (vidi Sliku 9. i uporedi percentile sa ekvivalentima na normalnoj krivi, kada bi razmaci bili jednaki). Tako će, razmak između sirovih skorova u domenu prosečnih skorova biti manji, nego što je razmak u domenu ekstremnih skorova. Odnosno, razlika u percentilima izgleda prividno povećana u domenu prosečnih skorova, i prividno smanjena u domenu ekstremnih skorova. Na primer, iako je distanca između 1. i 2. percentila ista kao i između 49. i 50. i iznosi 1 percentil, razlika u sirovim skorovima između 1. i 2. percentila je veća, nego između 49. i 50. percentila. Razlika od jednog sirovog skora u skoru između ispitanika koji se nalaze oko proseka može rezultovati razlici od nekoliko percentila, dok ta ista razlika u sirovim skorovima između ispitanika koji se nalaze na ekstremima skale može rezultovati pripadnosti čak istom percentilu. Na primer, moguće je da sirovi skorovi 3, 4 i 5 odgovaraju percentilu 2, dok sirovi skorovi 25, 26 i 27 odgovaraju percentilima 43, 44 i 45 (Slika 9).

Kao nedostatak navodi se i to što će uvek neko imati najviši, a neko najniži percentil, tj. ekstremne vrednosti uvek postoje. U slučaju testiranja veće populacije, to može biti problematično za interpretaciju – može se desiti da za dobijeni sirovi skor nemamo normu u okviru izrazito visokih ili izrazito niskih skorova. Kada se to dešava u okviru

izrazito visokih skorova, reč je o tzv. **efektu "plafona"** (eng. *ceiling effect*) koji predstavlja fenomen kada merni instrument više ne može da registruje stvarni rast, pa je više sirovih skorova smešteno na gornji ekstrem skale. Na primer, moguće je da na testu čiji je teorijski maksimum 30, neko ostvari taj teorijski maksimum, a u normama imamo samo vrednost 99. percentila koja odgovara sirovom skoru 28, te ne možemo precizno utvrditi kom percentilu pripada sirovi skor 30. Ovaj efekat nije ograničen samo na percentile i može se javiti i kod drugih vrsta normi, ali je kod percentila češći. Efekat "plafona" je čest prilikom primene testova sa zastarelim normama, a u slučaju testiranja inteligencije povezuje se sa Flinovim efektom. Moguće je, takođe, da veći broj ispitanika ostvaruje maksimalni teorijski skor, skorove oko teorijskog maksimuma ili generalno visoke skorove, te ćemo imati gomilanje skorova oko visokih vrednosti, tj. varijansa u ovom domenu skorova će nam biti mala. S obzirom na to, instrument neće biti dovoljno precizan u domenu visokih skorova.



Slika 9. Raspodela percentila.

S druge strane, kod izrazito niskih skorova može se registrovati tzv. **efekat "poda"** (eng. *floor effect, basement effect*) kada se ne može precizno odrediti neko izrazito nisko postignuće, kada veći broj ispitanika ostvaruje teorijski minimalni skor, skorove oko teorijskog minimuma ili skorove smeštene na donji ekstrem skale. Drugim rečima, preciznost merenja je niska u okviru niskih skorova. Ukoliko bismo test inteligencije za odrasle dali deci, dobili bismo ovaj efekat. U praksi, ovaj efekat se može sresti kada se klinički ili supklinički testovi daju pripadnicima nekliničke (opšte) populacije. Na primer, u slučaju zadavanja Skale depresivnosti, anksioznosti i stresa (DASS-21) na populaciji studenata, dobijeno je da najniži teorijski skor (skor nula) na ove tri skale ostvaruje redom 28%, 23,4% i 6,3% ispitanika (Jovanović, Gavrilov-Jerković, Žuljević i Brdarić, 2014). Ponovo, valja samo napomenuti da je efekat "poda" moguć i kod drugih vrsti normi, ali da je češći kod percentilnih normi. Prema nekim autorima, efekti "plafona" ili "poda" prisutni su kada minimum 15% ispitanika ostvaruje najviši ili najniži mogući skor (McHorney & Tarlov, 1995).

Postoji još jedan nedostatak percentilnih normi. Ukoliko je uzorak ispitanika mali, desiće se da svaki percentil nema svog parnjaka među sirovim skorom. Na primer, moguće je da na nekom testu sirovi skor 16 odgovara 23. percentilu, a da sirovi skor 17 odgovara 26. percentilu, pa tako ostajemo bez 24. i 25. percentila, tj. ovi percentili neće imati parnjaka među sirovim skorovima.

Kao oprez navodi se to da se percentilni rangovi ne mešaju sa procentom uspešnih odgovora na testu. Iako ovo nije česta zabuna, valjalo bi skrenuti pažnju na to. Naime, ukoliko neko ostvaruje skor 40 na testu čiji je teorijski maksimum 80, onda će on imati 40 x 100 / 80 = 50% tačno urađenih zadataka, a njegov percentilni rang će biti 67.

Nekada se može desiti da su ove dve brojke iste, ali percentilni rangovi se odnose na poređenje skora sa ostalim ispitanicima, a procenat tačnog odgovora na poređenje skora sa maksimalnim mogućem skorom na testu (bez obzira na to kakvo postignuće imaju ostali ispitanici). Uostalom, iz same formule za percentilni rang i procenat tačnih odgovora se može primetiti da je delilac različit: u slučaju percentilnog ranga je to veličina uzorka, a u slučaju procenta tačnih odgovora je to teorijski maksimum koji se može ostvariti na testu.

Standardni skorovi

Standnim skorovima transformišemo sirove skorove tako da oni uvek imaju istu vrednost za aritmetičku sredinu i za standardnu devijaciju. Njima se, u stvari, određuje odstupanje sirovog skora od aritmetičke sredine u funkciji varijabilnosti skorova, odnosno u funkciji standardne devijacije. Na taj način za svaki skor na bilo kom testu možemo tačno odrediti lokciju unutar distribucije skorova, a u odnosu na aritmetičku sredinu, bez obira na to kakvi su varijabilnost i opseg skorova na testu. Ovim postupkom dobijamo tzv. standardni skor (termin "standard" je od standardna devijacija), koji se još naziva i standardizacioni ili *z*-skor (termin "standardizacioni" je od operacije standardizacije).

Standardni skorovi se dobijaju primenom određene tansformacije sirovih skorova. Transformacija koja se primenjuje može biti linearna ili nelinearna. Linearnim transformacijama se ne menja oblik distribucije skorova, ali se menja skala izražavanja skorova ili jedinica mere. Dakle, kada se primeni linearna transformacija, odnosi između sirovih skorova bivaju isti kao odnosi između standardnih

skorova. Nelinearnim tranformacijama se menjaju i skala izražavanja skorova i oblik distribucije skorova.

Jedna od linearnih transformacija je upravo standardizacija i njome se dobija osnovni standardni ili *z*-skor, a na osnovu njega se mogu izvesti drugi standardni skorovi (npr. T-skor ili IQ-skor). Objasnićemo najpre *z*-skor:

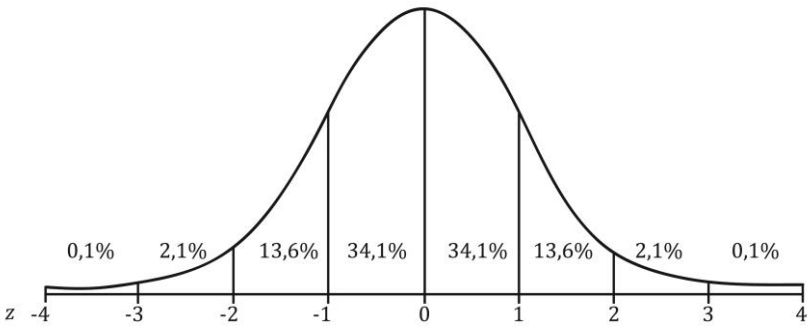
***z*-skor** je osnovni standardni skor i standardizacija je uobičajeno prva transformacija koja se primenjuje nad skorovima, kada je distribucija skorova normalna. Kao što je rečeno, osnovni pokazatelji distribucije su aritmetička sredina i standardna devijacija i standardni skorovi uvek imaju istu vrednost za artimetičku sredinu i za standardnu devijaciju. Kod *z*-skorova aritmetička sredina je 0, a standardna devijacija je 1. To nam omogućava da skorove izražavamo u terminima standardne devijacije. Kada je originalna distribucija skorova normalna, skorovi u termina standardne devijacije su jednaki duž celog kontinuuma distribucije, odnosno svaka standardna devijacija obuhvata isti procenat susednih skorova. Drugim rečima, intervali između skorova su jednaki duž celog kontinuuma skale, što čini da standardni skorovi pripadaju intervalnoj skali merenja. S obzirom na to da su intervali skorova jednaki, nema potrebe za prethodnim "peglanjem" distribucije (podsetićemo se, standardizacija je takva transformacija koja ne menja postojeće odnose između sirovih skorova).

Distribucija *z*-skorova je simetrična, pri čemu pozitivan *z*-skor ukazuje na to da dati skor ima vrednost iznad aritmetičke sredine, a negativan *z*-skor ukazuje na to da dati skor ima vrednost ispod aritmetičke sredine. Tako, znak *z*-skora ukazuje na njegovu poziciju od aritmetičke sredine, a vrednost na udaljenost od aritmetičke sredine. Na

primer, *z*-skor 1,5 ukazuje na to da je sirovi skor koji mu pripada veći od artimetičke sredine i da je za 1,5 *SD* jedinica veći od artimetičke sredine.

Iako teorijski raspon *z*-skorova nije ograničen, u praksi, 99% skorova na nekom testu nalazi se u intervalu od -3 do +3 *SD*. Imajući u vidu karakteristike normalne distribucije, skorove možemo grubo interpretirati kao ispodprosečne (ispod -1 *SD*), prosečne (u rasponu od -1 do +1 *SD*) ili iznadprosečne (preko +1 *SD*). Finiju podelu možemo dobiti ako ispodprosečne skorove podelimo na ispodprosečne (od -2 *SD* do -1 *SD*) i one veoma ispodprosečne (ispod -2 *SD*), i ako iznadprosečne podelimo na iznadprosečne (od +1 *SD* do +2 *SD*) i veoma iznadprosečne (preko +2 *SD*).

Kao što se može videti na Slici 10, raspon od -1 do +1 *SD* sadrži oko 68% skorova (tačnije, 68,26%), od -2 do +2 *SD* oko 95% skorova (tačnije, 95,44%) i od -3 do +3 *SD* oko 99% skorova (tačnije, 99,73%). Takođe, može se videti da se 50% skorova nalazi ispod i 50% iznad aritmetičke sredine, što je u skladu sa karakteristikom normalne distribucije da su joj aritmetička sredina i medijana iste. Kada dalje pogledamo distribuciju skorova, možemo primetiti da iznad 1 *SD* ima oko 16% skorova, iznad 2 *SD* oko 2% i iznad 3 *SD* oko 0,15%. Isti odnos važi i za vrednosti ispod -1 *SD*, -2 *SD* i -3 *SD*.



Slika 10. Raspodela *z*-skorova.

Uzmimo primer da je aritmetička sredina na nekom testu 40, a standardna devijacija 5. Ukoliko ispitanik ima skor na testu 46, njegov standardni skor možemo izračunati pomoću sledeće formule:

$$z = \frac{X - AS}{SD}$$

u kojoj je *X* sirovi skor, *AS* je aritmetička sredina testa i *SD* je standardna devijacija testa. Primenom formule izračunavamo ispitanikov *z*-skor na sledeći način: (46 – 40) / 5 = 1,2. U terminima standardne devijacije, ovaj skor je 1,2 *SD* iznad proseka. Da test ima drugačiju distribuciju skorova, npr. da je njegova aritmetička sredina 50 i standardna devijacija ista (5), onda bi sirovi skor 46 ogovarao *z*-skoru -0,8 [(46 – 50) / 5] i interpretirao bi se kao na donjoj granici proseka. Isto tako, da je aritmetička sredina testa 35 uz istu standarnu devijaciju (5), onda bi sirovi skor od 46 odgovarao *z*-skoru 2,2 [(46 – 35) / 5] i interpretirao bi se kao visoko iznadprosečni skor. Lako je uočiti da, ako sirovi skor odgovara aritmetičkoj sredini, njegov *z*-skor će biti nula.

Ukupni z-skor

Ponekad, kada imamo više testova u okviru baterije, informativno je i preporučljivo dobiti skor na celoj bateriji testova. Jedan od načina kako se to može uraditi je da se izračuna ukupni ili sumacioni *z*-skor (u priručnicima se on često označava sa z_S). To bi značilo da se za svaki test u bateriji izračuna z-skor, te da se dobijeni *z*-skorovi saberu. Na taj način se može dobiti IQ skor za celu bateriju testova.

U našoj praksi, ukupni z-skor je predviđen npr. u okviru baterije kognitivnih testova KOG 3, koja sadrži tri testa: IT-1 za procenu

perceptivnih sposobnosti, AL-4 za procenu verbalnih sposobnosti i S-1 za procenu spacijalnih sposobnosti (Momirović, Wolf i Džamonja, 1992). Tako, ako neki ispitanik postiže na ovim testovima redom *z*-skorove 0,04, 1,11 i 0,55, njegov ukupni *z*-skor će biti $z_S = z_1 + z_2 + z_3 = 0,04 + 1,11 + 0,55 = 1,70$. Na osnovu normi datih u priručniku može se očitati da ovaj ukupni *z*-skor odgovara IQ skoru od 109. Ponekad se, takođe, dešava da nemaju svi testovi u bateriji istu važnost u objašnjenju ukupnog skora. To se može videti na osnovu njihovih opterećenja na glavnom faktoru ili na osnovu regresionih koeficijana prilikom predviđanja nekog kriterijuma. U tom slučaju, pribegava se ponderisanju *z*-skorova pojedinačnih testova. Na primer, ako tri testa baterije imaju opterećenja na glavnom faktoru redom 0,85, 0,76 i 0,54, onda se umesto izraza $z_S = z_1 + z_2 + z_3$ može primeniti sledeći izraz $z_S = 0,85 * z_1 + 0,76 * z_2 + 0,54 * z_3$, što bi u datom primeru bilo 1,17. Ovakav primer računanja u našoj praksi može se naći u i slučaju Verbalne serije Borislava Stevanovića, koja se sastoji od nekoliko testova za procenu različitih verbalnih sposobnosti (Stevanović, Dragičević, Štajnberger i Bukvić, 1988).

Prednosti

Prednosti koje važe za *z*-skorove su iste one koje važe za sve vrste standardnih skorova. Prva prednost je ta što je njima rešen problem nejednakih intervala koji je prisutan kod percentila. U okviru standardnih skorova, udaljenost nekog skora od aritmetičke sredine je u funkciji varijabilnosti skorova, tj. u funkciji standardne devijacije. Drugim rečima, standardni skorovi potiču s intervalne skale. Ova karakteristika pruža još jednu prednost, a to je primena parametrijske statistike nad tim skorovima. Dakle, na standardnim skorovima može se

računati aritmetička sredina i standardna devijacija. Treća prednost je mogućnost poređenja rezultata s različitih testova, pod uslovom da se sirovi skorovi na tim testovima normalno distribuiraju. Standardnim skorovima se rezultati na različitim testovima prevode na istu skalu, bez obzira na to što oni imaju različite pokazatelje originalne distribucije (*AS* i *SD*) i opseg skorova.

Pored ovih opštih prednosti koje važe za sve standardne skorove, specifična prednost konkretno *z*-skorova je i u tome što omogućavaju direktan uvid u smer i nivo udaljenosti datog skora od proseka. Kao što je rečeno, ako neko ima *z*-skor 1,5, na osnovu predznaka odmah znamo da je njegov skor iznad proseka, a na osnovu veličine znamo da je udaljen 1,5 *SD* od proseka. Znajući da prosečne vrednosti obuhvataju raspon od -1 do +1, odmah možemo zaključiti i da je u pitanju iznadprosečni skor, a ne npr. visoko iznadprosečni. Takođe, *z*-skorovi omogućavaju dobijanje kombinacije skorova s više testova u ukupni, sumcioni *z*-skor skor (z_S), što odgovara skoru na bateriji testova.

Nedostaci

Nedostatak generalo svih standardnih skorova koji se dobijaju linearnom transformacijom je taj što oni nisu adekvatni u slučaju kada sirovi skorovi na testu ne tvore normalnu distribuciju. Dakle, primena *z*-skorova ograničena je na testovne skorove koji se distribuiraju normalno.

Specifični nedostatak *z*-skorova je da su nejasni laicima i otežavaju komunikaciju između korisnika testa i ispitanika. Postoje dva osnovna izvora nejasnoće tumačenja *z*-skorova kod laika. Najpre, veoma je teško objasniti nekome da ima negativan skor na testu, a po definiciji, polovina *z*-skovora ima negativni predznak. Zamislite situaciju u kojoj Vam dolazi učenik na razgovor u vezi sa svojim postignućem na testu

inteligencije, i Vi treba da mu saopštite da on na testu inteligencije ostvaruje standardni skor -1,0. Možete i sami pretpostaviti da će informacija da neko ima negativni skor na testu inteligencije negativno uticati na njega, u ovom slučaju na motivaciju učenika, pa i na njegovo samopoštovanje. Drugi izvor nejasnoće potiče od uskog raspona celih vrednosti *z*-skorova, što za posledicu ima skorove u decimalnom zapisu. Ponovo, zamislite situaciju u kojoj jedan učenik na nekom testu inteligencije ostvaruje standardni skor 0,65, dok njegov drug iz klupe ostvaruje skor 0,95. Da li će učenik koji ima standardni skor 0,65 pomisliti da je njegov nivo inteligencije manji od nivoa inteligencije njegovog druga iz klupe?

Izvedeni standardni skorovi

Imajući u vidu navedene nedostatke *z*-skorova, a da bi se olakšala interpretacija, predloženi su tzv. izvedeni standardni skorovi. Njihova prednost je u tome što nemaju negativne vrednosti i decimalne zapise, te su laicima lakši za razumevanje. Svi izvedeni standardni skorovi računaju se po sledećoj formuli:

$$\text{Izvedeni skor} = AS_{\text{nova}} + SD_{\text{nova}} * z$$

pri čemu različiti izvedeni standardni skorovi imaju različite novopredložene *AS* i *SD*. Postoji nekoliko vrsta izvedenih standardnih skorova, a među njima su svakako najpoznatiji T-skor i IQ-skor.

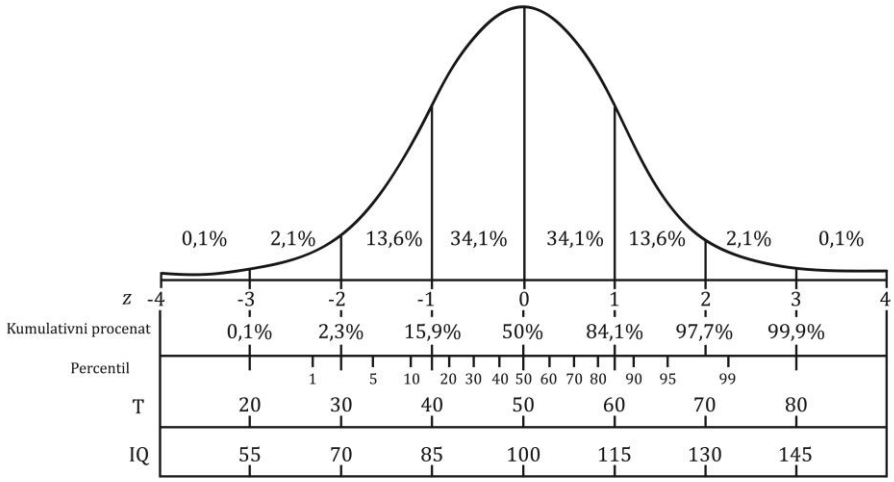
T-skor

T-skor je prvi predložio Mekol 1922. godine (videti u Chadha, 2009), a naziv je dobio po Torndajku. Originalno, Mekol je imao nameru

da razvije sistem za izražavanje mentalnog postignuća u jednakim jedinicama. Sugerisao je da se nasumični uzorak 12-godišnjaka testira kako bi se dobila distribucija njihovih skorova, i potom za svaki skor pripisao percentil kako bi se dobio percentilni rang. Na ovako uspostavljenom sistemu bi aritmetička sredina bila podešena da bude 50, što korespondira 50. percentilu, a standardna devojacija bi bila 10. Iako Mekolova namera nije bila da kreira sistem alternativni standardnim skorovima, ovaj sistem je potpuno isti sistemu standardnih skorova, uz razliku što je aritmetička sredina 50 (umesto nule) i standardna devijacija 10 (umesto 1). T-skor se lako može dobiti iz *z*-skora prema sledećoj formuli:

$$T = 50 + 10 * z.$$

Ukoliko, na primer, ispitanik ima *z*-skor -0,6, onda će njegov T-skor biti 50 + 10 * (-0,6) = 50 – 6 = 44. S obzirom na to da se 99% *z*-skorova nalazi u opsegu od -3 do +3, ako odredimo koji T-skor pripada ovim graničnim vrednostima, tj. *z*-skoru -3 i *z*-skoru +3, onda možemo dobiti raspon T-skorova koji obuhvata 99% skorova. Lako je izračunati da je taj raspon od 20 [50 + 10 * (-3) = 50 – 30 = 20] do 80 [50 + 10 * 3 = 50 + 30 = 80]. Isto tako, lako možemo zaključiti i da *z*-skor 0 odgovara T-skoru 50 (Slika 11). Na ovom mestu samo treba uočiti da raspon T-skorova može ići i od 1 do 100, što bi možda bilo intuitivno, ali bi taj raspon obuhvatao vrednosti od otprilike -5 do +5 *z*-skorova. Kako se većina skorova nalazi u rangu od -3 do +3 *z*-skorova, u praksi je raspon T-skorova najčešće od 20 do 80.



Slika 11. Uporedni prikaz raspodele percentila, *z*-skorova i izvedenih standarnih skorova (T i IQ).

Prilikom tumačenja T-skorova postoji nekoliko preporuka, odnosno nekoliko ponuđenih graničnih vrednosti za određivanje kategorije ispitanika (Tabela 8). **Prvi pristup** sledi logiku interpretacije *z*-skorova po kojoj se prosečne vrednosti nalaze u okviru raspona od -1 do +1 *SD*, tj. raspona od -1 do +1 *z*-skora, što odgovara rasponu T-skorova od 41 do 60. Prema **drugom pristupu**, opseg prosečnih T-skorova je nešto uži i kreće se od 45 do 55, što odgovara rasponu od -0,5 do +0,5 *SD*. Ne postoje jasni razlozi za i protiv jednog ili drugog pristupa prilikom tumačenja T-skorova, ali se pretpostavlja da autori pribegavaju drugom pristupu kada skorovi na testu nemaju dovoljno veliku varijabilnost.

Tabela 8

Pristupi za tumačenje kategorija ispitanika na osnovu T-skorova

Kategorija ispitanika	I pristup	II pristup
Veoma ispodprosečni skor	≤ 30	≤ 34
Ispodprosečni skor	31–40	35–44
Prosečni skor	41–60	45–55
Iznadprosečni skor	61–70	56–65
Veoma iznadprosečni skor	≥ 71	≥ 66

T-skorovi su uobičajeniji za prikaz rezultata na testovima tipične aktivnosti, kakvi su testovi ličnosti. Na primer, T-skorovi se koriste pri interpretaciji MMPI-a (npr. verzije MMPI-202 koja se kod nas koristi, videti Biro, 1995), inventara Velikih pet plus dva (Smederevac, Mitrović i Čolović, 2010), itd.

Prednosti

Sve prednosti koje važe za standardne skorove generalno, važe i za T-skorove. Specifična prednost T-skorova je u tome što su vrednosti pozitivne, te je laicima lakše za tumačenje u odnosu na napred pomenute *z*-skorove.

Nedostaci

Osnovni nedostatak koji važi za standardne skorove dobijene linearnom transformacijom, važe i za T-skorove – ograničeni su na primenu na testovnim skorovima koji tvore normalnu distribuciju.

Pored toga, kako je raspon T-skorova u praksi najčešće od 20 do 80, dešava se da se oni nekada greškom interpretiraju kao procenti, pa tako ako neki učenik postigne T-skor 60 na nekom testu postignuća, greškom se može protumačiti da je uradio tačno 60% zadataka na testu. No, ovo je veoma retka greška i otklanja se adekvatnim objašnjenjem interpretacije skorova. Takođe, ova potencijalna greška se može otkloniti i

tako što se T-skorovi neće koristiti prilikom prikazivanja postignuća na testovima sposobnosti, već će se koristiti neki drugi skorovi, npr. IQ-skorovi.

IQ-skor

IQ-skor pruža lokaciju sirovog skora na distribuciji čija je aritmetička sredina 100 i standardna devijacija 15. Samim tim, i IQ-skor se lako može dobiti iz *z*-skora prema sledećoj formuli:

$$IQ = 100 + 15 * z$$

U pomenutom primeru, ako ispitanik ima *z*-skor -0,6, onda će njegov IQ-skor biti 100 + 15 * (-0,6) = 100 – 9 = 91. Vodeći se istom logikom kao i za određenje raspona T-skorova, ukoliko odredimo kojim IQ-skorovima odgovara vrednost -3 i +3 *z*-skora, možemo zaključiti da je raspon IQ-skorova od 55 [100 + 15 * (-3) = 100 – 45 = 55] do 145 [100 + 15 * 3 = 100 + 45 = 145]. Lako možemo zaključiti i da *z*-skor 0 odgovara IQ-skoru 100 (videti Sliku 11). Uobičajeno je tumačenje da se prosečni IQ-skorovi nalaze u rasponu od 90 do 109 (Tabela 9). Može se primetiti da je ovaj raspon uži od raspona od -1 do +1 *z*-skora, tj. odgovara rasponu od -0,67 do +0,67 *z*-skorova. Razlog tome je što je pretpostavljeno da 50% populacije ima prosečni IQ skor, a 50% površine ispod normalne krive (s obe strane prosečne vrednosti tj. s obe strane od $z = 0$) odgovara *z*-skorovima -0,68 i 0,68. IQ vrednosti koje odgovaraju ovim *z*-skorovima iznose upravo 90 i 109, što je postavljeno kao interval IQ-skorova za prosečne vrednosti.

Tabela 9

Tumačenje kategorija ispitanika na osnovu IQ-skorova prema Vekslerovom testu inteligencije

Kategorija ispitanika	IQ
Mentalna zaostalost	69 i niže
Veoma ispodprosečni skor (granični)	70–79
Ispodprosečni skor	80–89
Prosečni skor	90–109
Iznadprosečni skor	110–119
Visoko iznadprosečni skor (superiorni)	120–129
Veoma visoko iznadprosečni skor (visoko superiorni)	130 i više

Valja napomenuti da postoje IQ-skorovi koji koriste drugačije standardne devijacije, npr. 16 ili 12, i ova neujednačenost može drastično uticati na interpretaciju skorova. Na primer, ako neko ostvaruje skor 32 na testu inteligencije čija su *AS* = 35 i *SD* = 4, njegov *z*-skor će biti -0,75. U slučaju da se IQ-skorovi računaju uz *SD* = 15, ovaj ispitanik bi imao IQ = 88,75 i okarakterisao bi se kao ispodprosečan. Ukoliko bi se IQ računao uz *SD* = 12, ovaj ispitanik bi imao IQ = 91 i okarakterisao bi se kao prosečan. Isto tako, nije moguće porediti postignuće na različitim testovima kod istog ispitanika ukoliko su ona računata na osnovu različito dobijenih IQ jedinica. Stoga treba voditi računa na koji način su izračunate IQ vrednosti i to dokumentovati prilikom pisanja izveštaja. U praksi se najčešće sreću IQ-skorovi čija je SD = 15, u skladu s tim kako ih je predložio Veksler u okviru Vekslerovog individualnog testa inteligencije (kod nas se koristi adaptacija ovog testa – VITI, namenjena ispitivanju inteligencije odraslih, više u Berger, Marković i Mitić, 1995, a u priremi je adaptacija WAIS-IV, više u Lazarević, Knežević, Mitić, & Jočić, 2018; u slučaju ispitivanja inteligencije kod dece, koristi se REVISK, Biro, 1997).

Izbor *AS* i *SD* za izvedene skorove je krajnje proizvoljan, ali su se neki skorovi odomaćili u upotrebi kao što su upravo opisani T-skorovi i IQ-

skorovi koji koriste *SD* = 15. Uobičajeno, IQ skorovi se računaju kada su norme potrebne za test sposobnosti, tj. za testove koji pripadaju prvoj paradigmi merenja, a T-skorovi se računaju kada su norme potrebne za test ličnosti, tj. za testove koji pripadaju drugoj paradigmi merenja.

Prednosti i nedostaci

Sve prednosti i nedostaci koje važe za standardne i T-skorove, važe i za IQ-skorove.

Normalizovani standardni skorovi

Normalizovani standardni skorovi koriste se kada postoje teorijska očekivanja da se skorovi normalno distribuiraju, ali iz nekog razloga se ne distribuiraju tako (npr. uzorak nije reprezentativan). Dakle, sam kontrukt koji je predmet merenja bi trebalo da se normalno distribuira, ali se skorovi na testu ne distribuiraju normalno. Normalizovani standardni skorovi su, u stvari, standardni skorovi koji su ekvivalentni sirovim skorovima da je distribucija skorova normalna (Cermak, 1989). Ovi skorovi se dobijaju tako što se prvo izvrši normalizacija kako bi se distribucija skorova transformisala u normalnu ("ispeglala"), pa se onda izvrši standardizacija.

Kada su uzorci manji (do 200 ispitanika), o normalnoj distribuciji možemo donositi zaključke na osnovu Šapiro-Vilkovog (Shapiro-Wilk) i Kolmogorov-Smirnovljevog testa (Kolmogorov-Smirnov). Ovi testovi polaze od pretpostavke da se empirijski skorovi ne razlikuju značajno od normalne distribucije, te ukoliko razlike skorova nisu značajne, time se potvrđuje da empirijski podaci ne odstupaju značajno od normalne distribucije. Međutim, mana ovih testova je u tome što su osetljivi na veličinu uzorka, pa s povećanjem uzorka postaju osetljivi i na male promene, te imaju tendenciju da budu značajni (Tabachnick & Fidell, 2013).

Kako su uzorci na kojima se računaju norme skoro uvek veći od 200, preporučuje se da se izračunaju vrednosti skjunisa i kurtozisa, odn. zakrivljenosti/asimetričnosti i spoljoštenosti/izduženosti distribucije. Ukoliko je distribucija normalna, skjunis i kurtozis ne bi trebalo da prelaze ± 1,5 (Tabachnick & Fidell, 2013), a prema nekim autorima ± 2 (Field, 2009). Ipak, svi autori preporučuju da se pogleda sam oblik distribucije i na osnovu vizuelne inspekcije odluči da li je distribucija normalna ili ne.

Ukoliko se zaključi da se skorovi ne distribuiraju normalno, kao što je rečno, oni treba najpre da se normalizuju. Postoji nekoliko načina za vršenje normalizacije skorova. Prvi način je da se na sirovim skorovima primeni neka od transformacija normalizacije. Na primer, u komercijalnom statističkom softveru IBM® SPSS® Statistics sofware[23] (u daljem tekstu – SPSS) dostupne su sledeće forumule za normalizaciju: Blomova formula (Blom), Takijeva formula (Tuckey), rankit i Van der Vardenova formula (Van der Waerden), a među njima se najčešće preporučuje rankit (Solomon & Sawilowsky, 2009). Nakon dobijanja normalizovanih skorova, oni se standardizuju, pa se može primeniti formula za izračunavanje izvedenih standardnih skorova.

Drugi način normalizacije je preko kumulativnih procenata (Urbina, 2004). Postupak se sastoji iz tri koraka. Prvi korak se odnosi na računanje procenata ispitanika za svaki sirovi skor, na osnovu kojih se pravi tabela s kumulativnim procentima, kao u primeru u Tabeli 10. Drugi korak odnosi se na konvertovanje procenata u proporcije, tako što se dobijeni procenti podele sa 100. Treći korak podrazumeva da se dobijene proporcije lociraju u okviru površine ispod normalne krive, kako bi se dobili odgovarajući *z*-skorovi.

[23] SPSS Inc. was acquired by IBM in October, 2009.

Tabela 10

Primer dobijanja normalizovanih standardnih skorova

Sirovi skor	*f*	%	1. korak kumulativni %	2. korak proporcija	3. korak norm. *z*-skor	T-skor
5	1	0,1	0,1	,00	-3,09	19
6	4	0,6	0,7	,01	-2,46	25
7	2	0,3	1,0	,01	-2,33	27
8	6	0,8	1,8	,02	-2,10	29
9	5	0,7	2,5	,03	-1,96	30
...						
25	32	4,4	92,1	,92	1,41	64
26	20	2,8	94,9	,95	1,64	66
27	16	2,2	97,1	,97	1,90	69
28	12	1,7	98,8	,99	2,26	73
29	9	1,2	100,0	,99*	2,33	73
Ukupno	723	100,0				

Napomena: *kako bi normalizovan standardizovani skor mogao da se izračuna, vrednosti u 2. koraku (proporcije) ne mogu biti tačno 0 ili 1, te je vrednost 1 (proporcija od 100) zamenjena sa ,99.

Postoji nekoliko načina kako da se proporcije prevedu u *z*-skorove. U statističkim udžbenicima se obično daju posebne tabele u kojima se ova vrednost očitava, i ovaj način može biti koristan ako nas interesuje skor pojedinačnog ispitanika. Dostupni su nam i onlajn kalkulatori[24] na osnovu kojih lako možemo za pojedinačnog ispitanika da dobijemo njegov *z*-skor. Međutim, ukoliko želimo da napravimo norme, onda će svakako bolja opcija biti da to izračunamo preko komande u statističkim softverima, za šta je potrebno da raspolažemo svim sirovim skorovima. Na osnovu formule za prevođenje kumulativnih proporcija u normalizovan skor, možemo dobiti vrednosti za svaki sirovi skor, a potom dalje računati standardne i izvedene standardne skorove (za detalje postupka u SPSS-u videti Priručnik na kraju ove knjige).

[24] https://www.zscorecalculator.com/

U literaturi se opsuje još jedan način dobijanja normalizovanih standardnih skorova preko percentila (Furr & Bacharach, 2014). Koraci su slični kao u prethodno opisanoj proceduri. U prvom koraku dobijaju se percentili iz svih sirovih skorova (umesto kumulativnih procenata kao u prethodnom primeru), pa se oni dele sa 100 da bi se dobile proporcije, i na kraju prevode u *z*-skorove na osnovu površine ispod krive. U oba slučaja, kada se dobiju *z*-skorovi, onda se mogu izračunati izvedeni standardni skorovi.

Važno je naglasiti da se svojstva normalizovanih standardnih skorova ne razlikuju od standardnih skorova koji su dobijeni linearnom transformacijom, ali bi uvek trebalo naglasiti da je reč o normalizovanim standardnim skorovima, kako bi se korisniku testa signaliziralo da potiču od skorova koji se ne distribuiraju normalno.

Prednosti i nedostaci

Sve prednosti i nedostaci koje važe za standardne skorove, važe i za normalizovane standardne skorove.

Stenajn skorovi

Stenajn skorovi su originalno nastali za potrebe vojske Sjedinjenih Američkih Država tokom Drugog svetskog rata. Termin stenajn (eng. *stanine*) je izvedenica od "standardnih devet" (eng. *standard nine*), pošto se skorovi izražavaju kao jednocifreni brojevi na skali od 1 do 9, pri čemu je skor 5 lociran u centru distribucije. Ovakav sistem omogućavao je lakše izračunavanje normi u odnosu na npr. T-skorove, što je bila prednost u doba kada su skorovi predloženi. U današnje vreme lakoća izračunavanja nije relevantna.

Postoje dva pristupa u tumačenju stenajn skorova. Prema **prvom pristupu**, skorovi se mogu interpretirati u odnosu na tri kategorije kao ispodprosečni, prosečni i iznadprosečni skorovi (Airasian, 2005). Prema **drugom pristupu**, skorovi se mogu interpretirati u odnosu na pet određenja (Domino & Domino, 2006), pri čemu su prosečni skorovi isti kao kod pristupa sa tri kategorije, ali je podela finija u domenu ekstrema. U slučaju niskih skorova, u okviru drugog pristupa tumačenja stenajn skorova, mogu se razlikovati oni koji imaju ispodprosečni skor i oni koji imaju veoma niski ili slabi skor, a u domenu visokih skorova mogu se razlikovati oni koji imaju iznadprosečni skor i oni koji imaju veoma visoki ili superiorni skor. Uporedni prikaz oba pristupa dat je u Tabeli 11.

Tabela 11

Pristupi za tumačenje kategorija ispitanika na osnovu stenajn skorova

Kategorija ispitanika	I pristup	II pristup
Veoma ispodprosečni ili slabi skor		1
Ispodprosečni skor	1, 2, 3	2, 3
Prosečni skor	4, 5, 6	4, 5, 6
Iznadprosečni skor	7, 8, 9	7, 8
Veoma iznadprosečni ili superiorni skor		9

Stenajn skorovi koriste, takođe, kumulativne procente u okviru distribucije skorova, pa tako, na primer, stenajn skoru 1 pripada, u stvari, 4% najnižih skorova, i tako redom (videti Tabelu 12 i Sliku 12). Na taj način stenajn skorovima se originalna distribucija prilagođava normalnoj, budući da najveći procenat ispitanika ima skorove oko prosečne vrednosti (npr. 19,74 tj. 20% ispitanika ima stenajn skor 5, videti Tabelu 12). To je i osnovna razlika između stenajn skorova i percentila: dok stenajn skorovi forsiraju normalnu distribuciju,

percentili forsiraju uniformnu distribuciju gde svaki percentil obuhvata 1% skorova (Mellenbergh, 2011). Ipak, stenajn skorovi i percentili se lako mogu prebaciti jedni u druge. Tako, ako neko ima skor koji odgovara 48. percentilu, on pada u interval od 41. do 60. percentila, što odgovara stenajn skoru 5.

Tabela 12

Konverzija stenajn skorova u percentile i z-skorove

	Stenajn	% rezultata	Kumulativni %	Opseg percentila (zaokruženo)	Opseg *z*-skorova
Ispod proseka	1	4,01	4,01	1–4	do -1,76
	2	6,55	10,56	5–11	-1,75 – -1,26
	3	12,10	22,66	12–23	-1,25 – -0,76
Prosek	4	17,47	40,13	24–40	-0,75 – -0,26
	5	19,74	59,87	41–60	-0,25 – 0,24
	6	17,47	77,34	61–77	0,25 – 0,74
Iznad proseka	7	12,10	89,44	78–89	0,75 – 1,24
	8	6,55	95,99	90–96	1,25 – 1,74
	9	4,01	100	97–99	1,75 i više

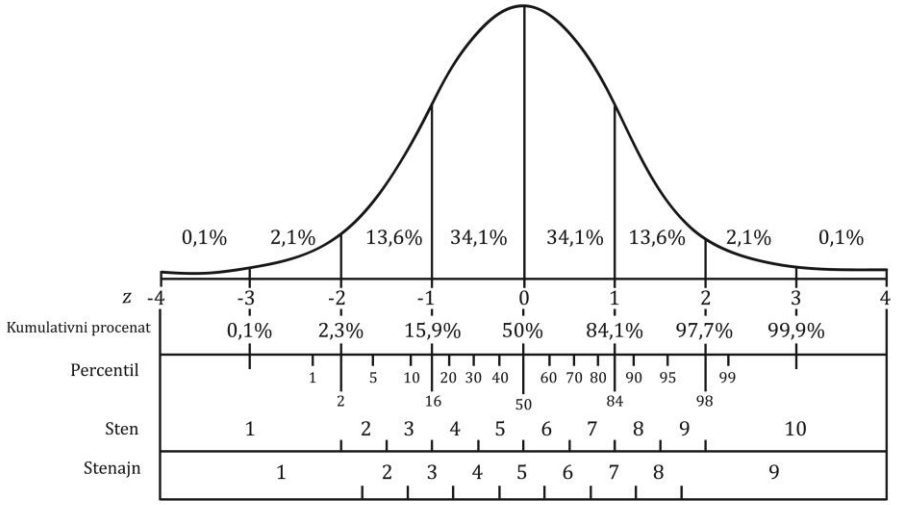
Napomena: % se odnosi na procenat rezultata ispod normalne krive.

Ukoliko je distribucija skorova normalna, onda se iz *z*-skorova lako može izvesti stenajn skor, kao izvedeni *z*-skor koji ima aritmetičku sredinu 5 i standardnu devijaciju 2:

$$\text{Stenajn} = z * 2 + 5$$

Izabrana je standardna devijacija 2 što je, u stvari, zaokruženo od 1,96 koliko iznosi kritična vrednost *z*-statistika za prihvatanje ili odbacivanje nulte hipoteze u okviru normalne distribucije za nivo značajnosti $p < ,05$ (tj. 95%, više na str. 134).

Kako raspon stenajn skorova može biti od 1 do 9, dolazi se do zaključka da svaka stenajn jedinica odgovara rasponu od 0,5 *SD*. U tom pogledu, lako se može zaključiti da 0 *z*-skor odgovara stenajn skoru 5, pri čemu bi raspon stenajn skora 5 odgovarao rasponu od -0,25 *SD* do +0,24 *SD*. Na taj način dolazimo do raspona prosečnih stenajn skorova (4-5-6) koji odgovara rasponu od -0,75 *SD* do +0,74 *SD*. Zaključujemo da je on nešto uži od uobičajenog raspona prosečnih skorova od -1 do +1 *SD*.



Slika 12. Raspodela stenajn i sten skorova u slučaju normalne distribucije.

Valja primetiti da na uporednom prikazu normi u slučaju normalne distribucije (Slika 12), 1. i 9. stenajn skor imaju širi opseg u odnosu na preostale stenajn skorove, isto kao i percentili na ekstremima. Drugim rečima, svi skorovi osim 1. i 9. obuhvataju 0,5 *SD*, dok skorovi 1 i 9 obuhvataju širi raspon. Iako skorovi 1 i 9 obuhvataju

širi raspon, svakako obuhvataju manji deo populacije (po 4,01%) u odnosu na preostale stenajn skorove.

Prednosti

Prednost stenajn skorova je jednostavnost interpretacije, budući da se interpretacija bazira na jednocifrenim brojevima (ima konačno mogućih 9 ishoda rezultata). Neki autori ističu da je izražavanje rezultata preko stenajn skorova ekonomičnije, budući da predstavlja sasvim dovoljan i smisleni opseg rezultata, u odnosu na precizniju procenu rezultata (Gredler, 1999). Prednost stenajn skorova ogleda se i u tome što su skorovi pogodni za situacije kada je potrebno na jednostavan i brz način klasifikovati ispitanike i kada je za odluku o klasifikaciji dovoljno poznavanje grupe kojoj ispitanik pripada. U praksi, stenajn skorovi se najčešće koriste prilikom trijaže, kada je dovoljna gruba, a brza procena.

Nedostaci

Nedostatak stenajn skorova je upravo u njegovoj jednostavnosti. Ishod ovih skorova omogućava suviše grubu klasifikaciju, tj. raspon skorova je suviše grub, u poređenju sa npr. percentilima (Nitko, 2004). Moguće je da dva ispitanika dobiju susedne stenajn skorove iako se njihovi sirovi skorovi razlikuju u skoro 1 *SD*. Na primer, neko može imati *z*-skor -0,75 koji pripada 4. stenajn skoru, dok druga osoba može imati *z*-skor 0,24 koji pripada 5. stenajn skoru. Drugim rečima, gubi se na preciznosti jer svaki stenajn skor obuhvata veći raspon sirovih skorova, tj. procenat skorova. Međutim, kao što je rečeno, neki autori (Gredler, 1999) upravo to ističu kao prednost i dovoljnu informativnost prilikom tumačenja skorova.

Sten skorovi

Sten skorovi nastali su iz potrebe da se opseg mogućih skorova zaokruži, te da umesto 9 skorova imamo 10. Sten skorovi (eng. *sten*) izvedenica su od "standardnih deset" (eng. *standard ten*). S obzirom na raspon skorova, prema pristupu u kojem se tumačenje skorova može vršiti na osnovu tri kategorije, skorovi se mogu interpretirati kao ispodprosečni, prosečni i iznadprosečni (Airasian, 2005, videti Tabelu 13). Iako u literaturi nema preporučenog drugog pristupa tumačenja skorova, i na sten skorove može se primeniti logika koja sledi iz tumačenja s pet kategorija, pri čemu se finija podela dobija kod ekstremnih skorova. Takođe, nekad se u praksi može sresti i treći pristup tumačenju skorova, po kojem svaka kategorija ispitanika obuhvata po dva sten skora. Ipak, prvi pristup se najčešće preporučuje prilikom tumačenja.

Tabela 13

Pristupi za tumačenje kategorija ispitanika na osnovu sten skorova

Kategorija ispitanika	I pristup	II pristup	III pristup
Veoma ispodprosečni ili slab skor		1	1, 2
Ispodprosečni skor	1, 2, 3	2, 3	3, 4
Prosečni skor	4, 5, 6, 7	4, 5, 6, 7	5, 6
Iznadprosečni skor	8, 9, 10	8, 9	7, 8
Veoma iznadprosečni ili superiorni skor		10	9, 10

Ukoliko je distribucija skorova normalna, sten skorove možemo dobiti iz standardnih na sledeći način:

$$\text{Sten} = z * 2 + 5{,}5.$$

Razlika u odnosu na stenajn skorove je u tome što je *AS* = 5,5 (kod stenajn skorova je *AS* = 5). Na osnovu navedene formule možemo zaključiti da je raspon prosečnih sten skorova od -1 do +0,99 *z*, što odgovara uobičajenom rasponu prosečnih skorova u terminima *z*-skorova. Detaljan prikaz konverzije sten skorova u percentile i *z*-skorove dat je u Tabeli 14.

Najpoznatiji test u kojem se primenjuju sten skorovi je Katelov test 16 faktora ličnosti (16PF: Cattell, Cattell, & Cattell, 2008).

Tabela 14

Konverzija sten skorova u percentile i z-skorove

	Sten	% rezultata	Kumulativni %	Opseg percentila (zaokruženo)	Opseg *z*-skorova
Ispod proseka	1	2,28	2,28	1–2	do -2,01
	2	4,40	6,68	3–7	-2,00 – -1,51
	3	9,17	15,85	8–16	-1,50 – -1,01
Prosek	4	15	30,85	17–31	-1,00 – -0,51
	5	19,15	50,00	32–50	-0,50 – -0,01
	6	19,15	69,15	51–69	0,00 – 0,49
	7	15	84,13	70–84	0,50 – 0,99
Iznad proseka	8	9,17	93,32	85–93	1,00 – 1,49
	9	4,40	97,72	94–98	1,50 – 1,99
	10	2,28	100	99	2,00 i više

Napomena: % se odnosi na procenat rezultata ispod normalne krive.

Prednosti

Sve prednosti koje važe za stenajn skorove, važe i za sten skorove, s tim što neki autori ističu da sten skorovi imaju jednu prednost u odnosu na stenajn – predstavljaju zaokružen raspon skorova koji je intuitivniji za interpretaciju (npr. Airasian, 2005).

Nedostaci

Svi nedostaci koji važe za stenajn, važe i za sten skorove.

C-skorovi

U starijim priručnicima mogu se naći tzv. C-skorovi koje je 1954. ponudio Džoj Paul Gilford. Ovi skorovi obuhvataju raspon od 11 podeka, tj. od 0 do 10 i dobijaju se na isti način kao i stenajn skorovi uz *AS* = 5 i *SD* = 2. Razlika u odnosu na stenajn je u tome što su ekstremi preciznije definisani. Naime, kod stenajn skorova skor 1 obuhvata 4% skorova, a kod C-skorova skor 0 obuhvata 1%, a skor 1 obuhvata 3% skorova (Tabela 15). Ipak, kasnije je i sam Gilford predložio da je bolje da raspon skorova bude od 1 do 11, nego od 0 do 10 kako bi se izbeglo da neko ima skor 0.

Kod nas u praksi se retko sreću ove vrste normi. Test koji ih sadrži, pored IQ-skrova, je Verbalna serija Borislava Stevanovića (Stevanović, Dragičević, Štajnberger i Bukvić, 1988).

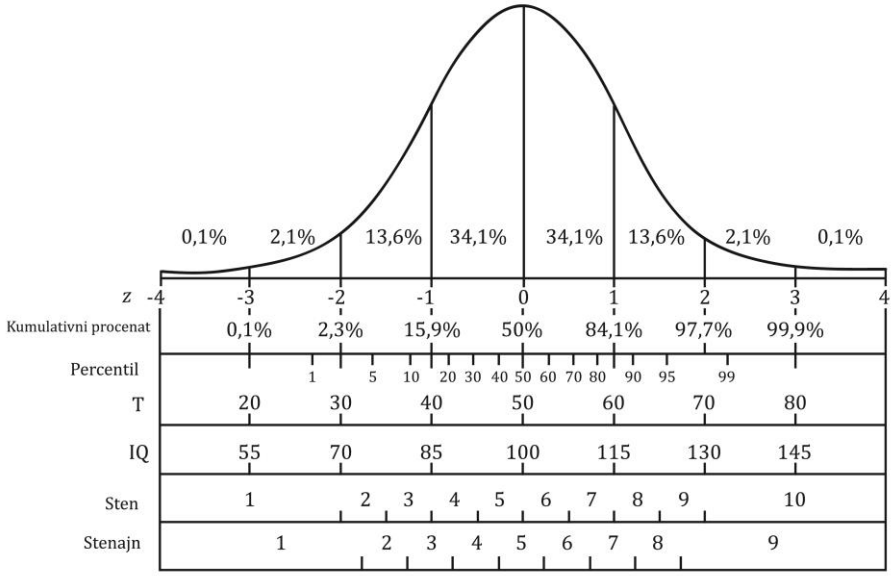
Tabela 15

Procentualna raspodela u okviru C-skorova

	C	% rezultata	Kumulativni %	Opseg percentila
Ispod proseka	0	1	1	1
	1	3	4	2–4
	2	7	11	5–11
	3	12	23	12–23
Prosek	4	17	40	24–40
	5	20	60	41–60
	6	17	77	61–77
Iznad proseka	7	12	89	78–89
	8	7	96	90–96
	9	3	99	97–99
	10	1	100	99

Napomena: % se odnosi na % rezultata ispod normalne krive.

Uporednim pregledom različitih vrsta normi koje se uobičajeno koriste (Slika 13) može se videti da se skorovi vrlo lako mogu prebacivati iz jednih normi u druge, pod pretpostavkom normalne distribucije. Tako, na primer, *z*-skor 1 *SD* odgovara 84,1 percentilu, T skoru 60, IQ skoru 115, sten skoru 8 i stenajn skoru 7.



Slika 13. Uporedni pregled vrsta normi.

Činioci koji utiču na izbor normi

Postoji veliki broj činilaca koji utiču na izbor normi, ali se među njima dva izdvajaju kao najvažnija – oblik distribucije skorova i potrebna preciznost procene.

Oblik distribucije skorova

Osnovni činilac koji utiče na izbor vrste normi je normalnost distribucije. Ukoliko je fenomen koji je predmet merenja takav da postoje

teorijska očekivanja da se ne distribuira normalno, onda treba koristiti percentilne norme ili sten i stenajn norme izvedene iz percentila. Ukoliko je fenomen takav da postoje teorijska očekivanja da se distribuira normalno, ali se iz nekog razloga podaci ne distribuiraju normalno, onda bi trebalo računati normalizovan standardni skor i izvedene skorove iz njega. Ukoliko je fenomen takav da postoje teorijska očekivanja da se distribuira normalno, i normalna distribucija jeste dobijena, onda se mogu koristiti bilo koje norme, ali najčešće se primenjuju standardni i izvedeni standardni skorovi.

Preciznost procene

U zavisnosti od namene procene, nekada nam je dovoljna samo gruba procena, a nekada nam je potrebna detaljnija i preciznija procena nečijeg skora. Ukoliko je test namenjen trijaži, brzoj i gruboj proceni, onda se preporučuje korišćenje sten ili stenajn skorova. Ukoliko je potrebna detaljnija, precizna i fina procena, npr. u cilju dijagnostike, onda se preporučuje korišćenje svih ostalih normi koje imaju širi raspon skorova.

Standardna greška

Svako merenje je merenje s greškom, i zbog te greške opaženi ili dobijeni skor će se razlikovati od pravog ili stvarnog skora. Greška merenja može poticati od sistematskih izvora, i tada se naziva sistematskom greškom, ili nesistematskih izvora, i tada se naziva nesistematskom greškom. U slučaju **sistematskih izvora**, odstupanje od pravog skora nije nasumično i ono ostaje konstantno pri ponovljenom merenju, ili se menja po određenim matematičkim zakonima. Važna odrednica slučajnih grešaka je da se s povećanjem uzorka one ne smanjuju. Sistematske greške mogu nastati usled pogrešno definisanog

predmeta merenja, nevalidnosti testa, grešaka u uzorkovanju (npr. sistematskih razlika u karakteristikama ispitanika koji su uključeni u uzorak i onih koji nisu), neizjednačenosti alternativnih ili paralelnih formi testa, interakcije istraživača i merenog fenomena, odgovaračkih usmerenja i sl. Dobra vest je što se one mogu identifikovati, kontrolisati ili se uvođenjem određenih korekcija mogu smanjiti.

U slučaju **nesistematskih izvora**, odstupanje od pravog skora je slučajno, bez pravilnosti i nastaje kao rezultat nezavisnih uticaja. Slučajne greške su posledica različitih uzroka koji se ne mogu kontrolisati, ni predvideti, ali s povećanjem uzorka i merenja izvori slučajnih grešaka se međusobno poništavaju. Primer uzorka nesistematske greške može poticati od nereprezentativnosti uzorka u smislu nedovoljne zastupljenosti neke od relevantnih kategorija ispitanika (ali ne i potpunog nedostajanja neke od relevantnih kategorija, jer bi to onda bila sistematska greška). Slučajne greške prate sva merenja, u krajnjem slučaju zato što naši instrumenti nisu savršeno pouzdani i precizni, tako da se one ne mogu nikada u potpunosti otkloniti. Ipak, dobra vest je što su uzorci slučajnih grešaka međusobno nezavisni, i što za slučajne greške važe zakoni normalne raspodele, te se pri velikim uzorcima i merenjima mogu odrediti granice unutar kojih se nalazi pravi skor. Upravo zbog ovih nesistematskih izvora grešaka merenja, rezultat jednog ispitanika na istom testu koji meri neku trajnu karakteristiku može da varira, kada se test ispitaniku zada više puta. Na primer, u jednoj situaciji testiranja, ispitanik je test radio ujutru, kada je bio odmoran, u drugoj popodne, kada je bio umorniji, potom, nekada je radio u prostoriji koja nije bila potpuno zvučno izolovana, a nekada je radio u zvučno izolovanoj prostoriji. Kada bismo uporedili skorove tog istog ispitanika u ovim različitim uslovima primene, sigurno ne bismo uvek dobili potpuno isti rezultat. Zbog toga

postoji potreba da se proceni u kom opsegu varira pravi skor ispitanika, kako bi se dobila precizna informacija o ispitanikovom pravom učinku ili izraženosti neke karakteristike. Na taj način bi se adekvatno izvršila interpretacija rezultata.

Opseg variranja skorova dobija se preko intervala poverenja. **Interval poverenja** (eng. *confidence interval*) predstavlja tačke na skali skorova koje definišu gornju i donju granicu intervala u kojem se, uz određenu verovatnoću, nalazi procenjeni parametar (AERA, APA, & NCME, 2006), odnosno pravi skor. Interval poverenja predstavlja tačnost ili preciznost izmerenog skora u odnosu na pravi skor, i najčešće se računa preko **standardne**[25] **greške merenja** (eng. *standard error of measurement* – *SEM* ili SE_m). Pre nego što objasnim SE_m, najpre ćemo se podsetiti načina izražavanja skorova prema Klasičnoj testnoj teoriji (KTT).

U okviru KTT, dobijeni ili opaženi skor ispitanika na testu (X) predstavlja zbir njegovog pravog skora (Xt) i skora nesistematske greške koja utiče na ispitanikove odgovore (Xe):

$$X = Xt + Xe.$$

Ilustrovaću primerom kako nesistematska greška može da utiče na opaženi skor. Pretpostavimo da je Miličin pravi skor na skali samopouzdanja 55 (iako nikada ne možemo znati koliki je nečiji pravi skor, zaista), a Lukin 70. Zaključićemo da Luka ima veće samopouzdanje od Milice. Međutim, Milica na upitniku samopouzdanja ostvaruje skor 65, a Luka 50. Kako? Moguće je, na primer, da je Milica pre nego što je

[25] U ovom slučaju reč "standardna" se, u stvari, odnosi na "prosečna".

popunjavala upitnik samopouzdanja radila test znanja iz psihometrije i bila prilično zadovoljna time kako ga je uradila, što je uticalo na njeno trenutno samopouzdanje. S druge strane, moguće je da je Luka pre nego što je popunjavao upitnik samopouzdanja dobio imejl o tome da se njegova prijava za master program koji je želeo da upiše odbija, te je bio razočaran zbog toga. U ovom primeru, Lukino trenutno razočaranje uticalo je na to da on ostvari niži skor, dok je Miličino trenutno zadovoljstvo uticalo na to da ona ostvari viši skor u odnosu na njen stabilni nivo samopouzdanja.

Šta se još može primetiti na osnovu opisanog primera? Najpre, može se primetiti da greška merenja nije ista za Milicu i Luku, te je u Miličinom slučaju ona +10[26], a u Lukinom slučaju -20[27]. Greška merenja, takođe, može da utiče na bilo koji skor, i to tako što neki skor može da poveća, a drugi na smanji. Odgovori ispitanika na testovima jesu nepredvidivi, ali opaženi skorovi neće upadljivo odstupati od pravih. Bila je puka slučajnost što je Milica pre popunjavanja upitnika samopouzdanja imala test iz psihometrije, a Luka pročitao imejl, a mogla je biti i obrnuta situacija. Tako je na celom uzorku ispitanika greška merenja slučajno povećala ili smanjila opažene skorove, što zapravo čini da je na ukupnom uzorku prosek grešaka merenja nula. Ova konstatacija odražava bazičnu pretpostavku o normalnosti raspodele koju tvore slučajne greške. Pored ovoga, greška merenja je nezavisna od visine pravog skora, tj. korelacija greške merenja i pravog skora je nula. To što je u jednoj situaciji greška merenja povećala opaženi skor, a u drugoj smanjila, nema veze s tim kakav je pravi skor. On uvek ostaje isti, dok opaženi skor varira.

[26] $Xe = X - Xt$, tj. $Xe = 65 - 55 = 10$
[27] $Xe = X - Xt$, tj. $Xe = 50 - 70 = -20$

Vratimo se sada na standardnu grešku merenja. Standardna greška merenja predstavlja prosek skorova greške koji utiču na opaženi skor. Što je veća standardna greška merenja, veća je i prosečna razlika između opaženog i pravog skora. Standardna greška merenja predstavlja, u stvari, standardnu devijaciju opaženih skorova koje bi ispitanik dobio kada bi bio ispitan beskonačnim brojem slučajno odabranih paralelnih testova (Fajgelj, 2013). Može se uočiti da je standardna greška merenja povezana s pouzdanošću, koja predstavlja odsustva greške merenja, odnosno preciznost merenja. Naime, visoko pouzdan test nam pruža visoko precizno procenjene skorove. Upravo u okviru izračunavanja standardne greške merenja (SE_m) uzima se u obzir i pouzdanost i to preko sledećeg obrasca:

$$SE_m = SD \sqrt{(1 - r_{tt})}$$

u kojem SD predstavlja standardnu devijaciju opaženih skorova, a r_{tt} pouzdanost izračunatu na bilo koji način.

Ukoliko želimo da izračunamo u kom opsegu se kreće, u ovom slučaju, Miličin pravi skor na upitniku samopouzdanja i ukoliko imamo da je SD = 8, a pouzdanost izračunata kao Kronbahov α koeficijent 80, dobićemo da je SE_m = 8 $\sqrt{}$ (1 – ,80) = 3,58. Sada možemo izračunati interval poverenja u kojem bi se kretao Miličin pravi skor na sledeći način:

$$95\%\text{-tni interval poverenja} = X \pm 1{,}96 * SE_m$$

gde je X opaženi skor, a SE_m standardna greška merenja.

U praksi se najčešće koristi 95%-tni interval poverenja, dobijen kao 95% pokrivenosti površine ispod normalne krive u rasponu od ± 2*SD* (tj. preciznije, ± 1,96*SD*), iako je moguće sresti i druge intervale poverenja (Tabela 16). Ovaj pristup računanja intervala poverenja oko SE_m naziva se **tradicionalnim pristupom** (eng. *traditional approach*, više u Charter & Feldt, 2000).

Tabela 16

Kritične vrednosti za određene intervale poverenja

Interval poverenja	Kritična vrednost
90%	1,65
95%	1,96
99%	2,58

U našem primeru bi 95%-tni interval poverenja bio 65 ± 1,96 * 3,58 = 65 ± 7,01, odnosno od 57,99 do 72,01. U praksi se ovaj pristup intervala poverenja tumači na sledeći način: s 95% sigurnosti možemo tvrditi da je Miličin pravi skor u rasponu od 57,99 do 72,01. Međutim, zapravo se misli da se očekuje da će intervali poverenja obuhvatiti 95% ispitanikovih pravih skorova kada se veliki broj ispitanika testira istim testom i kada se interval poverenja računa za svakog ispitanika. Iako nije psihometrijski potpuno ispravno, laicima će biti jasnije ako im se kaže: ukoliko bismo isti test dali Milici 100 puta, možemo očekivati da će u 95 puta Miličini skorovi biti u rasponu od 57,99 do 72,01.

Na osnovu formule za izračunavanje SE_m možemo videti da ona zavisi od varijabilnosti i pouzdanosti skorova. Naime, što je varijabilnost skorova manja, a pouzdanost veća, to će SE_m biti manja. Ukoliko bismo imali pouzdaniji test (npr. r_{tt} = ,90), a da je *SD* ostala ista, SE_m bi bila manja i iznosila bi 2,53, a 95%-tni interval poverenja 65 ± 4,96, odnosno pravi

skor bi bio u rasponu od 60,04 do 69,96. Primetićete da je raspon skorova u prvom slučaju bio oko 7,16 bodova, a u drugom 4,96. Iako je ovo hipotetički primer, dobar je i zbog još jednog zapažanja, a to je da je Miličin pravi skor 55 i da nije obuhvaćen rasponom od 60,04 do 69,96. Budući da je korišćen 95%-tni interval poverenja, postoji 5% rizika da interval ne sadrži pravi skor. Dakle, interval poverenja ne sadrži nužno pravi skor, već ga sadrži s određenom verovatnoćom. Pravi skor predstavlja prosek beskonačnog broja merenja istog ispitanika istim testom, i zato je moguće da je opaženi skor u jednom slučaju bio 50, u drugom 60 itd., ali prosek svih merenja bi trebao biti 55.

Iako je navedeni način računanja greške merenja i interpretacije skora na testu uobičajen u praksi, izvor je mnogih kontroverzi. Najpre, SE_m se računa na celom uzorku i primenjuje za sve individualne skorove, ali postavlja se pitanje da li je SE_m ista u celom kontinuumu predmeta merenja. Istraživanja su pokazala da nije, i da je SE_m najveća u domenu prosečnih skorova, a manja u domenu ekstremnih skorova (više u Harvill, 1991). Međutim, važnije je to što, zapravo, SE_m treba da se tumači kao varijansa greške oko pravog, a ne opaženog skora, budući da prestavlja procenu varijabilnosti opaženog skora kada je pravi skor konstanta. Da bismo to uradili, potrebno je da znamo pravi skor, a kad znamo pravi skor ni ne treba nam njegova procena. Ipak, autori, iako su svesni da se SE_m treba primeniti u odnosu na pravi skor, dozvoljavaju njenu primenu u odnosu na opaženi skor uz dva uslova: 1. da je pouzdanost testa (r_{tt}) dovoljno visoka (o nekim preporukama za Konbahov alfa koeficijent - α može se videti na str. 385), i 2. da se SE_m ne primenjuje u slučaju ekstremnih skorova (Harvill, 1991). Naime, kod ekstremnih skorova je mogućnost variranja skorova ograničena opsegom mogućih skorova, te npr. ako neko ostvaruje skor 58, a maksimalni mogući skor na skali je 60,

interval poverenja se ne može primeniti na opseg preko skora 60. Drugim rečima, za ekstremne skorove će standardna greška biti lažno niska. Međutim, ukoliko su navedena dva uslova ispunjena, onda SE_m oko opaženog skora daje dovoljno dobru aproksimaciju greške merenja oko pravog skora. Pored toga, argument za računanje SE_m oko opaženog skora leži i u tome što je za opaženi skor mala verovatnoća da će varirati u odnosu na pravi skor više od opsega skorova koji obuhvata interval poverenja. S obzirom na navedeno, može se pretpostaviti da će se pravi skor zaista naći u opsegu intervala poverenja oko opaženog skora (više u Charter & Feldt, 2001).

Alternativa računanju greške merenja oko pravog skora je **standardna greška ocene** (eng. *standard error of estimation – SEE* ili SE_e). Ova greška se izračunava na sledeći način:

$$SE_e = SD \sqrt{r_{tt}\,(1 - r_{tt})}$$

gde je SD standardna devijacija opaženih skorova, a r_{tt} pouzdanost izračunata na bilo koji način. U odnosu na SE_m, razlika je u tome što je pod korenom izraz množen sa r_{tt}, tako da je SE_e uvek manja od SE_m. Interval poverenja po ovoj koncepciji računa se na sledeći način:

$$95\%\text{-tni interval poverenja} = T \pm 1{,}96 * SE_e$$

i naziva se **pristupom baziranim na regresiji ili regresionim pristupom** (eng. *regression-based approach*, više u Charter & Feldt, 2000). Da bismo izračunali interval poverenja, potreban nam je ocenjeni pravi skor:

$$T = AS + r_{tt}(X - AS)$$

gde *AS* predstavlja artmetičku sredinu opaženih skorova, r_{tt} pouzdanost izračunatu na bilo koji način i *X* opaženi skor.

Valja primetiti da ovaj interval nije simetričan oko opaženog, već oko *ocenjenog* pravog skora. Iako je ovaj način računanja ispravan, jer se SE_e primenjuje oko pravog skora, u praksi nije zaživeo, te i dalje dominira tradicionalni pristup intervalima poverenja.

U našem primeru, SE_e bi iznosila $8\sqrt{,80(1 - ,80)} = 3,2$, i ukoliko bi *AS* bila 60, pravi Miličin skor bi iznosio $T = 60 + ,80(65 - 60) = 64$, prema čemu bi interval poverenja iznosio 64 ± 1,96 * 3,2, odnosno pravi skor bi se kretao u rasponu od 57,73 do 70,27. Tumačenje zasnovano na ovom pristupu bi glasilo: očekujemo da 95% osoba s istim opaženim skorom imaju pravi skor u rasponu od 57,73 do 70,27. U okviru ovog pristupa intervalima poverenja, ispravnije je koristiti tumačenje koje se u praksi često sreće prilikom tumačenja skorova na osnovu SE_m, i koje bi glasilo: očekujemo da 95% osoba s istim opaženim skorom kao Milica ima pravi skor u rasponu od 57,73 do 70,27. Ovde valja obratiti pažnju na različitu interprtaciju intervala poverenja u okviru tradicionalnog i regresionog pristupa. Dok je prema tradicionalnom pristupu svaki skor nezavistan od skorova drugih osoba, jer se interval poverenja odnosi na ponavljanje testiranja jedne osobe istim testom, kod regresionog pristupa je fokus na suppopulaciji s istim opaženim skorom, jer se interval poverenja odnosi na ponavljanje testiranja osoba s istim opaženim skorom istim testom. Međutim, kao što je već rečeno, u praksi, iako se češće koristi tradicionalni pristup (uz računanje SE_m), interpretacija koja se izvodi više odgovara regresionom pristupu (kada se računa SE_e), jer je interpretacija bazirana na regresionom pristupu lakša za razumevanje. Kao što je istaknuto, postoje argumenti koji opravdavaju upotrebu SE_m oko pravog skora.

Standardna greška razlike

Ukoliko dva testa imaju isti teorijski raspon skorova, npr. od 0 do 100, to ne znači da se skorovi na tim testovima mogu olako porediti. Uzmimo na primer situaciju da Milica na testu matematike postiže skor 55, a na testu srpskog jezika 70. Prostim poređenjem bismo greškom zaključili da Milici bolje ide srpski jezik od matematike. Moguće je da je, na primer, test iz matematike teži od testa iz srpskog jezika, te da je skor 55 na testu iz matematike, ustvari, iznadprosečan, dok je skor 70 na testu iz srpskog jezika prosečan. Pored toga, kako nijedan test nije savršeno pouzdan, možemo očekivati da opaženi skorovi variraju, te da je pravi Miličin skor na testu matematike 65, a na testu srpskog jezika 60. Kako onda da odredimo šta Milici ide bolje?

Da bi se izvršilo pravilno poređenje skorova s dva testa, koristi se **standardna greška razlike** (eng. *standard error of difference* – *SED*). Formula za izračunavanje *SED* glasi:

$$SED = \sqrt{SE_{m1}^2 + SE_{m2}^2}$$

gde SE_{m1} predstavlja standardnu grešku merenja prvog testa, a SE_{m2} standardnu grešku merenja drugog testa. Valja primetiti da će *SED* biti veća nego pojedinačne SE_m, jer je *SED* pod uticajem greške merenja oba testa. Da bi poređenje uopšte moglo da se izvrši, neophodan uslov je da skorovi s oba testa potiču s iste skale izražavanja skorova (npr. u Domino & Domino, 2006). To mogu biti IQ skorovi ili bilo koji drugi, a kako se svedu na istu skalu izražavanja skorova, to znači da će im standardna devijacija biti ista. Kada se u formuli za *SED*, SE_m zameni izrazom $SD\sqrt{(1 - r_{tt})}$, svođenjem izraza dobijamo:

$$SED = SD\sqrt{2 - r_{tt1} - r_{tt2}}$$

u kojem je *SD* standardna devijacija opaženih skorova, a r_{tt1} i r_{tt2} pouzdanosti prvog i drugog testa, izračunate na bilo koji način ali preko istog koeficijenta.

Uzmimo da u našem primeru test matematike ima pouzdanost ,95, a test srpskog jezika ,85, i da *SD* na oba testa iznosi 10. Dobićemo da je $SED = 10\sqrt{2 - ,95 - ,85} = 4,47$. Interval poverenja se računa na uobičajeni način, 1,96 * *SED* (95%-tni interval), što iznosi 1,96 * 4,47 = 8,76 ~ 9. Na osnovu intervala poverenja možemo zaključiti da, ako se skorovi na dva testa razlikuju 9 i više sirovih bodova, onda postoje značajne razlike među njima. Kako se Miličini skorovi na testovima razlikuju za 15 bodova, zaključujemo da postoji značajna razlika između njih, te da Milici zaista bolje ide srpski jezik nego matematika, bar u slučaju sadržaja ova dva testa.

Na osnovu *SED* možemo porediti skorove istog ispitanika na različitim testovima, kao i skorove različitih ispitanika na istim ili različitim testovima, sve dok su skorovi svedeni na istu skalu. Pogledajmo primer postignuća jednog ispitanika na različitim testovima znanja na Slici 14. Crnim poljima označeni su intervali poverenja za pojedinačne testove. Ukoliko se dva intervala poverenja preklapaju, onda razlike između skorova na tim testovima nisu značajne. Šta sve možemo zaključiti na osnovu Slike 14? Najpre, postignuća na fizici i matematici se ne razlikuju značajno, jer se intervali poverenja preklapaju za skorove na ovim testovima, i postignuće na ova dva testa je značajno manje u odnosu na postignuće na testovima iz srpskog i stranog jezika i istorije. Postignuće na testovima iz sprskog i stranog jezika i iz istorije se međusobno, takođe, ne razlikuju. Pored toga,

možemo zaključiti da je greška merenja najveća u slučaju testa iz istorije, jer je interval poverenja najširi za ovaj test (20), a najmanja je u slučaju testa iz stranog jezika (10) i iz fizike (10).

	...	40	45	50	55	60	65	70	75	80	85	90	95	100
matematika						■	■	■						
srpski jezik										■	■	■	■	
strani jezik									■	■	■			
istorija									■	■	■	■	■	
fizika				■	■	■								

Slika 14. Primer intervala poverenja za jednog ispitanika na različitim testovima koji su svedeni na istu skalu izražavanja skorova.

	...	30	33	36	39	42	45	48	51	54	57	60	63	...
matematika						■	■	■						
	...	40	45	50	55	60	65	70	75	80	85	90	95	100
srpski jezik										■	■	■	■	
	...	45	52	59	66	73	80	87	94	101	108	115	122	...
strani jezik									■	■	■			
	...	50	52	54	56	58	60	62	64	66	68	70	72	...
istorija									■	■	■	■	■	
	...	35	45	55	65	75	85	95	105					
fizika				■	■	■								

Slika 15. Primer intervala poverenja za jednog ispitanika na različitim testovima koji su svedeni na originalnu skalu izražavanja skorova.

Iako, da bi se izračunala standardna greška razlike, skorovi s različitih testova treba da se svedu na istu skalu, u praksi se skorovi vraćaju na originalnu skalu, kako bi se direktno izvršilo tumačenje razlika. Ovo svođenje na originalnu skalu se radi tako da se vodi računa o standardnoj devijaciji svakog testa, te u slučaju testa s većom standardnom devijacijom su skorovi "zgusnutiji" (primer matematike na Slici 15), a u slučaju testa s manjom standardnom devijacijom su skorovi "raširenији" (primer srpskog jezika na Slici 15).

Pitanja za razmišljanje

1. Koje su osnovne razlike između normativne i kriterijumske interpretacije?
2. Koja je razlika između standardizacionog uzorka, normativnog uzorka, referentne grupe i validacionog uzorka? Za koju od navedenih vrsta uzoraka nije karakteristično dobijanje normi?
3. Koje su prednosti i mane percentila u odnosu na standardne skorove.
4. U kojoj situaciji u praksi se koriste sten i stenajn skorovi?
5. Zašto je prilikom interpretacije skorova važno uzeti u obzir standardnu grešku merenja?

Zadatak

Izračunaj *z*, T, IQ, stenajn i sten skorove za ispitanika koji je ostvario sirovi skor 26 na testu AL-4 verbalnih sposobnosti, ako znamo da je *AS* = 20,50 i *SD* = 7,60. Skorovi na testu se distribuiraju normalno.

Rešenje

$z = (26 - 20{,}5) / 7{,}60 = 0{,}72$

Interpretacija: Ispitanik postiže prosečni skor (u rasponu je od -1 do 0,99).

$T = 0{,}72 * 10 + 50 = 57{,}2 \sim 57$

Interpretacija: Prema prvom pristupu tumačenja T-skorova, ispitanik postiže prosečni skor (u rasponu od 41 do 60), a prema drugom pristupu tumačenja T-skorova, ispitanik postiže iznadprosečni skor (u rasponu od 56 do 65). S obzirom na to, ukoliko drugačije nije naglašeno u priručniku ili nekom drugom dokumentu za tumačenje skorova, usvojili bismo prvi pristup za tumačenje skorova i skor okarakterisali kao prosečan.

$IQ = 0{,}72 * 15 + 100 = 110{,}8 \sim 111$

Interpretacija: Imajući u vidu da prosek obuhvata raspon od 90 do 109 IQ jedinica, skor bismo interpretirali kao iznadprosečan, i to na granici između prosečnog i iznadprosečnog skora.

$\text{Stenajn skor} = 0{,}72 * 2 + 5 = 6{,}44 \sim 6$

Interpretacija: Prema oba pristupa tumačenja stenajn skorova, skor bismo okarakterisali kao prosečan, i to na gornjoj granici proseka.

Sten skor = 0,72 * 2 + 5,5 = 6,94 ~ 7

Interpretacija: Prema prvom i drugom pristupu tumačenja sten skorova, skor 7 bismo okarakterisali kao prosečan, i to na gornjoj granici proseka, a prema trećem pristupu kao iznadprosečni.

Generalno, imajući u vidu sve norme, ovaj skor bismo okarakterisali kao skor na gornjoj granici proseka ili visoko prosečan skor.

Reference

Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2006). *Standardi za pedagoško i psihološko testiranje*. Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Barrett, P. (2011). *Percentiles and percentile ranks: Confused or what?* The technical whitepaper series, 3, 1–22. Preuzeto sa https://www.pbarrett.net/techpapers/percentiles.pdf

Bechger, T., Hemker, B., & Maris, G. (2009). *Over het gebruik van continue normering* [On the use of continuous norming]. Arnhem, The Netherlands: Cito.

Berger, J., Marković, M. i Mitić, M. (1995). *VITI Vekslerov individualni test inteligencije.* Beograd, Srbija: Centar za primenjenu psihologiju.

Biro, M. (1997). *REVISK Revidirana skala za merenje inteligencije dece po principima Wechslera.* Beograd, Srbija: Centar za primenjenu psihologiju.

Biro, M. (1995). *MMPI-202 Dijagnostička procena ličnosti*. Beograd, Srbija: Centar za primenjenu psihologiju.

Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2009). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences, 19*(3), 330–338.

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, *29*, 512–518.

Cattell, R. B., Cattell, A. K. S., & Cattell, H. E. P. (2015). *16 faktora ličnosti – 16PF (5. izdanje)*. Beograd, Srbija: Sinapsa edicije, za delo prevedeno na srpski jezik.

Cermak, S. (1989). Norms and scores. In L. J. Miller (Ed.), *Developing norm-referenced standardized tests* (pp. 91–123). Binghamton, NY: Haworth Press.

Chadha, N. K. (2009). *Applied psychometry*. New Delhi, India: SAGE Publications India Pvt Ltd.

Charter, R. A., & Feldt, L. S. (2000). Confidence Intervals for True Scores: Is There a Correct Approach? *Journal of Psychoeducational Assessment, 19*(4), 350–364.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to test and measurement* (4th ed.). Mountain View, CA: Mayfield.

De Vries, A. L. M. (2006). *The merit of ipsative measurement*. Unpublished doctoral dissertation. Maastricht University, the Netherland.

Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction.* New York, NY: Cambridge University Press.

Dutton, E., & van der Linden, D., & Lynn, R. (2016). A negative Flynn effect: A systematic literature review. *Intelligence, 59*, 163–169.

European Federation of Psychologists' Associations – EFPA (2013). *FPA review model for the description and evaluation of psychological*

tests: Test review form & notes (v.4.2.6). Preuzeto sa http://www.efpa.eu/download/650d0d4ecd407a51139ca44ee704fda4

Fajgelj, S. (2007). *Metode istraživanja ponašanja* (3. dopunjeno izdanje). Beograd, Srbija: Centar za primenjenu psihologiju.

Fajgelj, S. (2013). *Psihometrija. Metod i teorija psihološkog merenja* (IV dopunjeno izdanje). Beograd, Srbija: Centar za primenjenu psihologiju.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: SAGE Publications Ltd.

Furr, M. R., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Thousand Oaks, California: SAGE Publications Ltd.

Gorsuch, R. L. (1983, August). *The theory of continuous norming*. Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA.

Gredler, M. E. (1999). *Classroom assessment and learning*. New York, NY: Longman.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practice, 10,* 33–41.

Hedrih, V. (2018). *Kroskulturalna adaptacija psiholoških mernih instrumenata.* Niš, Srbija: Filozofski fakultet.

Ivić, I., Milinković, M., Rosandić, R. i Smiljanić, V. (1985). *Nova beogradska revizija Bine-Simonove skale.* Beograd, Srbija: Centar za primenjenu psihologiju.

Jackson, C. (2003). *Psihologijsko testiranje* (2. izdanje). Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Jerković, I. i Zotović, M. (2017). *Razvojna psihologija.* Beograd, Srbija: Centar za primenjenu psihologiju,

Jokela, M., Pekkarinen, T., Sarvimäki, M., Terviö, M., & Uusitalo, R. (2017). Secular rise in economically valuable personality traits. *PNAS, 114*(25), 6527–6532.

Jovanović, V., Gavrilov-Jerković, V., Žuljević, D. i Brdarić, D. (2014). Psihometrijska evaluacija Skale depresivnosti, anksioznosti i stresa–21 (DASS–21) na uzorku studenata u Srbiji. *Psihologija, 47*(1), 93–112.

Kodžopeljić, J. i Pekić, J. (2017). *Psihologija u nastavi: odabrane teme iz psihologije obrazovanja.* Novi Sad, Srbija: Filozofski fakultet.

Lazarević, LJ. B., Knežević, G., Mitić, M., & Đurić Jočić, D. (2018). Psychometric properties of the Serbian version of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). *Psihologija, 5*(3), 333–349.

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research, 4*, 293–307.

Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instrutional decision making.* Thousand Oaks, California: SAGE Publication, Inc.

Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: Development, analysis, and application of psychological and educational tests*. Amsterdam, The Netherlands: Eleven International Publishing.

Mitrović, D. i Trogrlić, A. (2014). *Psihologija polnih razlika i sličnosti*. Beograd, Srbija: Sinapsa edicije.

Momirović, K., Wolf, B. i Džamonja, Z. (1992). *KON 6 – Kibernetička baterija konativnih testova*. Beograd, Srbija: Centar za primenjenu psihologiju.

Nitko, A. J. (2004). *Educational assessment of students* (5th ed.). Upper Saddle River, NY: Merrill.

Oosterhof, A. (2001). *Classroom applications of educational measurement* (3rd ed.). Upper Saddle River, NJ: Merrill Education/Prentice Hall.

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science, 10*(3), 282–306.

Santrock, J. W. (2013). *Lifespan development* (14th ed.). New York, NY: McGraw Hill.

Schaie, K. W. i Willis, S. L. (2001). *Psihologija odrasle dobi i starenja* (4. izdanje). Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Smederevac, S., Mitrović, D. i Čolović, P. (2010). *Velikih pet plus dva, primena i interpretacija.* Beograd, Srbija: Centar za primenjenu psihologiju.

Solomon, S. R., & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods, 8*(2), 448–462.

Stevanović, B. (1937). *Merenje inteligencije sa uputstvom za upotrebu beogradske revizije Bine-Simonove skale.* Beograd, Srbija: Drag.

Stevanović, B., Dragićević, Č., Štajnberger, I. i Bukvić, A. (1988). *Verbalna serija prof. dr B. Stevanovića (priručnik)*. Beograd, Srbija: Institut za psihologiju i Republički zavod za zapošljavanje radnika SR Srbije.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360.

Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons Inc.

KRITERIJUMSKA INTERPRETACIJA

Kriterijumski način interpretacije prvi put je predložio Glaser (Glaser, 1963) i od tada je u širokoj upotrebi, posebno u obrazovanju. U okviru kriterijumske interpretacije na osnovu skora ispitanika utvrđuje se da li je ispitanik dostigao određeni kriterijum, standard, tj. granični skor. Koen i Sverdlik (Cohen & Swerdlik, 2009) definišu kriterijum kao standard na osnovu kojeg se može doneti sud ili odluka. Kod kriterijumske interpretacije skor se poredi sa kriterijumom i od interesa je da li je konkretan ispitanik zadovoljio kriterijum. Za razliku od normativne interpretacije, kod kriterijumske interpretacije skor se ne tumači u odnosu na postignuće drugih ljudi, te nam informacija o postignuću drugih ljudi nije potrebna.

U Standardima (AERA, APA, & NCME, 2006) se navodi da se rezultat u okviru kriterijumske interpretacije tumači u odnosu na neki funkcionalni nivo učinka. Dostizanje kriterijuma može podrazumevati ovladavanje određenim nivoom veštine, usvajanje određenog nivoa znanja koje je potrebno da bi se položio neki kurs, steklo određeno zvanje, dobila profesionalna dozvola, odredilo da li neko ima potrebne kompetencije za obavljanje konkretnog posla, kao i utvrđivanje prisustva neke patologije i slično. Neretko, dostizanje određenog kriterijuma uključuje demonstraciju ili kreiranje nekog produkta ili proizvoda.

Kada imamo kriterijumsku interpretaciju, nije nam važno da poznajemo distribuciju, ni varijabilnost skorova. Čak, skorovi na kriterijumskom testu često se ne distribuiraju normalno. Na primer, ukoliko želimo da ispitamo da li ima potrebe za uvođenjem određenih

nastavnih jedinica u okviru nekog predmeta, učenicima bismo zadali ulazni test za proveru predznanja. Kao kriterijum možemo postaviti 55% tačno urađenih pitanja, te ukoliko većina učenika ne prelazi ovaj kriterijum, zaključujemo da je učenicima potrebna edukacija iz određene nastavne jedinice. Ukoliko nastavne jedinice koje su predmet merenja testa nisu zastupljene u gradivu na nekim drugim predmetima, vrlo je verovatno da će većina učenika ostvariti niske skorove na testu, te da će distribucija biti pozitivno zakošena. S druge strane, ukoliko želimo da proverimo koliko su učenici naučili gradivo nakon odslušanog gradiva iz određenog predmeta, i ponovo postavimo kao kriterium 55% uspešno rešenih pitanja, vrlo je verovatno da će većina učenika imati visoke skorove, tj. da će distribucija biti negativno zakošena. U oba slučaja je vrlo verovatno da će postojati i učenici koji imaju upadljivo visoke ili upadljivo niske skorove, ali je suština da se skorovi ne distribuiraju normalno. Bez obzira na to što distribucije nisu normalne, mi se i dalje možemo pouzdati u rezultat i na osnovu njega doneti odluku o tome da li treba uvoditi određenu nastavnu jedinicu ili ne. U tom smislu, kriterijumski testovi mogu sadržati i lake i teške zadatke u neproporcionalnom odnosu, tj. dozvoljeno je da težine zadataka budu manje ujednačene (Kodžopeljić i Pekić, 2017). S druge strane, za normativne testove poželjno je da imaju najviše prosečno teških zadataka (npr. 50% prosečno teških, 25% lakših i 25% težih), ili da imaju zadatke različitih težina (npr. nekoliko zadataka čija je proporcija tačnih odgovora 10%, odnosno čija je težina 0,10, pa nekoliko čija je težina 0,20 i tako redom), ali da njihov prosek odgovara prosečno teškom zadatku.

Isti je slučaj i sa varijabilnošću skorova. U obe pomenute situacije očekujemo nisku varijabilnost jer imamo grupisanje oko

niskih skorova u prvom, odnosno oko visokih skorova u drugom primeru. Zamislimo idealnu situaciju u kojoj su svi učenici uspešno savladali gradivo i svi imaju veoma visoke skorove. Svakako da će u ovoj situaciji varijabilnost skorova biti minimalna, ili čak jednaka nuli u slučaju da su svi učenici ostvarili maksimalni mogući skor. No, bez obzira na to, mi i dalje možemo ispravno zaključiti da li je standard ispunjen i na osnovu toga doneti odluku u vezi s daljim planom delovanja.

Ovo ima važne implikacije za određenje nekih psihometrijskih karakteristika testa koji je zasnovan na kriterijumskoj interpretaciji, npr. za određenje pouzdanosti. Podsetićemo se, pouzdanost, uopšteno govoreći, predstavlja odnos prave varijanse skorova i ukupne varijanse (ukupna varijansa predstavlja zbir prave varijanse i varijansne greške). Pouzdanost zavisi od varijabilnosti testovnih skorova, i što je varijabilnost skorova veća, veća je i pouzdanost. Koa što je rečeno, kod kriterijumskih testova varijansa skorova može biti vrlo mala. Na primer, prilikom polaganja testa za dobijanje vozačke dozvole, svi kandidati obično imaju skor oko maksimalnog. S obzirom na to, postupci određivanja pouzdanosti preko interne konzistencije, poput Kronbahovog alfa koeficijenta, nisu uvek adekvatni kod kriterijumskih testova, već npr. test-retest ili split-half pouzdanost. Pored toga što kod kriterijumske interpretacije nije važna distribucija skorova, ni varijabilnost, takođe nije važno ni da test bude diskriminativan u celom opsegu karakteristike koja se meri, već samo u opsegu graničnih skorova. Ono što je najvažnije je utvrđivanje **validnosti testa**. Da bi interpretacija uopšte imala smisla, postavljeni kriterijum mora biti jasno definisan i baziran na realnim podacima, mora biti empirijski

utvrđen, kako bi se mogla doneti ispravna odluka na osnovu njegovog zadovoljenja.

Prednosti

Kao što je rečeno, kriterijumska interpretacija skora na testu najčešće se primenjuje u okviru obrazovanja. Njena osnovna prednost u ovom domenu je u tome što se **eliminiše efekat takmičenja**, budući da se ispitanici ne porede međusobno, već u odnosu na postavljeni kriterijum. Pretpostavlja se da ovakav način interpretacije deluje stimulišuće i motivišuće na učenike. Kakav bi bio ishod toga da se postiguće u obrazovanju interpretira na osnovu normi, a ne kriterijuma? Zamislite jedno odeljenje koje je "jako", u kojem se nalaze najbolji učenici. Ukoliko primenite normativnu interpretaciju i ukoliko se postignuće učenika poredi sa postignućem učenika iz odeljenja, polovina učenika će imati skor ispod proseka. Zamislite sada odeljenje koje ima slabije učenike po postignuću. I u njemu bi polovina učenika imala skor ispod proseka, a polovina skor iznad proseka. Međutim, učenik iz prvog odeljenja koji ima ispodprosečni skor može imati na istom testu isti bruto skor kao i učenik iz drugog odeljenja koji ima iznadprosečni skor. Šire posmatrano, na nivou postignuća svih učenika jednog razreda jedne škole, ili na nivou države, uvek bi polovina učenika imala ispodprosečni skor, a polovina iznadprosečni skor, što bi značilo da će uvek postojati učenici koji će "pasti" na testu, ma koliko blag kriterijum bio (npr. -2 *SD*). U slučaju kriterijumske interpretacije ovog problema nema – može se desiti da svi učenici zadovolje kriterijum, npr. polože maturski test i da to odražava realno stanje stvari.

Nedostaci

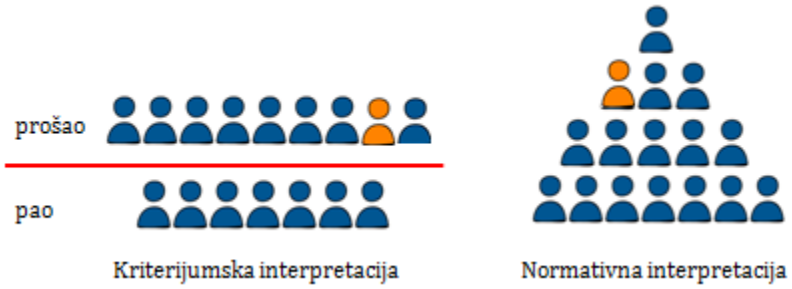
Bez obzira na pomenute prednosti, kriterijumskom načinu interpretacije upućene su i kritike. U prvom redu, kriterijumska interpretacija se kritikuje upravo zbog onoga što je razlikuje od normativne, a to je **zanemarivanje relativnog nivoa** rezultata na testu. Na taj način se gubi informacija o tome kakav je rezultat neke osobe u odnosu na druge osobe. U vezi s tim je i druga kritika koja se odnosi na **nemogućnost uvida u izuzetnost rezultata**. Uvid u iznadprosečne i izuzetne rezultate nam omogućava normativna interpretacija, te bi to bili rezultati, na primer, iznad 95. ili na 99. percentilu. Poseban problem predstavlja **odabir kriterijuma**, i ovaj problem je povezan s metodološkim nedostacima nacrta i/ili teorijskim, kao što je nejasno određenje kriterijuma. Ilustracije radi, zamislimo da ste konstruisali novi test inteligencije i da želite na osnovu njega da procenite uspešnost studiranja. Šta biste uzeli kao pokazatelj uspešnosti studiranja – prosečnu ocenu, dužinu studiranja, broj publikovanih saopštenja i radova...? Da li biste različiti ponder dali ovim kriterijuma i koliki? Potom, ukoliko zadržite sve navedene kriterijume i želite da od skupa zadataka u testu odabrete one koji su najviše povezani s kriterijumom, moguće je da će se desiti da su različiti zadaci povezani s različitim kriterijumima, te se postavlja pitanje kojim kriterijumima ćete dati prednost. Navedeni primeri predstavljaju dileme koje se javljaju prilikom nejasno definisanog kriterijuma.

Kombinovanje kriterijumske i normativne interpretacije

Važno je napomenuti da normativna i kriterijumska interpretacija ne moraju nužno isključivati jedna drugu. Postoje

situacije u kojima se one često dopunjuju. U najširem smislu, svi testovi su orijentisani na norme, jer svi podrazumevaju kontinuum predmeta merenja, a određivanje graničnog skora podrazumeva dihotomizaciju tog kontinuuma. Kao slikoviti primer može nam poslužiti crta ličnosti ekstraverzija. Ova crta, kao i bilo koja druga crta ličnosti, ponaša se dimenzionalno, tj. ljude možemo rasporediti duž kontinuuma ekstraverzije, pri čemu bi na jednom polu bili ekstremno introvertni, a na drugom ekstremno ektravertni. Međutim, ako na tom kontinuumu odredimo granični skor ispod kojeg se nalaze introverti, a iznad kojeg su ekstraverti, mi smo samo veštački ljude podelili u dve kategorije.

Najbolji primer kako se normativna i kriterijumska interpretacija dopunjavaju može se naći u obrazovanju. Na testovima znanja obično postoje zadaci koji predstavljaju tzv. kritično znanje na koje svi učenici moraju da znaju odgovor da bi položili test. Dakle, kriterijum predstavlja tačno rešenje na svim pitanjima kritičnog znanja. U ovom slučaju imamo kriterijumsku interpretaciju u odnosu na to da li je učenik savladao kritično znanje ili nije. Međutim, ukoliko pored informacije o ispunjenosti kriterijuma poznajemo još i distribuciju skorova svih učenika, onda postignuće svakog učenika možemo locirati i u odnosu na referentnu grupu čime dobijamo informaciju i o relativnom položaju učenika u grupi, tj. normativnu interpretaciju (vidi Sliku 16). U ovom primeru dobijamo dve informacije: 1. da li je učenik ispunio kriteriijum – da li je savladao kritično znanje i 2. kakvo je njegovo postignuće u odnosu na ostale učenike koji su radili test (npr. da li je izuzetno, prosečno i sl.).



Slika 16. Kriterijumska i normativna interpretacija.

Pored toga, nekad norme dobijene na specifičnoj populaciji mogu poslužiti za dobijanje kriterijuma. Na primer, ukoliko znamo kakav je prosek na nekom testu depresivnosti na uzorku iz opšte populacije i na uzorku pacijenata kojima je dijagnostikovan depresivni poremećaj, možemo odrediti kritični ili granični skor na testu koji bi ukazivao na prisustvo depresije. Možemo ići korak dalje, pa odrediti raspone skorova, odnosno nekoliko kritičnih skorova, i testom identifikovati one koji nemaju depresivni poremećaj, koji su vulnerabilni na depresiju, i koji imaju jasno prisustvo depresivnog poremećaja. Ukoliko se susedni rasponi skorova razlikuju po interpretaciji i ukoliko među njima postoje kvalitativno različite deskripcije, to predstavlja osnovu za kriterijumsku interpretaciju. Još jedan primer se može naći u oblasti razvojne psihologije. Prilikom definisanja normi, spomenute su i razvojne norme. Na osnovu poznavanja dinamike određenih razvojnih aspekata, postavljene su razvojne norme, a one mogu poslužiti za postavljanje kriterijuma. Na primer, u motornom razvoju može se očekivati da će dete od 16 meseci moći samostalno da pravi po nekoliko koraka napred, da se saginje ili čučne kako bi podiglo predmet sa poda i slično. Na osnovu toga šta sve dete može u određenom uzrastu, dobijaju se bodovi koji se potom tumače u odnosu na razvojne norme. Međutim,

na osnovu uvida u norme, moguće je postaviti kriterijum koji odgovara onome što je prosečno ili tipično ponašanje deteta na određenom uzrastu. Na primer, kriterijum za pravilan motorni razvoj može biti da dete samostalno hoda u 16. mesecu. U ovom slučaju bi se na osnovu ispunjenosti postavljenog kriterijuma tumačio razvoj deteta. Takođe, nekada se kao kriterijum može postaviti i nešto što bi se okarakterisalo kao ispodprosečan ili iznadprosečan skor, a ne nužno prosečan skor. S obzirom na to, odluke donete na osnovu kriterijumske interpretacije mogu biti i odluke donete na osnovu normi, iako norme nisu eksplicitno specifikovane (npr. nema interpetacije u terminima nekih normi kao što su razvojne norme ili IQ-skorovi i slično).

Vrste kriterijumske interpretacije

Postoje tri osnovne vrste kriterijumske interpretacije: one koje uključuju poređenje skora ispitanika sa graničnim skorom, interpretacije bazirane na tabelama ili graficima očekivanih vrednosti i interpretacije u odnosu na sadržaj ispitivanog područja (AERA, APA, & NCME, 2006). Među njima, najčešće se koristi interpretacija bazirana na graničnom skoru. U nastavku će biti objašnjena svaka od navedenih vrsta interpretacije.

Granični skor

Granični skor je specifična tačka na skali testovnih skorova u odnosu na koju se rezultati ispod i iznad te tačke različito interpretiraju i tretiraju, u smislu donošenja odluke (AERA, APA, & NCME, 2006). U nekim slučajevima dovoljan je jedan granični skor koji odvaja ispitanike koji su zadovoljili od onih koji nisu zadovoljili kriterijum (oni koji su položili ispit i oni koji to nisu). Na primer, ako se na nekom testu znanja

postavi kriterijum da je 55% tačnih odgovora potrebno za prolaz, svi studenti koji ostvaruju 54% tačnih odgovora ili manje, neće položiti test. To ujedno znači i da će se interpretacija postignuća za studenta koji ima 54% tačnih odgovora i studenta koji ima samo 20% – biti ista. S druge strane, interpretacija postignuća studenta koji ima 55% tačnih odgovora biće ista kao i interpretacija studenta koji ima 100% tačnih odgovora.

S obzirom na to da u nekim situacijama postoji potreba za finijom distinkcijom, ponuđeno je **više graničnih skorova**. Na taj način su definisani različiti nivoi onoga što je predmet merenja testa, čime se definiše i različita interpretacija skorova. Najčešći primer koji se koristi za to je u proceni nivoa poznavanja jezika, pa tako na testu jezika možemo imati četiri granična skora za određivanje početnog, srednjeg, višeg i ekspertskog nivoa znanja jezika. Pored toga, postoje situacije kada su postavljeni zahtevi ispunjenja kriterijuma visoki, te postoji potreba da se definiše i kategorija ispitanika koja ima uslovni prolaz ili nešto tome slično. Na primer, kada bismo na testu znanja postavili kriterijum 80% tačnosti za prolaz, dati šansu onima koji imaju postignuće nešto ispod graničnog skora. Tako bi svi koji su uradili tačno od 70 do 79% pitanja, iako su pali, imali priliku da se ponovo testiraju kroz određeni vremenski period. Još jedan primer je već promenut iz kliničke prakse, kada se na osnovu skora na testu definiše granični skor za određenje prisustva poremećaja, ali i drugi granični skor za određenje vulnerabilnosti za poremećaj. Iako u ovoj situaciji postoji jedan granični skor, sama činjenica da postoje dva moguća ishoda za one koji su pali ili koji nisu ispunili uslov za dijagnostikovanje poremećaja, čini da ovakvi testovi ipak imaju, tehnički, dva granična skora.

Takođe, nekada postoji potreba za uključivanjem **više kriterijuma**. Na primer, za dobijanje stipendije možemo kao kriterijume

ili uslove postaviti prosečnu ocenu preko 9,00 i objavljen rad na naučnoj konferenciji ili u časopisu. Ukoliko kandidat ispunjava oba uslova, dobiće stipendiju. Važno je naglasiti da postavljeni kriterijumi ne mogu biti kompenzacija jedan drugom. Tako, ukoliko imamo kandidata čiji je prosek 10,00 ali nema objavljen rad, takav kandidat, prema ovako postavljenim kriterijumima, ne može dobiti stipendiju. Isto tako, ukoliko su, na primer, kriterijumi za pilotsku obuku minimum prosečna opšta inteligencija i veoma dobar vid, nećemo primiti kandidata koji je iznadprosečno inteligentan ukoliko ima problema s vidom. Višestruki kriterijumi se mogu postaviti i hijerarhijski po važnosti, te se na osnovu jednog kriterijuma može napraviti početna selekcija ili trijaža, pa se na osnovu drugog kriterijuma može izvršiti konačan izbor kandidata koji su prošli početnu selekciju (što je česta praksa u okviru profesionalne selekcije kandidata).

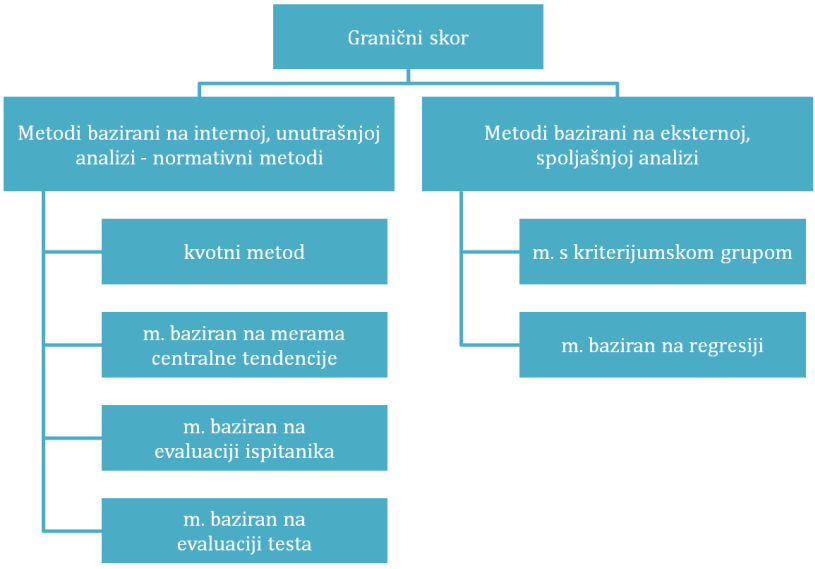
Granični skorovi mogu se odrediti kako za bruto rezultate, tako i za izvedene rezultate tj. za različite vrste normi. S obzirom na to, način izražavanja skorova nije prepreka za postavljanje kriterijumske interpretacije i već je bilo reči o tome kako se kriterijumska i normativna interpretacija međusobno ne isključuju.

Kako odrediti granični skor?

Postoji nekoliko načina određivanja graničnog skora na testu (Slika 17). U najširem smislu, mogu se razlikovati normativni metodi i metodi bazirani na eksternoj analizi. **Normativnim metodama** je obuhvaćen samo skor na testu, bez uvida u relacije tog skora sa nekim eksternim kriterijumom. Za normativne metode bi se moglo reći da su bazirane na internoj, unutrašnjoj analizi i u okviru ovih metoda ključna je *sadržinska validnost* testa. S obzirom na to da se sadržinska validnost

utvrđuje na osnovu intersubjektivne saglasnosti procenjivača, za utvrđivanje kriterijuma se često i angažuju procenjivači tzv. panelisti.

Važno je naglasiti da se normativni *metod* ne meša sa normativnom *interpretacijom*, iako sam naziv metoda nije najsrećnije odabran. Naime, normativni metod je metod određivanja kriterijuma, a naziva se normativnim jer nam je za njega potreban uvid u skorove svih ispitanika. Na osnovu normativnog metoda se postavlja kriterijum i skor ispitanika se tumači u odnosu na zadovoljenje kriterijuma. Iako nam uvid u sve skorove služi za određivanje kriterijuma, kriterijumska interpretacija se odnosi isključivo na ostvarenje standarda, a ne i na to da li je skor prosečan ili iznadprosečan. Podsetićemo se, normativna interpretacija je referetni okvir za tumačenje skorova, na osnovu kojeg se skor ispitanika tumači u odnosu na njegov položaj unutar referentne grupe, pa tako, neki skor možemo okarakterisati kao prosečni, iznadprosečni, na 65. percentilu i sl.



Slika 17. Podela metoda određivanja graničnog skora.

Za razliku od normativnih metoda, **metode bazirane na eksternoj analizi** obuhvataju relacije sa nekom spoljnom, eskternom varijablom. Samim tim, ovaj skup metoda baziran je na eksternoj, spoljašnjoj analizi. Druga, eksterna varijabla mora biti povezana sa onim što je predmet merenja testa za koji želimo da dobijemo granični skor. Ovaj način podrazumeva utvrđivanje *kriterijumske validnosti* testa i granični skor je uvek empirijski zasnovan. Prednost eksternih metoda je u tome što zahtevaju manje resursa – nema angažovanja procenjivača (iako je moguće), te je moguće za puno kraće vreme odrediti više graničnih skorova, ukoliko postoji potreba za tim. Nedostatak ovih metoda je upravo u oslanjanju isključivo na empirijske podatke, te ne postoji uvid u to da li je određeni granični skor adekvatan za primenu testa na drugim grupama ispitanika.

Kvotni metod

Prvi od normativnih metoda je tzv. kvotni metod koji je vrlo jednostavan i sastoji se u tome da se selektuje određeni procenat ispitanika ili broj najboljih, npr. 10% najboljih ili prvih 20 kandidata na listi. Kriterijum je, dakle, postavljena kvota. Prednost ovog metoda je u tome što se unapred zna broj onih koji će ispuniti kriterijum, ali je nedostatak što se na osnovu ovog kriterijuma ne garantuje da će svi kompetentni kandidati proći kriterijum.

Metod baziran na merama centralne tendencije

Za određivanje graničnog skora na osnovu ovog metoda potrebno je da znamo distribuciju skorova i neku od mera centralne tendencije. U ovom slučaju, granični skorovi se mogu odrediti na osnovu podele po medijani (eng. *median split*), tj. po središnjoj vrednosti distribucije skorova. Time dobijamo dve grupe ispitanika od

kojih je 50% ispod i 50% iznad te središnje vrednosti. Moguće je, takođe, granične skorove odrediti na osnovu podele po aritmetičkoj sredini. Ukoliko želimo više graničnih skorova, možemo uključiti i vrednost standardne devijacije, te podelu izvršiti po principu AS - 1SD i AS + 1SD, na osnovu čega se dobijaju tri grupe:

1. niskoprosečni (oni koji imaju skorove ispod AS - 1SD);
2. prosečni (oni koji imaju skorove u rangu od AS - 1SD do AS + 1SD);
3. iznadprosečni (oni koji imaju skorove iznad AS + 1SD).

Naravno, moguće je napraviti podelu i u pet grupa ukoliko se uzmu i granični skorovi od AS - 2SD i AS + 2SD.

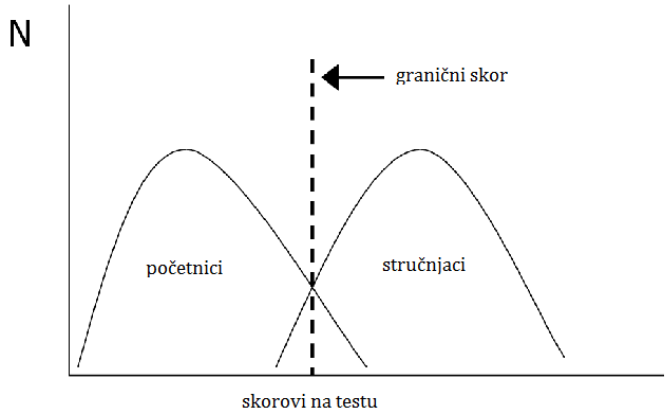
Ovaj način je već pomenut u okviru normativne interpretacije, prilikom određivanja kategorijalnih normi, te je u ovom slučaju teško napraviti razliku između normativne i kriterijumske interpretacije. Nevolja u razgraničavanju ovog metoda u odnosu na to da li je normativni ili kriterijumski potiče od toga što je kriterijum postavljen na distribuciji skorova svih ispitanika. Ovakav metod se najčešće koristi za potrebe istraživanja ili kada je dovoljna samo gruba procena nečijeg funkcionisanja (najčešće prilikom trijaže).

Metod baziran na evaluaciji ispitanika *(eng. method based on evaluation of examineers)*

U okviru ovog metoda možemo razlikovati metod baziran na kontrastiranju grupa i metod baziran na graničnoj grupi. U oba slučaja izbor članova grupe vrše procenjivači (**panelisti**), eksperti u datoj oblasti koja je predmet merenja testa, za razliku od metoda baziranih na merama centralne tendencije gde se članstvo u grupi vrši na osnovu prethodno utvrđene distribucije skorova. Za razliku od nezavisnih

procenjivača, panelisti donose odluku grupno, u okviru diskusije, gde iznose argumente za i protiv i zajedno prolaze kroz njih, obično u nekoliko faza. Pored toga što odluka o pripadnosti grupi zavisi od shvatanja panelista o tome šta bi za dati predmet merenja predstavljao nizak, prosečan, ili visok nivo znanja, kompetencije ili učinka, ta odluka može biti pristrasna ako se ocenjuju irelevantne karakteristike ispitanika (npr. ako se o znanju donosi odluka i na osnovu osobina ličnosti, opšteg utiska i drugih karakteristika). Stoga je vrlo važno da panelisti prođu sistematsku obuku u vezi s kriterijumima donošenja odluke o pripadnosti grupi u odnosu na to šta je predmet merenja testa.

Prvi od navedenih metoda bazranih na evaluaciji ispitanika je **metod baziran na kontrastiranju grupa** (eng. *method based on contrasting groups*, Bingham, 1937; Cizek, 2001; Zieky & Livingston, 1977). Ovaj metod sastoji se u tome da se ispitanici podele na grupe u odnosu na njihov nivo ekspertize, uobičajeno na grupu početnika i grupu stručnjaka u oblasti koja se procenjuje testom. Potom se obema gupama zada isti test za koji se želi odrediti granični skor i odrede distribucije skorova za svaku grupu. Granični skor predstavlja presek tih distribucija (Slika 18). Ukoliko su distribucije normalne, dovoljno je da poznajete *AS* i *SD*, te na osnovu softvera na sledećoj veb-adresi dobijete tačku preseka dve distribucije: http://psihologija.ff.uns.ac.rs/presek/



Slika 18. Određivanje graničnog skora preko metode bazirane na kontrastiranju grupa.

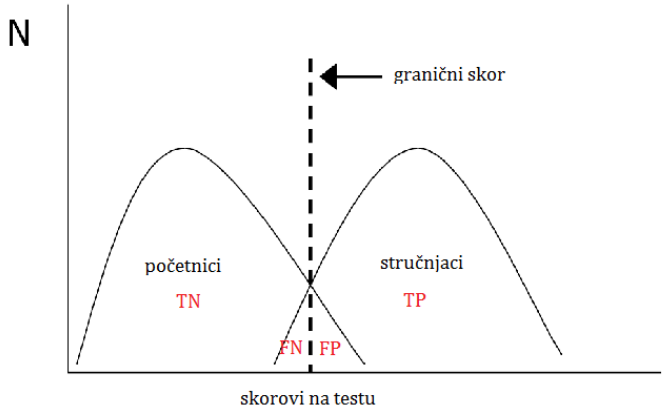
Prilikom određenja graničnog skora možemo se voditi i drugim principima, a ne samo time u kojoj tački se nalazi presek distribucija. Polaznu bazu za tumačenje nam pruža teorija detekcije signala, a da bismo bolje objasnili njene principe, razmotrimo moguće ishode podešenog graničnog skora kao na Slici 19. Moguće su četiri grupe ispitanika:

TP – tačno pozitivni (eng. ***T**rue **P**ositive*) – to su stručnjaci koji postižu skor iznad graničnog, tj. tačno prepoznati stručnjaci;

FP – falš ili lažno pozitivni (eng. ***F**alse **P**ositive*) – to su početnici koju imaju skor iznad graničnog, tj. koje je test prepoznao kao stručnjake;

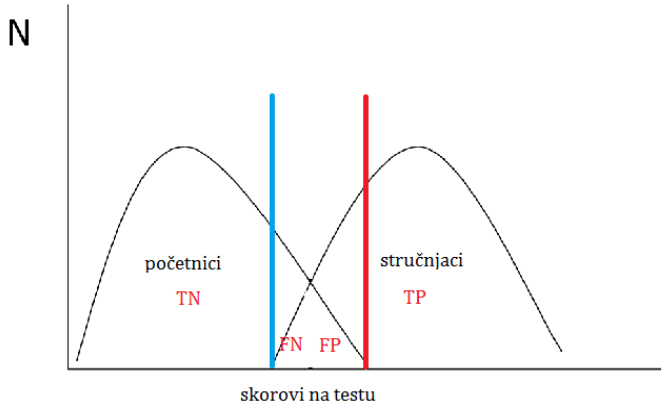
FN – falš ili lažno negativni (eng. ***F**alse **N**egative*) – to su stručnjaci koji imaju skor ispod graničnog, tj. koje je test prepoznao kao početnike;

TN – tačno negativni (eng. ***T**rue **N**egative*) – to su početnici koji imaju skor ispod graničnog, i tačno su prepoznati kao početnici.



Slika 19. Moguće grupe ispitanika na osnovu metoda baziranog na kontrastiranju grupa.

S obzirom na ovu podelu, granični skor možemo odrediti tako da smanjimo broj lažno pozitivnih, tj. da ga pooštrimo (crvena linija na Slici 20) ili da smanjimo broj lažno negativnih, tj. da ga ublažimo (plava linija na Slici 20, više u Sireci, Robin, & Patelis, 1999).



Slika 20. Mogući granični skorovi na osnovu metoda baziranog na kontrastiranju grupa.

Prednost ovog metoda je u tome što se granični skor određuje pre finalne primene testa, u pilot-studiji u kojoj učestvuju samo početnici i stručnjaci. Pored toga, kao prednost ističe se i to što se ovaj metod može primeniti na binarnim i politomnim ajtemima (Morgan & Michaelides, 2005). Ovaj metod se preporučuje i kada se neki već utvrđeni granični skor želi proveriti (zarad potvrđivanja istog ili uspostavljanja novog graničnog skora). Nedostatak predstavlja to što odluka o tome koji ispitanik će pripadati kojoj grupi (u našem primeru, stručnjacima ili laicima) može biti subjektivna, zasnovana na stavovima panelista. Kao što je već rečeno, sistematska obuka panelista može pomoći u eliminaciji pristrasnosti prilikom donošenja odluke o pripadnosti određenoj grupi ispitanika.

Drugi od metoda baziranih na evaluaciji ispitanika je **metod baziran na graničnoj grupi** (eng. *method based on bordeline group*, Zieky & Livingston, 1977). U okviru ovog metoda panelisti identifikuju tzv. graničnu grupu, koja ne pripada ni početnicima ni stručnjacima, već je negde između. Kada se selektuje ova grupa, njima se zada test, a medijana ili aritmetička sredina na tom testu se uzima kao granična vrednost. Iako deluje jednostavno, identifikovanje granične grupe je vrlo zahtevno, zahtevnije nego određivanje grupe početnika ili stručnjaka. Stoga je osnovni nedostatak ovog metoda upravo u selekciji ispitanika. Svi nedostaci pomenuti kod odlučivanja o pripadnosti grupama kod metoda baziranog na kontrastiranju grupa, važe i za ovaj metod.

Metod baziran na evaluaciji testa *(eng. method based on evaluation of the test)*

Ova porodica metoda je veoma raznovrsna i zasniva se, takođe, na proceni panelista. U ovom slučaju ne procenjuju se ispitanici, već stavke u testu i predloženi granični skor. Svaki metod ima svoje karakteristike, ali u osnovi, procenuju se ajtemi testa tako da se odredi proporcija isptanika s minimalnim ili graničnim kompetencijama koji bi tačno odgovorili na dati ajtem. Potom se procene panelista (koje predstavljaju proporcije) uprosečavaju po svakom pitanju, pa sumiraju i dobijeni skor predstavlja granični skor. Procena panelista vrši se u nekoliko etapa, pri čemu je samo prva procena nezavisna, a ostale se donose na osnovu diskusije sa ostalim panelistima. Iako se retko dešava da se kasnija procena drastično razlikuje od inicijalne, moguće su manje korekcije nakon određenih faza i diskusija između panelista, novih uvida i slično.

U ove metode spadaju Angoffov metod i modifikovani Angoffovi metodi, Nedelskyjev metod, Jeagerov metod, Ebelov metod, Body of Work, Bookmark metod itd. Svaki od njih primeren je za određeni format ajtema (npr. dihotomni, višestruki izbor), a npr. Bookmark metod je baziran na Teoriji odgovora na stavke, a ne Klasičnoj testnoj teoriji, i može da se koristi na testovima koji sadrže različite formate ajtema. Dok je karakteristika nekih metoda u tome što ne zahtevaju podatke ispitanika (Angoffove metode), karakteristika drugih je ta da su zasnovane na realnim podacima (Body of Work). Takođe, dok se u nekim metodama procena donosi na osnovu toga da li ispitanik s minimalnim kompetencijama može tačno odgovoriti na ajtem (Angoffove metode), u drugima se procenjuje da li ispitanici mogu tačno odgovoriti na ajtem, bez obzira na nivo kompetencija (Jeagerov

metod). Kao nedostatak ove grupe metoda navodi se to što je panelistima nekada teško da daju procene u terminima proporcija, te se među njihovim odgovorima često ponavljaju iste proporcije. Kao odgovor na ovaj nedostatak razvijeni su dihotomni metodi gde panelisti samo procenjuju sa "da" ili "ne" da li neko s minimalnim kompetencijama može potvrdno ili tačno odgovoriti na dati ajtem. Pored toga, kao nedostatak navodi se i to što panelisti mogu izgubiti uvid u opšte postignuće i učinak ispitanika zbog fokusa na pojedinačne ajteme (Morgan & Michaelides, 2005).

Metod s kriterijumskom grupom

Ovaj metod je prvi od metoda baziranih na eksternoj analizi. U okviru ovog metoda imamo eksternu, kriterijumsku varijablu nominalnog nivoa merenja. Za određivanje graničnog skora najčešće se primenjuje analiza tzv. ROC krive (eng. *Receiver Operating Charachteristic Curve*, više u Hanley & McNeil, 1982; Swets, Dawes, & Monahan, 2000). Pre nego što opišem ovu analizu, najpre ću objasniti logiku koja stoji u njenoj osnovi. Uzmimo, na primer, da smo sastavili novi test za procenu sklonosti ka zloupotrebi alkohola i razvijanju problema s alkoholom, i da smo za njega ponudili granični skor na osnovu kojeg se ispitanici mogu svrstati u one koji imaju i koji nemaju problema s upotrebom alkohola. Ovaj test smo dali dvema grupama: 1) osobama koje su na tretmanu lečenja od alkoholizma i 2) osobama iz opšte populacije koje nemaju problema s alkoholom, što je utvrđeno detaljnom dijagnostikom. Da bismo eliminisali sve potencijalne izvore grešaka u zaključivanju, ove dve grupe treba da budu ujednačene po relevantnim karakteristikama za probleme s alkoholom (pol, starost,

istorija bolesti). Moguća su četiri ishoda testiranja, kao u slučaju metoda s kontrastiranim grupama (Tabela 17):

TP – tačno pozitivni – to su osobe koje imaju problema s alkoholom, i koji su na testu tako i prepoznati (to su osobe na tretmanu koje imaju skor iznad graničnog);

FP – falš ili lažno pozitivni – to su osobe koje nemaju problema s alkoholom, ali koje je test prepoznao kao da imaju problema s alkoholom (to su osobe koje nisu na tretmanu i koje imaju skor iznad graničnog);

FN – falš ili lažno negativni – to su osobe koje imaju problema s alkoholom, ali koje test prepoznaje kao da nemaju problema (to su osobe na tretmanu koje imaju skor ispod graničnog);

TN – tačno negativni – to su osobe koje nemaju problema s alkoholom i koje test tako i prepoznaje (to su osobe koje nisu na tretmanu i imaju skor ispod graničnog).

Granični skor je dobro određen kada se njime tačno procenjuje ko ima, a ko nema problema s alkoholom, i kada nemamo lažno pozitivno i lažno negativno svrstanih ispitanika. Međutim, u praksi se to ređe dešava, te uvek imamo određeni broj ispitanika u sva četiri ishoda. Tačnost klasifikacije je onda stepen slaganja rezultata dobijenih preko graničnog skora i stvarnog stanja.

Tabela 17

Tabela za određivanje ishoda testiranja spram pripadnosti grupi

	Stvarno stanje		
Test	Ima problema s alkoholom	Nema problema s alkoholom	Ukupno
Pozitivan	TP	FP	TP + FP
Negativan	FN	TN	FN + TN
Ukupno	TP + FN	FP + TN	

Da bismo dobili tačnost klasifikacije, treba da vodimo računa o osetljivosti ili senzitivnosti (eng. *sensitivity*) testa i specifičnosti testa (eng. *specificity*). **Osetljivost testa** je karakteristika testa da pokaže pozitivan rezultat kod osoba koje stvarno poseduju mereno svojstvo – stopa stvarno pozitivnih. Osetljivost testa predstavlja proporciju tačno pozitivnih, tj. tačno klasifikovanih među osobama s problemima s alkoholom, u odnosu na ukupan broj onih koji stvarno imaju ovih problema:

$$\text{Osetljivost} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

Specifičnost testa je karakteristika testa da pokaže negativni rezultat kod osoba koje stvarno nemaju mereno svojstvo – stopa stvarno negativnih. Specifičnost testa predstavlja proporciju tačno negativnih, tj. tačno klasifikovaih među onima koji nemaju problema s alkoholom, u odnosu na ukupan broj onih koji stvarno nemaju problema s alkoholom:

$$\text{Specifičnost} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Uspešnost klasifikacije predstavlja zbir tačno pozitivnih i tačno negativnih, podeljen ukupnim brojem ispitanika u obe grupe i izražava se preko procenta tačnosti klasifikacije:

$$\text{Uspešnost klasifikacije} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

Zamenimo brojkama naš primer, kao u Tabeli 18. Lako možemo izračunati da je osetljivost testa 80 / 100 = 0,80, tj. 80%, a senzitivnost 75 / 100 = 0,75, tj. 75%. Uspešnost klasifikacije je (80+75) / (80 + 75 + 25 + 20) = 155 / 200 = 0,775, tj. 77,5%.

Tabela 18

Primer tabele za određivanje ishoda testiranja spram pripadnosti grupi uz baznu stopu 50%

Test	Stvarno stanje		Ukupno
	Ima problema s alkoholom	Nema problema s alkoholom	
Pozitivan	80	25	105
Negativan	20	75	95
Ukupno	100	100	200

Ukoliko bismo menjali granični skor, pa ga snizili/ublažili, povećali bismo osetljivost, ali snizili senzitivnost. Naime, snižavanjem graničnog skora bismo smanjili broj lažno negativnih i povećali broj tačno pozitivnih, ali bismo ujedno povećali broj lažno pozitivnih i smanjili broj tačno negativnih. U ovoj situaciji, iako je veća šansa da oni koji imaju problema s alkoholom zaista tako budu i klasifikovani, ujedno je i veća šansa da oni koji nemaju problema s alkoholom budu klasifikovani kao da imaju problema s alkoholom. S druge strane, ukoliko bismo pooštrili kriterijum, tj. povećali granični skor, onda bismo povećali senzitivnost, ali snizili osetljivost. Tada bi broj i lažno i tačno pozitivnih bio smanjen, a broj i lažno i tačno negativnih bi bio povećan. Na ovaj način, iako bismo smanjili broj tačno pozitivnih, bili bismo sigurniji da oni koji imaju skor iznad graničnog zaista imaju problema s alkoholom, jer je broj lažno pozitivnih smanjen.

Primećujemo da se uspešnost klasifikacije menja s menjanjem bazne stope prisutnosti fenomena koji se ispituje, na šta su ukazali još Mil i Rouzen (Meehl & Rosen, 1955). Naime, da bi se povećala snaga testa i zadovoljili metodološki standardi, istraživači uglavnom gledaju da broj ispitanika u grupama bude ujednačen, što odgovara baznoj stopi 50%. Međutim, fenomeni uglavnom nemaju ovako veliku baznu stopu, posebno u domenu kliničke psihologije. Uzmimo, na primer, da je osoba kojima je dijagnostikovan poremećaj upotrebe alkohola oko 15% u populaciji. Na uzorku od 200 ispitanika, 15% je 30 ispitanika. Kada bi se u nacrtu ispunila ova bazna stopa, a uz istu osetljivost i specifičnost kao u prethodnom slučaju, dobili bismo vrednosti kao u Tabeli 19.

Tabela 19

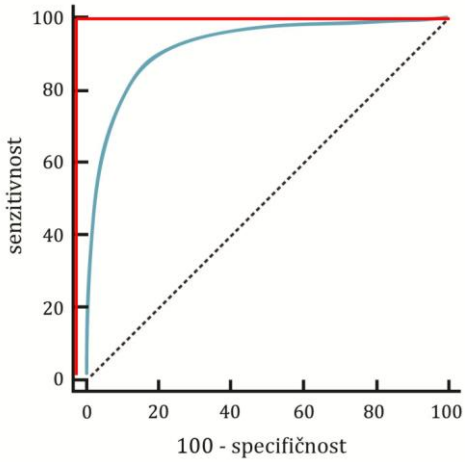
Primer tabele za određivanje ishoda testiranja spram pripadnosti grupi uz baznu stopu 15%

Test	Stvarno stanje		Ukupno
	Ima problema s alkoholom	Nema problema s alkoholom	
Pozitivan	24	42	66
Negativan	6	128	134
Ukupno	30	170	200

Na osnovu Tabele 19 možemo videti da ćemo kod pozitivnog testa verovatnije pogrešiti u odluci o tome da li neko ima problema s alkoholom, jer većina ispitanika koja ima iznadgranični skor zapravo nema problema s alkoholom (42 / 66 * 100 = 63,64%). Rezultati se drastično razlikuju kada je bazna stopa još niža, npr. oko 5%. Dakle, kada je populacija onih koji nemaju problema s alkoholom upadljivo

veća, veće su šanse da ćemo pogrešiti pri klasifikaciji upravo te populacije.

Na osnovu ovoga može se zaključiti da je pravi izazov naći optimalni granični skor tako da osetljivost i specifičnost budu najveće moguće. Upravo tome služi analiza ROC krive. ROC kriva je grafički prikaz osetljivosti i specifičnosti za svaki mogući granični skor, na čijoj se *x*-osi nalaze vrednosti specifičnosti oduzete od 1 (1 – specifičnost), tj. proporcija lažno pozitivnih, a na *y*-osi vrednosti osetljivosti (Slika 21). Na osnovu ove analize nalazi se najbolji odnos osetljivosti i specifičnosti, tj. situacija da nam je broj tačno pozitivnih i tačno negativnih najviši mogući. Na osnovu ROC krive izračunava se površina ispod krive (eng. *area under the curve* – AUC) koja je pokazatelj stepena diskriminativnosti među grupama. U našem primeru, AUC bi predstavljala verovatnoću da će slučajno odabrana osoba koja ima problema s alkoholom imati viši skor na testu, nego što će to imati slučajno odabrana osoba koja nema problema s alkoholom. AUC može imati vrednost od 0 do 1, pri čemu vrednost 1 predstavlja savršenu diskriminativnost među grupama, tj. savršenu predikciju na osnovu graničnog skora. Na Slici 21, kada je AUC = 1, to bi bila savršena senzitivnost i specifičnost i ROC kriva bi zapravo bila podudarna s *y*-osom i gornjom granicom vrednosti na *x*-osi, kako bi pokrivenost površine ispod krive bila 100% (na Slici 21 to je predstavljeno crvenom ROC krivom). Vrednost AUC = 0,5 predstavlja predikciju na nivou slučaja (50–50%), koja nam je praktično neupotrebljiva. U našem primeru, to bi bila situacija u kojoj postoji 50% šanse da osoba na osnovu graničnog skora bude okarakterisana kao da ima problema s alkoholom. Na Slici 21 je AUC = 0,5 predstavljena dijagonalnom crnom linijom.



Slika 21. Primer ROC krive.

U praksi nikada nemamo savršenu senzitivnost i specifičnost, te se ROC kriva najčešće nalazi između dijagonale i gornjeg levog ugla. Što test ima bolju specifičnost i senzitivnost, to će ROC kriva biti bliža levom gornjem uglu, a što im goru specifičnost i senzitivnost, to će biti bliža dijagonali[28]. Dakle, što je veća površina ispod krive, to je test diskriminativniji (bolje razlikuje dve grupe) i veća je šansa da se među tim potencijalnim graničnim skorovima na ROC krivoj nalaze oni koji će najbolje razlikovati grupe, sa što manjim brojem lažno pozitivnih i negativnih. U primeru na Slici 21 plavom linijom je predstavljena empirijska, dobijena ROC kriva i može se zaključiti da je površina ispod nje prilična, tj. da je u pitanju test dobre diskriminativnosti.

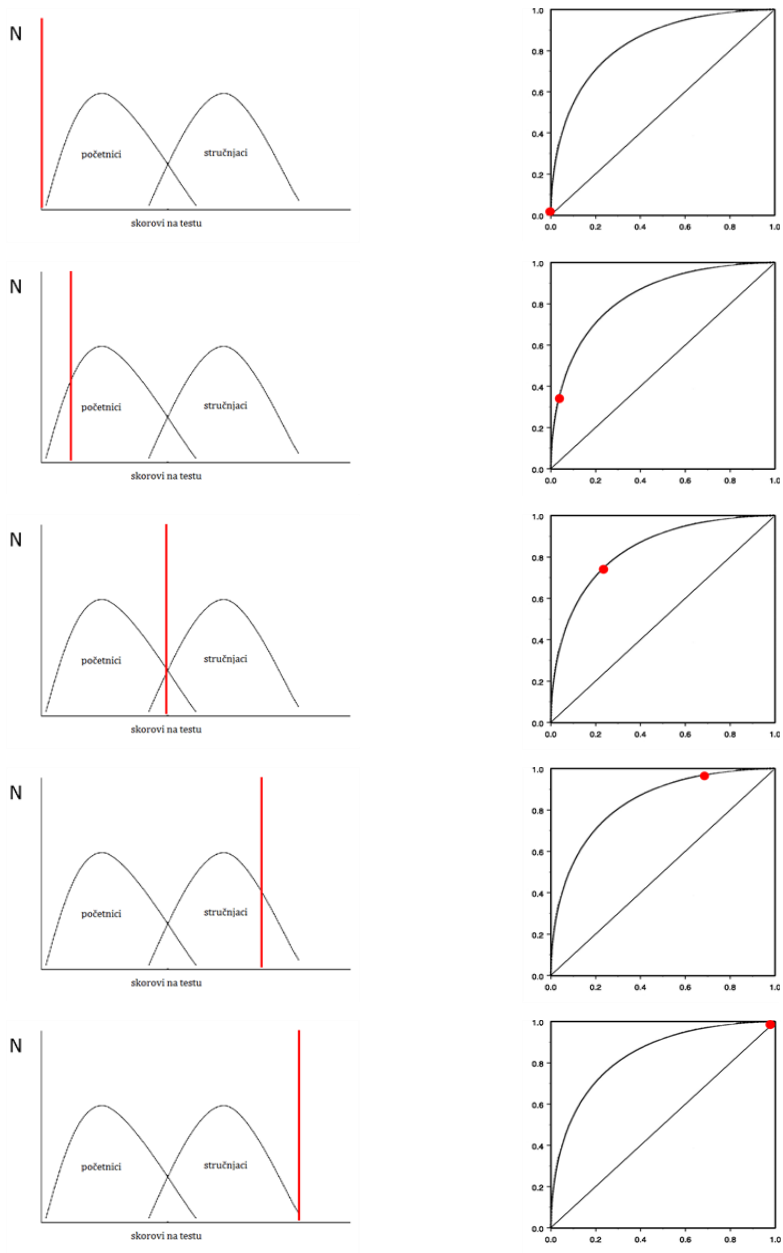
U literaturi se mogu naći različite preporuke za tumačenje veličine AUC. Referirajući na Koenove preporuke u vezi s veličinom efekta za visinu korelacije, neki autori (Dahle, Schneider, & Ziethen, 2007) preporučuju sledeće kritične vrednosti AUC u cilju određivanja jačine

[28]Animirani edukativni prikaz može se naći na: http://www.navan.name/roc/

predikcije: vrednosti ispod 0,64 (r < ,24) upućuju na malu predikciju, vrednosti od 0,64 do 0,71 (r ≥ ,24) upućuju na umerenu predikciju i vrednosti od 0,72 pa naviše (r ≥ ,37) upućuju na dobru predikciju. U zavisnosti od predmeta merenja, ovi kriterijumi mogu biti stroži, pa tako, na primer, u slučaju određenja prediktivnosti graničnog skora za procenu rizika za ispoljavanje nasilja, autori (Douglas, Webster, Hart, Eaves, & Ogloff, 2001) za umerenu predikciju preporučuju vrednosti od 0,70 do 0,75, a preko 0,75 za dobru.

Podsetićemo se, za kriterijumsku interpretaciju test ne mora biti diskriminativan u celom opsegu predmeta merenja, već samo u domenu graničnog skora, a analiza ROC krive nam pomaže u tome da odredimo koji je granični skor najbolji. Na Slici 22 dati su primeri određivanja graničnog skora u zavisnosti od variranja kriterijuma, tj. od variranja senzitivnosti i specifičnosti. Najoptimalnije određenje kriterijuma, kada se grupe maksimalno razlikuju, predstavlja središnji primer na Slici 22. Pomoću analize ROC krive se preporučuje vrednost skora na testu koja odgovara ovom najoptimalnijem određenju kriterijuma.

Analiza ROC krive ima važne prednosti u određenju graničnog skora, ali takođe, i određene metodološke nedostatke. Osnovni nedostatak je u tome što AUC veoma zavisi od varijabilnosti skorova (npr. Hanson, 2008), te se kao alternativa preporučuje količnik verovatnoće ili odnos šansi (eng. *odds ratio* – *OR*) koji se dobija kao rezultat logističke regresione analize. Analiza ROC krive primenjuje se kada je kriterijum varijabla nominalnog nivoa merenja, mada se u tom slučaju može primeniti i binarna logistička regresija (regresija u kojoj je kriterijum nominalna, binarna varijabla), koja pripada posebnoj grupi metoda eksterne analize. S obzirom na to, analiza ROC krive se ne može primeniti u situaciji kada je kriterjum varijabla višeg nivoa merenja od nominalnog.



Slika 22. Određivanje graničnog skora variranjem senzitivnosti i specifičnosti.

Napomena: *N* – broj ispitanika. Na *x*-osi ROC krive nalazi se vrednost 1 – specifičnost, a na *y*-osi nalazi se vrednost senzitivnosti.

Metod baziran na regresionoj analizi

Metod baziran na regresionoj analizi predstavlja skup različitih vrsta regresionih analiza u zavisnosti od nivoa merenja kriterijuma. S obzirom na nivo merenja kriterijuma, mogu se primeniti logistička regresiona analiza (kada je kriterijum nominalna, grupišuća varijabla), ordinalna regresiona analiza (kada je kriterijum ordinalna varijabla), ili linearna regresiona analiza (kada je kriterijum intervalna ili racio varijabla). U okviru regresione analize procenjuje se predikcija kriterijuma na osnovu skora na testu koji predstavlja prediktor. Kao optimalni granični skor na testu postavlja se onaj na kojem se nalazi zadovoljavajući ili unapred dati nivo kriterijuma.

Budući da se količnik verovatnoće iz logističke regresione analize preporučuje umesto AUC (npr. Hanson, 2008), detaljnije će biti opisan. Logistička regresiona analiza omogućava predikciju kategorijalnih ishoda – kriterijuma, kojih može biti dva (npr. prošao/pao, ima dijagnozu/nema dijagnozu) i tada se radi binarna logistička regresija, ili više (npr. ima dijagnozu/vulnerabilan/nema dijagnozu) i tada se radi multinominalna logistička regresija, pri čemu prediktori mogu biti različitog nivoa merenja. U okviru logističke regresije se, pored osetljivosti, specifičnosti i procenta tačno klasifikovanih ispitanika, dobija i količnik verovatnoće (u IBM SPSS Statistic software-u označen kao Exp(*B*) od eksponencirani logistički koeficijent). Količnik verovatnoće predstavlja promenu verovatnoće pripadanja jednoj kategoriji ishoda kada se vrednost određenog prediktora poveća za jednu mernu jedinicu, kada se ostali prediktori u modelu, ako ih ima, drže pod kontrolom, tj. kada su konstante (Tabachnick & Fidell, 2013). Drugim rečima, to je šansa da ispitanik pripada određenoj kategoriji ishoda kada se prediktor poveća za jednu mernu jedinicu. Na primer, ukoliko na osnovu skora na testu znanja

želimo da predvidimo da li je neko stručnjak ili laik (pri čemu smo stručnjake kodirali kao 1, a laike kao 0), ako je količnik verovatnoće 8,5, to bi značilo da što je veći skor na testu, to je veća verovatnoća da je ispitanik stručnjak. Za svaki dodatni bod na testu, verovatnoća da je ispitanik stručnjak raste 8,5 puta. Ukoliko je količnik verovatnoće manji od 1, onda važi obrnuto tumačenje. Na primer, da nam je test loš i da pogrešno klasifikuje stručnjake i laike, da smo dobili da je količnik verovatnoće 0,50, to bi značilo sledeće – što je skor na testu veći, to je manja verovatnoća da je ispitanik stručnjak. Tačnije, za svaki dodatni bod na testu verovatnoća da je osoba stručnjak opada za 0,5. Ukoliko je odnos verovatnoće 1, to znači da nema značajne povezanosti između prediktora i kriterijuma.

Već je pomenuto da regresioni model može obuhvatati nekoliko prediktora, npr. nekoliko testova na osnovu kojih se predviđa neki ishod, odnosno kriterijum. Kako ima više prediktora u modelu, koristi se višestruka regresiona analiza, pa tako možemo imati logističku višestruku analizu, linearnu višestruku analizu itd. U ovom slučaju je važno da prediktori ne koreliraju međusobno tako da su jedan drugom redundantni, ali da budu povezani sa kriterijumom. Drugim rečima, važno je da nema multikolinearnosti među prediktorima, ali da svaki od njih pokaže inkrementalnu validnost. Inkrementalna validnost se odnosi na količinu informacija ili objašnjenja kriterijuma koju ostvaruje svaki prediktor, povrh onoga što je već objašnjeno drugim prediktorima. Dakle, najbolja situacija je da svaki prediktor objašnjava drugačiji izvor varijabiliteta kriterijuma. Takođe, ne treba veći broj prediktora mešati sa zahtevom za višestrukim kriterijumima. Kod metoda višestruke regresije postoji jedan kriterijum, a prediktori se mogu međusobno nadopunjavati, te ukoliko neko ima nizak skor na jednom testu koji je prediktor, a visok

na drugom, moguće je da će ispuniti kriterijum. Podsetićemo se, kod višestrukih kriterijuma svaki kriterijum mora biti ispunjen i ne postoji mogućnost nadopunjavanja kriterijuma.

Jedan ili više graničnih skorova?

Da li će se autori testa odlučiti za jedan ili više graničnih skorova zavisi najviše od predmeta merenja testa. Na primer, u situacijama visokog uloga, kao što su operacija srca ili sletanje aviona, samo jedna greška je dovoljna da dovede do katastrofalnih posledica. Ovakve situacije su "sve ili ništa", te tehnički gledano granični skor ni ne postoji, jer se samo kompletna uspešnost na testu ocenjuje kao ispunjen kriterijum. Uobičajeno, kod testova visokog uloga postoji samo jedan granični skor. Takav je slučaj, na primer, kod prijema za srednju školu, fakultet, posao, dobijanje stipendije i sl. Međutim, postoje situacije kada ima prostora da se definišu različiti nivoi uspešnosti na testu i to označi većim brojem graničnih skorova. Na primer, u obrazovanju prilikom polaganja nekog testa znanja postoje jasni kriterijumi za prolaznost na testu, ali može se definisati i tzv. uslovni prolaz. Uslovni prolaz odnosi se na učenike ili studente koji nisu zadovoljnili kriterijum, ali su vrlo blizu tome (ako je na testu od 100 poena za prolaz potrebno minimum 55, uslovni prolaz može obuhvatiti skorove od 50 do 54). U tom slučaju, postoje dva granična skora – 50 i 55. U obrazovanju je, takođe, testovima znanja tokom školovanja unapred definisano koja količina rešenih zadataka i pitanja podrazumeva određeni nivo znanja (odličan, vrlodobar, dobar...), tj. postoji više graničnih skorova, ali na kraju školovanja ishod može biti samo prošao–pao i tada postoji samo jedan granični skor. Dakle, da li će postojati jedan ili više graničnih skorova, ili graničnog skora uopšte neće biti, zavisi od više činilaca, a prvenstveno od predmeta merenja i od toga

da li je reč o testu visokog ili niskog uloga. Naime, testovi visokog uloga češće imaju jedan granični skor.

Nedostaci graničnih skorova

Osnovni nedostatak graničnih skorova je u tome što vrlo retko postoji toliko velika razlika između ispitanika koji ima skor tek nešto ispod graničnog i onog ko ima skor tek nešto iznad graničnog. Naime, većina fenomena koji se ispituju testovima su kontinuirani, a postojanje graničnih skorova veštački pravi podelu na kategorije. Svakako da interpretacija na osnovu graničnog skora nama pomaže u donošenju odluka, ali treba imati na umu da će greška klasifikacije biti visoka za osobe koji imaju skorove bliske graničnom skoru.

Na kraju, važno je istaći da se odluka o graničnom skoru mora doneti na osnovu namene testa i konteksta donošenja odluka. Na primer, u situaciji prijema za fakultet bolje je imati niži standard i dozvoliti većini da upiše fakultet, iako neće svi prihvaćeni kandidati završiti fakultet. Međutim, u situaciji kada se selektuju kandidati za neki rizičan posao, npr. za pilota ili astronauta, bolje je imati visok standard i odabrati kandidate koji će s većom sigurnošću biti uspešni u zadatku, iako to znači da neće svi kandidati koji mogu biti uspešni, biti i primljeni.

Takođe, od toga koliko je kriterijum dobro definisan i izmeren zavisi i upotrebljivost graničnog skora. U praksi u obrazovanju, na primer, kriterijum je obično ocena učenika ili studenta, a ista ocena može podrazumevati različiti nivo znanja od odeljenja do odeljenja, od škole do škole, ili od fakulteta do fakulteta (Camara, Kimmel, Scheuneman, & Sawtell, 2003).

Tabele i grafikoni očekivanih vrednosti

Kao što je rečeno, za testove orijentisane na kriterijum najvažnija je validnost. Često postoji potreba da se validnost nekog testa predstavi drugima na jednostavniji i konkretniji način koji ne uključuje (bar ne u prvom planu) statističke informacije poput koeficijenta korelacije i slično. Na taj način je olakšana komunikacija između ljudi različite struke, kao i sa onima koji rešavaju test. Naime, statističke informacije su suviše apstraktne laicima, kao i onim osobama koje nemaju nikakvu edukaciju iz oblasti statistike. Osim toga, pokazano je da se preferiraju drugi načini predstavljanja rezultata (predstavljanje u vidu anegdote ili vizuelno) u odnosu na predstavljanje rezultata preko nekih statističkih pokazatelja, kao što je npr. koeficijent korelacije. Na primer, u jednom istraživanju su ispitanici imali zadatak da pročitaju tekst o nekoj izmišljenoj bolesti i da odaberu jedan od dva tretmana, pri čemu je efikasnost jednog tretmama predstavljena u statističkim terminima, a drugog u terminima anegdote (Freymuth & Ronan, 2004). Rezultati pokazuju da ispitanici preferiraju tretman predstavljen u vidu anegdote, čak iako je jasno negativan. Potom, ispitanicima je dato da učestvuju u istom zadatku, ali ovog puta su statistički podaci predstavljeni vizuelno, u vidu grafikona. U ovakvom zadatku ispitanici su preferirali tretman koji je zaista efektivniji, nezavisno od toga kako je ta efikasnost predstavljena (vizuelno ili preko anegdote). Na osnovu rezultata može se zaključiti da se nestatistički način prikazivanja rezultata preferira, i ne samo to, već i da laici daju tačnije procene ukoliko se rezultati ne prikazuju u statističkim terminima.

S obzirom na potrebu za olakšanom komunikacijom, razvijeni su alternativni načini prikazivanja rezultata o validnosti testa, i to kriterijumske validnosti. Dva osnovna načina prikaza su preko tabela i grafikona očekivanih vrednosti. Ove tabele i grafikoni pokazuju unakrsnu

distribuciju skorova na testu za jednu ili više grupa osoba, u odnosu na izraženost kriterijuma (Lawshe & Bolda, 1958; Urbina, 2014). Drugim rečima, one pokazuju verovatnoću različitih ishoda za osobe koje postižu određene rezultate na testu. Kako se to uobičajeno definiše, tabele i grafikoni očekivanih vrednosti pokazuju verovatnoću uspeha u odnosu na kriterijum u datom intervalu skorova na testu (Lawshe & Bolda, 1958) ili procenat osoba u svakom intervalu skorova na testu koji će biti uspešan u odnosu na dati kriterijum (Cascio, 1977). Njima je prikazana korelacija između skora na testu i kriterijuma. Ukoliko imamo jedan graničan skor, tj. binarni kriterijum (npr. prošao–pao), onda bismo za svaki interval skorova na testu izračunali frekvencu ili procenat ispitanika koji zadovoljava kriterijum (tj. onih koji su prošli). Naravno, tabele i grafikoni očekivanih vrednosti imaju smisla samo ako postoji značajna korelacija između skora na testu i kriterijuma.

Tabela 20

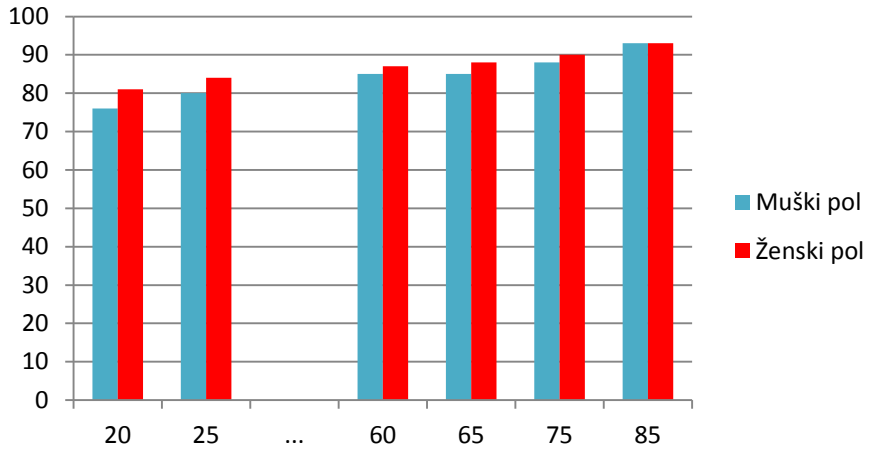
Teorijski primer tabele očekivanih vrednosti za dužinu života

Aktuelne godine starosti	Očekivana dužina života	
	Muški pol	Ženski pol
20	76	81
25	80	84
...	...	...
60	85	87
65	85	88
75	88	90
85	93	93

Najčešća primena tabela i grafikona očekivanih vrednosti je u proceni dužine života. Uzmimo izmišljeni primer u Tabeli 20 u kojoj su prikazane prosečne očekivane godine života za muški i ženski pol na osnovu aktuelnih godina starosti, ukoliko aktuelna stopa mortaliteta

ostane nepromenjena tokom životnog veka. Ukoliko smo osoba ženskog pola i trenutno imamo 25 godina, prema proceni bi trebalo da živimo 84 godine. Međutim, ukoliko trenutno imamo 75 godina, doživećemo 90. rođendan. Tabela očekivanih vrednosti nam govori da što je viša godina starosti ili inicijalna vrednost, to je duži i životni vek, što je vrlo logično. Naravno, proračun koji stoji u osnovi ovog zaključka je strogo statistički, i primer je vrlo jednostavan, jer ne uzima u obzir ostale parametre kvaliteta života, kao što su npr. zdravstveni status i životni stil, niti stopu mortaliteta u datoj državi[29].

Grafikoni očekivanih vrednosti pokazuju istu stvar kao i tabele. Uzmimo primer u Tabeli 18 i predstavimo ga grafički (Slika 23).

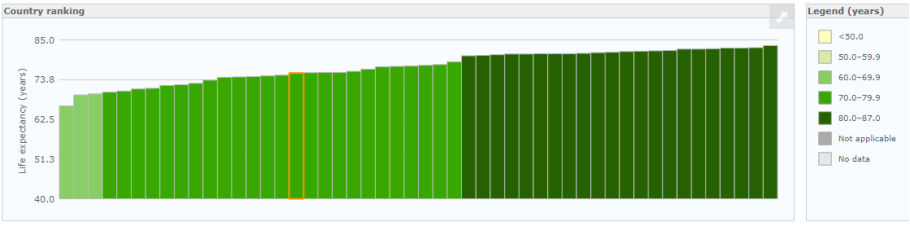


Slika 23. Teorijski primer grafika očekivanih vrednosti za dužinu života.

Pogledajmo sada empirijski primer očekivane dužine života predstavljen na Slici 24. Podaci prikazani na Slici 24 generisani su na

[29] Za detaljniji proračun očekivanih godina života, možete posetiti sajt https://www.livingto100.com/calculator

osnovu izveštaja Svetske zdravstvene organizacije (World Helath Organization – WHO, 2018[30]). Narandžastim je uokvirena Srbija, pri čemu je prosečna starost stanovnika u Srbiji 76,3, čime Srbija spada u grupu zemalja čija je prosečna starost od 70 do 79 godina, zajedno sa npr. Vijetnamom i Lebanom koje imaju istu prosečnu starost stanovnika. Prema ovom izveštaju, zemlja sa najvišom prosečnom dužinom života stanovnika je Japan (84,2), pa potom Švajcarska (83,3), a sa najnižom prosečnom dužinom života je Lesoto (52,9), pa potom Centralnoafrička Republika (53).



Slika 24. Grafik očekivanih vrednosti: očekivana dužina života stanovnika država Evrope (WHO, 2018).

Postoji debata u vezi s tim da li je bolje koristiti frekvence ili verovatnoće prilikom prikazivanja rezultata, odnosno treba voditi računa o tome šta je ljudima razumljivije. Na osnovu istraživanja donošenja odluka, Kaneman i Tverski (Kahneman & Tversky, 1972; Tversky & Kahneman, 1973) su zaključili da je teže donošenje odluka kada je informacija predstavljena u vidu verovatnoća. Nakon njih, veliki broj istraživanja pokazao je da je proces suđenja uspešniji kada se statističke

[30]http://gamapserver.who.int/gho/interactive_charts/mbd/life_expectancy/atlas.html

informacije prezentuju u vidu frekvenci (npr. Brase, 2002; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Jedno od objašnjenja zašto se bolje snalazimo s frekvencama je zbog toga što su informacije u našem okruženju tokom evolucije predstavljene u vidu frekvenci, te su nam one bliže svakodnevnom iskustvu (npr. Hoffrage et al., 2000). Na primer, ako smo od 20 odlazaka u lov 12 puta bili uspešni, bolje ćemo proceniti koliko ćemo biti uspešni u lovu u terminima 12 od 20, nego u terminima 60% uspešnosti ili verovatnoće od 0,60. Neka istraživanja, međutim, pokazuju da prikazivanje frekvenci ne vodi nužno boljem suđenju, i da postoje i druge karakteristike koje utiču na proces suđenja i donošenja odluka u vezi sa statistički prezentovanim informacijama, kao što je npr. razvijenost numeričkih sposobnosti (npr. Hoffrage et al., 2000). Ipak, treba imati u vidu da rezultati nisu dosledni (npr. Macchi, 2000) i da postoji potreba za sprovođenjem meta-analitičke studije u ovoj oblasti.

Kako napraviti tabelu očekivanih vrednosti?

Uzmimo situaciju da na osnovu postignuća na klasifikacionom ispitu za upis na fakultet želimo da odredimo uspeh studenta na kraju studija. Zarad jednostavnijeg prikaza, postignuće na klasifikacionom ispitu će biti izraženo preko percentilnih kvartila. Napravićemo tabelu u kojoj će u redovima biti percentilni kvartili dobijeni na klasifikacionom ispitu, a u kolonama opsezi prosečnih ocena na završetku studija (Tabela 21). U ćelije ove tabele ćemo uneti broj studenata (frekvence) koji odgovara datim ukrštenim vrednostima postignuća na ispitu i finalne prosečne ocene. Potom ćemo izračunati sume frekvenci po redovima i kolonama.

Tabela 21

Tabela kontingencije: prosečna ocena na kraju studija na osnovu postignuća na klasifikacionom ispitu

Klasifikacioni ispit Percentilni kvartil	Prosečna ocena na kraju studija 6–6,99	7–7,99	8–8,99	9–10	Suma
1–25	12	10	2		24
26–50	1	5			6
51–75	2	4	29		35
76–99			2	10	12
Suma	15	19	33	10	77

Nakon toga, treba da izračunamo procenat studenata u svakoj ćeliji za svaki red. To se radi tako što se najpre izračunaju proporcije za svaku ćeliju, pa one pomnože sa 100 kako bi se dobili procenti. Na primer, u redu koji odgovara drugom kvartilu (26–50) nalazi se 1 student u prvom i 5 u drugom redu. Da bi se dobile proporcije za svaku ćeliju, svaku ćeliju, tj. broj studenata u ćeliji, treba podeliti sumom za taj red (u našem primeru suma iznosi 6). Dakle, za ćeliju u prvom redu će biti 1 / 6 = 0,17, tj. 17%, za ćeliju u drugom redu će biti 5 / 6 = 0,83, tj. 83% i tako redom (Tabela 22). Dobijeni procenti ukazuju na verovatnoću ostvarivanja određenog proseka na osnovu postignuća na klasifikacionom ispitu. Helmstater (Helmstadter, 1964) je istakao da tabele očekivanih vrednosti ukazuju na verovatnoću da će student s datim skorom na testu da se ponaša u nekoj narednoj situaciji na određeni način. Tako, za studenta koji je na klasifikacionom ispitu imao postignuće u četvrtom kvartilu (76–99), tj. iznadprosečno, najverovatnije je da će imati prosek preko 9 na kraju studija.

Tabela 22

Tabela kontingencije: verovatnoća ostvarivanja određene prosečne ocene na kraju studija na osnovu postignuća na klasifikacionom ispitu

Klasifikacioni ispit	Prosečna ocena na kraju studija				Suma
Percentilni kvartil	6–6,99	7–7,99	8–8,99	9–10	
1–25	50%	42%	8%		24
26–50	17%	83%			6
51–75	6%	11%	83%		35
76–99			17%	83%	12
Suma	15	19	33	10	76

Lako se može uočiti da je za izvođenje zaključka kako na osnovu tabele, tako i na osnovu grafikona očekivanih vrednosti, potrebno sprovesti longitudinalno istraživanje kako bi se ispitala kriterijumska validnost. Treba napomenuti da ovi procenti nisu "garancija", već su u pitanju samo verovatnoće. Naša odluka o ispitaniku bazirana je na slučajevima koji se nalaze u okviru određene ćelije tabele, te ukoliko je broj tih slučajeva mali, treba biti oprezan prilikom donošenja odluke. Svakako, od kvaliteta sprovedenog istraživanja zavisi i upotrebna vrednost informacija dobijenih preko tabela i grafika očekivanih vrednosti.

Interpretacija u odnosu na sadržaj ispitivanog područja

Posebna vrsta interpretacije odnosi se na ocenu kvaliteta pokazanog znanja, veštine ili učinka. U ovom slučaju sam skor, ako i postoji, nije važan, već se o zadovoljenju kriterijuma zaključuje na osnovu opšteg, holističkog utiska procenjivača, a u vezi sa predmetom merenja testa. Primer za ovu vrstu interpretacije može biti kada se, na primer, odluka o tome da li nekog studenta treba primiti na studije Akademije umetnosti donosi na osnovu njegovog portfolija.

Dijagnostikovanje na osnovu neke od klasifikacije mentalnih poremećaja (DSM-V ili ICD-10), takođe, predstavlja ovu vrstu interpretacije.

Iako deluje da je zaključivanje o ispunjenosti kriterijuma više subjektivna procena, pristrasnosti u određivanju ispunjenosti kriterijuma se smanjuju ukoliko je kriterijum jasno definisan i ukoliko postoji komisija koja procenjuje ispunjenost kriterijuma (bili to panelisti ili nezavisni procenjivači).

Vrste testova u odnosu na kriterijum

Iako je rečeno da su kriterijumska i normativna interpretacija vrste interpretacije, a ne vrsta testova, postoje testovi za koje je karakteristična primena kriterijumske interpretacije, te će oni biti detaljnije opisani. Naime, postoje tri osnovne vrste testova u odnosu na to šta je kriterijum. Kriterijum se najčešće odnosi na domen, sadržaj ili područje s jedne strane, i na učinak s druge strane, te u skladu s tim, testovi se dele na testove postignuća i na testove učinka. Pored toga, kriterijum se može odnositi na sadašnji ili na budući učinak ili postignuće, tj. predikovan učinak ili postignuće. Ako se odnosi na sadašnji, u pitaju je test postignuća ili test učinka, a ako se odnosi na budući učinak ili postignuće, u pitanju je test podobnosti.

Testovi postignuća

Testovi postignuća (eng. *achievement tests*) orijentisani su na sadržaj/domen/područje i njima se određuje posedovanje minimalnog potrebnog znanja iz nekog domena ili nivo znanja. Dakle, kod ovih testova akcenat je na usvojenoj količini znanja i onome što osoba aktuelno može i ostvaruje. Zbog toga se često za ove testove kaže da

"gledaju u prošlost i sadašnjost" (npr. u Fajgelj, 2013). Kronbah (Cronbach, 1990) u definiciji testova postignuća ističe da se njima meri postiguće koje je rezultat neke obuke ili edukacije. Prema Kronbahovom određenju, podučavanje u domenu onoga što je predmet merenja testa se eksplicitno podrazumeva. Po pravilu, testovi postignuća pripadaju testovima maksimalne aktivnosti, tj. prve paradigme merenja.

Skor na ovim testovima interpretira se u terminima predmeta merenja, odnosno sadržaja testa. Na primer, ako procenjujemo učenikovo znanje istorije XX veka, prilikom interpretacije njegovog skora ćemo zaključiti koliko učenik zna o istoriji XX veka, ali ne možemo zaključiti koliko zna o istoriji IX veka ili koliko će znati naredne lekcije koje će izučavati na časovima Istorije. Na ovaj način postoji direktna veza između predmeta merenja testa i kriterijuma – skup indikatora koji je obuhvaćen predmetom merenja testa je, u stvari, podskup mogućih indikatora koje obuhvata kriterijum.

Kao što je rečeno, validnost je ključna kod testova orijentisanih na kriterijum, a u ovom slučaju ključna je *sadržinska validnost*. Ukoliko je u fokusu procena postignuća I ukoliko želimo da procenimo celokupni nivo znanja učenika na kraju osnovnoškolskog školovanja, test za proveru znanja bi trebalo da ima podjednak broj pitanja iz svake oblasti znanja – Srpskog jezika i književnosti, Matematike, Fizike, Hemije, Biologije, Geografije, Istorije, pri čemu bi pitanja trebalo da sadrže najvažnije i ključne pojmove i fenomene, dakle one koje su reprezent znanja date oblasti. Ukoliko želimo da procenimo znanje samo u oblasti istorije, onda test treba da sadrži pitanja iz svake relevantne oblasti istorije. Dakle, testom mora biti obuhvaćen reprezentativni skup indikatora za domen koji je predmet merenja.

Planiranje testa postignuća podrazumeva raščlanjivanje predmeta merenja na relevantne oblasti/deskriptore, ali i specifikovanje ishoda učenja za svaku selektovanu oblast. Na primer, ukoliko je domen vokabular, jedna od oblasti za procenu postignuća može biti znanje reči. Ishodi učenja koji se postavljaju za ovu oblast mogu biti: 1. pravilno definisanje reči, 2. razumevanje značenja reči, 3. primena znanja reči u pisanom tekstu. Tako, za svaki ovaj ishod treba da se sastave posebni zadaci i pitanja. Na primer, za prvi ishod pitanje može da glasi: Šta znači reč skitnica? Za drugi ishod, pitnje može da glasi: Koja od navedenih reči je najbliža značenju reči "skitnica"?, a ponuđeni odgovori mogu biti: šetač, lutalica, beskućnik, boem, pas. Za treći ishod, zadatak bi bio da se sastavi rečenica u kojoj će se upotrebiti reč "skitnica" u smislenom kontekstu.

Najčešća primena testova postignuća je prilikom evaluacije programa, npr. obrazovnog plana i programa i postignuća u obrazovanju, uspešnosti treninga ili obuke i sl. Potom, ovi testovi se koriste i za potrebe klasifikacije, npr. dobijanja sertifikata, u cilju selekcije i sl. Postoji posebna vrsta testova koja se vezuje za testove postignuća, najčešće u obrazovanju, a to su **testovi ovladavanja** (eng. *mastery tests*). U širem smislu, testovi ovladavanja služe za evaluaciju procesa učenja u specifičnom polju ekspertize i njima se meri stepen vladanja određenim znanjem ili veštinama. U užem smislu, kako se navodi i u Standardima (AERA, APA, & NCME, 2006) u kojima se ovaj test izjednačava sa testom znanja, u nekoj tehničkoj primeni, ovo je test namenjen proceni dostignuća propisanog nivoa znanja u nekom području za pojedine ispitanike.

Kako se određuje kriterijum?

Postoji nekoliko načina za određivanje kriterijuma kod testova postignuća. Najčešći od njih podrazumeva da grupa eksperata odredi obim i primenu znanja, kao i kriterijum na osnovu kojeg bi se donela odluka o nivou usvojenog znanja. Dakle, grupa eksperata određuje sadržaj testa i kritični nivo ili nivoe znanja. Među eskpertima treba da postoji saglasnost u vezi s navedenim, kako bi se osigurala objektivnost testa. Da bismo bili sigurni da će svi ispitanici na isti način shvatiti uputstvo, pitanja i zadatke, i da će ocenjivanje testa biti objektivno itd., testovi moraju biti strukturisani.

Prikaz rezultata

Najčešći prikaz rezultata je sirov/bruto skor koji može biti iskazan u procentima kao broj tačnih odgovora. U odnosu na postavljeni kriterijum, skor se tumači kao zadovoljavajući ili nezadovoljavajući ili prošao–pao i sl. Takođe, rezultat se može prikazati i kao bilo koji izvedeni skor, npr. percentilni rang, i tumačiti u skladu sa postavljenim kriterijumom (npr. svi kandidati koji imaju postignuće na 55. percentilu ili više, zadovoljili su kriterijum).

Testovi učinka

Testovi učinka (engl. *performance tests*) koriste se za složenije, zahtevnije i realističnije predikcije i njima se procenjuju veštine i kompetencije ispitanika. Veštine i kompetencije se procenjuju preko kvaliteta manifestacije uspešnosti obavljanja nekog posla ili kvaliteta napravljenog proizvoda ili produkta. Dakle, o učinku se zaključuje na osnovu nekog ispoljenog ponašanja koje je rezultat usvojene veštine, u uslovima koji simuliraju realno okruženje. Isto kao u slučaju testova

postignuća, i za ove testove je važna *sadržinska validnost*. Ukoliko, na primer, želimo da procenimo da li je neko usvojio veštinu vajanja, daćemo mu zadatak da izvaja određenu skulpturu i na osnovu uvida u proces vajanja i sličnosti orignalne sa izvajanom skulpturom možemo proceniti da li kandidat ima potrebne veštine vajanja. Pritom, važno je da zadatak bude reprezentativan, tj. skulptura koja se postavlja kao kriterijum treba da bude takva da se na osnovu posmatranja načina njenog vajanja i samog finalnog produkta može učiti da li je veština vajanja pravilno usvojena. Drugim rečima, neko ko poznaje neke osnovne veštine vajanja ne bi trebalo da uspešnije reprodukuje skulpturu u odnosu na nekoga ko poznaje i napredne veštine vajanja. Drugi primer može biti iz domena veština poznavanja rada u određenim statističkim softverima, npr. u SPSS-u. Da bismo procenili koliko neko poznaje rad u SPSS-u moramo mu dati raznovrsne zadatke – da izračuna osnovne deskriptivne podatke za neku varijablu, ali i da zna kako da primeni određene parametrijske i neparametrijske testove. Ukoliko bi zadatak sadržao samo desktiptivni opis varijabli, ne bismo bili sigurni u to da li kandidat ima potrebne veštine ili ne, jer u stvari, naš test nije sadržinski validan.

Postoji posebna vrsta testova koja je karakteristična za grupu testova učinka, a to su **testovi obučivosti**. U okviru njih se demonstrira način obavljanja nekog zadatka, nakon čega kandidati imaju vežbu, a potom se procenjuje koliko su uspešno izveli dati zadatak ili se procenjuje razlika u kvalitetu obavljanja zadatka između prvog i drugog pokušaja (Jackson, 2003).

Najčešća primena testova učinka je u cilju klasifikacije, npr. u profesionalnoj selekciji, ili u cilju raspoređivanja, ali primenjuju se u

cilju samospoznaje i ličnog razvoja, npr. prilikom ispitivanja darovitosti (da li je neko za muzičku školu), savetovanja i sl.

Razlike testova postignuća i testova učinka

Iako su i testovi postignuća i testovi učinka orijentisani na sadržaj/domen/područje, ipak ne mogu zameniti jedan drugog. Testom učinka se direktno može izmeriti primena usvojenog znanja, ali ti testovi se uglavnom svode na nekoliko reprezentativno izdvojenih problemskih zadataka. S druge strane, testovima postignuća može se obuhvatiti širi spektar oblasti područja tj. širi spektar deskriptora, te u tom smislu oba testa mogu biti informativna. Pored toga, u nekim situacijama je dovoljno zadati samo test učnika koji se sastoji od jednog zadatka (npr. prilikom procene veštine vajanja), a u drugim je ekonomičnije i adekvatnije zadati samo test postignuća (npr. prilikom procene uspešnosti savladavanja nekog obrazovnog programa).

Kako se određuje kriterijum?

Utvrđivanje ispunjenosti kriterijuma u praksi je češće subjektivna, nego objektivna procena. Kao i u slučaju testova postignuća, kod testova učinka grupa eksperata određuje sadržaj testa i kriterijum na osnovu kog će se doneti odluka u vezi s tim da li je demonstrirani učinak uspešan ili ne. Najčešći primer koji se navodi je ocenjivanje performansa skijaških skakača ili klizača na ledu, koji ocenjuje žiri koji se sastoji od nekoliko eksperata. Međutim, koliko god kriterijumi bili definisani, u procenu je uvek uključena subjektivna procena svakog člana žirija i opšti utisak o celokupnom izvođenju neke aktivnosti. Izuzetak od subjektivne procene su testovi u kojima se procenjuje brzina rada, broj grešaka, broj tačno reprodukovanih proizvoda u jedinici vremena i dr., što se može objektivno oceniti.

Prikaz rezultata

Uobičajeni metod ocenjivanja učinka je preko rejting skale ili preko rubrika skorovanja u kojima se definišu i ilustruju pravila za ocenjivanje kvaliteta učinka.

Testovi podobnosti

Testovi podobnosti su testovi orijentisani na kriterijum u tradicionalnom psihometrijskom smislu. Testovi podobnosti (engl. *aptitude test*) mere specifične sposobnosti, veštine ili talente i služe predikciji budućeg učinka. Njima se određuje da li ispitanik poseduje određeni nivo kompetencije, kako bi u budućnosti mogao uspešno da obavlja neku delatnost. Neki autori navode da se testovima podobnosti meri sposobnost učenja, tj. spremnost za učenje nove veštine, usvajanje znanja i slično (npr. Santrock, 2008). Predmet testova podobnosti su posebne akademske, profesionalne ili radne sposobnosti, veštine ili talenti, tako da i ovi testovi češće pripadaju testovima maksimalne aktivnosti. Ipak, iako se ređe koriste u praksi, i testovi tipične aktivnosti mogu biti testovi podobnosti, npr. procena afektivne vezanosti i empatije za ispitivanje podobnosti za pomagačke profesije.

Podobnost predstavlja potencijal ili kapacitet za uspešno obavljanje neke delatnosti, te se stoga za testove podobnosti često kaže da "gledaju u budućnost" (npr. u Fajgelj, 2013). Za razliku od testova postignuća i učinka koji mere šta osoba aktuelno može, testovima podobnosti se meri šta će osoba moći. Na primer, ovim testovima se procenjuje koliko bi neko bio uspešan u poslu, na studijama i slično. Druga razlika između testova podobnosti i testova postignuća ili učinka, takođe, povezana je sa predmetom merenja. Naime, Koen i Sverdlik (Cohen & Swerdlik 2009) naglašavaju da je u testovima podobnosti

fokus na životnom iskustvu i onome što je naučeno preko neformalnih izvora, dok je u testovima postignuća fokus na proceni naučenog preko relativno strukturisanog i sistematskog izvora, kao što je formalno obrazovanje. Treća razlika odnosi se na to što testovi postignuća najčešće mere jednu stvar (npr. znanje iz istorije), dok testovi podobnosti obuhvataju širi opseg predmeta merenja (npr. podobnost za fakultet, videti Macklem, 1990). Na osnovu navedenog, moglo bi se zaključiti da je veza između predmeta merenja i sadržaja testa *manje direktna* kod testova podobnosti.

Takođe, treba razlikovati testove podobnosti od testova inteligencije i sposobnosti. U literaturi se često za inteligenciju i podobnost kaže da se mogu opisati kao "dvojajčani blizanci" (npr. Laak, Gokhale, & Desai, 2013). Kao takvi, trebalo bi da se i tretiraju drugačije. U odnosu na testove inteligencije koji mere opšti ili *g* faktor, testovi podobnosti (ukoliko su namenjeni proceni sposobnosti), mere specifične sposobnosti, kao što su npr. verbalne, numeričke, manuelne sposobnosti, apstraktno rezonovanje i sl.

U praksi, posebno kada je u pitanju profesionalna selekcija, često se donosi odluka o tome da li primeniti test sposobnosti ili ličnosti ili neki treći test, ili njihovu kombinaciju. Do sada najopsežnije meta-analitičko istraživanje ukazuje na to da je *g* faktor najbolji prediktor uspešnosti u poslu i treninga za posao, a da se poboljšanje predikcije dobija dodavajem rezultata na merama ličnosti (prvenstveno merama integriteta i savesnosti), merama uzorka posla ili podataka sa strukturisanog intervjua (Schmidt & Hunter, 1998).

Praktičari se, međutim, suočavaju sa još jednom odlukom u vezi sa kojom nema tako jasnih nalaza. Ta odluka odnosi se na to da li je bolje primeniti test opšteg, tj. *g* faktora ili nekih specifičnih sposobnosti.

Prethodna istraživanja ukazuju na to da specifične sposobnosti ne doprinose predikciji uspešnosti na poslu (npr., Ree, Earles, & Teachout, 1994; Brown, Le, & Schmidt, 2006). Međutim, jedno meta-analitičko istraživanje pokazuje visoku povezanost između opšte mentalne sposobnosti, ali i specifičnih numeričkih i verbalnih sposobnosti sa uspešnošću na treningu za posao i sticanju znanja, dok je manja povezanost dobijena u slučaju specifičnih memorijskih testova, iako je ta povezanost i dalje visoka (rang korigovanih korelacija je od ,50 do ,70; više u Ones, Viswesvaran, & Dilchert, 2004). Isti obrazac povezanosti dobijen je i u slučaju predikcije ukupnog učinka na poslu, s tim što je intenzitet korigovanih korelacija nešto niži i u rangu od ,35 do ,55. Dakle, rezultati ovog istraživanja sugerišu da i test opšte sposobnosti i testovi određenih specifičnih sposobnosti dobro rade posao predikcije. Reanaliza meta-analitičkih studija pokazuje da se, u zavisnosti od primenjene analize, različiti značaj pridaje testovima specifične sposobnosti u proceni uspešnosti na poslu (Kell & Lang, 2017). Tako, ukoliko se primeni hijerarhijska regresiona analiza u kojoj se ispituje doprinos specifičnih sposobnosti povrh opšte sposobnosti, rezultati ukazuju na mali, praktično zanemarljivi doprinos specifičnih sposobnosti. Međutim, ukoliko se primeni analiza relativnog značaja, u kojoj nema unapred postavljenih pretpostavki o važnosti neke mere, rezultati pokazuju da specifične sposobnosti bolje predikuju uspešnost na poslu. Koje će se specifične sposobnosti istaći kao bolji prediktori zavisi od vrste posla i operacionalizacije kriterijuma, tj. uspešnosti na poslu. Na osnovu ovih rezultata može se zaključiti da testovi specifičnih sposobnosti imaju prednost nad testovima opšte sposobnosti, ako su specifične sposobnosti pažljivo odabrane i u uskoj vezi s onim što će biti opis posla kandidata. Pored toga, raznovrsnost predmeta merenja

primenjenih testova specifičnih sposobnosti, pod uslovom da su pažljivo odabrani, može nam pružati bolji uvid u prednosti i slabosti kandidata.

Na ovom mestu je važno uočiti razliku između onoga što se može navežbati i onoga što se, bez obzira na kvalitet treninga i obuke, ne može navežbati do uspešnog i kvalitetnog izvođenja. U tom pogledu, za podobnosti je usvojeno stanovište da su one više nasledne i urođene i da se stvaranjem adekvatnih uslova mogu manifestovati u vidu uspešnog obavljanja delatnosti (Sternberg, 2001). Drugim rečima, prema usvojenom stanovištu, podrazumevaju se individualne razlike u podobnostima i, iako neka obuka može smanjiti razliku između onih koji su podobni za neku delatnost i onih koji to nisu, ta razlika će uvek postojati. Najbolji teorijski okvir za razumevanje podobnosti pruža Katelova i Hornova teorija inteligencije (npr. u Cattell, 1963; Gardner, 1999). Prema ovoj teoriji razlikuju se fluidna i kristalizovana inteligencija. Fluidna inteligencija definiše se kao opšta sposobnosti uočavanja odnosa između elemenata. Ona je većim delom nasledna, tj. manje je osetljiva i promenljiva u odnosu na obuku i trening. S druge strane, kristalizovana inteligencija je stečena obrazovanjem i drugim socijalnim i kulturološkim uticajima. Odnosi se na rešavanje problema potpomognuto znanjem i iskustvom. Pretpostavka je da podobnosti predstavljaju domen fluidne inteligencije, a npr. ono što se meri testovima postignuća, kao što je znanje, predstavlja domen kristalizovane inteligencije.

S obzirom na to da se testovima podobnosti predviđa budući učinak, za njih je posebno važna *kriterijumska validnost*. Testovi podobnosti se zato nekada nazivaju i prognostičkim testovima. Skor na testovima podobnosti ukazuje na to koji nivo kriterijuma se može očekivati od ispitanika u budućnosti. Kao što je rečeno, za razliku od

testova postignuća i učinka, u testovima podobnosti test i kriterijum ne moraju biti istog sadržaja, odnosno veza između predmeta merenja i testa nije nužno direktna. Na primer, ukoliko želimo da procenimo da li će neko biti uspešan u poslu menadžera, to možemo uraditi na osnovu procene kandidatovih kognitivnih sposobnosti, veština u donošenju odluka, komunikacijskih veština i rada u timu, karakterističnog profila ličnosti i sl. Dakle, sadržaj datih testova ne mora nužno obuhvatati karakteristike i situacije menadžerskog posla, ali na osnovu rezultata na tim testovima treba se proceniti da li bi kandidat bio uspešan u ovom poslu. U pitanju je procena budućeg ponašanja tj. učinka na osnovu drugih, ne nužno istih zadataka kao što su zadaci koji se postavljaju na osnovu zahteva posla. Moguće je kandidatu dati niz hipotetičkih situacija koje su karakteristične za menadžerski posao, i proceniti kako bi se kandidat snašao u njima. U ovom slučaju imamo simulaciju realnog okruženja menadžerskog posla koja je direktnije povezana sa kriterijum, ali i dalje ne sasvim direktno, jer su u pitanju hipotetičke situacije. U svim navedenim situacijama se procenjuje mogućnost ili potencijal kandidata da se snađe u menadžerskoj ulozi i da je obavlja na traženom nivou. Teorijski posmatrano, ako kandidatu obezbedimo probni rad u kojem može demonstrirati menadžerske veštine, procena tih veština više ne bi bila vezana za test podobnosti, nego za test učinka. O problemima razlikovanja ovih testova biće kasnije više reči.

Važno je naglasiti da rezultat na testu podobnosti nije nužno vezan za prethodnu obuku ili ovladavanje, kao što je to slučaj kod testova postignuća. Naprotiv, testovi podobnosti se mogu koristiti za procenu da li je potrebno nekoga poslati na dalje usavršavanje i obuku. Na primer, ukoliko se proceni da je neki učenik talentovan za matematiku, može se preporučiti kao polaznik seminara matematike u

Istraživačkoj stanici Petnica ili se može poslati u letnju školu matematike da proširi svoje znanje. Dakle, testovima podobnosti se vrši procena mogućnosti nekoga da nauči i usvoji neke nove sadržaje ili razvije određenu veštinu u određenom području obrazovanja ili obuke. Praktično gledano, testovi podobnosti doprinose proceni da li vredi ulagati u nekoga, i u okviru koje oblasti, odnosno za šta.

Najčešća primena testova podobnosti je slična primeni testova učinka, u cilju klasifikacije (npr. selekcija, raspoređivanje), kao i samospoznaje i ličnog razvoja (npr. talenta i darovitosti), savetovanja i sl.

Problemi razlikovanja testova postignuća i podobnosti

Nekada je teško napraviti razliku između testova postignuća i podobnosti, te je ta razlika više stvar stepena. Naime, aktuelno postignuće je dobar prediktor budućeg postignuća koje podrazumeva neke napredne veštine i znanja (Macklem, 1990). Na primer, poznavanje rada u nekom programskom jeziku može biti mera podobnosti za usavršavanje programerskih veština i usvajanje znanja iz drugog programskog jezika. Na osnovu navedenog može se primetiti da jedan isti test u jednoj situaciji može biti test postignuća, a u drugoj test podobnosti, čime se granica između ovih testova gubi. U ovom slučaju, ono što pravi razliku između testova je **namena testiranja**. Ukoliko je namena, odonosno cilj testiranja procena znanja, u pitanju je test postignuća, a ukoliko je cilj procena potencijala za usvajanje novih sadržaja iz date oblasti, u pitanju je test podobnosti. Na osnovu toga može se zaključiti da podela na testove postignuća i podobnosti, barem u jednom delu, zavisi od toga da li je onaj ko rešava test prethodno pohađao neku formalnu obuku ili obrazovanje u vezi s onim što je

predmet merenja testa ili ne (Cohen & Swerdlik, 2009). Ukoliko jeste i ukoliko se testom meri stepen savladanog znanja i veština tokom obuke i obrazovanja, u pitanju je test postignuća. Santrok (Santrock, 2008) navodi primer testa na kvalifikacionom ispitu. Ukoliko na osnovu učenikovog rezultata na kvalifikacionom ispitu za upis na fakultet predviđamo njegov uspeh u godinama školovanja koje slede, onda je u pitanju test podobnosti. Ukoliko na osnovu pomenutih rezultata želimo da dobijemo informaciju o tome koliko učenik poznaje određeno gradivo, onda je u pitanju test postignuća.

Razlikovanje postignuća i podobnosti slikovito opisuje Kerol (Carroll, 1993), na osnovu zamišljenog longitudinalnog istraživanja u kojem su ispitanici podvrgnuti nekoj obuci. Njihovo postignuće i podobnost procenjuju se pre i posle obuke. Ono što se očekuje je da se skor na testu podobnosti ne menja nakon obuke, ali da se skor na testu postignuća menja nakon obuke. Ovakav rezultat išao bi u prilog tome da podobnosti ne zavise od treninga i obuke, već da su u pitanju više nasledne sposobnosti. Očekuje se, takođe, da nema značajne povezanosti između skora na testu postignuća i podobnosti pre obuke, kada još nije došlo do učenja. Značajna povezanost između skora na testu podobnosti u prvom merenju i skora na testu postignuća očekuje se nakon obuke. Drugim rečima, podobnost doprinosi boljem postignuću nakon obuke. Može se primetiti da su efekti obuke i podobnosti na postignuće međusobno nezavisni, a podobnost bi trebalo da doprinese boljem postignuću povrh obuke, tj. da ima inkrementalnu validnost u odnosu na obuku.

Testovi postignuća i podobnosti se **različito interpretiraju** s obzirom na to da imaju različite namene. Uobičajeno, testovi postignuća imaju uži predmet merenja koji je više povezan s formalnim

obrazovanjem, tako da se interpretacija skora na njima zadržava na domenu onoga što je testirano (npr. na osnovu skora na testu poznavanja rada u jednom programskom jeziku procenjuje se koliko osoba zna rad u tom, konkretnom programskom jeziku). S druge strane, testovima podobnosti predviđa se širi skup varijabli, tako da je i interpretacija skora na njima šira (npr. na osnovu skora na testu poznavanja rada u jednom programskom jeziku procenjuje se da li osoba može brzo da savlada rad na novim softverima koji zahtevaju poznavanje tog ili drugog programskog jezika, da li može da pruži unapređenje postojećeg softvera i sl.). Može se zaključiti da su testovi postignuća specifičniji i povezaniji s nastavnim programima i srodnim područjima u odnosu na testove podobnosti kojima se obično mere opšte sposobnosti. Ovo ima važne implikacije za zaključke izvedene na osnovu njih ukoliko se obe vrste testova koriste za predikciju; testovi postignuća mogu biti korisni prediktor budućeg učinka i postignuća u uže određenom području, ali ne i izvan tog područja, dok testovi podobnosti mogu koristiti kao opštiji prediktori budućeg učinka i postignuća (Cohen, Manion, & Morrison, 2007).

Kako se određuje kriterijum?

Kriterijum se utvrđuje empirijskim putem, najčešće preko regresione analize, a može se prikazati slikovito preko tabela i grafikona očekivanih vrednosti. U svakom slučaju, to znači da, pre nego što odlučimo da primenimo neki test podobnosti, treba da imamo rezultate koji dokazuju povezanost skora na testu i kriterijuma. U slučaju standardizovanih testova podobnosti, utvrđivanje kriterijumske validnosti je sastavni deo standardizacije testa. Ukoliko je reč o testu koji nije standardizovan, već se koristi interno, i u toj situaciji mora

postojati neki rezultat o povezanosti skora na testu sa kriterijumom koji ukazuje na to da taj test podobnosti dobro radi posao. U ovoj situaciji pretpostavlja se da postoji interna baza podataka na osnovu koje se zaključuje o kriterijumskoj validnosti testa, ali i opasnost da je ona dobijena na ograničenom opsegu varijable (npr. samo na kandidatima koji su prošli selekciju).

Prikaz rezultata

Rezultati testova podobnosti obično se prikazuju u terminima ispunjenosti kriterijuma preko graničnog skora. Iako se ovi testovi češće koriste u svrhe selekcije, te se češće interpretiraju u skladu sa ispunjenošću kriterijuma, oni se takođe mogu interpretirati i normativno, kako i bilo koji drugi test sposobnosti ili ličnosti.

Pitanja za razmišljanje

1. Koja je psihometrijska karakteristika najvažnija za kriterijumski orijentisane testove?
2. U čemu je razlika između internih i eksternih metoda za određivanje graničnog skora?
3. Ko su panelisti? Kako se panelisti razlikuju od eksperata procenjivača?
4. Čemu služe tabele i grafikoni očekivanih vrednosti? Zašto postoji potreba da se informacije dobijene preko statističkih pokazatelja prikažu preko tabela ili grafikona očekivanih vrednosti?
5. Koja je razlika između testova postignuća, učinka i podobnosti? Navedi primer iz prakse.

Reference

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2006). *Standardi za pedagoško i psihološko testiranje*. Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Bingham, W. V. (1937). *Aptitudes and aptitude testing*. Oxford, England: Harpers.

Brase, G. L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making, 15*(5), 381–401.

Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment, 14*, 87–100.

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. (2003). *Whose grades are inflated? College Board Research Report No. 2003–4.* New York, NY: College Board.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Cascio, W. F. (1977). Formal education and police officer performance. *Journal of Police Science & Administration, 5*(1), 89–96.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22.

Cizek, G. J. (2001). *Setting performance standards: Concepts, methods, and perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007). *Metode istraživanja u obrazovanju* (5. izdanje). Jasterbarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to test and measurement* (4th ed.). Mountain View, CA: Mayfield.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York, NY: Harper Collins Publisher.

Dahle, K.-P., Schneider, V., & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose [Actuarial instruments for the prediction of criminal reoffences]. *Forensische Psychiatrie*, *Psychologie*, *Kriminologie*, *1*, 15–26.

Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (2001). *HCR-20 violence risk management companion guide.* Burnaby, Canada: Simon Fraser University, Mental Health, Law, and Policy Institute and University of South Florida, Louis de la Parte Florida Mental Health Institute.

Fajgelj, S. (2013). *Psihometrija. Metod i teorija psihološkog merenja* (IV dopunjeno izdanje). Beograd, Srbija: Centar za primenjenu psihologiju.

Freymuth, A. K., & Ronan, G. F. (2004). Modeling patient decision-making: The role of base-rate and anecdotal information. *Journal of Clinical Psychology in Medical Settings, 11,* 211–216.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.* New York, NY: Basic Books.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*(8), 519–521.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene*, *15*(1), 15–17.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*(5500), 2261–2262.

Jackson, C. (2003). *Psihologijsko testiranje* (2. izdanje). Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454.

Kell, H. J., & Lang, J. W. B. (2017). Specific abilities in the workplace: More important than *g*? *Journal of Intelligence, 5*(2), 13. https://doi.org/10.3390/jintelligence5020013

Kodžopeljić, J. i Pekić, J. (2017). *Psihologija u nastavi: odabrane teme iz psihologije obrazovanja.* Novi Sad, Srbija: Filozofski fakultet.

Laak, J. J. F., Gokhale, M., & Desai, D. (2013). *Understanding psychological assessment: A primer on the global assessment of the client's behavior in educational and organizational setting.* New Delhi, India: SAGE Publications India Pvt Ltd.

Lawshe, C. H., & Bolda, R. A. (1958). Expectancy charts: I. their use and empirical development. *Personnel Psychology, 11*(3), 353–365.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes, 82*, 217–236.

Macklem, G. L., (1990). *Measuring aptitude. ERIC digest*. American Institutes for Research, Washington, DC. Preuzeto sa https://eric.ed.gov/?id=ED328608

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52* (3), 194–216.

Morgan, D. L., & Michaelides, M. P. (2005). *Setting cut scores for college placement (College Board Research Report No. 2005–9)*. New York, NY: The College Board.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2004). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431–468). Thousand Oaks, CA: SAGE Publications.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79* (4), 518–524.

Santrock, J. W. (2008). *Educational psychology* (3rd ed.). New York, NY: McGraw Hill.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin, 124*(2), 262–274.

Sireci, S. G., Robin, F., & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education, 12*(3), 301–325.

Sternberg, R. J. (2001). Why schools should teach for wisdom: The balance theory of wisdom in educational settings. *Educational Psychologist, 36*(4), 227–245.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons Inc.

World Helth Organization (2018). *WHO methods and data sources for life tables 1990–2016*. Global Health Estimates Technical Paper WHO/HIS/IER/GHE/2018.2.

Zieky, M. J., & Livingston, S. A. (1977). *Basic skills assessment manual for setting standards*. Princeton, NJ: Educational Testing Service.

U okviru interpretacije orijentisane na osobu od interesa je samo skor konkretnog ispitanika na različitim testovima (tzv. interpretacija orijentisana na test) ili na istom testu u različitim okolnostima (tzv. interpretacija orijentisana na okolnost). Kod ove vrste interpretacije ispitanik je sam sebi referetni okvir za poređenje, i ne zanima nas ni kakvo je postignuće ostalih ispitanika koji rade isti test, niti da li je ispitanik zadovoljio standard. Drugim rečima, u okviru ove interpretacije vršimo idiografsku procenu koja je usmerena na jedinstvenost i neponovljivost u iskustvu i karakteristikama pojedinca. Za razliku od idiografske procene, nomotetska procena se odnosi na utvrđivanje zajedničkih karakteristika pripadnika jedne populacije, i ona je karakteristična za interpretaciju orijentisanu na norme.

Ipsativni skorovi

Interpretacija orijentisana na osobu najčešće se povezuje sa tzv. **ipsativnim skorovima**. Ipsativni skorovi govore o poziciji jedne crte u odnosu na ostale crte kod jednog ispitanika, odnosno njima se rangiraju atributi jednog ispitanika u odnosu na izraženost. Da bi se to ostvarilo, neophodno je da postoji procena za jednog ispitanika na više atributa. Ipsativni skorovi su međusobno zavisni, te skor na jednoj crti zavisi od skorova na ostalim crtama kod istog ispitanika. Oni ukazuju na *relativnu* vrednost jednog atributa u odnosu na druge atribute, te je poređenje atributa *intra*-individualno (Brown, 2010). Na taj način, odgovori ispitanika se distribuiraju oko aritmetičke sredine tog ispitanika. Ipsativni skorovi ne zavise od skorova ostalih ispitanika i nisu komparabilni

između ispitanika. Drugim rečima, dovoljan nam je protokol jednog ispitanika za određivanje ipsativnih skorova i ništa drugo. Za razliku od ipsativnih skorova, kod **normativnih skorova** se skor jednog ispitanika poredi sa skorovima drugih ispitanika iz referentne grupe. Na taj način dobijamo informaciju o *apsolutnoj* izraženosti atributa u kontekstu populacije, te je poređenje atributa *inter*-individualno. Za interpretaciju normativnih skorova nam je potrebno da znamo i skorove drugih ispitanika iz referentne grupe i distribuciju tih skorova. U ovom slučaju se skorovi distribuiraju oko aritmetičke sredine referentne grupe. Skale za koje se računa normativni skor obično sadrže samo jedan atribut, koji ispitanik procenjuje da li je izražen kod njega ili ne, ili u kom je stepenu izražen.

Uzmimo, na primer, Raju čiji su rezultati na tri testa znanja prikazani u Tabeli 23. Na osnovu interpretacije orijentisane na osobu bismo poredili Rajine skorove na tri testa, i pod uslovom da su skorovi prethodno svedeni na istu skalu izražavanja skorova, mogli bismo zaključiti da Raji najbolje ide strani jezik, pa srpski jezik, a da mu najgore ide matematika. S druge strane, ukoliko vršimo normativnu interpretaciju, poredili bismo Rajin rezultat na jednom testu s rezultatima ostalih ispitanika na tom istom testu, te bismo tako zaključili da mu strani jezik ide bolje nego ostalim ispitanicima, srpski jezik umereno dobro, dok mu matematika ide kao i Gaji, pri čemu obojica imaju niže postignuće od Vlaje. Iako se normativnom interpretacijom najčešće vrši inter-individualna procena, valja napomenuti da je njome moguće vršiti i intra-individualnu procenu. Na primer, ukoliko imamo norme za svaki test, možemo zaključiti da li nekom učeniku bolje ide srpski jezik ili matematika, npr. na kom testu postiže prosečni, a na kom iznadprosečni skor.

Tabela 23

Primer interpretacije orijentisane na osobu i normativne interpretacije

Ispitanik	Srpski jezik	Matematika	Strani jezik
Raja	10	5	15
Gaja	5	5	10
Vlaja	20	10	5

Napomena: Crvena linija predstavlja smer interpretacije orijentisane na osobu, a plava smer normativne interpretacije.

Vrste ipsativnih skorova

Upotreba ipsativnih skorova datira još od 1928. godine, a prvi ih je predložio Marston (videti u Martinussen, Richardsen, & Vårum 2001). Diferencijaciju različitih ipsativnih mera prvi je predložio Katel 1944, ali njegova kategorizacija nije više u upotrebi zbog terminoloških nejasnoća. Do danas, najkoherentnija podela je ona koju je predložio Hiks (Hicks, 1970), a prema kojoj se mogu razlikovati čisto ipsativni skorovi, kvazi-ipsativni ili parcijalno ipsativni skorovi i semi-parcijalno ipsativni skorovi, pri čemu ćemo mi ovde napraviti razliku samo između čisto ipsativnih i kvazi-ipsativnih skorova budući da se oni najčešće i koriste.

Čisto ipsativni skorovi za jednog ispitanika imaju svojstvo da im je suma skorova uvek konstanta (Clemans, 1966). Ipsativni skorovi se najčešće mogu dobiti preko ajtema prisilnog izbora, iako se ovi ajtemi mogu skorovati i na drugačiji način. Na primer, ako imamo sledeći par ajtema:

A. Uživam da radim s drugim ljudima.

B. Pažljiv sam s detaljima.

pri čemu je zadatak ispitanika da se opredeli za onu alternativu koja ga bolje opisuje. Izbor jedne opcije znači automatsku eliminaciju druge. Kako opcija A meri ekstraverziju, a opcija B savesnost, neko ko odabere opciju A imaće viši skor na ekstraverziji, ali istovremeno i manji na savesnosti.

Na ovom mestu samo treba skrenuti pažnju na to da postoji i format koji nije čisto prisilni izbor, iako veoma podseća na njega:

A. Volim divlje žurke bez ograničenja.
B. Više mi se dopadaju tihe zabave.

U okviru ovog pitanja, opcije A i B su suprotni polovi istog konstrukta – traženja uzbuđenja. U ovom primeru imamo, u stvari, bipolarnu skalu samo prezentovanu na drugačiji način u odnosu na primer, klasičnu binarnu skalu sa odovorima "da" i "ne" ili skalu Likertovog tipa. Ovakvim pitanjima se dobijaju, zapravo, normativni skorovi. Za razliku od njih, kod čisto ipsativnih skorova alternative koje se nude kao odgovori moraju pripadati *različitim* skalama, odnosno konstruktima.

Format prisilnog izbora ne mora biti ograničen samo na dve alternative/parove odgovora. U praksi se mogu naći trijade, tetrade i sl., u okviru kojih je ukupni skor na ajtemu uvek isti, ali ne nose svi odgovori isti broj poena. Na primer, u trijadnom pitanju imamo tri ponuđena odgovora, a zadatak ispitanika je da rangira odgovore po tome koliko ga najbolje opisuju, pri čemu rang 1 treba dati ajtemu koji ga najbolje opisuje, a rang 3 ajtemu koji ga najlošije opisuje:

	Ispitanikov rang	Skorovanje
A. marljiv	1	2
B. smiren	3	0
C. timski igrač	2	1

U ovom slučaju bismo prvom izboru ("marljiv") dodelili maksimalnih 2 poena (tj. 2 poena ide skali koja sadrži ajtem "marljiv"), drugom izboru bismo dodelili 1 poen, i trećem izboru – 0 poena. Svi ispitanici bi na istom ovom ajtemu dobili 3 poena ukupno, ali bi se skorovi razlikovali po njihovim izborima, tj. odgovorima. Važno je naglasiti da kod ovakvih trijada, tetrada i sl., svaki odgovor mora da se rangira kako bi se dobili čisto ipsativni skorovi. Naime, čisto ipsativni skorovi se dobijaju kada ispitanici rangiraju sve ponuđene odgovore, kada se svi odgovori/rangovi skoruju i kada se ponuđeni odgovori, koji predstavljaju sve procenjene varijable, međusobno porede (Hicks, 1970).

Primer čisto ipsativnih skorova može biti i uz drugačiju instrukciju, npr. ukoliko ispitanik treba da označi odgovor koji najviše liči na njega i koji najmanje liči na njega, a da jedan ostavi prazan, da ga preskoči. U ovom slučaju, skorovanje bi bilo sledeće:

	Najviše liči na mene	Najmanje liči na mene	Skorovanje
A. marljiv	√		2
B. smiren		√	0
C. timski igrač			1

Ukupni skor koji bi se dobio na ovom pitanju je 3 i takav bi bio za svako pitanje i za sve ispitanike, bez obzira na to kakav se izbor odgovora napravi.

Kvazi-ipsativni skorovi uključuju mere koje ne zadovoljavaju sve kriterijume da budu čisto ipsativni skorovi, odnosno koje uključuju neku od sledećih karakteristika (Hicks, 1970):

1. suma atributa može da varira između ispitanika u određenom rangu skorova;
2. skorovi ne tvore istu konstantu za sve ispitanike;
3. povišenje skora na jednom atributu ne uzorkuje sniženje skora na drugom atributu.

Razlozi odstupanja od kriterijuma za dobijanje čisto ipsativnih mera mogu biti sledeći (Hicks, 1970):

1. ispitanici vrše samo parcijalno rangiranje (npr. označavaju samo ajtem koji najmanje liči na njih i/ili najviše liči na njih) umesto rangiranje svih odgovora;
2. skale imaju različiti broj ajtema;
3. ne skoruju se svi odgovori ispitanika (npr. odovori koji nisu od interesa za merenje nekog svojstva);
4. skale se različito skoruju za ispitanike s različitim karakteristikama, ili uključuju različite transformacije u zavisnosti od karakteristika ispitanika;
5. odgovori se različito ponderišu (npr. u jednom pitanju s dva ponuđena odgovora, jedan odabrani odgovor može da nosi 1 bod, a drugi 2 boda, a u drugom pitanju oba da nose po jedan bod);
6. instrument sadrži i normativni deo.

Kvazi-ipsativni skorovi imaju psihometrijske karakteristike slične i čisto ipsativnim i normativnim skorovima. Naime, na osnovu njih se mogu porediti ispitanici ili grupe (što je odlika normativnih skorova), ali su skale ipak na neki način međusobno zavisne (što je odlika ipsativnih skorova).

Ipsativni skorovi se mogu dobiti ne samo preko formata ajtema, već i preko transformacije, tzv. **ipsatizacije**. Ova transformacija podrazumeva da se od opaženih skorova jednog ispitanika (AS_O) oduzme aritmetička sredina na svim skorovima tog ispitanika (AS_T) i podeli standardnom devijacijom na svim skorovima tog ispitanika (SD_T):

$$\text{ipsativni skor} = (AS_O - AS_T) \,/\, SD_T.$$

Izraz $AS_O - AS_T$ predstavlja, zapravo, centriranje skorova na aritmetičku sredinu i ovim centriranim skorovima se iskazuje odstupanje od artimetičke sredine svih rezultata jednog ispitanika. Praktično, to znači da se od pojedinačnog skora na jednom ajtemu ili skali oduzme aritmetička sredina dobijena na svim ajtemima ili skalama, za svakog ispitanika posebno. Centrirani skorovi imaju aritmetičku sredinu jednaku 0, a standardnu devijaciju jednaku standardnoj devijaciji izvornih opaženih skorova. Dakle, centriranje je postupak svođenja skorova na odnos koji imaju spram aritmetičke sredine, te će skorovi koji su niži od aritmetičke sredine biti negativni, a skorovi viši od aritmetičke sredine će biti pozitivni. Ipsatizacija onda predstavlja centrirane skorove koji su podeljeni standardnom devijacijom svih skorova jednog ispitanika.

Uzimo Rajine skorove iz Tabele 23. Aritmetička sredina koju on postiže na tri testa iznosi (10 + 5 + 15) / 3 = 10, te bi centriranjem

skorovi bili redom 0, -5 i 5 (primetićete da je njihov zbir jednak 0). Ono što nam treba za transformisane ipsativne skorove je standardna devijacija koju ostvaruje Raja na svim njegovim skorovima. Ona se izračunava na sledeći način:

$$SD = \sqrt{\frac{\Sigma\,(X - AS)^2}{N - 1}}$$

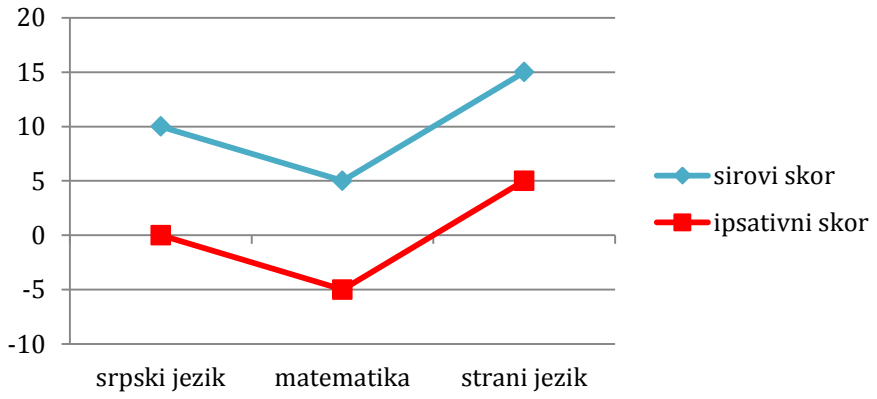
gde izraz $\Sigma\,(X - AS)^2$ predstavlja sumu kvadriranih odstupanja svake pojedinačne vrednosti (X) od aritmetičke sredine (AS), a N predstavlja veličinu uzorka. U našem primeru, SD je:

$$SD = \sqrt{\frac{\Sigma\,(0 - 5 + 5)^2}{3 - 1}} = 3{,}87$$

te je Rajini transformisani ipsativni skor za prvi test srpskog jezika (10 – 10) / 3,87 = 0, za drugi test matematike (5 – 10) / 3,87 = –1,29, i za treći test stranog jezika (15 – 10) / 3,87 = 1,29.

Kako je zbir ipsativnih skorova za svakog ispitanika nula, ispitanici se ne mogu međusobno porediti jer bi to značilo da svi imaju podjednako znanje. Praktično, ovi skorovi bi nam govorili o tome da svi imaju podjednako ukupno znanje, ali ne i posebna znanja iz pojedinih predmeta. Zato se za svakog pojedinačnog ispitanika procenjuje sklop odgovora, odnosno skorova na skalama ili testovima, čime se dobija uvid u profil ispitanika. Profil ispitanika koji se na ovaj način dobije tumači se u odnosu na prosek tog konkretnog ispitanika, a ne norme. Na Slici 25 dat je prikaz Rajinih normativnih i ipsativnih skorova. Ukoliko je prosek dobijen na referentnoj grupi, na primer 15, gledajući normativne

skorove (npr. sirove skorove) možemo zaključiti da Raja ima prosečno postignuće na stranom jeziku, ali ispodprosečno na matematici i srpskom jeziku. Uvidom u ipsativne skorove možemo zaključiti da Raji strani jezik ide najbolje, a matematika najgore poređenjem ova tri predmeta. Pritom, prilikom tumačenja ipsativnih skorova ne dobijamo informaciju o tome da li Rajino postignuće na testu stranog jezika predstavlja izvanredan rezultat, niti da li Rajino postignuće na testu matematike predstavlja nizak rezultat u kontekstu postignuća ostalih učenika. Dakle, ipsativnim skorovima se dobija informacija o preferencijama ispitanika, njegovom profilu na osnovu nekoliko atributa, ali se ne dobija informacija o izraženosti tih preferencija, generalnom kvantitetu preferencija.



Slika 25. Primer normativnih i ipsativnih skorova.

Kada imamo matricu sa sirovim skorovima, uobičajeno su u redovima ispitanici, a u kolonama varijable. Za izvršenje ipsatizacije potrebno je da matricu transponujemo, te da joj zamenimo kolone i redove. Takva transponovana matrica će u redovima imati varijable, a u kolonama ispitanike. Nad takvom matricom onda treba da izvršimo

standardizaciju. Potom se takva matrica ponovo transponuje u prvobitni oblik, pri čemu ona sada sadrži skorove dobijene ipsatizacijom. Kada bismo hteli da izračunamo sumacioni skor na ovim skorovima, on bi bio nula. Istraživači često kada rade faktorsku analizu urade analizu i na sirovim skorovima, i na ipsativnim skorovima, kako bi proverili stabilnost faktorskog rešenja. Ideja koja stoji u osnovi ipsatizacije je u tome da se time eliminišu odgovaračke pristrasnosti poput slaganja, preferencije ranga odgovora i slično.

Prednosti i nedostaci

Ipsativni skorovi primenu nalaze u ispitivanju motiva, vrednosti, profesionalne orijentacije, kognitivnih stilova, kao i u praćenju napretka, promena i razvoja. Koriste se u praksi i zbog pretpostavke da se njima eliminišu odgovaračke pristrasnosti, npr. slaganje, neslaganje, socijalna poželjnost i dr. Naime, zbog prisilnog izbora ne postoji mogućnost da se dobiju visoki skorovi na svim ispitivanim konstruktima, a takođe je davanje socijalno poželjnih odgovora otežano kada se u paru nalaze dva socijalno poželjna ili socijalno nepoželjna atributa (Mccloy, Heggested, & Reeve, 2005). Takođe, ipsativni skorovi se koriste i kako bi se eliminisali efekti blagih ili strogih procenjivača. Međutim, istraživanja pokazuju da se njima ne mogu eliminisati odgovaračke pristrasnosti, ali se mogu smanjiti, jer format ajtema na osnovu kojih se dobijaju ipsativni skorovi to otežava (npr. kao što je pomenuto, kada treba odabrati jedan između dva podjednako socijalno poželjna atributa). Otud, nije iznenađujuće što je kognitivna sposobnost pozitivan prediktor uspešne distorzije odgovora na pitanjima s prisilnim izborom (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006). Iako rezultati nisu jednoznačni,

većina autora ima stav da se prisilnim izborom ipak ne može smanjiti davanje socijalno poželjnih odgovora (više u Dilchert & Ones, 2012).

Zbog pretpostavke da se formatom prisilnog izbora može eliminisati ili smanjiti socijalna poželjnost, ovaj format se dosta koristi u profesionalnoj selekciji (30% kompanija ih koristi, videti u Burns & Christiansen, 2011). U jednoj meta-analitičkoj studiji je dobijeno da kvazi-ipsativni skorovi osobina ličnosti ostvaruju prednost u odnosu na čisto ipsativne i normativne skorove u predikciji radnog učnika (Salgado, Anderson, & Tauriz, 2015). Takođe, što je više ajtema ili skala koje se ipsativno skoruju (20, posebno preko 30), to se skorovi više približavaju normativnim (Baron, 1996). Neka domaća istraživanja pokazuju da su korelacije ipsativnih skorova, dobijenih ipsativnom transformacijom, i normativnih skorova – srednjeg intenziteta ili visoke (Knežević, 2014). Ovaj rezultat upućuje na zaključak da distorzija normativnih skorova nije velika. U istom istraživanju se može primetiti da postoje razlike između selekcionih i standardnih uslova zadavanja testa u korelacijama normativnih i ipsativnih skorova u slučaju nekoliko osobina ličnosti, te su npr. korelacije manje za savesnost u selekcionoj situaciji. Ove informacije mogu biti važne za izbor mera u procesu selekcije, te bi trebalo odabarati one mere i testove čiji su normativni skorovi sličniji ipsativnim skorovima.

Jedan od nedostataka ipsativnih skorova je već pomenut, a odnosi se na interpretaciju skorova. Naime, ipsativnim skorovima se dobija uvid u profil ispitanika, tj. odnos procenjenih atributa kod jednog ispitanika, ali ne i informacija o kvantitetu tih atributa. Na primer, ipsativnim skorovima možemo proceniti da li neko više voli „Milku", „Galeb" ili „Najlepše želje", ali ne i koliko generalno voli ove čokolade, da li uopšte voli čokolade i slično. Zbog toga što je suma ipsativnih skorova

konstanta, osobe koje imaju ekstremno nizak i ekstremno visok skor će imati isti ukupni skor, a moguće i iste profile. Iako neki autori isitiču da je verovatnoća da neko ima ekstremni skor veoma mala i da se sa povećanjem skala (30 i više) ova verovatnoća smanjuje (Baron, 1996; Brown, 2010), činjenica je da je praktična implikacija ovih nalaza upitna.

Ipsativni skorovi se ne mogu koristiti za *inter*-individualno poređenje. Kako nema uvida u položaj ispitanika u odnosu na druge, ukupni skorovi dobijeni preko ipsativnih skorova se ne preporučuju prilikom trijaže i selekcije (Heggestad, Morrison, Reeve, & McCloy, 2006). Prethodna istraživanja pokazala su da se percentilne norme i poredak ispitanika razlikuju na ipsativnim skorovima dobijenim preko prisilnog izbora i na normativnim skorovima dobijenim preko Likertove skale za isti test (npr. Cornwell & Dunlap, 1991). Drugim rečima, normativna interpretacija se ne može primeniti na ipsativnim skorovima.

Nedostatak čisto i kvazi-ipsativnih skorova je u tome što se nad njima ne mogu primeniti standardni statistički postupci, posebno faktorska analiza. Naime, kod ipsativnih skorova, skor na jednom atributu se može predvideti na osnovu skorova na ostalim atributima, što rezultira negativnim korelacijama između skorova. Budući da su mere međusobno zavisne, i varijanse grešaka su međusobno zavisne. Zbog ove međuzavisnosti ipsativnih skorova, nad njima se ne mogu primeniti iste psihometrijske metode kao nad normativnim skorovima (Cornwall & Dunlap, 1991). S obzirom na to, određivanje pouzdanosti i validnosti je problematično, te su ove karakteristike precenjene. Ovo je posebno očito u situacijama kada postoji mali broj ajtema ili skala u upitniku. Kao rešenje se nudi posebno razvijen metod faktorizacije ipsativnih skorova (tzv. Q-sort metoda) ili modeli bazirani na Teoriji

ajtemskog odgovora na osnovu kojih se dobijaju normativni skorovi iz ipsativnih (npr. Chernyshenko et al., 2009).

Pored navedenih psihometrijskih nedostataka, priroda formata prisilnog izbora je takva da neodlučne dovodi u veštačku situaciju izbora, kao i one koje ne poseduju ni jedan od ponuđenih atributa ili one koji poseduju oba atributa u istoj meri (isto se može zameriti i pitanjima s rangiranjem). S druge strane, neki autori ističu da je prednost prisilnog izbora u tome što se procena vrši u odnosu na neki referentni okvir (šta se od ponuđenih atributa više preferira) i što se ne mogu odabrati sve opcije, što je bliže životnom iskustvu (Meade, 2004).

Pitanja za razmišljanje

1. Koja je razlika između ipsativnih i normativnih skorova?
2. Koje vrste ipsativnih skorova postoje i u čemu je njihova razlika? Navedi primer.
3. Da li se ipsativnim skorovima mogu eliminisati odgovaračke pristrasnosti i koje? Objasniti.

Reference

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*(1), 49–56.

Brown, A. (2010). *How Item Response Theory can solve problems of ipsative data.* Unpublished doctoral dissertation. Barcelona, Spain: Faculty of Psychology, University of Barcelona. Preuzeto sa https://www.tesisenred.net/bitstream/handle/10803/80006/ANNA_BROWN_PhD_THESIS.pdf?sequence=1

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*(4), 358–372.

Chernyshenko, O. S., Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R., & Tuttle, M. D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: Empirical comparisons with other formats. *Human Performance, 22*, 105–127.

Cornwell, J. M., & Dunlap, W. P. (1991). On the questionable soundness of factoring ipsative data: A response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology, 67*(2), 89–100.

Clemans, W. V. (1966). *An analytical and empirical examination of some properties of ipsative measures (Psychometric Monograph, No. 14)*. Richmond, VA: Psychometric Society. Preuzeto sa https://www.psychometricsociety.org/sites/default/files/pdf/MN14.pdf

Dilchert, S., & Ones, D. S. (2012). Application of preventive strategies. In M. Ziegled, C. Maccann, & R. D. Roberts (Eds.), *New perspective on faking in personality assessment* (pp. 177–200). New York, NY: Oxford University Press.

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*(1), 9–24.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*(3), 167–184.

Knežević, G. (2014). *Profesionalna selekcija: elementi, izazovi i jedno praktično rešenje.* Beograd, Srbija: Centar za primenjenu psihologiju.

Martinussen, M., Richardsen, A. M., Vårum, H. W. (2001). Validation of an ipsative personality measure (DISCUS). *Scandinavian Journal of Psychology, 42*(5), 411–416.

Mccloy, R. A., Heggestad, E. D., & Reeve, C. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods, 8*(2), 222–248.

Meade, A. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*(4), 531–552.

Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology, 88*(4), 797–834.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance, 19*(3), 175–199.

ODGOVARAČKE PRISTRASNOSTI

Osnovna pretpostavka u vezi sa odgovaranjem na ajteme testa je da je ajtem stimulus koji izaziva određeno ponašanje, misli i osećanja kod ispitanika, u skladu s kojima ispitanik daje odgovore na ajteme. Međutim, odavno je već poznato da na odgovaranje na ajteme ne utiče samo ajtemski sadržaj, već i ostale karakteristike testa (npr. format, medijum) i uslovi testiranja (npr. prisutnost distraktora, da li je u pitanju selekciono testiranje itd.). Opisijući tzv. dogmu psihometrijskog merenja, Fajgelj (2013) navodi da u odgovaranju na ajteme učestvuje čitava ličnost. To znači da je odgovor na ajtem i odraz konstelacije osobina ličnosti ispitanika, trenutne motivacije, stavova prema predmetu merenja i testiranju, pripadnosti određenoj kulturi i slično. Navedeni faktori koji utiču na odgovaranje ispitanika, a ne potiču od samog predmeta merenja testa, predstavljaju greške merenja i jedan su od glavnih izvora ugrožavanja validnosti testovnih skorova. Među ovim greškama, najčešće su istraživane odgovaračke pristrasnosti ili usmerenja (eng. *response bias*).

Odgovaračke pristrasnosti odnose se na sistematsku tendenciju odgovaranja na različite ajteme koja ne zavisi od ajtemskog kontksta, tj. od onoga šta bi ajtem trebalo da meri (Paulhus, 1991). Odgovaračke pristrasnosti mogu biti odgovarački setovi ili stilovi. **Odgovarački set** (eng. *response set*) je situaciono određen i predstavlja trenutnu reakciju na zahteve situacije, npr. ukoliko postoji vremenski pritisak, javno prezentovanje rezultata ili povećana motivacija da se osoba prikaže u socijalno poželjnom svetlu. Takođe, nekad ajtemski kontekst može izazvati odgovarački set, kao što je ajtemski format ili sadržaj prethodnog

ajtema. Kronbah (Cronbach, 1946) je definisao odgovarački set kao bilo koju tendenciju davanja sistematski drugačijih odgovora na isti ajtem u zavisnosti od toga u kom formatu je ajtem prikazan. Na primer, za ajtem s binarnim formatom prezentovanja, možemo od ispitanika dobiti odgovor "ne", ali ako se isti ajtem prezentuje u vidu petostepene Likertove skale, onda možemo dobiti veće slaganje sa tvrdnjom (npr. podeok 4). Za razliku od odgovaračkog seta, **odgovarački stil** (eng. *response style*) je stabilna tendencija koja se ispoljava nezavisno od situacije – dosledno kroz različite situacije, vreme i domene procene, i na različitim upitničkim merama.

Važno je napomenuti da prema sadašnjem shvatanju odgovaračkih pristrasnosti, bilo koja vrsta pristrasnosti može se manifestovati i kao stil i kao set. Na pirmer, davanje socijalno poželjnih odgovora može biti rezultat nekritičke i nerealne slike o sebi (odgovarački stil), ali i trenutne motivacije da se na prijemu za posao osoba prikaže u što boljem svetlu (odgovarački set). Ipak, za neke odgovaračke pristrasnosti postoji veće slaganje u vezi s tim da li su više stil ili set. Istraživanja dosledno pokazuju vremensku stabilnost sklonosti ka povlađivanju, neslaganju, preferenciji ekstremnih odgovora i preferenciji srednjeg odgovora, te se ove odgovaračke pristrasnosti često nazivaju stilovima (npr. Furr & Bacharach, 2014; Weijters, Geuens, & Schillewaert, 2010). Valja pomenuti da ne pokazuju svi rezultati dosledno da su navedene odgovaračke pristrasnosti stabilne tendencije, te se u nekim istraživanjima dobija da su neke od njih tek umereno stabilne, a ne dosledno stabilne (npr. povlađivanje, videti Danner, Aichholzer, & Rammstedt, 2015).

Među istraživačima ne postoji jasno slaganje u vezi s tim koliki je uticaj odgovaračkih pristrasnosti na opažene skorove. Dok jedini ističu da

je taj uticaj neznatan i mit, drugi upozoravaju na to da su svi skorovi na neki način kontaminirani odgovaračkim pristrasnostima (više u Podsakoff, MacKenzie, & Podsakoff, 2012). U jednom kros-kulturalnom istraživanju različitih stavova potrošača, u kojem su iz sirovih ajtemskih skorova izračunate mere različitih odgovaračkih pristrasnosti, pokazano je da se, u proseku, 8% varijanse opaženih skorova može pripisati odgovaračkim pristrasnostima (više u Baumgartner & Steenkamp, 2001). Pritom, raspon varijanse koji se može pripisati odgovaračkim pristrasnostima varira od 0 do 29%, i zavisi od predmeta merenja upitnika. Iako autori nisu direktno testirali efekat predmeta merenja na odgovaračke pristrasnosti, može se primetiti da je najveći udeo odgovaračkih pristrasnosti u slučaju pitanja o zdravstvenoj svesti, a najmanje u slučaju traženja promena i informacija u funkciji kupovine. Ovaj nalaz implicira da nisu sve mere podjednako podložne odgovaračkim pristrasnostima, što takođe treba imati u vidu kada se želi ispitati ili kontrolisati njihov efekat.

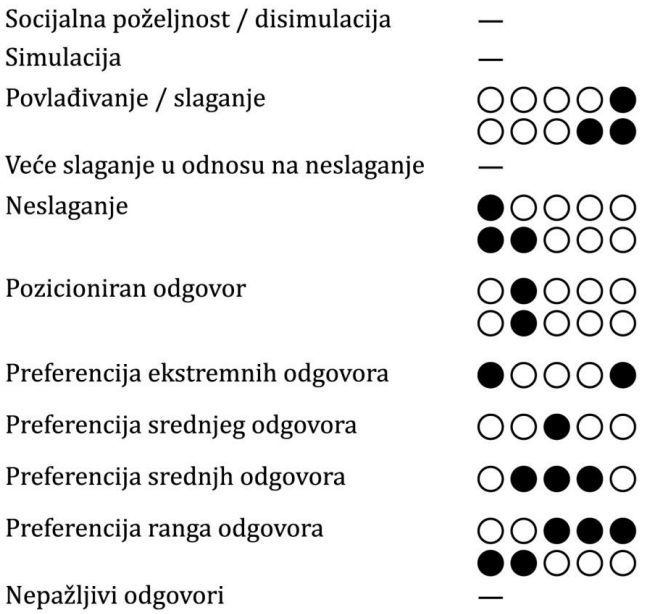
Iako se čini da je procenat varijanse skorova koji se može pripisati odgovaračkim pristrasnostima mali, odgovaračka usmerenja mogu značajno da utiču na dva aspekta merenja. Prvi se odnosi na **skorove ispitanika**, pri čemu odgovaračke pristrasnosti mogu da utiču na izraženost tih skorova – mogu da smanje ili povećaju opaženi skor u odnosu na pravi. Drugim rečima, menja se rang ispitanika, pa samim tim i donošenje odluke o pojedincu. Međutim, najveći problem predstavlja to što se ne može jasno odvojiti sistematska varijansa koja potiče od odgovaračkih pristrasnosti od sistematske varijanse koja potiče od predmeta merenja (konstrukta), posebno kada je reč o odgovaračkim pristrasnostima kao stabilnim tendencijama, tj. stilovima.

Drugi aspekt odnosi se na efekat odgovaračkih pristrasnosti na **odnose između varijabli**, tako što korelaciju između varijabli mogu povećati ili smanjiti u odnosu na stvarnu visinu korelacije. Ukoliko je korelacija između dva konstrukta pozitivna, kao i korelacija između odgovaračkih pristrasnosti na mere oba konstrukta, korelacija između skorova na tim dvema merama će biti veštački veća. Takođe, ukoliko je korelacija između dva konstrukata negativna, kao i korelacija između odgovaračkih pristrasnosti, korelacija između konstrukata će biti veštački veća. U slučaju da je korelacija između konstrukata negativna, a da odgovaračke pristrasnosti pozitivno koreliraju, korelacija između mera će biti niža, čak je moguće i da promeni znak. Isto važi i u slučaju kada su korelacije između konstrukata pozitivne, a između odgovaračkih pristrasnosti negativne (Baumgartner & Steenkamp, 2001). Kako je većina psihometrijskih karakteristika testa bazirana na korelacijama, kao i većina multivarijatnih statističkih postupaka obrade podataka, može se pretpostaviti da odgovaračke pristrasnosti mogu uticati na psihometrijske karakteristike, pa samim tim i na zaključke o konstruktima i njihovim relacijama. S obzirom na navedene razloge, većina autora se slaže da odgovaračke pristrasnosti mogu ugroziti interpretaciju rezultata i da bi ih trebalo kontrolisati kad god je to moguće (Podsakoff et al., 2012).

Vrste odgovaračkih pristrasnosti

Prvi sistematičan pregled odgovaračkih pristrasnosti ponudio je Kronbah (Cronbach, 1946). Vremenom, listu odgovaračkih pristrasnosti proširili su i mnogi drugi istraživači, a najčešće se može razlikovati 11 vrsta odgovaračkih pristrasnosti. Njihov pregled dat je na Slici 26, kao i primer manifestovanja na petostepenoj skali Likertovog tipa u

slučajevima gde je to moguće utvrditi na nivou pojedinačnog ajtema. U nastavku biće detaljnije objašnjena svaka od navedenih odgovaračkih pristrasnosti.



Slika 26. Pregled odgovaračkih pristrasnosti.

Socijalna poželjnost/disimulacija i simulacija

Definicija

Socijalno poželjno odgovaranje ili disimulacija (eng. *socially desirable responding*, *faking good, disimulation*) jedna je od najistraživanijih odgovaračkih pristrasnosti. Socijalna poželjnost odnosi se na tendenciju da se osoba prikaže u što boljem "socijalnom svetlu" i javlja se u testovima koji pripadaju drugoj paradigmi merenja, tj. u testovima ličnosti. Neki autori u definiciji socijalne poželjnosti

naglašavaju prihvatljivost ponašanja u određenom kulturološkom kontekstu (npr. Crowne & Marlowe, 1964), budući da jedno isto ponašanje može biti prihvatljivo u jednoj kulturi, a neprihvatljivo u drugoj. Obrnuto od disimulacije, simulacija (eng. *simulation*, *malingering, faking bad*) ili davanje socijalno nepoželjnih odgovora predstavlja tendenciju osobe da se prikaže u socijalno nepoželjnom svetlu, da naglašava patološke karakteristike i simptome ili da se prikaže kao zla, opaka, žestoka osoba i sl.

Od samog početka interesovanja za socijalnu poželjnost, ona je definisana kao namerno i svesno ponašanje, što bi je svrstalo u odgovarački set (Edwards, 1957). Kasnija istraživanja su, međutim, pokazala da se socijalna poželjnost teško može odvojiti od varijanse ličnosti, odnosno da obuhvata nesvesna i automatska ponašanja, te da se može shvatiti i kao stil (npr. Paulhus, 2002). Prema savremenom modelu, socijalna poželjnost je višedimenzionalna i sadrži aspekt koji se odnosi na nesvesno ponašanje, kao i aspekt koji se odnosi na svesno ponašanje (Paulhus, 2018). Aspekt socijalne poželjnosti, kao i bilo koje druge pristrasnosti, koji se odnosi na *namerno i svesno* predstavljanje drugačije slike o sebi, zajedničkim terminom se naziva **lažiranjem** (eng. *faking*). Prema Cigleru i saradnicima (Ziegler, MacCann, & Roberts, 2012) lažiranje se shvata odgovaračkim setom i definiše kao promocija selfa na način koji će osobi pomoći da ostvari određeni lični cilj. Lažiranje se javlja kada je ovaj odgovarački set aktiviran situacionim zahtevima i ličnim karakteristikama da se produkuje sistematska razlika u skorovima na testu koja niije rezultat predmeta merenja. Drugim rečima, lažiranje predstavlja interakciju između situacije i osobe. U zavisnosti od toga kako osoba doživljava situaciju i u zavisnosti od nekih osobina ličnosti, osoba će odlučiti da li će lažirati odgovore ili ne.

Lažiranje je prepoznato kao ozbiljan problem u psihološkom testiranju, posebno u okviru forenzičke prakse i prilikom profesionalne selekcije. Prema nekim podacima, od 7,3% do 27% opšte populacije lažira na testovima, dok je u forenzičkoj populaciji taj procenat veći i kreće se od 31% do 45% (više u Furr & Bacharach, 2014). Motivacija za lažiranjem može biti različita, ali je najčešće podstaknuta nekim eksternim nagradama, kao što su, na primer, dobijanje novčane nadoknade, invalidske ili prevremene penzije, ili izbegavanje ili smanjenje zatvorske kazne za krivično delo, izbegavanje vojne dužnosti itd. Za razliku od disimulacije, simulacija ne mora nužno predstavljati manipulativni čin, već može predstavljati i apel za pomoć ili negativistički stav (Domino & Domino, 2006).

Dva su osnovna načina na koji osoba može da se prikaže u socijalno poželjnom svetlu, tj. da disimulira:

1. naglašavanje socijalno poželjnih karakteristika i ponašanja, poput iskrenosti, altruizma i sl., ili dobrog mentalnog zdravlja i prilagođenosti;
2. poricanje nepoželjnih karakteristika i ponašanja, uključujući ona koja u iskustvu ima većina ljudi (npr. svako je u životu barem nekada slagao, uvredio drugoga, pravio se da zna nešto što zapravo ne zna i sl.).

S druge strane, simuliranje se može manifestovati kao:

1. naglašavanje negativnih simpotoma, tegoba, problema u funkcionisanju i generalno lošeg mentalnog zdravlja, bolesti i loše adaptacije, ali i negativnih karakteristika kao što su opakost, oholost, zloba, agresija i sl.;
2. naglašavanje onih simptoma koji odražavaju ispitanikovo shvatanje koncepta mentalne bolesti ili neprilagođenosti.

Socijalna poželjnost je više istraživana u odnosu na simulaciju, te ima više i modela i tehnika njene detekcije. Jedan od prvih autora koji je ukazao na problem socijalne poželjnosti je Edvards koji je i konstruisao skalu za procenu socijalne poželjnosti tzv. Edvardsovu skalu socijalne poželjnosti (Edwards, 1957). On je preuzeo 150 ajtema iz različitih skala MMPI-ja (L, F, K itd.) i dao procenjivačima da kategorišu ajteme kao socijalno poželjne ili nepoželjne. U svoju finalnu verziju skale uvrstio je 79 ajtema kod kojih je postojalo potpuno slaganje među procenjivačima, a potom je na osnovu diskriminativnosti odabrao 39 ajtema. Ubrzo, Kraun i Marlou (Crowne & Marlowe, 1960) su kritikovali Edvardsovu skalu jer je nastala na kliničkom testu, te je nejasno da li neko ko ima nizak skor na skali ima nisku socijalnu poželjnost ili odsustvo psihopatoloških simpotoma. Oni su socijalnu poželjnost definisali šire, u kontekstu socijalno prihvatljivog i normativno odobrenog ponašanja u određenoj kulturi, i ponudili svoju skalu koja ne obuhvata psihopatološke fenomene, tzv. Marlou-Kraun skalu socijalne poželjnosti (Crowne & Marlowe, 1960, 1964). Kasnije su i drugi istraživači redefnisali operacinalizaciju socijalne poželjnosti, a skale socijalne poželjnosti su se sve češće nalazile u okviru inventara ličnosti kao kontrolne skale. Iako istraživanje socijalne poželjnosti ima dugu istoriju, njena aktuelnost ne zastareva kako u pogledu modela kojima se ona objašnjava, upitničkih i drugih mera procene, tako i u pogledu načina kontrolisanja ove odgovaračke pristrasnosti.

Dimenzije

Ideja o dimenzionalnosti socijalne poželjnosti nastala je još na samom početku interesovanja za ovaj fenomen, ali je najbolje uobličena u Paulusovim radovima. Ipak, konceptualizacija dimenzija socijalne

poželjnosti je bila prilično dimanimčna, i sam Paulus je predložio nekoliko modela, svaki put revidirajući i usavršavajući prethodni. U svom početnom modelu, istraživanjem zajedničke latentne strukture nekoliko skala socijalne poželjnosti, Paulus (Paulhus, 1984) je zaključio da se mogu razlikovati dve dimenzije:

1. upravljanje utiskom (eng. *impression management*), koje se odnosi na svesnu i namernu tendenciju prikazivanja sebe u pozitivnom svetlu;
2. samozavaravanje (eng. *self-deception*), koje se odnosi na iskrenu i nesvesnu tendenciju prikazivanja sebe u pozitivnom svetlu.

Kako se **upravljanje utiskom** odnosi na svesno lažiranje i zavisi od namene testiranja, predstavlja suštinu socijalne poželjnosti i upravo je ova odgovaračka pristrasnost glavni izvor ugrožavanja validosti upitnika. Drugim rečima, upravljanje utiskom je onaj deo neželjene varijanse skorova koji onemogućava donošenje objektivne odluke o ispitaniku ili konstruktu. U ovom kontekstu, upravljanje utiskom se shvata kao *odgovarački set*, tj. kao tendencija koja zavisi od situacije (npr. da li je u pitanju test visokog ili niskog uloga, koliko smo motivisani u konkretnoj situaciji da se prikažemo u povoljnom svetlu i slično). S druge strane, **samozavaravanje** se smatra *odgovaračkim stilom*, tj. stabilnom tendencijom ispitanika koja ne zavisi od uslova i namene testiranja. Samozavaravanje predstavlja nedostatak kritičkog uvida u sopstvene karakteristike koji je usko povezan s narcizmom (Paulhus, 1994). Naime, kod samozavaravanja osoba zaista veruje da je takva kakvom se predstavlja i ima idealizovanu sliku o sebi. S obzirom na to, samozavaravanje zapravo i nije socijalno poželjno odgovaranje, već predstavlja iskren iskaz onoga u šta sam ispitanik veruje, čime se dobija dodatna informacija o njegovoj ličnosti. Prema Paulusovom shvatanju,

samozavaravanje predstalja pozitivnu iluziju o sebi, tj. crtu samoosnaživanja, koja je adaptivna kada treba da se podigne samopouzdanje i ostvari pozitivna prva impresija, ali ima negativne posledice po interpersonalne relacije (Paulhus, 1994).

Paulus je konstruisao Uravnoteženi inventar socijalno poželjnog odgovaranja (Balanced Inventory of Desirable Responding – BIDR: Paulhus, 1984, 1991, za srpsku adaptaciju videti Knežević, Radović, & Peruničić, 2008) koji predstavlja trenutno najkorišćeniju meru socijalne poželjnosti (Holden & Passey, 2009). Primer ajtema iz skale upravljanja utiskom je "Nikad ne prikrivam svoje greške", a samozavaravanja "Moj prvi utisak o ljudima obično bude tačan". Ove dve skale generalno nisko koreliraju (Paulhus, 1991). Prema Paulusovom shvatanju, visoki rezultati na skalama socijalne poželjnosti ne znače sami po sebi odstupanje od realnosti ukoliko nema dokaza za to. Stoga je on uveo da se prilikom bodovanja inventara za svaki ekstremni odgovor (kada se šestostepena skala od 0 do 6 prebaci u skalu od 1 do 7, to bi bili odgovori 6 i 7) doda još jedan bod. Pokazano je da je skala upravljanja utiskom osetljiva na uslove zadavanja tj. instrukcije, i da se skorovi na ovoj skali menjaju kada se od ispitanika traži da se predstavi u socijalno poželjnom svetlu, socijalno negativnom svetlu ili iskreno (Paulhus, Bruce, & Trapnell, 1995). Iako je Paulus napustio ideju o ovakvoj dvofaktorskoj strukturi socijalne poželjnosti, BIDR se i dalje uveliko koristi.

Ipak, ovom inventaru, pa samim tim i modelu, upućene su i kritike koje se odnose na nepotpunost dvodimenzionalnog modela, na to da je svesna komponenta socijalne poželjnosti kompleksnija i višedimenzionalna, kao i da skorovi na svesnoj i nesvesnoj komponenti nisu međusobno nezavisni, kao što je to originalno dobijeno (više u Holden, Book, Edwards, Wasylkiw, & Starzyk, 2003). Pored toga, iako

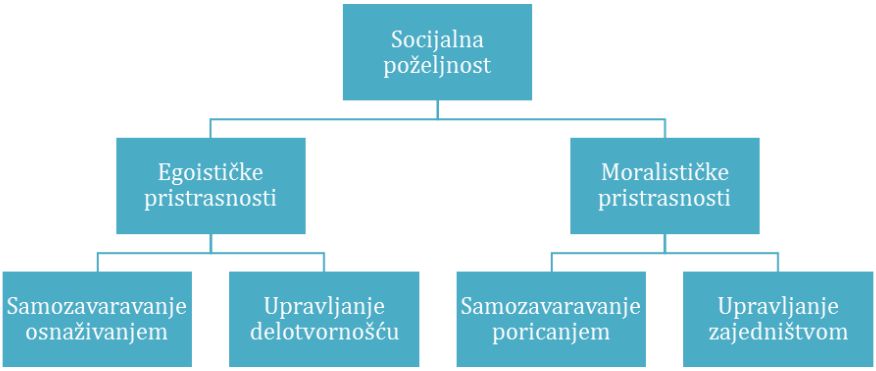
samozavaravanje nije trebalo da predstavlja odgovarački set, pokazano je da se i samozavaravanje povećava u situaciji kada je motivacija da se prikažemo u povoljnom svetlu visoka (Galić i Jerneić, 2006). Drugim rečima, deluje kao da je samozavaravanje osetljivo na promenu situacije i konteksta testiranja. Još jedna važna kritika je da je varijansa upravljanja utiskom stabilna, te da predstavlja sistematsku varijansu skorova povezanu sa individualnim razlikama. Naime, u mnogim istraživanjima je pokazano da je upravljanje utiskom povezano sa osobinama ličnosti u situaciji anonimnog odgovaranja, pre svega sa prijatnošću i savesnošću modela Velikih pet (npr. Puls & Stemmler, 2003). Čak i sam Paulus u kasnijim radovima ističe kao ključni problem skala socijalne poželjnosti nemogućnost da se razlikuje stvarna izraženost neke crte ličnosti od socijane poželjnosti (više u Paulhus, 2002). Rezultati navedenih istraživanja ne podržavaju početne pretpostavke modela u vezi s tim koji aspekt socijalne poželjnosti je odgovarački stil, a koji odgovarački set, čak, pokazuju suprotno od početnih pretpostavki.

S obzirom na to, Paulus je revidirao model socijalne poželjnosti (Paulhus, 2002). Novi model i dalje podržava da postoji svesna i nesvena komponenta socijalne poželjnosti, samo što one više nisu ključna distinkcija faktora socijalne poželjnosti. Ključna razlika između faktora socijalne poželjnosti je njihov domen. Inspiraciju za ovaj model Paulus nalazi u Viginsovim radovima u kojima je takođe predložena dvofaktorska struktura socijalne poželjnosti (Wiggins, 1966). Predložena dva osnovna faktora su:

1. **egoističke ili Alfa pristrasnosti** (eng. *egoistic, Alpha biases*), koje se odnose na tendenciju davanja preterano pozitivnih samoopisa u domenu socijalnog i intelektualnog statusa (fenomen “superheroja”);

2. **moralističke ili Gama pristrasnosti** (eng. *moralistic, Gamma biases*), koje se odnose na tendenciju negiranja socijalno devijantnih impulsa i prikazivanje sebe kao pouzdane i altruistične osobe koja veoma poštuje socijalne norme (fenomen "sveca", Paulhus, 2002; Paulhus & John, 1998).

Oba faktora sadrže *i svesnu i nesvesnu* komponentu, te je model je usložnjen. Drugim rečima, oba aspekta sadrže i samozavaravanje i upravljanje utiskom. Najpre, u domenu samozavaravanja mogu se razlikovati dva aspekta, a to su isticanje pozitivnih karakteristika (samozavaravanje osnaživanjem) i negiranje ili poricanje negativnih karakteristika (samozavaravanje poricanjem, Paulhus & Reid, 1991). Potom, u domenu upravljanja utiskom mogu se razlikovati takođe dva aspekta, pri čemu se jedan odnosi na upravljanje delotvornošću – prikazivanje sebe kao dominantne, kompetentne, sposobne, inteligente osobe, a drugi se odnosi na upravljanje zajedništvom – prikazivanje sebe kao altruistične, pouzdane, moralne, poštene i iskrene osobe (Paulhus & John, 1998). Opisana dva aspekta upravljanja utiskom mogu se dovesti u vezu s različitim fundamentalnim vrednostima, a to su upravo delotvornost (eng. *agency*) ili dominacija i zajedništvo (eng. *community*) ili afilijacija. Egoističke pristrasnosti obuhvataju samozavaravanje osnaživanjem i upravljanje delotvornošću, dok moralističke pristrasnosti obuhvataju samozavaravanje poricanjem i upravljanje zajedništvom (Slika 27). Obe vrste pristrasnosti odnose se na samo-favorizovanje, ali na različiti način i u različitim domenima. Značaj ovog modela je u tome što se po prvi put uvodi razlika socijalne poželjnosti u odnosu na domen (egoistički i moralistički).

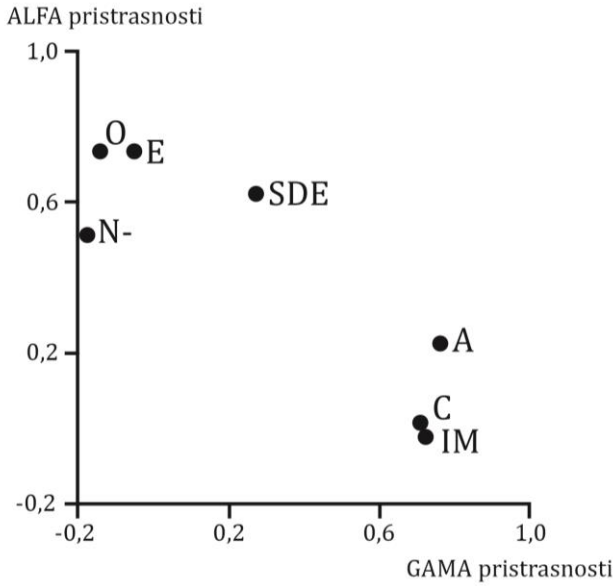


Slika 27. Paulusov model socijalne poželjnosti (Paulhus, 2002).

Obe vrste pristrasnosti javljaju se u anonimnim uslovima, dakle kada ne postoji motivacija za prikazivanjem u pozitivnom svetlu, i obe su osetljive na instrukciju, tj. menjaju se u zavisnosti od instrukcije. Na ovaj način je *izbrisana granica između stila i seta*, ali moglo bi se pretpostaviti da, ako neko ima tendenciju da se prikaže u socijalno poželjnom svetlu, u nekim situacijama ta tendencija može više da dođe do izražaja (npr. u situaciji selekcije), ali će se vrlo retko desiti da se neko, ko nema ovu tendenciju, prikaže u socijalno poželjnom svetlu u posebnoj situaciji. Dakle, situacija može da oblikuje ovu tendenciju, ali samo u smislu da je još više naglasi.

S obzirom na popularnost BIDR inventara i nedostatak upitnika kojim bi se operacionalizovale dimenzije novog modela, ili neadekvatne psihometrijske karakteristike predloženih mera, u istraživanjima egoističkih/Alfa i moralističkih/Gama pristrasnosti najčešće se koristi postojeći BIDR inventar. Tada se skala upravljanja utiskom poistovećuje s egoističkim, a skala samozavaravanja s moralističkim pristrasnostima. Istraživanja su pokazala da su egoističke pristrasnosti usko povezane s narcizmom, a da su moralističke pristrasnosti, zapravo, povezane sa drugim skalama socijalne poželjnosti, kao što su Marlou-Kraun skala, L

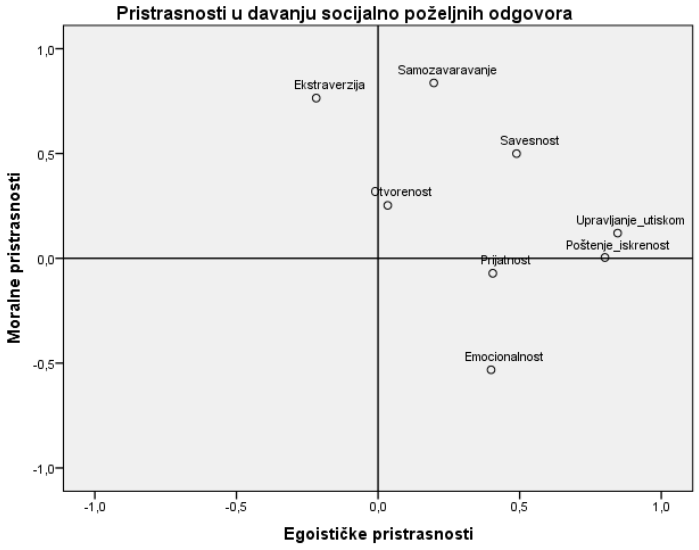
skala iz Ajzenkovog EPQ-a i L skala iz MMPI-a (Paulhus, 2002). Na osnovu ovog rezultata može se zaključiti da je egoistički domen socijalne poželjnosti zanemaren u postojećim skalama socijalne poželjnosti, te da se u njima favorizuje samo moralistički domen. Kada su ova dva aspekta socijalne poželjnosti pozicionirana u prostor bazičnih osobina ličnosti modela Velikih pet, egoističke pristrasnosti su pokazale pozitivnu povezanost sa ekstraverzijom i otvorenošću ka iskustvu, a negativnu sa neuroticizmom, dok su moralističke pristrasnosti pokazale pozitivnu povezanost sa prijatnošću i savesnošću (Paulhus, 2002; Pauls & Stemmler, 2003, videti Sliku 28).



Slika 28. Opterećenja na faktorima Alfa i Gama pristrasnosti na osnovu zajedničke faktorske strukture osobina ličnosti modela Velikih pet i dimenzija socijalne poželjnosti (adaptirano iz Pauls & Stemmler, 2003).

Napomena: N = neuroticizam (minus označava niski neuroticizam), E = ekstraverzija, O = otvorenost, A = prijatnost, C = savesnost, SDE = samozavaravanje, IM = upravljanje utiskom.

U radu Drobnjaković, Cmiljanović, Dinić i Janičić (2016) su ispitivane tendencije samozavaravanja i upravljanja utiskom u kontekstu šestofaktorskog HEXACO modela ličnosti. U ovom radu je dobijeno da je dominantni korelat upravljanja utiskom osobina poštenje-skromnost (ili poštenje-iskrenost) koja predstavlja moralni domen ličnosti, a potom slede savesnost i prijatnost. S druge strane, dominantni korelat samozavaravanja je ekstraverzija, a potom savesnost i (negativna) emocionalnost. Zajednička faktorska struktura dimenzija HEXACO modela i skala BIDR može se videti na Slici 29, gde se uočava da se dva izdvojena faktora mogu poistovetiti sa egoističkim i moralističkim pristrasnostima. Na osnovu navedenih nalaza može se zaključiti da ove dve vrste pristrasnosti imaju svoje distinktivne korelate, te da je njihovo razlikovanje opravdano.



Slika 29. Zajednička faktorska struktura osobina ličnosti HEXACO modela i dimenzija socijalne poželjnosti (preuzeto iz Drobnjaković i sar., 2016).

Vremenom je Paulus razvio novi instrument za procenu socijalne poželjnosti koji sadrži četiri skale u skladu s modelom – Sveobuhvatni inventar socijalno poželjnog odgovaranja (Comprehensive Inventory od Desirable Responding – CIDR; Paulhus, 2006). Primeri ajtema mogu se videti u Tabeli 24. Slično kao i u slučaju bodovanja BIDR, u ovom inventaru se prilikom bodovanja uzimaju u obzir samo ekstremni odgovori (poslednja dva odgovora na pozitivnom polu). Međutim, ovaj upitnik još nije zaživeo u upotrebi kao BIDR.

Tabela 24

Primeri ajtema Sveobuhvatnog inventara socijalno poželjnog odgovaranja

Skala	Ajtem
Samozavaravanje osnaživanjem	Potpuno verujem svom sudu.
Upravljanje delotvornošću	Mogu vladati situacijom kad god to poželim.
Samozavaravanje poricanjem	Nikada nisam učinio nešto čega se sramim.
Upravljanje zajedništvom	Nikada ne bacam otpatke po ulici.

Napomena: Ajtemi su preuzeti iz hrvatske adaptacije CIDR (Jerneić, Galić i Parmač, 2007).

Istraživanja su pokazala da su skorovi na skalama iz domena upravljanja utiskom osetljivi na instrukciju, i to specifičnu za domen procene (Galić, Jerneić i Belavić, 2009; Galić & Jerneić, 2013). Naime, u uslovima kada ispitanici treba da se prikažu kao najbolji kandidati za menadžersku poziciju, dobijaju se povišeni skorovi na upravljanju delotvornosti, a u uslovima kada ispitanici treba da se prikažu kao najbolji kandidati za poziciju učitelja, dobijeni su viši skorovi na upravljanju zajedništvom, što je u skladu sa očekivanjima. Međutim, nesvesne komponente se takođe menjaju s instrukcijama, što nije u skladu s pretpostavkama modela jer bi one trebalo da odražavaju stabilne tendencije (Galić i sar., 2009). Takođe, relacije skala

odgovaračkih pristrasnosti sa osobinama ličnosti modela Velikih pet nisu u potpunosti u skladu s pretpostavkama. Na primer, u uslovu iskrenog odgovaranja, skale upravljanja utiskom povezane su sa osobinama ličnosti (Galić i sar., 2009), što ne bi trebalo s obzirom na pretpostavku da upravljanje utiskom predstavlja svesnu komponentu koja se menja pod različitim uslovima. Otuda ona ne bi trebala biti povezana sa stabilnim tendencijama kao što su osobine ličnosti. Ovi i slični nalazi upućuju na probleme u vezi sa podelom na svesne i nesvesne komponente, što je bila zamerka i originalnom Paulusovom modelu.

Ipak, Paulus nastavlja s revizijom modela socijalne poželjnosti, i u svom najnovijem modelu ističe ponovo dva njena aspekta, ali sada usmerena upravo na delotvornost i zajedništvo (Paulhus, 2018). Dakle, osnovne pristrasnosti nisu više egoističke i moralističke, već su prethodno podređeni aspekti ovih pristrasnosti sada u prvom planu. Ponovo, inspiraciju nalazi u Viginsovim radovima, u kojima delotovornost i zajedništvo predstavljaju dva velika klastera osobina, motiva i obrazaca ponašanja (Wiggins, 1991). U skladu sa dvema osama Viginsovog cirkumpleks modela, dva aspekta socijalne poželjnosti koja Paulus predlaže su:

1. **delotvornost ili dominacija** (eng. *agency social desirability*), koja se manifestuje kao isticanje dominacije, samopouzdanja, kompetencija, produktivnosti, inteligencije, kreativnosti;
2. **zajedništvo ili afilijacija** (eng. *communion social desirability*), koja se manifestuje kao isticanje onih karakteristika koje doprinose harmoničnim interpersonalnim odnosima i dobrobiti zajednici, npr. kooperativnost, iskrenost, ljubaznost, žrtvovanje zbog drugih, činjenje dobih dela i sl.

U svim Paulusovim modelima menjaju se nazivi komponenti socijalne poželjnosti, ali se čini da je njihov sadržaj uglavnom relativno nepromenjen. O tome svedoči i upotreba BIDR inventara u ispitivanju koncepata novih, revidiranih Paulusovih modela. Razlog ove prakse možemo naći u tome što su samozavaravanje, egoističke pristrasnosti i delotvornost međusobno bliske s jedne strane, a upravljanje utiskom, moralističke pristrasnosti i zajedništvo s druge strane. Kako je Paulus najnoviju koncepcija socijalne poželjnosti predložio tek nedavno, ostaje na istraživačima i samom autoru da testiraju pretpostavke modela, a veći je izazov napraviti upitničku meru. Na osnovu činjenice da je iz novog modela izbačena dimenzija svesno-nesvesno, postavlja se pitanje na koji način će se onda detektovati socijalna poželjnost i može li se bilo koji njen aspekt uopšte odvojiti od osobina ličnosti. Ovaj problem još više usložnjavaju nalazi koji pokazuju da je socijalna poželjnost povezana sa dobrim mentalnim zdravljem i blagostanjem (Brajša-Žganec, Ivanović, & Lipovčan, 2011), te da su osobe koje imaju visok skor na socijalnoj poželjnosti bolje adaptirane, druželjubivije i otvorenije za iskustva (McCrae & Costa, 1983). Povrh toga, u jednom sistematičnom pregledu upotrebe skala socijalne poželjnosti u kliničkim istraživanjima dobijeno je da je upravljanje utiskom pozitivno povezano sa nepoželjnim ponašanjima kao što su upotreba alkohola i partnersko nasilje, dok se samozavaravanje pokazalo kao protektivni faktor u odnosu na depresiju, beznadežnost i rizik od suicida (Perinelli & Gremigni, 2016).

Za razliku od socijalne poželjnosti, za simulaciju ne postoje istraživanja o njenoj strukturi ili dimenzijama. Simulacija je najčešće ispitivana u domenu kliničke psihologije kada je osoba motivisana da se prikaže kao neuračunljiva zarad oslobođanja ili smanjenja zatvorske

kazne, ili kada naglašava simptome kako bi dobila prevremenu ili invalidsku penziju i druge benefiti.

Problemi

Socijalna poželjnost se ističe kao glavni izvor ugrožavanja validnosti skorova na testovima ličnosti, posebno u uslovima selekcije. Na primer, pokazano je da 30 do 50% kandidata lažira na testovima prilikom apliciranja za posao (Griffith, Chmielowski, & Yoshita, 2007). Međutim, izgleda da nisu sve vrste validnosti ugrožene ovom odgovaračkom pristrasnošću. Naime, pokazano je da ovakvo iskrivljavanje odgovora ne remeti faktorsku strukturu upitnika (Ones, Viswesvaran, & Reiss, 1996), odnosno da su konstrukt i sadržinska validnost očuvane. Takođe, Ones i Visvesveren (Ones & Viswesvaran, 1998) iznose stanovište da je socijalna poželjnost lažni alarm za uzbunu u selekcionoj situaciji, te da je kontrola socijalne poželjnosti bezrazložan napor. Ipak, socijalna poželjnost utiče na visinu skorova, pa samim tim i rang kandidata, tj. na odluku u selekcijskom postupku (npr. Rothstein & Goffin, 2006). Drugi problem odnosi se na relacije sa drugim varijablama, pa tako socijalna poželjnost može da maskira stvarne relacije između varijabli, da se ponaša kao supresor ili moderator.

Detekcija

Detekcija lažiranja ne uključuje nužno samo detekciju socijalne poželjnosti ili simulacije, već i bilo koji nedosledni odgovor koji može upućivati na obmanjivanje od strane ispitanika koji radi test. Detekcija lažiranja može se vršiti na nekoliko metodološki različitih načina. Prvi od njih podrazumeva primenu posebno konstruisanih **skala za procenu socijalne poželjnosti**. Na ovaj način možemo detektovati koji ispitanik je davao socijalno poželjne odgovore. Tada njegov upitnik možemo smatrati

nevalidnim i neverodostojnim, ili skorove tog ispitanika uzeti s rezervom. U Tabeli 25 dat je pregled skala procene socijalne poželjnosti, pri čemu su neke od njih razvijene kao samostalne skale, a neke su deo inventara ličnosti. U kliničkim istraživanjima najčešće se koristi Marlou-Kaunova skala, pa potom Paulusov BIDR (Perinelli & Gremigni, 2016). Za neke od njih su utvrđene norme, dok ih druge nemaju i više se koriste za istraživačke svrhe. Neke od ovih skala su već uvrštene u inventare ličnosti (npr. L skala u okviru EPQ), dok su druge nastale zasebno (npr. BIDR). Primena ovih skala zasnovana je na pretpostavci da ako neko na njima ostvaruje visok skor, tj. ako ima tendenciju davanja iskrivljenih odgovora, onda će se ova tendencija javiti i na drugim upitnicima. Tipična skala socijalne poželjnosti sastoji se od niza tvrdnji koje podrazumevaju ekstremno moralna ponašanja (socijalno poželjno ponašanje), koje je u praksi vrlo retko, npr. "Nikad nisam slagao u životu".

Međutim, stavovi istraživača su vrlo oprečni u pogledu upotrebljivosti skala socijalne poželjnosti za detekciju socijalne poželjnosti, iako većina autora naginje ka tome da se ove skale ipak ne preporučuju (više u Peinelli & Gremigni, 2016). Pored toga, istraživanja pokazuju da skale socijalne poželjnosti uopšte ne obuhvataju socijalnu poželjnost u nekim oblastima primene (npr. Griffith & Peterson, 2008).

Tabela 25

Pregled skala disimulacije i simulacije

Originalni naziv skale	Prevod na srpski	Autor(i)	Broj ajtema
Edwards Social Desirability Scale	Edvardsova skala socijalne poželjnosti	Edwards, 1957	39
Marlowe-Crowne Social Desirability Scale – MC-SDS ili MC	Marlou-Kaunova skala socijalne poželjnosti	Crowne & Marlowe, 1960	33
Multidimensional Social Desirability Inventory	Multidimenzionalni inventar socijalne poželjnosti	Jacobson, Kellogg, Cauce, & Slavin, 1977	68
Jackson Scale	Džeksonova skala	Jackson, 1984	20
Balanced Inventory of Desirable Responding – BIDR	Uravnoteženi inventar socijalno poželjnog odgovaranja	Paulhus, 1984, 1991	40
Comprehensive Inventory od Desirable Responding – CIDR	Sveobuhvatni inventar socijalno poželjnog odgovaranja	Paulhus, 2006	80
Lay – L, Infrequency – F, Defansiveness – K (MMPI-2)	Skala laži – L, Skala retkih odgovora – F, Skala odbrambenih mehanizama – K	Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989	15, 64, 30
Well-being, Good Impression, Communality (CPI260)	Blagostanje, Dobra impresija, Zajedništvo	Gough & Bradley, 2005	20, 27, 22
Lay – L (EPQ-R)	Skala laži	Eysenck & Eysenck, 1975	21
Unlikely Virtues (MPQ)	Malo verovatne vrline	Tellegen, 1982	14
Negative Impression, Positive Impression (PAI)	Negativna imparesija, Pozitivna impresija	Morey, 1991	9, 9
Impression management (16PF)	Upravljanje impresijom	Cattell, Cattell, & Cattell, 1993	12

Kako je osnovna kritika skala socijalne poželjnosti da skor na njoj može odražavati iskreno mišljenje osobe o sebi, te da je teško razlučiti koji su odgovori pod uticajem lažiranja, a koji ne (Paulhus, 2002), predloženi su drugi načini detekcije lažiranja. Jedan od njih podrazumeva isto konstrukciju posebnih ajtema kojima se meri **precenjivanje znanja** (eng. *overclaiming*). Ovi ajtemi mere tendenciju da se tvrdi znanje u vezi sa nekim sadržajem koji realno *ne* postoji. Termin originalno potiče od Filipsa i Klensija (Phillips & Clancy, 1972) koji su prvi upotrebili ovu meru prilikom anketiranja potrošača. Oni su pitali potrošače da procene koliko su im poznati određeni proizvodi, međutim, nijedan od tih proizvoda zapravo nije postojao. Svaka procena da je ispitanik upoznat s takvim proizvodom sugeriše na precenjivanje znanja, i to namerno, tj. sigurni smo da ispitanik laže kada potvrdno odgovara na ovakve stavke. Ova tendencija poistovećuje se sa samoosnaživanjem (eng. *self-enhacement*). Samoosnaživanje predstavlja motivaciju osobe da se orijentiše na svoje pozitivne strane, da održava ili uvećava pozitivno viđenje sebe, više nego što bi se to moglo zaključiti na osnovu objektivnih pokazatelja, kao što su skorovi na testovima, mišljenje drugih i sl. (Sedikides & Alicke, 2018). Ovaj motiv je posebno prisutan u situacijama koje se percipiraju kao ugrožavajuće po samopoštovanje. Prema teorijskim pretpostavkama, precenjivanje znanja predstavlja meru samoosnaživanja delotvornosti u domenu egoističkih pristrasnosti, jer se odnosi na precenjivanje znanja, kompetencija, sposobnosti itd. (Paulhus, Harms, Bruce, & Lysy, 2003).

Najpoznatiju meru precenjivanja znanja ponudio je Paulus. Originalno, Upitnik precenjenog znanja (Over-Claiming Questionnaire – OCQ: Paulhus & Bruce, 1990) sadržao je 150 ajtema koji predstavljaju pojmove raspoređene u deset kategorija: istorijske ličnosti i događaji,

umetnost, jezik, knjige i poeme, autori i likovi iz književnih dela, društvene nauke i pravo, fizičke nauke, biološke nauke, popularna kultura i trenutni potrošački proizvodi. Zadatak ispitanika je bio da na sedmostepenoj skali proceni koliko im je poznat svaki termin, od 0 = *nikad nisam čuo* do 6 = *veoma dobro znam*. Od 150 pojmova, 30 je bilo lažnih (nepostojećih), i to u okviru svake kategorije pojmova (npr. megafrenija). Kasnije je predložena i kraća skala od 80 ajtema od kojih je 16 lažno (Paulhus & Dubois, 2014). Kako ajtemi skale precenjenog znanja liče na ajteme iz testova opšte kulture i informisanosti, iako postoje neka opšta mesta znanja koja važe za sve kulture i koja su vanvremenska, preporučuje se da se u svakoj kulturi konsturišu posebne skale koje će sadržati ajteme specifično vezane za određenu kulturu u istoriju date kulture. Kod nas postoji preliminarna verzija skale precenjivanja znanja koja obuhvata 150 ajtema (30 izmišljenih) raspoređenih u sličnih 10 kategorija znanja, ali u kojima se nalaze npr. sprski autori, pisci itd. (Kašiković, Jerotijević, Novaković, Mijatović i Žeželj, 2013).

Kao referentni okvir za skorovanje predložena je teorija detekcije signala, a njene osnovne postavke izložene su prilikom objašnjenja metoda za odabir kriterijuma (npr. metod baziran na kontrastiranju grupa, str. 162). Postoje dve mere koje se mogu dobiti na osnovu skora na OCQ upitniku. Prva je mera tačnosti znanja (eng. *knowledge accuracy*), indeks tačnosti ili OCQ tačnost, koja se može izračunati kao:

$$\text{OCQ tačnost} = p\text{H} - p\text{FA}$$

gde je *p*H proporcija biranja ajtema koji se odnosi na tačni pojam (H od *hits*), a *p*FA je proporcija biranja ajtema koji sadrži lažni pojam (FA je od *false alarm*). Alternativno, može se izračunati razlika između standardizovanih ovih vrednosti (označava se kao *d'*):

$$d' = zp\text{H} - z\text{p}\text{FA}.$$

Osoba koja zaista zna nije osoba koja pokazuje familijarnost sa većim brojem pojmova, već ona koja pokazuje najveću sposobnost razlikovanja stvarnih od lažnih pojmova.

Odgovaračka pristrasnost koja se odnosi na slaganje predstavlja tendenciju osobe da češće bira odgovor "da, poznat mi je ovaj pojam" u odnosu na odgovor "ne, nepoznat mi je ovaj pojam". Kako ova pristrasnost utiče na odgovaranje na sve pojmove, i tačne i lažne, predložena je mera odgovaračke pristrasnosti, indeks pristrasnosti ili OCQ pristrasnost. Ova mera predstavlja sumu proporcije biranja ajtema koji sadrže stvarne pojmove (*p*H) i proporcije biranja ajtema koji sadrže lažne pojmove (*p*FA), podeljenu sa dva:

$$\text{OCQ pristrasnost} = (p\text{H} + p\text{FA}) / 2.$$

Alternativno, proporcije se mogu zameniti njihovim standardizovanim vrednostima kada dobijemo tzv. lokaciju kriterijuma (označava se kao *c*):

$$c = (zp\text{H} + z\text{p}\text{FA}) / 2.$$

Dakle, OCQ pristrasnost meri jačinu procene familijarnosti kod ispitanika.

Kako bi ovi indeksi mogli da se izračunaju, potrebno je dihotomizirati odgovore na pojedinačnim ajtemima, budući da oni u originalnom obliku sadrže šestostepenu skalu Likertovog tipa za odgovaranje. Paulus i saradnici su predložili više pragova, tačnije pet, za svaki prelaz s jedne na narednu kategoriju odgovora (Paulhus et al., 2003). Tako, svi odgovori 0 bi bili u jednoj kategoriji, a odgovori od 1 do 6 u drugoj kategoriji, potom u sledećoj, posebnoj dihotomizaciji bi se napravile nove varijable u kojoj bi svi odgovori 0 i 1 bili u jednoj kategoriji, a odgovori od 2 do 6 u drugoj, i tako redom za sve ajteme. Ukupno, za svaki ajtem bi postojalo 6 novih varijabli s navedenim načinima dihotomizacije i 6 indeksa bi se izračunalo, a konačna mera bi bila prosek tih 6 indeksa.

U prethodnom istraživanjuje dobijeno da je korelacija ova dva indeksa niska (,21) i da oni imaju različite korelate: OCQ tačnost je povezana sa inteligencijom, a OCQ pristrasnost sa narcizmom i samozavaravanjem, dok nijedna od OCQ mera nije povezana s upravljanjem utiskom (Paulhus et al., 2003), čime je potvrđeno da je precenjivanje znanja povezano s nesvesnim pristrasnostima. Iako je korelacija između njih niska, svakako se preporučuje da se skor na OCQ pristrasnosti parcijalizuje u odnosu na OCQ tačnost, kako bi se dobila "čistija" mera samoosnaživanja, koja ne uključuje i varijansu stvarno tačnih odgovora. Na ovaj način bi parcijalizovana mera predstavljala *uverenje* ispitanika da poseduje bogato i raznovrsno zanje. U istom istraživanju je pokazano da OCQ pristrasnost varira u zavisnosti od uslova, pa ukoliko se u uputstvu naglasi da postoje izmišljeni pojmovi, smanjuje se pristranost. Iako u neutralnim uslovima, tj. uslovu iskrenog

odgovaranja, OCQ pristrasnost nije povezana sa skalom upravljanja utiskom, u uslovu u kojem je ispitanicima data instrukcija da se prikažu u što boljem svetlu, OCQ pristrasnost se povećava i ponaša kao skala upravljanja utiskom. Međutim, u drugim istraživanjima je dobijena značajna veza između precenjivanja znanja i upravljanja utiskom i u neutralnim uslovima, čime je dovedeno u pitanje da li je precenjivanje znanja svesno ili nesvesno (npr. Bensch, Paulhus, Stankov, & Ziegler, 2017; Bing, Kluemper, Davison, Taylor, & Novicevic, 2011; Kam, Risavy, & Perunovic, 2015; Tonković, Galić, & Jerneić, 2011). Kako preterano znanje odgovara egoističkim pristrasnostima, odnosno upravljanju delotvornošću, ne iznenađuje to što korelira sa osobinama ličnosti koje su povezane sa ovom vrstom pristrasnosti, tj. sa ekstraverzijom i otvorenošću za iskustva iz modela Velikih pet (Bing et al., 2011). Neka istraživanja, pak, pokazuju da preterano znanje nije značajno povezano sa osobinama ličnosti (Bensch et al., 2017). Ipak, korelacije koje preterano znanje ostvaruje sa osobinama ličnosti, ako su i značajne, upadljivo su niže u odnosu na one koje ostvaruju skale socijalne poželjnosti te se može zaključiti da preterano znanje ne obuhvata supstantivnu varijansu stabilnih karakteristika. Kako prema teorijskim pretpostavkama precenjivanje znanja treba da spada u domen egoističkih pristrasnosti, pretpostavlja se da će ono ostvariti više korelacije sa samozavaravanjem u odnosu na upravljanje utiskom. Dok je u nekim istraživanjima to potvrđeno (Paulhus et al., 2003), druga pokazuju da precenjivanje znanja značajno korelira sa oba aspekta socijalne poželjnosti, pri čemu čak nešto više sa upravljanjem utiskom (Bensch et al., 2017; Bing et al., 2011). Bez obzira na korelacije koje ostvaruje, u zajedničkoj faktorskoj analizi se precenjivanje znanja i socijalna poželjnost jasno izdvajaju kao zasebni faktori (Bing et al.,

2011), ili precenjivanje znanja ne ostvaruje značajno opterećenje na faktorima socijalne poželjnosti (Bensch et al., 2017), što ide u prilog njihovoj distinktivnosti.

Sledeći način detekcije lažiranja je preko **lažnih ajtema** (eng. *bogus items*). Lažni ajtemi, kao i ajtemi za precenjeno znanje, sadrže nepostojeće objekte, pojmove, sposobnosti koji su uže definisani i konstruišu se za svaki domen ili posao posebno (Anderson, Warner, & Spencer, 1984). Na primer, ukoliko želimo da proverimo da li neko zaista poznaje posao programera, pitaćemo ga da li ima veštinu "transkribovanja memo-podataka", a ukoliko želimo da proverimo da li poznaje posao ekonomskog analitičara, pitaćemo da ga li zna "izometrijsku analizu". Ukoliko kandidat odgovori potvrdno, znamo da laže jer takve veštine i tehnike ne postoje. Za razliku od lažnih stavki, stavke u upitniku precenjenog znanja odnose se na opšte znanje, slično domenu opšte kulture i informisanosti. Upitnik precenjenog znanja se stoga može zadati većem broju kandidata, tj. kandidatima za različite vrste poslova, dok se lažni ajtemi mogu zadati samo određenim kandidatima, za konkretan posao. Pošto je teško sastaviti lažne ajteme za svaku vrstu posla, njih obično bude manje u okviru seta testova. S druge strane, veće su mogućnosti za sastavljanje ajtema za procenu precenjenog znanja jer je konstrukt šire definisan. U jednom istraživanju je pokazano da lažni ajtemi, ajtemi precenjenog znanja i ajtemi skale socijalne poželjnosti tvore tri zasebna faktora, pri čemu nema značajnih sekundarnih opterećenja ajtema, tako da se može napraviti jasna razlika između sadržaja ekstrahovanih faktora (Bing et al., 2011). Drugim rečima, preterano znanje i lažni ajtemi predstavljaju zasebne konstrukte. Pored toga, i lažni ajtemi, kao i mera preteranog znanja,

pokazuju upadljivo niže korelacije sa osobinama ličnosti, za razliku od skala socijalne poželjnosti (Bing et al., 2011).

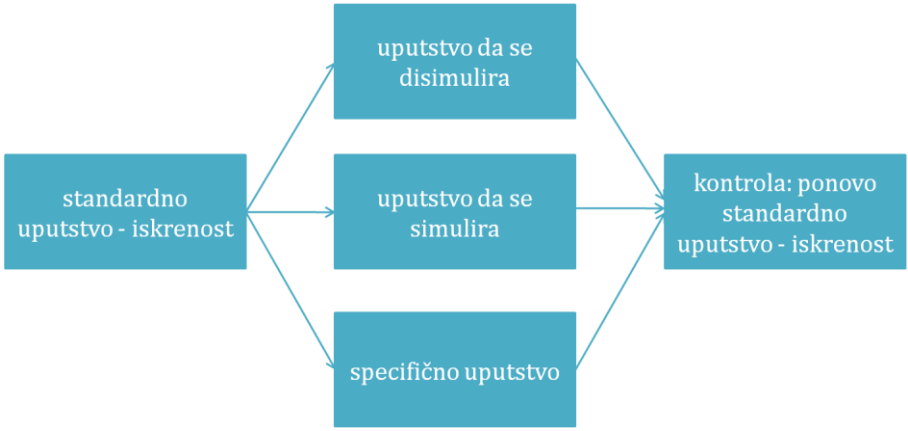
Ipak, dešava se da ispitanici u velikoj meri (35–45% njih u ispitanim uzorcima) potvrdno odgovaraju bar na jednu lažnu stavku, čak i u slučaju iskrenog odgovaranja (20% ispitanika, više u Burns & Christiansen, 2011). Ovakvi rezultati navode na zaključak da odgovaranje na lažene ajteme nije samo pod uticajem lažiranja, već se može javiti i usled generalizacije svog znanja i iskustva (npr. Okanović, 2013). Zbog toga je potrebno da se tendencija lažiranja meri preko više lažnih ajtema. Iako se čini da su ajtemi precenjenog znanja više imuni na generalizaciju prethodnog radnog iskustva, i jedni i drugi imaju svoje polje primene i implikacije (Burns & Christiansen, 2011).

Lažni ajtemi mogu se koristiti i za detekciju lažiranja psihopatologije. U ovom slučaju imamo dva pristupa prilikom konstrukcije ajtema. Prvi se odnosi na sadržaj uverenja, ponašanja, misli i osećanja koja ne mogu biti istinita jer ne postoje (npr. "Bog mi je pokazao istinu o duguljastim trapezoidima"). Drugi se odnosi na sadržaj ajtema koji može ukazivati na patologiju, ali koji odražava suviše specifičan način manifestovanja patologije koji nije karakterističan za poremećaje ("Vidim obojene trouglove", više u Rogers, Bagby, & Gillis, 1992). Ispitanici koji većinom potvrdno odgovaraju na ove ajteme, najverovatnije lažiraju, tj. simuliraju.

Kao još jedna opcija detekcije lažiranja, predložene su **skale retkih odgovora**. Pretpostavka u vezi s ovim skalama je da one sadrže tvrdnje na koje će veoma retko ko iz opšte populacije odgovoriti potvrdno, npr. "Valjalo bi odbaciti sve zakone" ili "Nikada ne izbegavam svoje dužnosti". Za razliku od lažnih ajtema, tvrdnje u skalama retkih odgovora su istinite, ali mogu biti istinite za veoma mali broj ljudi. Na

ovaj način se može detektovati disimulacija, simulacija, ali i nepažljivo odgovaranje. Problemi s ovim skalama su isti kao i problemi sa skalama socijalne poželjnosti – slaganje s ovim tvrdnjama stvarno može odražavati iskreni odgovor i nejasno je kako odrediti granični skor. Na primer, ako se granični skor postavi tako da smanji broj falš pozitiva (pogrešno označenih ispitanika kao onih koji su skloni lažiranju, a koji su zapravo odgovarali iskreno), smanjiće se i broj onih koji su skloni lažiranju. Jedan od načina za rešavanje ovog problema je da se ajtemi retkih odgovora zadaju, ali da njihov sadržaj ne bude vezan za sadržaj koji nam je ključan prilikom cilja testiranja (Dilchert & Ones, 2012). Na primer, ako nam je u cilju da selektujemo kandidate u odnosu na savesnost, skale retkih tvrdnji bi trebalo da sadrže ajteme koji se odnose, na primer na otvorenost, ili koji se uopšte ne odnose na ličnost već na stavove koji ne odražavaju odnos prema radu i slično.

Pored opisanih načina, postoji i način koji ne podrazumeva uvrštavanje dodatnih skala, već je dovoljan samo test koji je od interesa. Uobičajeni nacrt u ovom slučaju sastoji se u tome da se ispitanicima jedan isti test zada pod nekoliko instrukcija da bi se proverilo na koji način skorovi variraju u zavisnosti od instrukcije kojom se, u stvari, menja motivacija ispitanika. Kako se jednoj ili različitim grupama ispitanika zadaju različite instrukcije, ovakav metodološki pristup podseća na igru igranja uloga. Pošto se u okviru ovog načina porede skorovi na istom testu, ovaj pristup je nazvan pristupom **diferencijalnih skorova**. Postoji nekoliko varijanti ovog nacrta, a najčešće se koriste dva. Prvi nacrt je nacrt s **jednom grupom** ispitanika kojoj se zada isti test uz dve ili više različitih instrukcija (Slika 30).



Slika 30. Nacrt s jednom grupom za računanje diferencijanih skorova.

Najpre se ispitanicima zada test uz standardnu instrukciju, da na pitanja odgovore iskreno. Nakon određenog vremenskog perioda (ne manje od 2 nedelje) im se ponovo zada isti test ali uz drugu instrukciju. Uobičajeno se uzorak ispitanika nasumice podeli na dva poduzorka, te se jednom poduzorku zada instrukcija da se prikažu u boljem socijalnom svetlu, a drugom da se prikažu u lošem svetlu.

Primer instrukcije za prikazivanje u socijalno poželjnom svetlu:
Zamislite da konkurišete za posao koji vas jako privlači i koji biste jako želeli da dobijete i da vam samo od rezultata na ovim testovima zavisi da li ćete ga dobiti ili ne. Podesite svoje rezultate tako da ga sigurno dobijete, dakle predstavite se u najboljem mogućem svetlu.

Primer instrukcije za prikazivanje u socijalno nepoželjnom svetlu:
Zamislite da ste nezaposleni i da vam je agencija za zapošljavanje našla neki posao koji vas nikako ne privlači i koji ne želite da dobijete. Pošto bi otvoreno odbijanje tog posla dovelo do ukidanja socijalne pomoći (vašeg jedinog izvora prihoda) vaš jedini izlaz je da se na testu ličnosti, od čijeg rezultata isključivo zavisi da li ćete dobiti taj posao dobiti ili ne, prikažete u što je moguće lošijem svetlu. Dakle, namestite svoje odgovore tako da se prikažete u što je moguće lošijem svetlu (preuzeto iz Mentus, 2014).

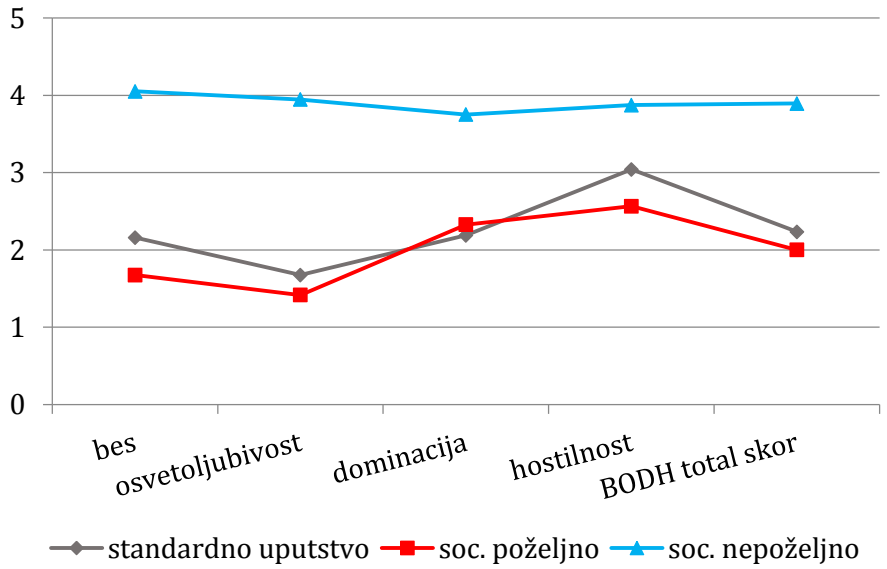
Umesto ovih instrukcija može se zadati i neka specifičnija instrukcija, npr. da ispitanici odgovaraju kao da su pacijenti s paranoidnim poremećajem. U tom slučaju dobijamo procene ispitanika u vezi s tim kako oni shvataju paranoidni poremećaj i kako misle da se on manifestuje. Nakon određenog vremenskog perioda, ispitanicima se ponovo zada isti test, s opozitnom instrukcijom – oni koji su imali zadatak da odgovaraju tako da se prikažu u što boljem socijalnom svetlu, sada bi imali zadatak da se prikažu u socijalno nepoželjnom svetlu. Na kraju, nakon određenog vremenskog perioda, može se ponovo zadati standardna instrukcija, zarad kontrole.

Analizom razlika između skorova pod različitim instrukcijama možemo detektovati koji su dobri ajtemi za detekciju lažiranja. Dobri detektori lažiranja bi bili ajtemi čiji se skorovi značajno razlikuju u zavisnosti od instrukcije. Na primer, ukoliko na neki ajtem pozitivno odgovori samo 15% ispitanika pod strandardnom instrukcijom, ali 60% pod instrukcijom da se prikaže u što boljem svetlu, onda je taj ajtem dobar detektor socijalne poželjnosti. Analiza se može raditi i na nivou skorova na skali, a ne samo ajtemskih skorova. Na ovaj način može se utvrditi koliko je neka skala generalno podložna lažiranju. Kako bismo dobili uvid u individualne razlike u lažiranju, treba uporediti rang ispitanika pod različitim instrukcijama. Ukoliko postoje individualne razlike u lažiranju, onda će se rang ispitanika menjati, npr. da neko ima 50. rang pod standardnom instrukcijom, ali 87. pod instrukcijom da se prikaže u socijalno poželjnom svetlu.

U jednom preliminarnom istraživanju na Upitniku agresivnosti BODH koji meri četiri facete agresivnosti, primenjen je nacrt s jednom grupom u cilju ispitivanja osetljivosti skorova na različite vrste instrukcija – prikazivanje u socijalno poželjnom svetlu (indukovano je

upravljanje zajedništvom preko zamišljanja prijave na posao za pomagačku profesiju) i socijalno nepoželjnom svetlu (da osoba zamisli da je osuđenik za krivično delo koje uključuje nasilje, videti Dinić & Radević, 2019). Budući da je agresivnost socijalno nepoželjna osobina, očekivalo se da će skorovi na ovom upitniku biti značajno niži kod prikazivanja osoba u socijalno poželjnom svetlu, potom umereni u situaciji standardnog uputstva (iskrenosti), i najviši u slučaju prikazivanja u socijalno nepoželjnom svetlu. Rezultati su to i potvrdili (Slika 31), te je dobijeno da se skorovi značajno razlikuju na svim dimenzijama ovog upitnika, pri čemu je ostvarena velika veličina efekta (η_p[231] se kreće od 0,67 u slučaju hostilnosti do 0,88 u slučaju osvetoljubivosti, mada je primetno da su veličine efekta upadljivo niže za razlike između standardnog uputstva i prikazivanja u socijalno poželjnom svetlu). Jedina situacija u kojoj nema značajnih razlika je u skorovima na dominaciji u slučaju standardnog uputstva i prikazivanja u socijalno poželjnom svetlu. Takođe, i na ukupnom skoru su dobijene značajne razlike, u očekivanom smeru. Pored toga, korelacija između ranga ispitanika na ukupnom skoru pod uslovima različitih instrukcija je bila niska (ρ = ,24 između standardne instrukcije i prikazivanja u socijalno poželjnom svetlu; ρ = –,11 između standardne instrukcije i prikazivanja u socijalno nepoželjnom svetlu; i ρ = –,19 između prikazivanja u socijalno poželjnom i nepoželjnom svetlu), što ukazuje na poremećaj redosleda ranga ispitanika u zavisnosti od instrukcije. Drugim rečima, upravljanje utiskom utiče na rang ispitanika.

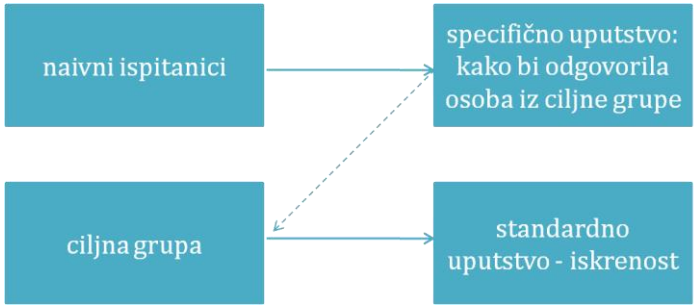
[231] Veličina efekta izražena kao η_p^2 se tumači na sledeći način: vrednosti ,01 ukazuju na malu veličinu efekta, ,06 na srednju i ,14 na veliku (Cohen, 1988).



Slika 31. Promene u skorovima na Upitniku agresivnosti BODH u zavisnosti od različitih instrukcija (preuzeto iz Dinić & Radević, 2019).

Drugi nacrt je nacrt s **različitim grupama** ispitanika kojima se isti test zadaje pod različitim instrukcijama. Bitno je da od dve grupe ispitanika jednu grupu čine naivni ispitanici iz opšte populacije, a drugu ciljnu grupu, npr. osobe s paranoidnim poremećajem (Slika 32). U navedenom primeru cilj je da izdvojimo stavke kojima možemo detektovati lažiranje paranoidnog poremećaja. Grupi iz opšte populacije se test zadaje uz specifično uputstvo, da odgovaraju kao osobe ciljne grupe, tj. kao tipična osoba s paranoidnim poremećajem. Na ovaj način dobijamo laičke predstave i shvatanja o paranoidnom poremećaju i manifestaciji istog. S druge strane, osobama s paranoidnim poremećajem zadamo isti test pod standardnom instrukcijom, dakle, da odgovaraju iskreno. Ukoliko je neki ajtem visoko biran od strane naivnih ispitanika, a nisko od strane ciljne grupe, taj ajtem je dobar detektor

lažiranja. Drugim rečima, ajtemi koji dobro razlikuju grupe ispitanika su dobri ajtemi za detekciju lažiranja. U slučaju lažiranja patologije, pokazano je da se takvi ajtemi obično odnose na halucinacije koje su mnogo manje prisutne u kliničkoj populaciji nego što to naivni ispitanici iz opšte populacije misle. Takođe, ukoliko je neki ajtem nisko biran od strane naivnih ispitanika, a visoko od strane ciljne grupe, i on je dobar detektor lažiranja. Takvi su "suptilniji" ajtemi psihopatologije, dakle koji su prediktivno validni, ali imaju nisku prividnu ili fejs validnost. Uobičajeno su to ajtemi koji se odnose na zaravljeni afekat, što je svojstveno mnogim kliničkim poremećajima, ali očigledno da naivni ispitanici ne percipiraju ovu karakteristiku kao karakteristiku psihopatološkog poremećaja, bar ne u tolikom stepenu u kojem se ona zaista javlja.



Slika 32. Nacrt s različitim grupama za računanje diferencijalnih skorova.

Nov pristup detekciji lažiranja pružaju kognitivni modeli u kojima se lažiranje detektuje preko vremena potrebnog za davanje odgovora – **vremena latencije**. U svim ovim modelima postoje kognitivne šeme koje predstavljaju organizaciju iskustava, osećanja, misli i ponašanja u integralnu celinu. Šeme su kognitivna gerenalizacija

koja vodi obradu informacija vezanih za pojedinca, proizašlu iz njegovog iskustva. Šeme pospešuju i ubrzavaju obradu karakteristika koje su s njom povezane, a usporavaju obradu karakteristika koje su neusaglašene sa šemom. Tako, osobe koje su sklone hostilnosti imaju razvijenu šemu hostilnosti po kojoj neutralne i nejasne signale iz okruženja tumače kao hostilne, jer se to uklapa u njihovu šemu, te na osnovu šeme generišu i hositlne reakcije na te signale. Pritom, generisanje hostilne reakcije u nekoj situaciji je brže nego generisanje prijatne reakcije, jer je hostilna reakcija u skladu sa šemom. U slučaju lažiranja, polazi se od toga da ispitanici brže odgovaraju ako su iskreni, tj. ako je odgovor koji bi dali u skladu s njihovom šemom, a sporije ukoliko lažu. Ipak, postoje modeli koji pretpostavljaju suprotan obrazac, da ispitanici koji imaju tendenciju da lažiraju daju brže odgovore na ajteme koji su socijalno poželjni jer upravljanje utiskom zahteva samo semantičku evaluaciju koja je manje kompleksna od evaluacije sopstvenog selfa.

Suprotstavljene modele pokušali su da pomire Holden i saradnici (Holden, Kroner, Fakken, & Popham, 1992) i predlože Model usvojene šeme (eng. *Adopted Schema Model*). Prema ovom modelu, ispitanici prilikom odgovaranja porede tvrdnje u upitniku sa svojom kognitivnom šemom, i ukoliko su iskreni, odgovaraju u skladu sa njihovom šemom. S druge strane, ispitanici koji su motivisani da iskrive svoje odgovore, aktiviraju šemu iskrivljavanja koja deluje jednako kao i šema o sebi kod iskrenih ispitanika. Ukoliko je aktivirana šema socijalne poželjnosti, nepoželjan odgovor, koji nije u skladu sa šemom, iziskivaće duže vreme obrade, nego odgovor koji upućuje na socijalno poželjno ponašanje. Kod iskrenih ispitanika ne bi trebalo da postoji razlika u vremenu odgovaranja na ajteme socijalne poželjnosti, jer su svi odgovori koje su

takvi ispitanici dali u skladu sa šemom koju imaju o sebi. Primena ovog modela u detekciji lažiranja na upitniku ličnosti pokazuje da se ovaj efekat dobija samo za ekstraverziju. Naime, grupi ispitanika koja je dobila instrukciju da se prikaže u socijalno poželjnom svetlu trebalo je duže vremena da prihvati odgovore koji se ne odnose na ekstraverziju, koja se smatra socijalno poželjnom osobinom (Parmač, Galić i Jerneić, 2009). U istom istraživaju je dobijeno da je ispitanicima koji su trebali da se prikažu u socijalno poželjnom svetlu na skali upravljanja utiskom trebalo duže vremena da prihvate odgovore koji nisu socijalno poželjni, u odnosu na socijano poželjne odgovore, dok značajnih razlika nije bilo u odgovaranju na skali samozavaravanja. Dalja istraživanja su pokazala da postoji niz faktora koji mogu uticati na vreme latencije odgovora, pri čemu neki proizilaze iz metodoloških nedostataka (greška uzorkovanja, nepouzdane mere, brzina čitanja), ali i drugih faktora poput upoznatosti sa poslom za koji se aplicira (više u Robie et al., 2000). S obzirom na to, iako vreme latencije pruža plodno tlo za detekciju lažiranja, čini se da ima više otvorenih pitanja nego datih odgovora kada bi se odluka bazirala na ovom pristupu.

Rešenja

Postoji nekoliko načina kako se socijalna poželjnost može eliminisati, smanjiti ili držati pod kontrolom, i oni se dele na postupke koji nastaju u toku konstrukcije ili zadavanja testa – *a priori*, i na postupke nakon prikupljanja podataka – *post hoc*.

Većina autora se slaže da je najbolji način prevencije socijalne poželjnosti *a priori*, u toku **konstrukcije upitnika**. S obzirom na to, predloženo je nekoliko načina kako se to može ostvariti. Najpre, u toku same konstrukcije upitnika može se **detektovati koji su ajtemi**

podložni socijalnoj poželjnosti, te se revidirati u konačnoj verziji upitnika. Detekcija takvih ajtema se uobičajeno može izvršiti na dva načina – preko procena eksperata o tome koliko je ajtem podložan davanju socijalno poželjnih odgovora i preko utvrđivanja korelacija između pojedinačnih ajtema i mera socijalne poželjnosti (neke od skala socijalne poželjnosti date su u Tabeli 25). Međutim, nema puno empirijske podrške ovim načinima eliminacije socijalne poželjnosti, čak neki nalazi upućuju na zaključak da je odnos ajtemskih skorova i skorova na skalama socijalne poželjnosti nelinearan (Kuncel & Tellegen, 2009). Ovaj rezultat dovodi u pitanje zaključivanje o pojedinačnom ajtemu, pa i skorovima upitnika, kao podložnim socijalnoj poželjnosti na osnovu njegove proste, linearne korelacije sa skalom socijalne poželjnosti. Stoga se preporučuje da se najpre ispita linearnost odnosa između mere socijalne poželjnosti i ajtema ili ukupnih skorova na upitniku, pa tek onda primeni odgovarajući koeficijent korelacije.

Drugi način u okviru *a prirori* pristupa odnosi se na to da se ajtemi konstruišu tako da se **zamaskira predmet merenja**, tj. da se smanji prividna ili fejs validnost (eng. *face validity*). To znači da treba konstruisati tzv. "suptilne" ajteme čiji sadržaj ne otkriva jasno na koji se konsturkt odnosi, odnosno takve ajteme u kojima će se **neutralizovati** sadržaj socijalne poželjnosti koliko god je to moguće. Na primer, ukoliko želimo da ispitamo kakav stav ima neko prema Romima (pozitivan ili negativan), suviše bi bilo očigledno ako ga pitamo koliko se slaže sa tvrdnjom "Romi su manje inteligentni od nas". Ukoliko imamo osobu koja ima negativan stav prema Romima, a želi da se prikaže u socijalno poželjnom svetlu, najverovatnije će odabrati odgovor 1 = *uopšte se ne slažem*, eventualno odgovor 2 = *delimično se ne slažem*, jer je ajtem prilično "težak", tj. pretpostavlja se da bi distribucija odgovora bila

prilično pozitivno zakošena, ka nižim skorovima. Međutim, ako tvrdnju "ublažimo" i formulišemo kao "Smetalo bi mi kada bi moj član porodice doveo Roma na večeru", veća je verovatnoća da će varijansa na ovom ajtemu biti veća, nego u prethodnom slučaju. Takođe, pošto ajtem nije toliko "težak"i fokus je prebačen, na neki način, na drugu osobu ili člana porodice, može se očekivati da će oni koji imaju negativan stav i ujedno tendenciju ka davanju socijalno poželjnih odgovora, pre birati odgovore 3 = *podjednako se i slažem i ne slažem*, 4 = *delimično se slažem*, ili čak i 5 = *u potpunosti se slažem*. Isto tako, ajtem za procenu druželjubivosti "Neprijateljski sam raspoložen prema drugima" može se preformulisati u "Ponekad sam manje prijatan nego drugi ljudi". Iako predloženi ajtem nije savršen, ispitanici koji zaista nisu druželjubivi, a imaju potrebu za prikazivanjem u socijalno poželjnom svetlu, pre će priznati svoju nedruželjubivost kod ovako formulisanog ajtema.

Maskiranje predmeta merenja pokazalo se kao dobra taktika kod testova koji pripadaju testovima niskog uloga. U jednom istraživanju su konstruisani novi ajtemi za operacionalizaciju modela ličnosti Velikih pet, po uzoru na Goldbergov IPIP-100 (Bäckstörm & Björklund, 2013). Konstrukcija novih ajtema je obuhvatala neutralisanje predmeta merenja u odnosu na socijalnu poželjnost (npr. umesto ajtema "Volim da pomažem drugima", predložen je ajtem "Imam potrebu da pomažem drugima"). Rezultati pokazuju da je efekat socijalne poželjnosti na ovaj način smanjen, tj. korelacije između skala novokonstruisanog inventara i skala socijalne poželjnosti je niža, a takođe su i niže međusobne korelacije između skala.

Postoji još nekoliko načina za maskiranje predmeta merenja. Jedan od njih je da se na setu upitnika **ne prikaže naziv upitnika**. Na primer, praksa je da se umesto Maslak inventara izgaranja taj upitnik

nazove Upitnik ljudskih resursa, s pretpostavkom da će eliminisati potencijalne pristrasnosti koje reč "izgaranje" može da aktivira.

Potom, mogu se ubaciti **filer ajtemi**. To su ajtemi koji ne mere ono što je predmet merenja, već su ubačeni u test da sadržaj predmeta merenja ne bi bio očigledan. Na primer, ukoliko želimo da ispitamo da li je neko sklon nasilju, bilo bi suviše očigledno ako svi ajtemi skale sadrže indikatore nasilnog ponašanja, te se onda mogu ubaciti ajtemi poput "Imam različite hobije" ili "Nemam puno slobodnog vremena". Iako istraživanja nisu dosledno potvrdila efektivnost upotrebe filer ajtema, ovi ajtemi nemaju efekta na pouzdanost i validnost, ali produžavaju test (više u Domino & Domino, 2006). Ipak, ukoliko se oni koriste preporuka je da se nikada ne stave na sam početak upitnika i da ih bude dovoljno u upitniku. Poznati upitnik koji sadrži filer ajteme je Koeficijent empatije (eng. Empathy Quotient – EQ; Baron-Cohen & Wheelwright, 2004) koji od 60 ajtema sadži 20 filer ajtema.

Postoji još jedan način kako se može zamaskirati predmet merenja, a to je preko pažljivo osmišljenog **redosleda ajtema**. Kako na odgovaranje na ajtem utiče ajtemski kontekst, tj. sadržaj okolnih ajtema, ukoliko u testu imamo poređane ajteme koji se odnose na isti konstrukt, ispitaniku će biti jasnije šta je predmet merenja. Kako bismo to otežali i potencijalno smanjili lažiranje, ajtemi jednog konstrukta ne bi smeli da budu grupisani. U jednom istraživanju su ajtemi NEO-PI-R-a zadati onako kako su originalno raspoređeni u inventaru (svaki konstrukt sadrži ajteme koji su u jednakim intervalima raspoređeni u testu, npr. za skalu neuroticizma ajtemi se nalaze na 1., 6., 11. poziciji itd.) i grupisano (prvih 48 ajtema se odnosi na jedan konstrukt, pa sledećih na drugi itd.) i dati pod iskrenom instrukcijom i pod instrukcijom da se ispitanici prikažu u socijalno pozitivnom svetlu (McFarland, Ryan, & Ellis, 2002).

Rezultati pokazuju da su skorovi viši na savesnosti i emocionalnoj stabilnosti pod instrukcijom davanja socijalno poželjnih odgovora u odnosu na instrukciju davanja iskrenih odgovora, ali da redosled ajtema takođe pravi razliku. Naime, pod instrukcijom davanja socijalno poželjnih odgovora skorovi su bili viši kada su ajtemi bili grupisani, u odnosu na standardni redosled ajtema. Iako je kasnije pokazano da je ovaj efekat dobijen samo za pozitivno formulisane ajteme, opšta preporuka je da se ajtemi istog konstrukta distribuiraju na neki razuman i podjednak intverval, a ne da se grupišu ili čak randomizirano prikazuju u testu (više u Dilchert & Ones, 2012). Ovo je važno još iz jednog razloga. Naime, ako se ajtemi istog konstrukta grupišu u testu, time se povećava pouzdanost skale jer su ajtemi u većoj međusobnoj korelaciji, ali takođe i korelacija između ajtema može biti visoka i može da upućuje na redundatnost ajtema. Tačnije, pouzdanost može biti veštački povećana kada su ajtemi grupisani, te je u tom slučaju preporučljivo proveriti prosečnu korelaciju između ajtema. Ukoliko je ona visoka (preko ,50, više u Clark & Watson, 1995), to je signal da je pouzdanost precenjena.

Sumirano, maskiranje predmeta merenja se može uraditi na nekoliko načina, kao što su neutralisanje socijalne poželjnosti iz sadržaja ajtema, neprikazivanje naziva upitnika, ubacivanje filer ajtema i pažljivo osmišljen redosled ajtema, sa dovoljno razmaka između stavki koje se odnose na isti predmet merenja.

Treći način redukcije socijalne poželjnosti u toku konstrukcije upitnika je korišćenje **ajtema s prisilnim izborom** (eng. *forced-choice items*) u kojima su ponuđeni odgovori podjednako socijalno poželjni. Ovakav način se preporučuje kod testova s visokim ulogom. Najkorišćeniji test koji sadži ovakve ajteme i koji se uveliko koristi u

procesu selekcije kod nas je Plučikov Indeks profila emocija – PIE (Plutchik & Kellerman, 1964). Ovaj test meri određene osnovne osobine ličnosti i konflikt u ličnosti. Ajtemi ovog testa se sastoje od ponuđena dva prideva koja se odnose na opis ličnosti, pri čemu je zadatak ispitanika da odabere onaj pridev u paru koji ga bolje opisuje, npr. da li je više "sramežljiv" ili "potišten". PIE sadrži i skalu pristrasnosti kojom se meri sklonost ka davanju socijalno poželjnih odgovora, odn. biranje socijalno poželjne opcije od zadate dve u paru, što je očito kada je jedna od opcija socijalno nepoželjna, a druga socijalno poželjna osobina. Dilema se, međutim, javlja kada su uparena dva socijalno poželjno odgovora (npr. "društven" i "srdačan") ili dva socijalno nepoželjna odgovora (npr. "nagao" i "sklon prepiranju"). U jednom domaćem istraživanju je pokazano da je test koji sadrži ajteme prisilnog izbora manje podložan lažiranju od testa koji sadrži ajteme Likertovog tipa za merenje sličnih koncepata (Okanović, 2013). Generalni je stav istraživača da ajtemi s prisilnom izborom ne mogu eliminisati socijalnu poželjnost, ali mogu učiniti da se teže lažira.

Postoji nekoliko problema vezanih za ovaj format odgovaranja. Naime, i ajtemi s prisilnim izborom mogu biti podložni davanju socijalno poželjnih odgovora. U testovima su oni konstruisani mahom tako da se odnose na neka opšta načela i shvatanja društva o tome šta je generalno socijalno poželjno, dok u praksi, kada ispitanici lažiraju, oni lažiraju u skladu sa specifičnim zahtevima kao što su karakteristike specifičnog posla. Na primer, ako imamo par atributa odlučan-ljubazan, lako je pretpostaviti šta bi bilo bolje zaokružiti ako se prijavljujemo za posao menadžera. Međutim, šta bismo zaokružili ako se prijavljujemo za posao nastavnika? Drugi problem proističe iz toga što ajtemi s prisilnim izborom podrazumevaju ipsativnu skalu, tj. verovatnoća biranja jedne

alternative je zavisna i obrnuto proporcionalna verovatnoći biranja druge alterntive, što čini da skorovi na skalama budu međusobno negativni. Na kraju, rezultati ukazuju na to da ipsativne skale ne doprinose većoj kriterijumskoj validnosti, te u najboljem slučaju doprinose isto koliko i druge skale (više u Dilchert & Ones, 2012).

Četvrti *a priori* način odnosi se na situaciju kada je upitnik već konstruisan, a davanje socijalno poželjnog odgovaranja može da se spreči prilikom **administracije upitnika**. Ovaj pristup podrazumeva da se ispitanici unapred obeshrabre da daju socijalno poželjne ili nepoželjne odgovore. Paće i Borman (Pace & Borman, 2006) su ponudili pregled različitih **upozorenja,** tj. uputstava:

1. upozorenje da se lažiranje može detektovati;
2. upozorenje da će oni koji lažiraju snositi posledice;
3. upozorenje da se najbolja odluka može doneti samo ako ispitanici odgovaraju iskreno, npr. da će se najbolja radna pozicija naći samo ukoliko ispitanici iskreno odgovaraju na pitanja o svojim karakteristikama;
4. objašnjenje svrhe testiranja i naglašavanje da se samo iskrenim odgovaranjem mogu ostvariti ciljevi ispitivanja, unapređenje organizacije i slično;
5. pozivanje na moralnost.

U jednoj mini meta-analitičkoj studiji pokazano je da upozorenje daje željene rezultate, tj. da iako kod motivisanih ispitanika postoji tendencija davanja socijalno poželjnih odgovora, izraženost socijalno poželjnih karakteristika je manja kada postoji upozorenje o negativnim posledicama u slučaju detekcije lažiranja, kao što je diskvalifikacija (Dwight & Donovan, 2003). Ipak, sistematičnih istraživanja u vezi s efektima različitih upozorenja nema, a rezultati pomenute meta-analize

pokazuju da različita upozorenja imaju različiti efekat na količinu lažiranja. Međutim, bilo kakvo menjanje instrukcije narušava standardizovanost postupka zadavanja testa, te ukoliko ovakva instrukcija već nije predviđena priručnikom i postupkom zadavanja testa, ne može se podrazumevati da je interpretacija skorova ista kao i kada ove instrukcije nema. U ovom slučaju, ispravno bi bilo utvrditi psihometrijske karakteristike, pre svega validnost skorova za upitnik pod datom instrukcijom.

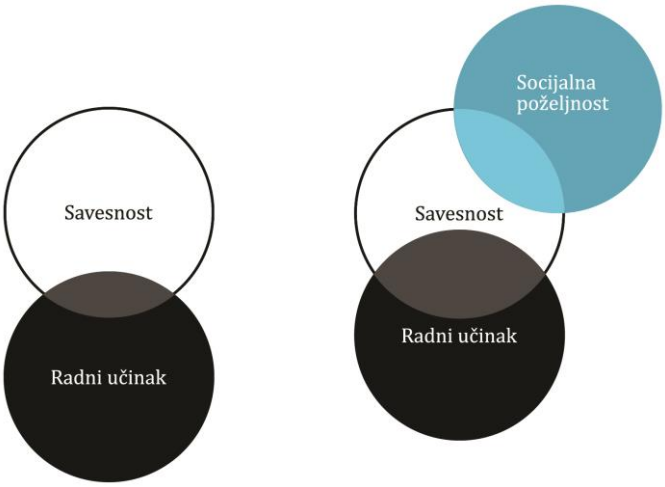
Sledeći načini kontrole socijalne poželjnosti odnose se na *post hoc* detekciju ove pristrasnosti, nakon što su testovi već zadati (više u Rothstein & Goffin, 2006). Uz set testova, zada se i skala socijalne poželjnosti (precenjenog zanja ili retke tvrdnje), te se mogu **ukloniti kandidati** koji imaju skor preko graničnog na skali socijalne poželjnosti. Međutim, ukoliko je varijansa socijalne poželjnosti značajno povezana s varijansom osobina ličnosti na osnovu kojih se vrši selekcija, ovakva praksa može rezultirati eliminacijom dobrih kandidata. Stoga se eliminisanje kandidata na osnovu visokog skora na socijalnoj poželjnosti veoma retko praktikuje i u tim situacijama uvek se savetuje da se takav kandidat pozove na intervju, ukoliko već ima rezultate koji upućuju na to da je dobar kandidat za dati posao. U istraživačkoj praksi, može se proveriti da li se rezultati menjaju kada se analiza radi na celom uzorku i na redukovanom, kada se eliminišu ispitanici koji pokazuju visok skor na socijalnoj poželjnosti. Ukoliko postoji razlika, bolje je zadržati rezultate na redukovanom uzorku, a u postupku objasniti način redukcije uzorka.

Drugi način koji uključuje skale socijalne poželjnosti je **korekcija skorova** na testu, koji je od interesa u odnosu na rezultat na skali socijalne poželjnosti. Uobičajeno se varijansa skorova na datom

upitniku parcijalizuje u odnosu na varijansu koja potiče od skora na skali lažiranja, i sačuva kao rezidualizovani skor. U nekim situacijama su razrađene procedure korekcije skorova preko posebne formule. Ipak, ovaj pristup se nije pokazao kao obećavajući. Rezultati ranijih istraživaja pokazuju da se korekcijom skorova niti popravlja procena ispitanikovih iskrenih odgovora, niti unapređuje kriterijumska validnost testova ličnosti i situaciji selekcije (npr. Ones et al., 1996). Kao što je već pomenuto, socijalna poželjnost je povezana sa osobinama ličnosti i parcijalizacijom skorova socijalne poželjnosti uklanja se i sistematska varijansa osobina ličnosti, te je upitno koja varijansa ostaje nakon ove parcijalizacije, tj. koji je njen izvor – prava varijansa ili varijansa greške merenja, a koja ne potiče od odgovaračkih pristrasnosti. Naime, ukoliko postoji veza između socijalne poželjnosti i osobina ličnosti, uklanjanjem varijanse socijalne poželjnosti vrlo je moguće da se od skorova koji mere osobine ličnosti ostati greška merenja.

Sličan ishod se dobija i kada se skor na socijalnoj poželjnosti tretira kao **supresor** varijabla. Supresor je varijabla koja je značajno povezana sa prediktorom, ali nije sa kriterijumom. Njenim uvođenjem u prediktivni model povećava se prediktivnost, tj. doprinos prediktora, jer supresor varijabla uklanja varijansu greške iz prediktora. Drugim rečima, supresor demaskira vezu između prediktora i kriterijuma. Na primer, u cilju ispitivanja prediktivnosti radnog učinka na osnovu savesnosti, dobili smo da je ta predikcija slaba (Slika 33, levo). Pritom, sumnjamo da kandidati daju socijalno poželjne odgovore na dimenziji savesnosti i želimo da proverimo da li je zaista socijalna poželjnost slab prediktor radne uspešnosti, ili su skorovi kontaminirani zbog davanja socijalno poželjnih odgovora. U ovom slučaju, uz savesnost kandidatima dajemo i meru socijalne poželjnosti i nju uvodimo u prediktivni model.

Kako socijalna poželjnost nije povezana sa radnim učinkom, a jeste sa savesnošću, ispunjava uslov da bude supresor, te se ispostavlja da je doprinos savesnosti u ovom modelu veći, nego kada nema socijalne poželjnosti (Slika 33, desno). Doprinos savesnosti je veći jer socijalna požejnost iz savesnosti eliminiše ono što se smatra varijansom greške.



Slika 33. Socijalna poželjnost kao supresor varijabla.

Nažalost, tretiranje socijalne poželjnosti kao supresora nije se pokazalo efektivno, tj. uvođenje socijalne poželjnosti ne doprinosi predikciji radnog učinka na osnovu osobina ličnosti (Li & Bagger, 2006). Međutim, neka istraživanja ukazuju na to da se skala precenjenih znanja ponaša kao supresor, za razliku od skala socijalne poželjnosti, kada je u pitaju kriterijum koji se povezan sa znanjem, kao što je školsko postignuće (Bing et al., 2011).

S druge strane, na isti način možemo proveriti prediktivnu vrednost socijalne poželjnosti, ukoliko nam je to od interesa. U ovom slučaju bi osobine ličnosti bile supresor, a prediktor socijalna poželjnost.

Istraživanje upotrebe skala socijalne poželjnosti u kliničkim istraživanjima pokazuju da su one povezane s različitim kliničkim ishodima (prekomereno jedenje, aleksitimija, blagostanje), ali kada se kontroliše efekat osobina ličnosti, ova povezanost se gubi (Perinelli & Gremigni, 2016). Ovo je jedan od načina za proveru prediktivne i inkrementalne validnosti socijalne poželjnosti. Na ovaj način možemo proveriti da li ona predstavlja supstantivnu varijansu ili ne.

Nevolja je ako je socijalna poželjnost povezana sa kriterijumom, te se u tom slučaju ne može tretirati kao supresor. U slučaju radnog učinka, istraživanja jasno pokazuju da socijalna poželjnost nije povezana s njim, uključujući različite mere radnog učinka (Ones et al., 1996). Međutim, u pogledu relacija sa indikatorima mentalnog zdravlja, nalazi nisu tako jednoznačni, te je pokazano da je socijalna poželjnost značajan prediktor negativnog afekta povrh varijanse koju objašnjavaju osobine ličnosti (Brajša-Žganec et al., 2011).

Drugi tretman socijalne poželjnosti je da se ona uvede kao **moderator**. Moderator je varijabla koja nije uzrokovana prediktorom, ali koja menja odnos između prediktora i kriterijuma. Na primer, moguće je da je veza prediktora i kriterijuma značajna kada je moderator nisko izražen, a da nije značajna ili čak opozitnog smera kada je moderator visoko izražen. Moderacija se ispituje preko efekta interakcije prediktora i moderatora na kriterijum. Ukoliko je interakcija značajna, to znači da je moderacija značajna, te da veza prediktora i kriterijuma zavisi od nivoa moderatora. U jednom istraživanju testirana je supresorska i moderatorska uloga socijalne poželjnosti u odnosu između osobina ličnosti koje su merene samoprocenom i procenom od strane drugih (Borkenau & Ostendorf, 1992). Rezultati su pokazali da se socijalna poželjnost ne ponaša kao supresor, ali da se ponaša kao

moderator samo u slučaju relacija između mera neuroticizma. Naime, kod ispitanika koji ostvaruju više skorove na socijalnoj poželjnosti veza neurotizma preko samoprocene i procene od strane drugih je niža, pa samim tim je i validnost samprocene niža.

Kako se i u ispitivanju supresije i moderacije primenjuje regresiona analiza, treba voditi računa da nema multikolinearnosti, odnosno da socijalna poželjnost ne korelira suviše visoko sa prediktorom. Multikolinearnost ne narušava ukupnu prediktivnost, ali narušava doprinose pojedinačnih prediktora. Ukoliko se socijalna poželjnost ne ponaša kao supresor, može se ispitati njena potencijalna moderatorska uloga, ali ukoliko postoji multikolinearnost, onda nema pomoći.

Povlađivanje/slaganje i neslaganje

Definicija

Povlađivanje ili slaganje (eng. *acquiescence*, *yea-saying*, *positivity*) se u testovima ličnosti manifestuje kao slaganje sa tvrdnjom bez obzira na njen sadržaj i značenje. Na binarnim ili dihotomnim stavkama u testovima ličnosti bi se povlađivanje manifestovalo kao biranje odgovora "da" ili "tačno", dok bi se na stavkama sa uređenim kategorijama, kakva je skala Likertovog tipa, povlađivanje manifestovalo kao biranje krajnjih kategorija odgovora na pozitivnom polu skale, poput 4 = *slažem se* i 5 = *u potpunosti se slažem*. Postoje različiti pristupi merenja povlađivanja, i dok neki autori kao indikator povlađivanja uzimaju u obzir samo krajnji odgovor (5 = *u potpunosti se slažem*), drugi uzimaju u obzir sve odgovore iznad srednjeg podeoka (npr. 4 = *slažem se* i 5 = *u potpunosti se slažem* na petostepenoj skali Likertovog tipa). U svakom slučaju, povlađivanje se izražava kao frekvenca biranja

kategorija odgovora koje upućuju na slaganje na skupu heterogenih ajtema, koji ne dele isti predmet merenja. Drugi način merenja povlađivanja je frekvenca biranja odgovora koji upućuju na slaganje na ajtemima i koji su pozitivno i koji su negativno formulisani, a odnose se na isti predmet merenja, tj. homogeni su. Kao poseban slučaj merenja navodi se učestalost slaganja sa tvrdnjama koje su jasni sintaksički i semantički opoziti (npr. "Volim sladoled" i "Ne volim sladoled").

S druge strane, neslaganje (eng. *disacquiescence, dissentience, nay-saying*, *negativity, opposition*) u testovima ličnosti predstavlja tendenciju osobe da se ne slaže sa tvrdnjom, bez obzira na njen sadržaj. Na binarnim ajtemima u testovima ličnosti, neslaganje bi se manifestovalo kao biranje odgovora "ne" ili "netačno", a na stavkama s uređenim kategorijama kao biranje odgovora 2 = *ne slažem se* i 1 = *uopšte se ne slažem*. I u ovom slučaju postoje različiti pristupi merenja neslaganja, pri čemu se kao indikator neslaganja može uzeti samo biranje krajnje kategorije odgovora koja upućuje na neslaganje (1 = *uopšte se ne slažem*) ili svih odgovora ispod srednjeg (npr. 2 = *ne slažem se* i 1 = *uopšte se ne slažem* na petostepenoj skali Likertovog tipa). Neslaganje se meri na isti način kao i slaganje, samo se uzima u obzir frekvenca biranja krajnje ili krajnjih kategorija koje upućuju na neslaganje.

U literaturi se pominje još jedna odgovaračka pristranost – veće slaganje u odnosu na neslaganje (eng. *net aquiescence*, *directional bias*), koja se dobija tako što se od skora na slaganju oduzme skor na neslaganju. Često se ova mera može dobiti i kao prosečan odgovor na skupu heterogenih ajtema.

Istraživače generalno zanima koji je uzrok odgovaračkih pristrasnosti. Dok je u slučaju davanja socijalne poželjnosti to jasno – u

pitanju je motivacija za prikazivanjem sebe u socijalno pozitivom svetlu, u slučaju slaganja postoje različita stanovišta (više u Knowles & Nathan, 1997). Prema jednom stanovištu, slaganje je odraz socijalne motivacije, iste kao i prilikom davanja socijalno poželjnih odgovora. U ovom slučaju ispitanik želi da se predstavi kao saradljiv, da bude prihvaćen i da izbegne neodobravanje, negativnu sliku o sebi i sl. Prema drugom stanovištu, slaganje je više odraz kognitivnih procesa, nego socijalne motivacije. Pritom, nekoliko kognitivnih procesa može biti uključeno, npr. razumevanje, pretraga memorije i slično. U jednom istraživanju je pokazano da ne postoji značajna relacija između slaganja i socijalne poželjnosti, već da je slaganje povezano sa rigidnom mentalnom organizacijom, kognitivnom jednostavnošću i netolerancijom na alternative (Knowles & Nathan, 1997). Drugim rečima, pokazano je da je slaganje pre karakteristika kognitivnog stila, a ne socijalne motivacije.

U testovima sposobnosti u kojima su uglavnom prisutna pitanja s višestrukim odgovorom, povlađivanje ili neslaganje ne postoji kao takvo, već se ispoljava kao biranje odgovora na određenom mestu u redosledu odgovora, što se naziva **pozicioniranim odgovorom** (eng. *position or order response bias*). Ukoliko je to biranje prvog ponuđenog odgovora (npr. odgovora pod a ili pod 1), pandan tome u testovima ličnosti bi bilo slaganje, ukoliko je to biranje poslednjeg ponuđenog odgovara, pandan tome bi bilo neslaganje, a ukoliko je to biranje središnjih ponuđenih odgovora (npr. b i c u slučaju kada postoje četiri ponuđena odgovora: a, b, c, i d), pandan tome bi bilo biranje srednjih odgovora. Jedan od načina za eliminisanje ovog efekta je da se različitim ispitanicima daju različiti redosledi odgovora u pitanjima s višestrukim ili alternativnim izborom.

Dimenzije

Povlađivanje se smatra dvodimenzionalnim konstruktom (Bentler, Jackson, & Messick, 1971), pri čemu se mogu razlikovati slaganje (eng. *agreement*) i prihvatanje (eng. *acceptance*). **Slaganje** predstavlja ono što se uobičajeno smatra povlađivanjem, a to je slaganje sa tvrdnjom bez obzira na njen sadržaj, što uključuje slaganje kako sa pozitivno, tako i sa negativno formulisanim tvrdnjama. **Prihvatanje** se odnosi na prihvatanje tvrdnji, ponovo, kako pozitivno, tako i negativno formulisanih, ali ne i njihove negacije. Na primer, osoba koja bi pokazivala tendenciju ka slaganju bi se složila sa tvrdnjama "Ja sam druželjubiva osoba", "Ja sam povučena osoba", kao i sa tvrdnjama "Ja nisam druželjubiva osoba" i "Ja nisam povučena osoba". S druge strane, osoba koja bi pokazivala tendenciju ka prihvatanju bi se složila samo sa tvrdnjama "Ja sam druželjubiva osoba" i "Ja sam povučena osoba". Bentler i saradnici (Bentler et al., 1971) su pokazali da su ove dve dimenzije povlađivanja međusobno nezavisne, što je potvrđeno i u kasnijim istraživanjima (npr. Danner & Rammstedt, 2016). Razlikovanje ovih dimenzija potkrepljuju i rezultati o njihovim različitim korelatima, pri čemu slaganje objašnjava upadljivo više varijanse ajtema za procenu znanja, dok prihvatanje objašnjava upadljivo više varijanse ajtema za procenu ličnosti, a potom i stavova (Danner & Rammstedt, 2016). Iako istraživanja ukazuju na potrebu za razlikovanjem slaganja i prihvatanja, najčešće se u istraživanjima ispituje samo slaganje.

U vezi sa neslaganjem, nema istraživanja koja ispituju dimenzionalnost ove odgovaračke pristrasnosti.

Problemi

Povlađivanje otežava razlikovanje onih ispitanika koji zaista imaju visok nivo crte od onih koji su skloni slaganju sa tvrdnjama. Važna posledica povlađivanja ogleda se u tome da će korelacija između dva upitnika biti veštački veća, ukoliko je povlađivanje prisutno. Naime, ukoliko neki ispitanik ima tendenciju da se slaže sa tvrdnjama, pretpostavka je da će se dosledno slagati sa tvrdnjama na oba upitnika – imaće visok skor na oba upitnika, što povećava korelaciju između ta dva upitnika. Isto važi i za tendenciju neslaganja, tj. zbog nje može doći do veštačkog povećanja korelacije između dva upitnika iako ta korelacija ne bi zaista bila tolika da je primenjena neka druga vrsta metode prikupljanja podataka, format ajtema i slično. Povlađivanje, takođe, utiče i na karakteristike stavki, pa tako lažni ajtem može biti više validan, nego pravi ajtem, i obrnuto, pravi ajtem može biti manje validan u odnosu na lažni.

Detekcija i rešenja

Kao rešenje eliminacije slaganja najčešće se preporučuje konstrukcija semantički **balansiranih skala**, u okviru kojih će postojati ujednačen broj pozitivno i negativno formulisanih tvrdnji. Na ovaj način neko ko ima tendenciju slaganja ne može dobiti visok skor, odnosno neko ko ima tendenciju neslaganja ne može dobiti nizak skor. Postoje saveti kako bi trebalo konstruisati obrnuto formulisane stavke, ali možda je najjednostavniji način zamisliti skup indikatora koji bi bili karakteristični za opozitni pol konsturkta koji je predmet merenja, i potom konstruisati autentično negativno formulusane stavke. Na primer, pozitivno formulisana stavka za procenu ekstraverzije bi bila "Ja sam društvena osoba", a za procenu opozitnog pola, tj. introverzije bi

bila "Ja sam povučena osoba". Loša taktika je dodavanje samo reči "ne" ispred pozitivno formulisane tvrdnje ili prebacivanje pozitivne tvrdnje u negaciju (npr. "Ja nisam društvena osoba"). Prilikom konsturkcije testa, istraživači često nemaju jasnu predstavu o tome šta predstavlja negativni pol neke crte, ili u teoriji ne postoji jasno određenje negativnog pola, što otežava konstrukciju negativno formulisnaih stavki. U tim situacijama je svakako poželjno sprovesti pilot-studiju kako bi se odabrale adekvatne negativno formulisane stavke.

Kada postoje balansirane skale, efekat povlađivanja može se proveriti ukoliko se izračuna poseban skor za pozitivno i za negativno formulisane ajteme, pa potom testiraju **razlike u skorovima pozitivno i negativno formulisanih ajtema**. Ukoliko među ovim skorovima nema značajne razlike, i ukoliko su oba visoka, onda se može sumnjati na to je ispitanik bio sklon povlađivanju. Ukoliko se dobije rezultat da nema značajne razlike među skorovima i ukoliko su oba skora niska, onda se može sumnjati u to da je ispitanik bio sklon neslagaju. Pogledajmo na primeru u Tabeli 26 kako bi se napravio izračun. Ukoliko su ispitanici odgovarali na skalu od 6 ajtema s priloženom Likertovom petostepenom skalom, pri čemu su 3 ajtema bila pozitivno, a 3 negativno formulisana, Miloš na pozitivno formulisanim stavkama ostvaruje AS_p = 4,67 (SD_p = 0,67), a na negativno formulisanim stavkama AS_n = 4,33 (SD_n = 0,67). Razlika između ova dva skora nije značajna ($t(2) = 0{,}71$, $p > {,}05$)[32], i oba skora su visoka, što upućuje na zaključak da je Miloš bio sklon slaganju s tvrdnjama, bar kada je dati test u pitanju. U Lukinom slučaju, takođe, nema značajnih razlika u skorovima ($t(2) = 0{,}71$, $p > {,}05$), a kako su oba

[32] Iako je u ovom slučaju primerenije raditi neprametrijski test zbog malog uzorka, primenom Mann-Whitney-evog *U* testa je takođe dobijena neznačajna razlika, te je zarad uobičajenosti prikaza zadržan *t*-test.

skora niska, može se zaključiti da je Luka bio sklon neslaganju s tvrdnjama.

Tabela 26

Primer ajtemskih skorova za ispitanika koji pokazuje slaganje i koji pokazuje neslaganje s tvrdnjama

	Br. stavke	Odgovori	
		Miloš	Luka
Pozitivno formulisane stavke	1	4	1
	3	5	2
	5	5	2
		AS = 4,67	*AS* = 1,67
Negativno formulisane stavke	2	4	2
	4	4	1
	6	5	1
		AS = 4,33	*AS* = 1,33

Balansirane skale ne mogu u potpunosti eliminisati slaganje ili neslaganje, ali mogu smanjiti ove pristrasnosti. Međutim, u praksi, skale često nisu uopšte balansirane, ili sadrže vrlo malo negativno formulisanih stavki. Na primer, analizom 200 skala koje se koriste u marketinškim istraživanjima, Baumgartner i Stinkemp (Baumgartner & Steenkamp, 2001) su ustanovili da je samo 9% skala banasirano, a čak 43% nema uopšte negativno formulisanih ajtema. Pored toga, često se dešava i da negativno formulisane stavke prosto nisu dobro konstruisane, te se stvaraju metodski faktori ili se na drugi način narušava validnost testa kada se skala balansira.

Drugi način provere prisutnosti povlađivanja ili neslaganja je preko izračunavanja **korelacija između heterogenih ajtema**, za koje se pretpostavlja da ne bi trebalo biti povezani. Ukoliko takvi ajtemi značajno i pozitivno koreliraju, možemo sumnjati na prisutnost

povlađivanja ili neslaganja, u zavisnosti od toga da li su ajtemski skorovi visoki ili niski. Osnovni problem kod ovog načina provere je u tome što nekad ne postoji jasno teorijsko stanovište u vezi s tim koji ajtemi ne bi trebalo značajno biti povezani. Takođe, ako smo ograničeni brojem primenjenih ajtema i testova, moguće je da u zadatom setu nemamo ajteme koje bismo mogli iskoristiti za ovu proveru, tj. za koje postoji teorijsko utemeljenje da su heterogeni – nepovezani.

Treći način provere je dodavanje **parova kontrolnih ajtema** koji ne bi ulazili u skor, ali na osnovu kojih bi se proverilo da li je ispitanik sklon slaganju ili neslaganju. Ovi kontrolni ajtemi mogu imati veze sa predmetom merenja testa, a i ne moraju, i sastoje se od para pozitivno i negativno formulisanih ajtema koji su evidentno opozitni, npr. "Srećan sam" i "Nisam srećan". Ukoliko se ispitak slaže s obe tvrdnje, može se sumnjati na to da je sklon slaganju. Međutim, već je spomenuto da dodavanje samo odrične rečce "ne" ne čini negativno formlisani ajtem dobrim. U narednom odeljku će biti istaknuti i drugi problemi u vezi s negativnom formlusanim tvrdnjama.

Problemi negativno formlusanih tvrdnji

Pisanje negativno formulisanih tvrdnji je zahtevan posao, i postoji nekoliko problema u vezi s ovim tvrdnjama. Prvi od njih tiče se statusa negativno formulisanih stavki. Naime, često se dešava da ove stavke mogu promeniti predmet merenja, ili na drugi način ugroziti konstrukt validnost testa. Nekada negativno formulisane stavke zaista mere drugi konstrukt u odnosu na pozitivno formulisane stavke, što ima svoja teorijska opravdanja. Primer za to je Inventar pozitivnog i negativnog afekta – PANAS koji meri pozitivan i negativan afekat (Watson, Clark, & Tellegen, 1988). Ajtemi skale pozitivnog afekta sadrže

samo pozitivne deskriptore ovog stanja (npr. "oduševljeno", "ponosno", "usredsređeno"). Ajtemi skale negativnog afekta, takođe, sadrže samo pozitivne deskriptore ovog stanja (npr. "nesrećno", "neprijateljski raspoloženo", "uplašeno"). Međutim, afekat meren ovom skalom nije bipolarni jednodimenzionalni konstrukt na kojem na jednom polu imamo negativan, a na drugom polu pozitivan afekat, već su ova dva aspekta afekta unipolarni jednodimenzionalni konstrukti koji međusobno ne koreliraju značajno. Drugim rečima, moguće je da osoba u isto vreme doživljava visok i pozitivan i negativan afekat (npr. usled strahopoštovanja prema nekome), ili ujedno nizak i pozitivni i negativni afekat (npr. usled ravnodušnosti). Ovaj primer ilustruje kako isprva negativno formulisane tvrdnje, koje su trebale da se odnose na negativni pol pozitivnog aspekta, zapravo tvore zaseban faktor negativnog afekta, ali koji ima svoje teorijske i praktične implikacije i koji nije artefakt metoda.

Međutim, mnogi autori su pokazali da negativno formulisani ajtemi ugrožavaju konstrukt validnost jer tvore poseban metodski faktor koji narušava faktorsku strukturu testa. Formiranje zasebnog faktora koji okuplja negativo formulisane ajteme je moguća posledica toga što ovi ajtemi imaju generalno niže težine, tj. ajtemske skorove (Weems, Onwuegbuzie, Schreiber, & Eggers, 2003). Zbog toga su njihove distribucije međusobno sličnije, što za posledicu ima više korelacije između njih. Na primer, u ispitivanju strukture agresivnosti dobijeno je da se izdvaja jedan faktor koji okuplja ajteme koji se odnose na opozitni pol agresivnosti, tj. prijatnost, te je ustavnovljeno da je reč o metodskom faktoru (Dinić, Mitrović i Smederevac, 2014). Kao primer izdvajanja metodskog faktora često se spominje struktura Rozenbergove skale samopoštovanja. Ova skala sadrži 10 ajtema, od kojih su 5 formulisana u

negativnom smeru i istraživanja pokazuju da ovi ajtemi tvore zaseban metodski faktor (DiStefano & Motl, 2009). Naime, često se dešava da se faktorskom analizom izdvoji jedan faktor koji okuplja pozitivno, a drugi koji okuplja negativno formulisane stavke. Ovakvi faktori su metodski faktori, a ne osobinski. Drugim rečima, izolovani faktori nisu facete jednog te istog konstrukta. S obzirom na to, pogrešno se može zaključiti da je konstrukt dvodimenzionalan.

Jedan od načina kako se može proveriti da li su dva izdvojena faktora koja okupljaju pozitivno i negativno formulisane ajteme, u stvari, metodski faktori, jeste uvid u njihove korelacije. Ukoliko su te korelacije visoke, najverovatnije su u pitanju metodski faktori koji se odnose na isti konstrukt. Kao kritična vrednost za određivanje visine korelacije može poslužiti preporuka koju je propisala Evropska federacija psiholoških udruženja (European Federation of Psychologists' Associations – EFPA, 2013) u vezi s minimalnom korelacijom za određivanje konstrukt validnosti. Ukoliko dve varijable koreliraju ,55 i više, najverovatije su u pitanju isti konstrukti (ili facete istog konstrukta). Ukoliko su te korelacije niske, to znači da izdvojeni faktori mere različite kontrukte, tj. da negativno formulisani ajtemi menjaju predmet merenja, pošto ne dele zajedničku varijansu sa pozitivno formulisanim ajtemima (kao što je slučaj sa skalama PANAS-a). U ovom slučaju treba dodatno proveriti, preko utvrđivanja različitih vrsta validnosti, da li su izdvojeni faktori zaista različiti konstukti, da li imaju različite korelate i da li im se razlikuje interpretacija skorova.

Pored pomenutih problema u vezi sa negativno formulisanim ajtemima, pokazano je da njihovo prisustvo u skali smanjuje pouzdanost interne konzistencije (DeVellis, 1991). Naime, korelacija između negativno i pozitivno formulisanih ajtema je niža, nego korelacija

između dva pozitivno ili dva negativno formulisana ajtema, što utiče na pouzdanost skale. Takođe, u jednom istraživanju anksioznosti, negativno formulisani ajtemi nakon rekodiranja imaju niže ajtemske skorove nego pozitivno formulisani ajtemi (Weems et al., 2003). Dodatni problem je to što je odgovaranje na negativno formulisane tvrdnje kognitivno zahtevnije, tj. zahteva razvijeniju verbalnu sposobnost (Marsh, 1996). Tako će osobe koje imaju manje razvijenu verbalnu sposobnost ili koje su manje obrazovane imati više poteškoća u odgovaranju na negativno formulisane tvrdnje.

Neki autori ističu da je konstrukcijom balansiranih skala samo zamenjen stari problem novim, mada je time izbegnut ozbiljniji problem (Fur & Bacharach, 2014). Taj novi, manje ozbiljan problem, odnosi se na to da kada imamo balansirane skale onda teško možemo razlikovati ispitanika koji ima tendenciju ka slaganju od ispitanika koji ima umereni nivo merene crte. Ipak, iako se negativno formulisani ajtemi preporučuju kako bi se eliminisao ili smanjio efekat odgovaračkih pristrasnosti, ukoliko oni narušavaju konstrukt validnost testa, bolje je da se uklone iz upitnika. Alternativna opcija je da se rezultati prikažu posebno za pozitivno i posebno za negativno formulisane ajteme, ili posebno za pozitivno formulisane ajteme i za ukupni skor (koji sadrži i pozitivno i negativno formulisane ajteme).

Treba napomenuti da pojedini autori definišu odgovaranje na negativno formulisane ajteme kao odovarački stil, odnosno kao predispoziciju za interpretiranje i biranje ajtema na osnovu njegove afektivne valence (DiStefano & Motl, 2009). Pokazano je da se osobe s višim nivoom samopoštovanja, savesnosti i bihejvioralne aktivacije ređe slažu sa negativno formulisanim stavkama, a osobe s višim nivoom

bihejvioralne inhibicije se češće slažu s negativno formulisanim tvrdnjama (DiStefano & Motl, 2009; Quilty, Oakman, & Risko, 2006).

Preferiranje ekstremnih ili srednjih odgovora

Definicija

Preferiranje ekstremnih odgovora (eng. *extreme response style*, *extremity*) je tendencija biranja krajnjih odgovora na skali procene i to na oba pola skale (najčešće su to 1 = *uopšte se ne slažem* i 5 = *u potpunosti se slažem*), uz izbegavanje biranja srednjih odgovora (npr. 2, 3 i 4) bez obzira na sadržaj stavke. Obrnuto, preferencija srednjeg odgovora (eng. *mid-point response style*) odnosi se na učestalo biranje srednjeg podeoka na skali (3 = *podjednako se i slažem i ne slažem*), a izbegavanje biranja ekstremnih podeoka. Može se razlikovati i preferencija srednjih odgovora (eng. *mild response style*), koja podrazumeva širi izbor kategorija odgovora (npr. 2, 3 i 4). Preferencija srednjeg odgovora uobičajeno se smatra opozitnim obrascem odgovaranja u odnosu na preferenciju ekstremnih kategorija odgovora. Iako je u većini istraživanja potvrđena negativna veza između ove dve pristrasnosti (npr. Baumgartner & Steenkamp, 2001), u nekim istraživanjima se ističe da ove dve pristrasnosti nisu nužno međusobno suprotstavljene (npr. Kieruj & Moors, 2010). Iako je moguće da se ove pristrasnosti odraze i u testovima sposobnosti, tj. kod pitanja s višestruktim izborom kao biranje prve ili poslednje opcije, ili biranja srednjih opcija, ove pristrasnosti se tretiraju kao posebna vrsta odgovaračkih pristrasnosti – pozicionirani odgovor, o čemu je već bilo reči. S obzirom na navedeno, preferencija ekstremnih i srednjih odgovora će se isključivo vezati za odgovaranje na testovima ličnosti.

Dimenzije

Za ove odgovaračke pristrasnosti nije ispitivana dimenzionalnost.

Problemi

U slučaju ispoljavanja preferencije ekstremnih ili srednjih odgovora dešava se to da dve osobe koje imaju isti pravi skor, nemaju isti opaženi skor. Na primer, ukoliko je pravi skor dve osobe 20, a jedna od njih ima tendenciju davanja ekstremnih odgovora, onda će njen opaženi skor biti, na primer 30, dok će osobi koja nema ovu tendenciju opaženi skor biti 20. Obrnuto, usled ove odgovaračke pristrasnosti može se desiti da dve osobe koje nemaju isti pravi skor, imaju isti opaženi skor. Na primer, Miloš i Luka imaju opaženi skor 30, ali kako Miloš ima sklonost ka preferiranju ekstremnih odgovora, njegov pravi skor bi bio zapravo 20, a ne 30, kao što je Lukin koji nema ovu tendenciju. Isti problem se javlja i u slučaju preferiranja srednjih odgovora, gde se može desiti da osoba ima sistematski niži ili viši opaženi skor u odnosu na pravi. Sve navedene situacije važe kada u testu nema obrnuto formulisanih stavki ili kada postoji disbalans pozitivno i negativno formulisanih stavki.

Većina autora se, međutim, slaže da preferencija ekstremnih kategorija odgovora značajno ugrožava opažene skorove, čak i kada su skale balansirane (više u Baumgartner & Steenkamp, 2001). Naime, kada se ajtemi rekodiraju, ova odgovaračka pristrasnost čini da su skorovi na skalama zakošeni (skjunis je visok), što za posledicu ima veću, ali razuđenu varijansu koja doprinosi nižim koeficijentima korelacije između tih skorova i drugih skorova koji nisu bili podložni ovoj pristrasnosti. S druge strane, ukoliko neko dosledno pokazuje

sklonost ka preferenciji ekstremnih odgovora kroz različite testove, ajtemske distribucije će biti međusobno slične, što za posledicu ima veštački više korelacije između ajtema, pa samim tim i skorova s različitih testova. Dakle, visoke korelacije između testova nisu posledica predmeta merenja, već odgovaračkih pristrasnosti. Preferencija ekstremnih kategorija odgovora takođe ugrožava i validnost. Na primer, u jednom istraživanju je pokazano da se faktorska struktura istog konstrukta razlikuje na uzorku ispitanika koji pokazuju nisku i koji pokazuju visoku tendenciju biranja ekstremnih odgovora (Arce-Ferrer & Ketterer, 2003).

Na ovom mestu treba pomenuti još jedan problem u vezi s preferiranjem srednjeg odgovora. Naime, često se postavlja pitanje da li **broj podeoka na skali** procene treba biti paran ili neparan. Parnim brojem podeoka prisiljavamo ispitanike da se odluče da li su više "za" ili "protiv", čime gubimo informaciju o postojanju neutralnog stava. Neka istraživanja pokazuju da se ispitanici, kada se ne ponudi srednji podeok, više slažu s tvrdnjama (Velez & Ashworth, 2007). Takođe, neka istraživanja pokazuju da se koeficijenti validnosti ne poboljšavaju kada ispitanike koji su birali srednji odgovor svrstamo u neku od preostalih kategorija odgovora kojima se iskazuje smer stava ili mišljenja (tzv. tehnikom granjanja, eng. *branching*, više u Malhotra, Krosnick, & Thomas, 2009). S druge strane, ukoliko se ponudi neparan broj podeoka, pretpostavka je da srednji odgovor odražava neutralni stav ili slaganje, tj. da zaista odražava stav "podjednako se i ne slažem i slažem". Neutralni stav podrazumeva da osoba zapravo ima neki stav o objektu merenja. Međutim, kao što je rečeno, biranje središnjeg odgovora može biti odraz odgovaračke pristrasnosti, a ne neutralnog stava. Tako, neko ko zapravo ima stav može birati srednji odgovor jer mu je tako lakše, tj.

nije motivisan da se udubljuje u sadržaj pitanja. Isto tako je moguće da preferencija srednjeg odgovora odražava nemanje stava ili mišljenja, što se naziva skrivenim "ne znam". U istraživanjima se dobija rezultat da većina ispitanika koja bira srednji odgovor (preko 50%), zaparavo nema mišljenje ili stav o predmetu merenja, tj. da je skriveno "ne znam" najčešći razlog biranja srednjeg odgovora, a ne neutralni stav (Baka, Figgou, & Triga, 2012; Nadler, Weston, & Voyles, 2015; Sturgis, Roberts, & Smith, 2012). Zanimljivo je da se ovaj odnos ne menja previše ni kada se ponudi odgovor "ne znam", tj. uvođenje ovog odgovora ne rešava problem. Naravno da uvek treba voditi računa o predmetu merenja, ali je opšta preporuka da se srednji odgovor uvrsti u skalu jer neutralna mišljenja i stavovi postoje i relevantni su za psihološku procenu (Malhotra et al., 2009; Sturgis et al., 2012). Ipak, da bi se rešio problem biranja srednjeg odgovora koji ne potiče od neutralnog stava, preporuka je da se ovaj podeok adekvatno definiše, npr. preko opisa primera ponašanja, stava ili mišljenja koji bi bio karakterističan za neutralni nivo.

Detekcija i rešenja

Preferiranje ekstremnih odgovora izražava se kao **frekvenca biranja krajnjih podeoka na heterogenim ajtemima**. Kao dodatni pokazatelj ove tendencije, mada ne tako dobar, može se uzeti **prosečni odgovor na skupu heterogenih ajtema**, te ukoliko skala sadrži balansirane ajteme, prosečni skor nekoga ko pokazuje preferenciju ekstremnih odgovora će biti oko srednjeg podeoka na skali. Ovaj pokazatelj nije dobar jer će i osoba koja preferira srednji odgovor takođe imati prosečni skor oko proseka na skali procene, kao i osoba koja preferira ekstremne odgovore. Ono što pruža uvid u to da li osoba

preferira ekstremne ili srednje odgovore je standrdna devijacija skorova za pojedinačne ispitanike. Dok osobe koje imaju tendenciju biranja srednjeg odgovora imaju nisku varijabilnost skorova i na neki način daju uniformne odgovore, osobe koje imaju tendenciju biranja ekstremnih odgovora će imati veću varijabilnost skorova. Ipak, nedostatak ovog načina detekcije je što se njime detektuje samo širina sklopa odgovora, te se nema precizan uvid u konkretne odgovaračke pristrasnosti. S obzirom na to, razvijeni su različiti indeksi kojima se meri učestalost preferencije konkretnog odgovora, a ne nužno ekstremnog, ali koji mogu onda dati uvid u konkretne odgovaračke pristrasnosti. Takvi indeksi su, na primer, indeks disperzije odgovora i indeks entropije odgovora (više u Peterson, Rhi-Perez, & Alnaum, 2014).

Preferencija središnjih odgovora se može eliminisati tako što skala za odgovaranje sadrži paran broj podeoka, npr. 4. Međutim, istraživanja pokazuju da su ovakve skale manje informativne, te da eliminacija srednjeg odgovora vodi u povišenje skorova i da se osobe više slažu s pozitivnim tvrdnjama (više u Nadler et al., 2015). Nešto više o optimalnom broju podoka na skali će biti reči kasnije, kod izvora odgovaračkih pristrasnosti.

Nepažljivi odgovori

Nepažljivi odgovori (eng. *careless responses*) obuhvataju širi dijapazon obrazaca odgovaranja, te mogu biti manifestovani kao random ili nasumično odgovaranje (eng. *random responding*), cik-cak odgovaranje, preskakanje random ajtema pri odgovaranju ili preskakanje ajtema po nekoj šemi, biranje uvek istog odgovora ili podeoka na skali, nekonzistentni odgovori (eng. *inconsistent responses*) itd. Oni mogu odražavati iskrenu grešku, poput pogrešno pročitanog

pitanja, pogrešno zaokruženog odgovora na listu za odgovore, promašenog rednog broja stavke i slično, ali mogu biti i rezultat nerazumevanja ajtema, odraz niske motivacije, bunta ili sabotiranja test situacije, ili akt pasivne agresije.

Dimenzije

Nepažljivo odgovaranje se najčešće deli na dva tipa. Prvi tip se odnosi na slučajno ili nasumično odgovaranje, npr. kada neko slučajno prekoči neki ajtem ili odgovori na ajtem bez čitanja sadržaja ajtema ili razumevanja istog. Drugi tip se odnosi na neslučajno ili namerno nevalidno odgovaranje, kao što su cik-cak odgovaranja, preskakanje svakog 5. ajtema, itd.

Problemi

Oba tipa nepažljivog odgovaranja predstavljaju varijansu greške koja utiče na smanjenje korelacija između ajtema i merenih konstrukata, na smanjenje pouzdanosti i na rezultate faktorske analize, a neslučajno odgovaranje može imati i nepredvidive efekte na korelacije između ajtema.

Ranija istraživanja upućuju na različite bazne stope nepažljivog odgovaranja, ali se ona kreće od 3,5% u slučaju dobrovoljačkog uzorka (Johnson, 2005) do 10,6% u slučaju uzorka studenata koji su popunjavali upitnike zarad dobijanja bodova u sklopu predispitnih obaveza (Kurtz & Parish, 2001). Iako se čini da je stopa nepažljivog odgovaranja niska, nepažljivo odgovaranje se javlja čak i u situaciji kada su ispitanici visoko motivisani, kao prilikom apliciranja za posao (5% u Ehlers, Greene-Shortridge, Weekley, & Zajack, 2009). Posebnu bojazan od nepažljivog odgovaranja predstavljaju onlajn ankete i upitnici, pod pretpostavkom da fizička udaljenost i izostanak personalizovanog i

direktnog kontakta sa administratorima utiču na manju odgovornost kod ispitanika, što se manifestuje i kao nepažljivo odgovaranje (Johnson, 2005). Pored nedostatka socijalne razmene sa ispitanikom, postoje i nekontrolisani uslovi prilikom onlajn učešća u istraživanju gde je vrlo verovatno da će ispitanikova pažnja biti podeljena. To dovodi do nepažljivog odgovaranja, a i zahtevi pred savremenim čovekom su takvi da je on sve skloniji multitaskingu.

Detekcija

Postoji nekoliko načina detekcije nepažljivog odgovaranja koji se mogu podeliti na *a priori* i *post hoc* načine. *A priori* načini odnose se na **posebno konstruisana pitanja** koja mogu biti lažni ajtemi, ili očigledno tačne ili netačne tvrdnje. Za razlliku od lažnih ajtema, netačne tvrdnje su prilično očigledno netačne (npr. "Rođen sam u gradu Boga-Boho", "Ne razumem ni reč srpskog" ili "Trenutno koristim računar" kod onlajn popunjavanja upitnika). Ovi ajtemi se obično stave na kraj nekog od upitnika i preporučljivo je da ih bude jedan do tri na svakih 50 do 100 ajtema (Meade & Craig, 2012). Takođe, u ove ajteme spadaju i direktivni ajtemi, s jasnom instrukcijom šta treba da se uradi (npr. "Na sledeće pitanje zaokružite broj 2"). Iako se ovakvi ajtemi preporučuju, neki autori upozoravaju na to da ih ispitanici mogu shvatiti kao neozbiljne i da na njih veliki broj ispitanika netačno odgovara (Huang, Curran, Keeney, Poposki, & DeShon, 2012).

U ovaj način spadaju i posebno konstruisane skale za detekciju nekonzistentnih odgovora, koje se obično stavljaju na kraj seta upitnika. Nekonzistentni odgovori su pominjani prilikom detekcije preferencije ekstremnih odgovora, a to su opozitne tvrdnje. Ukoliko neki ispitanik

odgovori potvrdno na dve opozitne tvrdnje, to može biti i signal da nije pažljivo odgovorio na pitanja.

Post hoc načini odnose se na različite indekse i metode utvrđivanja nepažljivog ponašanja. Postoji nekoliko indeksa kojima se može identifikovati nepažljivo ponašanje. Jednu grupu čine **indeksi konzistencije** kojim se utvrđuje sličnost odgovaranja na slične ajteme, pri čemu sličnost ajtema može biti potpuna (dva potpuno identična ajtema), utvrđena na osnovu teorijskih pretpostavki (dva ajtema pripadaju istom konstruktu) ili empirijski, na osnovu korelacija (‚60, prema Meade & Craig, 2012). Ukoliko je indeks konzistencije u slučaju sličnih ajtema nizak, te postoji, zapravo, nekonzistentan obrazac odgovaranja, može se sumnjati na nepažljivo odgovaranje. S druge strane, može se računati indeks konzistencije za heterogene ajteme, te ukoliko je on visok (preko –‚60), može se sumnjati na nepažljivo odgovaranje. Ove dve vrste indeksa konzistencije nazivaju se još i indeksima psihometrijskih sinonima, odnosno antonima. Postoji još jedna varijanta indeksa konzistencije, indeks dugačkog niza, kada se računa konzistentnost biranja određene kategorije odgovora na određenom broju sledujućih heterogenih ajtema. Dakle, prema ovim indeksima, učestala preferencija jedne kategorije odgovora upućuje na nepažljivo odgovaranje.

Pored opisanih načina, može se računati i **indeks par-nepar**, u okviru kojeg se računa korelacija između skora na parnim i na neparnim ajtemima, pa se ona koriguje za dužinu testa preko Spirman-Braunove formule. Logika je slična kao i kod računanja split-half pouzdanosti. Ukoliko je ovaj indeks tj. korelacija niska (ispod ‚30), zaključuje se da su odgovori nekonzistentni – ispitanik nije pažljivo odgovarao na pitanja.

Drugi način detekcije nepažljivih odgovora je preko **detekcije autlajera**. Autlajeri su netipični odgovori, koji odstupaju od modela odgovaranja, obično od normalne distribucije. Dok su ekstremni odgovori krajanje tačke na distribuciji (npr. nalaze se na 3,33 *SD*), autlajeri su odgovori koji štrče i koji veoma odstupaju od normalne distribucije (npr. nalaze se na 5 *SD)*. Autlajeri mogu biti univarijatni, tj. na jednom ajtemu ili jednoj skali, i multivarijanti, tj. na kombinaciji nekoliko ajtema ili skala. Prilikom detekcije nepažljivog odgovaranja, preporučuje se da se detektuju multivarijatni autlajeri, a ne univarijatni. Jedan od načina detekcije multivarijatnih autlajera je preko Mahalanobisovih distanci (više u Tabachnick & Fidell, 2013).

Treći način detekcije nepažljivih odgovora je preko **vremena latencije** potrebnog za odgovaranje. Postoji nelinearna veza između vremena potrebnog za odgovor i kvaliteta odgovora, pri čemu se pretpostavlja da veoma brzi i veoma spori odgovori ukazuju na nepažljivo odgovaranje. Graničnu vrednost sporih ili brzih odgovora je najbolje utvrditi empirijski, na konkretnim podacima. Ipak, neki autori se oslanjaju na svoju ekspertizu prilikom određivajna granične vrednosti, pa su tako Huang i sararadnici (Huang et al., 2012) odredili da je minimalno 2 sekunde potrebno za odgovaranje na jedan ajtem iz njihove baze, te se sve ispod toga može smatrati nepažljivim odgovaranjem.

Poredivši nekoliko načina detekcije nepažljivog odgovaranja, Mid i Kreg (Meade & Craig, 2012) preporučuju najpre detekciju preko veoma kratkih perioda latencije i preko samoprocene ispitanika o kvalitetu njegovih podataka, pa potom preko upotrebe direktivnih pitaja, i na kraju preko izračunavanja različitih indeksa i eliminacije na osnovu Mahalanobisovih distanci. Ipak, efikasnost ovih poslednjih

metoda zavisi od uslova primene, pa je tako efikasnost metoda eliminacije multivarijatnih autlajera ograničena na slučaj kada postoji uniformna random distribucija nepažljivih odgovora ili normalna distribucija, ali samo za neke nepažljive odgovore, dok ovaj način nije efikasan u slučaju normalne distribucije svih nepažljivih odgovora (Meade & Craig, 2012). S druge strane, indeksi konzistencije su se pokazali kao generalno efikasni načini detekcije nepažljivog odgovaranja u većini uslova (npr. pod uslovom uniformne ili normalne distribucije nepažljivog odgovaranja, više u Meade & Craig, 2012). Ipak, svi navedeni metodi važe za dugački set upitnika koji sadrži 300 i više pitanja. U slučaju kraćih skala i setova, do 100 ajtema, preporučuje se vreme latencije (i to potrebno po stanici upitnika, a ne po ajtemu), kao i učestalost preferencije određene kategorije odgovora (indeks dugačkog niza). Ostali metodi nisu dali visoku senzitivnost i specifičnost (Niessen, Meijer, & Tendeiro, 2016). Alternativno, preporučuju se metodi bazirani na Teoriji odgovora na stavke na osnovu kojih se mogu detektovati misfitujući (nesaglasni) odgovori na nivou ispitanika, ali njihov opis prevazilazi cilj ove knjige.

Rešenja

Najpre, ukoliko prilike to dozvoljavaju, instrukcijom se može smanjiti nepažljivo ponašanje. U jednom istraživanju je u uputstvu navedeno: "Vaši podaci će biti potpuno poverljivi, ali na svakoj strani upitnika će biti pitanje da unesete Vaše ime i prezime, kako bismo kasnije mogli da spojimo Vaše podatke" (Meade & Craig, 2012). Rezultati ovog istraživanja pokazuju da se manja stopa nepažljivog odgovaranja dobija u opisanom slučaju, nego u uslovima potpune

anonimnosti. Ipak, druga istraživanja ne idu u prilog efikasnosti opisane prakse (Ong & Weiss, 2000).

Potom, u onlajn istraživanju se na lažne ili direktivne ajteme može ubaciti fidbek ukoliko ispitanik odgovori drugačije od očekivanog, te se u tom fidbeku ispitaniku može skrenuti pažnja na to da pažljivije odgovara. Prednost ovog načina je u tome što se fidbek daje odmah po odgovoru, tj. u toku samog rada na upitniku. Takođe, ne treba preterivati s dužinom seta upitnika. Za pažljivog ispitanika 15 minuta je sasvim dovoljno da odgovori na 50 ajtema nekog testa ličnosti (Meade & Craig, 2012), te u skladu s tim treba proceniti koliko je ajtema optimalno u setu upitnika. Na kraju, ispitaniku se može postaviti pitanje da sam proceni da li su njegovi podaci upotrebljivi, npr. "Prema Vašem iskrenom mišljenju, da li se možemo osloniti na Vaše podatke".

Izvori odgovaračkih pristrasnosti

S obzirom na to da se odgovaračke pristrasnosti mogu tretirati kao setovi ili stilovi, izvori odgovaračkih pristrasnosti mogu se podeliti u dve grupe (Weijters, 2006):

1. izvori koji potiču od samog **testa i uslova testiranja**, u koje spadaju format ajtema (npr. skala Likertovog tipa, koliko kategorija odgovora ima skala procene), tehnika prikupljanja podataka, tj. medijum (npr. upitnik ili intervju, telefonska anketa ili uživo anketiranje), postojanje kognitivnog opterećenja, efekat ispitivača (npr. da li je vešt intervjuer ili ne, koliko je srdačan i pristupačan), jezik testa (da li je maternji ili ne) i predmet merenja testa (npr. da li je predmet merenja nešto do čega je ispitaniku jako stalo i u šta je involviran ili nešto u vezi s čim nema izgrađen i jasan stav ili mu nije važno).

2. izvori koji potiču od **ispitanika** u koje spadaju socio-demografske karakteristike kao što su pol, starost, nivo obrazovanja, rasa, pripadnost kulturi, potom visina primanja i ekonomske prilike, osobine ličnosti, sposobnosti, motivacija i slično.

Iako ova podela obuhvata dva generalna izvora odgovaračkih pristrasnosti, navedeni izvori se ne mogu posmatrati izlovano jedan od drugog. Takođe, rezultati istraživanja u vezi sa efektima ovih izvora na odgovaračke pristrasnosti su nedosledni, i dok su u nekim istraživanjima efekti potvrđeni, u drugima nisu, ili su, pak, različitog smera. Postoji nekoliko meta-analitičkih studija u kojima su ispitivane neke od odgovaračkih pristrasnosti (npr. Batchelor & Miao, 2016) koje će biti spomenute, ali opšti je utisak da nedostaju sveobuhvatne studije koje se bave problematikom izvora odgovaračkih pristrasnosti na sistematičan način.

Izvori koji potiču od testa i uslova testiranja

Ajtemi testa su, u stvari, stimulusi koji izazivaju određena ponašanja, osećanja, misli i na koje ispitanici reaguju. Kao takvi, oni mogu uzrokovati odgovaračke pristrasnosti.

Format ajtema

Javljanje većine odgovaračkih pristrasnosti je moguće samo u slučaju određenog formata ajtema, pre svega skale procene. Istraživanja dosledno pokazuju, uključujući i meta-analitičku studiju, da se preferencija ekstremnih odgovora povećava kada se povećava broj stepeni procene u okviru skale i kada se povećava broj ajtema u testu (Batchelor & Miao, 2016). Preferencija srednjeg odgovora se, takođe,

povećava kada skala procene sadrži više podeoka, npr. 9 i više (Kieruj & Moors, 2010). Zanimljivo je da neki nalazi pokazuju da se preferencija ekstremnih odgovora smanjuje kada se svaki podeok na skali definiše (npr. 1 = *uopšte se ne slažem*, 2 = *delimično se ne slažem*...), umesto da se definišu samo ekstremni podeoci (Weijters et al., 2010).

Prema mišljenju pojedinih autora, povećanje broja podeoka povezano je s većom težinom zadatka, te će ispitanici koji nisu voljni da ulože veće kognitivne resurse u odgovaranje na stavke, pribegavati odgovaračkim pristrasnostima kao prečicama da formulišu i daju odgovor kojim će biti zadovoljni (Krosnick, 1991). Takođe, povećanje broja podeoka utiče na to da ispitanik bira samo određeni raspon odgovora, koji smatra optimalnim (Weathers, Sharma, & Niedrich, 2005). Na neki način, ispitanik sam daje referentni okvir podeocima i određuje koji podeoci će biti ekstremi u konkretnom slučaju. Oba objašnjenja se odnose na odgovaračke setove, kada broj podeoka na skali procene direktno utiče na način odgovaranja i odgovaračke pristrasnosti.

Zanimljivo je i to da su skorovi viši ukoliko skala procene počinje od stepena najvećeg slaganja (npr. od 5 do 1), umesto od stepena najmanjeg slaganja (od 1 do 5, videti Hartley & Betts, 2013).

Koji broj podeoka je optimalan?

Postoje različite preporuke u vezi s tim koji broj podeoka je optimalan kako bi udeo odgovaračkih pristrasnosti bio što manji, i prema nekim autorima izbor broja podeoka zavisi od svrhe korišćenja instrumenta (Weijters et al., 2010). Ukoliko je cilj da se dobiju direktne procene skorova, poput sumacionih skorova ili procenata, preporuka je da se koristi skala sa 5 ili 7 podeoka s deifnisanim svakim podeokom.

Naime, definisanje značenja svakog podeoka čini skalu interpretabilnijom, što poboljšava njenu validnost i pouzdanost (Krosnick, 1999). Ukoliko je cilj utvrđivanje povezanosti između varijabli ili linearnih modela, korišćenje petostepene skale s definisanim samo ektremnim podeocima se čini boljom opcijom, jer ovakve skale pokazuju bolju saglasnost s linearnim modelima (Weijters et al., 2010). Praktično, ukoliko je cilj testiranja klasifikacija ili samospoznaja i lični razvoj, pre treba koristiti definisane podeoke, a ukoliko se testiranje sprovodi u naučnoistraživačke svrhe, može se koristiti definisanje samo ekstremnih podeoka na skali procene.

Nejasni i složeni ajtemi

Kompleksnost ajtema može poticati od kompleksnosti samog konstrukta i njegove apstraktnosti, ali nekada je ajtem složen zbog drugih, više jezičkih karakteristika. Na primer, ajtem može biti složen ukoliko se jednim ajtemom pitaju dve stvari (npr. "Volim da čitam i idem u pozorište."), ukoliko se koristi dvostruka negacija, ili može biti nejasan ukoliko se koriste reči koje imaju višestruko značenje, izrazi poput "često" i "ponekad", ukoliko se koristi žargon ili kolokvijalizmi, ili ukoliko se koriste niskofrekventne reči. Nažalost, nekada ovakva pitanja služe za manipulaciju odgovorom. Na primer, pitanje postavljeno na referendumu u Republici Makedinoji 2018. godine je glasilo: "Da li ste za članstvo u EU i NATO, uz prihvatanje sporazuma o imenu između Republike Makedonije i Grčke". Naime, neko može da bude za članstvo u Evropskoj uniji, ali ne i za NATO, a neko može da bude za članstvo u Evropskoj uniji i NATO, ali ne i za prihvatanje sporazuma o promeni imena. Kako će onda odgovoriti? Navedeni složeni i nejasni ajtemi čine da je ispitanik nesiguran u vezi s tim kako da odgovori, da stvara

implicitne pretpostavke o tome šta je predmet merenja, te češće pribegava odgovaračkim usmerenjima kao ekonomičnom načinu odgovora na zahteve test-situacije. U prethodnim istraživanjima pokazano je da ovakvi ajtemi povećavaju verovatnoću povlađivanja ili neslaganja (npr. Krosnick & Presser, 2010). Da bi se izbegle odgovaračke pristrasnosti, pitanja moraju biti jasna i nedvosmislena, jezik prilagođen ispitanicima koji su ciljna grupa, i svi podeoci na skali definisani (a ne samo krajnji podeoci). Ukoliko je zaista složeni predmet merenja u pitanju, on se ispitanicima najpre treba definisati i objasniti preko primera.

Medijum testa

Pokazano je da u telefonskim anketama ispitanici više pokazuju tendenciju slaganja s tvrdnjama i preferenciju ekstremnih odgovora, a manje preferenciju srednjeg odgovora, u odnosu na anketiranje licem-u-lice, papir-olovka anketama ili onjaln (npr. Weijters et al., 2010). Takođe, u intervjuima (telefonskim ili licem-u-lice) se češće daju socijalno poželjni odgovori, pokazuje tendencija slaganja sa tvrdnjom i manje se ispitanici otvaraju u vezi sa osetljivijim temama ili informacijama, dok su opozitni obrasci zabeleženi kod papir-olovka ili kompjuterski podržanih upitnika (Browling, 2005). Pretpostavka je da je sumnja u anonimnost veća prilikom ličnog kontakta, te treba uložiti dodatne argumente u cilju osiguravanja anonimnosti u skladu sa etičkim principima. Ovakvi rezultati upućuju na oprez prilikom kombinovanja podataka istih mera koje su dobijene različitim medijumima procene.

Kognitivno opterećenje

Kognitivno opterećenje je posledica različitih faktora, među kojima su situacioni, kao što je prisutnost različitih distraktora (buke,

nedovoljnog svetla, generalno neadekvatnih ulova za rad), faktori koji potiču od stanja i motivacije ispitanika (npr. ako je umoran, gladan ili žedan, ako nije motivisan da se dodatno angažuje u odgovaranju na pitanja), faktori koji potiču od samog testa (npr. ukoliko se zadaje velika i zahtevna baterija testova, ukoliko su ajtemi kompleksni, nejasni ili teški, ukoliko sadrže dvojnu negaciju ili se jednim pitanjem pitaju dve stvari).

Ukoliko je situacija testiranja takva da opterećuje naše kognitivne resurse, pretpostavka je da će ispitanici pre pribegavati odgovaračkim pristrasnostima kao lakšem načinu za generisanje svog odgovora. U literaturi se to naziva fenomenom "zadovoljstva odgovorom" (eng. "*satisfice*"), koji je rezultat manjeg ulaganja truda, promišljanja, razumevanja, suđenja o pitanju i slično (Krosnick, 1991). Rezultati prethodnih istraživanja pokazuju da su u situaciji većeg kognitivnog opterećenja ispitanici skloniji slaganju s tvrdnjama (Knowles & Condon, 1999). S obzirom na to, preporučuje se da, kad god je to moguće, ispitanik sam bira vreme rada testa, ili mu se omogući neko vreme za odmor pre rada testa, ili pauze tokom rada. Takođe, testiranje ne bi trebalo da traje duže od sat vremena.

Dužina testa

Već je pomenuto da dugački testovi mogu biti izvor kognitivnog opterećenja, usled čega se javlja pribegavanje odgovaračkim pristrasnostima u većoj meri. Istraživanja su pokazala da se slaganje i nasumično odgovaranje češće javljaju na kraju dugačkog upitnika, kada su ispitanici generalno umorniji (više u MacKenzie & Podsakoff, 2012). Čak i visoko motivisani ispitanici, kao što su kandidati za posao, mogu

odgovarati nepažljivo na dužim testovima, što je povezano sa umorom i kognitivnim opterećenjem (Meade & Craig, 2012).

Vremensko organičenje

Kada postoji vremensko ograničenje rada testa ili bilo kakav drugi vremenski pritisak, veća je verovatnoća javljanja odgovaračkih pristrasnosti, npr. slaganja (npr. Furr & Bacharach, 2014; Krosnick & Presser, 2010).

Ispitivač

Ispitivač, takođe, može da utiče na odgovaračke pristrasnosti. Na primer, pokazano je da vešti intervjueri smanjuju grešku merenja, smanjuju neodgovaranje na ajteme i davanje socijalno poželjnih odgovora, dok efekti na slaganje nisu jednoznačni (više u Vaerenbergh & Thomas, 2013). Svakako da se intervjueri s više iskustva preferiraju za anketare, ali treba imati na umu da i uz njih odgovaračke pristrasnosti mogu da se jave.

Jezik testa

Adaptacija testa na drugi jezik je posebna procedura i čak i kad je ona sprovedena adekvatno i kada su dobijeni dokazi o psihološkoj ekvivalentnosti dve jezičke verzije testa, odgovaračke pristrasnosti se mogu javiti u zavisnosti od toga da li ispitanik radi test na maternjem ili sekundarnom jeziku. Tako, dobijeno je da su slaganje i preferencija ekstremnih odgovora veći, a preferencija srednjeg odgovora manja kada se test na radi na maternjem jeziku, umesto na sekundarnom (npr. Harzing, 2006). Kada se uzmu u obzir prednosti i nedostaci, generalno, primena testa na maternjem jeziku ispitanika pokazuje više prednosti.

Predmet merenja

Ponekad se dešava da sam predmet merenja utiče na pojavu odgovaračkih pristrasnosti. Na primer, može se pretpostaviti da su upitnici koji mere neke socijalno nepoželjne karakteristike, kao što su agresivnost, psihopatija i slično, podložniji davanju socijalno poželjnih odgovora. Ukoliko predmet merenja testa ili neka pojedina pitanja nisu ispitaniku važna ili nema jasno izgrađen stav o istim, manje će se angažovati da odgovori na ta pitanja, te će pre pribegavati odgovoračkim pristrasnostima (Gibbons, Zellner, & Rudek, 1999). S druge strane, ukoliko je ispitanik veoma zainteresovan za temu koja je predmet merenja testa, i jedva čeka da izrazi svoje mišljenje, manja je verovatnoća da će pribegavati odgovaračkim pristrasnostima. Takođe, moguće je da neke teme izazivaju nelagodu kod ispitanika (npr. pitanja o seksualnom ponašanju), a neka izazivaju strah (npr. pitanja o političkoj situaciji i liderima), te će ispitanik svesno davati ili socijalno poželjne odgovore ili neutralne.

Postoje oprečni rezultati u vezi s tim da li povlađivaje zavisi od predmeta merenja testa ili ne. Dok neki istraživači ističu da se povlađivanje dosledno javlja i u upitnicima ličnosti i u upitnicima stavova (Danner et al., 2015), drugi ističu da je povlađivanje veći problem u upitnicima stavova (Paulhus, 1991). Kao jedan od razloga zašto povlađivanje može biti veći problem u upitnicima stavova, navodi se što su ajtemi u tim upitnicima kompleksiji, dok su ajtemi u upitnicima ličnosti uglavnom pridevi ili proste rečenice.

Izvori koji potiču od ispitanika

Izvori koji potiču od ispitanika odnose se na socio-demografske karakteristike ispitanika, kognitivne sposobnosti, osobine ličnosti i

pripadnost određenoj kulturi. Prethodna istraživanja utvrdila su da je efekat pojedinačnih socio-demografskih karakteristika na odgovaračke pristrasnosti generalno nizak (do 8%, Weijters et al., 2010). Takođe, neki autori nude objašnjenje da su efekti socio-demografskih karakteristika ispitanika pre odraz efekata osobina ličnosti (npr. Moors, 2008), a istraživanja pokazuju da efekat pojedinih osobina ličnosti na odgovaračke pristrasnosti nije zanemarljiv (npr. Paulhus, 2002).

Pol

Prema jednoj meta-analitičkoj studiji, polne razlike u preferenciji ekstremnih odgovora postoje u korist žena, ali su praktično zanemarljive (Koenov d = 0,09[33], Batchelor & Miao, 2016). U slučaju ostalih odgovaračkih pristrasnosti nema meta-analitičkih studija, a nalazi pojedinačih studija su vrlo nekonzistenti. Na primer, u nekim istraživanjima je dobijeno da su žene sklonije slaganju (Weijters et al., 2010) i preferenciji srednjih odgovora (Harzing, 2006), dok u drugim, uglavnom starijim istraživanjima, nisu dobijeni značajni efekti pola (npr. Marin, Gamba, & Marin, 1992; Grimm & Church, 1999).

Starost

Odnos preferencije ekstremnih odgovora i starosti je izgleda nelinearan, tačnije kurvilinearan (obrnuto U), tj. ove pristrasnosti rastu do ranih 20-ih godina, nakon čega opadaju (Batchelor & Miao, 2016). U slučaju slaganja, većina istraživanja je potvrdila pozitivnu vezu sa starošću, dok nisu nađene značajne veze sa neslaganjem i preferencijom srednjih odgovora (više u Weijters et al., 2010).

[33] Veličina efekta izražena preko Koenovog d statistika se tumači na sledeći način: od 0,2 do 0,5 je mala veličina efekta, od 0,5 do 0,8 srednja, i preko 0,8 velika (Cohen, 1988).

Ekonomski status

Većina istraživanja potvrđuje negativnu vezu između slaganja i preferencije ekstremnih odgovora s ekonomskih statusom (npr. Meisenburg & Williams, 2008).

Obrazovanje

Prema meta-analitičkoj studiji, godine obrazovanja nisu povezane s preferencijom ekstremnih odgovora, ali varijansa u okviru godina obrazovanja je veoma jak moderator (Batchelor & Miao, 2016). Naime, dobijeno je da nema značajne povezanosti sa preferencijom ekstremnih odgovora u uzorcima u kojima obrazovanje ima malu varijansu, ali da je veza značajno negativna u uzorcima s velikom varijansom varijable obrazovanja, te se može očekivati da je ova pristrasnost veća kod manje obrazovanih. U drugim istraživanjima je potvrđena negativna veza i između nivoa obrazovanja i drugih pristrasnosti kao što su slaganje i preferencija srednjih odgovora, dok neslaganje nije bilo značajno povezano sa nivoom obrazovanja (Weijters et al., 2010).

Inteligencija

Inteligencija je pozitivno povezana s obrazovanjem, te su efekti odgovaračkih pristrasnosti koji se dobijaju u odnosu na inteligenciju, podudarni s onima koji se dobijaju u odnosu na obrazovanje. Meta-analitička studija je pokazala da je inteligencija negativno povezana sa preferiranjem ekstremnih odgovora (Batchelor & Miao, 2016). Takođe, niže izražene kognitivne sposobnosti, posebno verbalne, rigidna mentalna organizacija i niska tolerancija na kognitivnu kompleksnost su povezane sa slaganjem (npr. Gudjonsson & Young, 2001; Knowles & Nathan, 1997). Kognitivne sposobnosti su povezane i sa davanjem

socijalno poželjnih odgovora, te ispitanici s višim kognitivnim sposobnostima, a posebno emocionalnom inteligencijom, češće biraju socijalno poželjne odgovore jer lakše mogu da prepoznaju zahteve situacije i očekivanja u vezi s datom situacijom (npr. Snell, Sydell, & Lueke, 1999). Ipak, rezultati nisu tako jednoznačni u pogledu relacija inteligencije i davanja socijalno poželjnih odgovora, te je u nekim istraživanjima dobijena niska negativna veza u situaciji selekcije, ali je pokazano da, kada se parcijalizuje efekat prijatnosti, korelacija biva neznačajna, što ide u prilog stanovištu da je socijalna poželjnost supstantivna varijansa osobina ličnosti (Knežević, 2014).

Osobine ličnosti

Već je bilo reči o tome da socijalna poželjnost predstavlja supstantivnu varijansu osobina ličnosti, pre svega savesnosti, prijatnosti i emocionalne stabilnosti (Connelly & Chang, 2016; McCrae & Costa, 1983). Postavlja se pitanje da li su osobine ličnosti povezane i sa drugim pristrasnostima. U jednoj metodološki zanimljivoj studiji pokazano je da su osobine ličnosti značajno povezane sa odgovaračkim pristrasnostima, pre svega sa bazičnom osobinom prijatnost (He, Bartram, Inceoglu, & van de Vijver, 2014). U ovoj studiji je ispitivana povezanost odgovaračkih pristrasnosti i osobina ličnosti merenih preko ajtema s prisilnim izborom i Likertovom skalom. S obzirom na to da kod ajtema s prisilnim izborom nije moguće ekstrahovati odgovaračke pristrasnosti koje se odnose na slaganje i preferenciju ekstremnih ili srednjeg odgovora, bilo koja značajna korelacija između osobina ličnosti merenih preko ovih ajtema i odgovaračkih pristrasnosti dobijenih na osnovu ajtema s Likertovom skalom, upućuje na zaključak da odgovaračke pristrasnosti sadrže smislenu varijansu individualnih razlika. Tako,

dobijeni rezultat pokazuje da je prijatnost pozitivno povezana sa preferencijom srednjeg odgovora, a negativno s preferencijom ekstremnih odgovora i, neočekivano, negativno sa socijalnom poželjnošću. U istom istraživanju je dobijen rezultat da je savesnost pozitivno povezana sa slaganjem i preferencijom ekstremnih odgovora, a negativno sa preferencijom srednjeg odgovora, dok neuroticizam, ekstraverzija i otvorenost nisu značajno bile povezane sa odgovaračkim pristrasnostima. U drugom istraživanju je potvrđena pozitivna veza između preferencije ekstremnih odgovora i savesnosti, ali je dobijena i pozitivna veza sa ekstraverzijom (Austin, Deary, & Egan, 2006). Ekstraverzija, impulsivnost i traženje uzbuđenja su, takođe, pozitivno povezane sa slaganjem, a suprotne karakteristike, poput introverzije, s tendencijom neslaganja (Baumgartner & Steenkamp, 2001). Knežević (2014) je na uzorku u uslovima selekcije zapazio da se socijalna poželjnost povezuje s adaptivnim i funkcionalnim karakteristikama ličnosti, kao što su viša savesnost, prijatnost i emocionalna stabilnost, dok se slaganje povezuje s karakteristikama koje su predispozicija za razvoj psihopatologije, kao što su dezintegracija i neuroticizam, uz višu otvorenost.

U istraživanjima u kojima su korišćene druge karakteristike ličnosti koje nisu bazične, dobijeno je da rigidne, anksiozne i osobe koje imaju nisku toleranciju na neizvesnost i nejasnoću, preferiraju ekstremne odgovore (Baumgartner & Steenkamp, 2001; Vaerenbergh & Tomas, 2013). S druge strane, nisko samopouzdanje i percipirana samoefikasnost povezane su s preferencijom srednjeg odgovora, slično kao što su nesigurnost i konformizam povezani sa slaganjem (Baumgartner & Steenkamp, 2001; Vaerenbergh & Tomas, 2013).

Na osnovu navedenih nalaza većina autora zaključuje da je teško, ako ne i nemoguće ekstrahovati varijansu odgovaračkih pristrasnosti iz upitika, jer se na taj način kontaminiraju same upitničke mere i skorovi na njima (više u Kieruj & Moors, 2010). Naime, ukoliko se ukloni varijansa odgovaračkih pristrasnosti iz upitnika ličnosti, postavja se pitanje šta predstavlja preostala varijansa i vrlo je moguće da ona predstavlja grešku merenja.

Motivacija

Motivacija je već spomenuta kao faktor koji može uticati na odgovaračke pristrasnosti. Generalno, ukoliko ispitanici nisu motivisani da odgovaraju, ukoliko nisu zainteresovani za samu temu ili nemaju jasno izgrađen stav o predmetu merenja, ukoliko nisu na neki način kognitivno pobuđeni da odgovaraju na konkretna pitanja, ili misle da njihovo učešće ništa neće doprineti ili promeniti, ili odgovaraju samo da bi ispunili neku obavezu ili se prosto demotivišu kada vide da je baterija predugačka, veća je verovatnoća da će pribeći odgovaračkim pristrasnostima (Baumgartner & Steenkamp, 2001). Ispitanici će češće birati opcije "ne znam" ili "nemam mišljenje" kada nisu motivisani da učestvuju u istraživanju i pred kraj dugačkog upitnika ili baterije, a ređe će birati ove opcije kada su motivisani, zainteresovani za temu, razumeju je, ili kada su podstaknuti da odgovore na pitanja (više u MacKenzie & Podsakoff, 2012). Da bi se ispitanici motivisali da iskreno odgovore na pitanja i da im pažljivo pristupe, treba ih podstaći na to. Motivacija kod ispitanika se može pobuditi tako što će im se objasniti svrha istraživanja, zašto je važno da oni učestvuju u ispitivanju i koji će doprinos dobiti od njihovog učešća. Instrukcije poput "važno nam je Vaše mišljenje", "cenimo Vaše iskreno mišljenje", "recite nam šta

mislite" mogu takođe da deluju stimulušuće. Davanje fidbeka o rezultatima ispitivanja je takođe motivišuće za ispitanike, i to treba praktikovati kad je to moguće.

Rasa

Rezultati meta-analitičke studije pokazuju da najmanju preferenciju ka biranju ekstremnih odgovora imaju Azijati, pa potom belci, hispanci i najveću preferenciju ka ovim pristrasnostima pokazuju Afroamerikanci, iako su dobijene veličine efekta ovih razlika male ili praktično zanemarljive (Batchelor & Miao, 2016). Ovaj nalaz potencijalno može ukazivati na veću sklonost ka odgovaračkim pristrasnostima kod pripadnika manjina.

Kultura

U jednom istraživanju stavova u 11 zemalja zaključeno je su variranja u odgovaračkim pristrasnostima između zemalja manja, u odnosu na variranja u zavisnosti od predmeta merenja skala (Baumgartner & Steenkamp, 2017). Ipak, u istoj studiji pokazano je da su odgovaračke pristrasnosti najprisutnije kod Grka i Portugalaca. U drugoj studiji dobijeno je da Amerikanci pokazuju veću preferenciju biranja ekstremnih odgovora u odnosu na Azijate (iz Japana i sa Tajvana), Kanađane i Francuze, a manju u odnosu na Meksikance i Australijance (Batchelor & Miao, 2016). S obzirom na to, prilikom kros-kulturalne adaptacije upitnika uvek se preporučuje da se vodi računa i o razlikama u odgovaračkim pristrasnostima, te da se one uvedu kao metodski faktori pri ispitivanju ekvivalencije dve jezičke verzije upitnika.

Opšte tehnike prevazilaženja

Napraviti "čistu" meru koja ne sadrži grešku merenja, nemoguć je posao. Svako merenje je merenje s greškom, a pitanje je samo kolika je ta greška, da li je možemo detektovati i kontrolisati. Odgovaračke pristrasnosti imaju različiti status, te se u jednom slučaju smatraju greškom (npr. Edwards, 1957), a u drugom važnim izvorom varijanse koji nam može pomoći u proceni (npr. Cronbach, 1946). Jedan od argumenata koji se može iskoristiti za razrešenje statusa konkretne odgovaračke pristrasnosti je njena **prediktivna validnost**. Naime, ukoliko odgovaračka pristrasnost doprinosi predikciji neke važne ishodne varijable, svakako se ne može smatrati greškom. Na primer, u jednom istraživanju dobijen je rezultat da se na osnovu nasumičnog odgovaranja kod muških srednjoškolaca može značajno predvideti manja verovatnoća odlaska na fakultet, manje postignuće na maturskom ispitu i veća verovatnoća da se učenik percipira kao delinkvent (Lanning, 1989). Međutim, ukoliko se varijansa odgovaračkih pristranosti preklapa sa varijansom opaženih skorova, onda se odgovaračke pristrasnosti smatraju greškom jer narušavaju razumevanje konstrukta, tj. konstrukt validnost. Dakle, da bi se odgovaračka pristrasnost smatrala važnim izvorom varijanse koji ima interpretabilnu vrednost, potrebno je da objašnjava ishod mimo varijanse ishoda koju objašnjavaju opaženi skorovi na nekoj meri. Ipak, bez obzira na dokaze o jedinstvenoj prediktivnoj validnosti odgovaračkih pristrasnosti, češće se na njih gleda kao na nepoželjnu varijansu koju treba eliminisati iz skorova ispitanika.

Pored specifičnih načina detekcije i eliminisanja pojedinačnih odgovaračkih pristrasnosti, neke opšte preporuke kako bi one mogle da se tretiraju su sledeće (Domino & Domino, 2006):

1. u testovni materijal treba uvrstiti posebne mere za procenu sklonosti ka lažiranju i socijalnoj poželjnosti (npr. lažne tvrdnje), kontrolne ajteme za procenu pažnje ispitanika ili izračunati skorove za odgovaračke pristrasnosti, te proveriti njihove korelacije sa merama od interesa;
2. kad god je to moguće, treba uporediti rezultate dobijene različitim metodama prikupljanja podataka (upitnik, intervju, posmatranje), tj. treba uporediti rezultat na upitniku s rezultatom dobijenim drugom metodom;
3. kad god je to moguće, treba uraditi kontrolni intervju nakon testiranja, posebno s kandidatima kod kojih postoji sumnja da su lažirali ili neslaganje rezultata dobijenih različitim metodama procene; intervjuom se, pored provere dobijenih rezultata, mogu dobiti i važne informacije o ponašanju ispitanika;
4. preliminarno bi trebalo proveriti koliko je upitnik podložan odgovaračkim pristrasnostima, tj. da li se skorovi istog ispitanika menjaju u zavisnosti od instrukcije;
5. ukoliko to ne narušava proceduru zadavanja testa, može se dati upozorenje o tome da se lažiranje može detektovati, čak i strože upozorenje da će ispitanici snositi negativne posledice ukoliko se kod njih detektuje lažiranje. Prethodna istraživanja potvrđuju da ova upozorenja doprinose iskrenijem odgovaranju na testu (Parmač Kovačić, Galić, & Jerneić, 2014). Pritom, bitno je da upozorenje deluje ubedljivo.

Savremena praksa nalaže da se odgovaračke pristrasnosti uvek detektuju i uvrste u model. Razvijeno je nekoliko modela kojima se mogu definisati faktori odgovaračkih pristrasnosti i proceniti u okvru

modela merenja. Oni su uglavnom bazirani ili na pristupu konfirmatorne faktorske analize ili Teorije odgovora na stavke (više u Falk & Cai, 2016; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), međutim njihov opis prevazilazi ciljeve ove knjige.

Pitanja za razmišljanje

1. Koja je razlika između odgovaračkih stilova i setova? Navedi primer u kojem se jedna ista odgovaračka pristrasnost može manifestovati i kao stil i kako set?
2. Na koja dva najvažnija aspekta merenja utiču odgovaračke pristrasnosti?
3. Na koje sve načine se davanje socijalno poželjnih odgovora može smanjiti? Po Vašem mišljenju, koji je među njima najbolji i zašto?
4. Zašto su negativno formulisane tvrdnje problematične?
5. Koji su izvori odgovaračkih pristrasnosti koji potiču od testa i uslova primene testa s jedne strane, i ispitanika s druge strane, povezani sa slaganjem?

Reference

Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69*, 571–580.

Arce-Ferrer, A. J., & Ketterer, J. J. (2003). The effect of scale tailoring for cross-cultural application on scale reliability and construct validity. *Educational and Psychological Measurement, 63*(3), 484–501.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO–FFI items. *Personality and Individual Differences, 40*(6), 1235–1245.

Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: Symptoms, diagnosis and prescribed cure. *Scandinavian Journal of Psychology, 54*, 152–159.

Baka, A., Figgou, L., & Triga, V. (2012). "Neither agree, nor disagree": A critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance, 5*(3), 244–263.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 3*4(2), 163–175.

Batchelor, J. H., & Miao, C. (2016). Extreme response style: A meta-analysis. *Journal of Organizational Psychology, 16*(2), 51–62.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143–156.

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2017). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment.* Online first. https://doi.org/10.1177/1073191117700268

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*(3), 186–204.

Bing, M. N., Kluemper, D., Kristl Davison, H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes, 116*(1), 148–162.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6*(3), 199–214.

Brajša-Žganec, A., Ivanović, D., & Lipovčan, LJ. K. (2011). Personality traits and social desirability as predictors of subjective well-being. *Psychological Topics, 20*(2), 261–276.

Browling, A. (2005). Mode of questionnaire asministration can have serious effects on data quality. *Journal of Public Health, 27*(3), 281–291.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*(4), 358–372.

Cattell, R. B., Cattell, A. K. S., & Cattell, H. E. P. (1993). *Sixteen Personality Factor Quiesstionnaire* (5. ed). Champaign, IL: Institute for Personality and Ability Testing, Inc.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319.

Connelly, B. S., & Chang, L. (2016). A meta-analytic multitrait multirater separatio of substance and style in social desirability scales. *Journal of Personality, 84*(3), 319–334.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6,* 475–494.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: John Wiley & Sons.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354.

Danner, D. & Rammstedt, B. (2016). Facets of acquiescence: Agreeing with negations is not the same as accepting inconsistency. *Journal of Research in Personality, 65*, 120–129.

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: relevance, domain specificity, and stability. *Journal of Research in Personality, 57*, 119–130.

DeVallis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Dinić, B., & Radević, L. (2019). *Faking aggressiveness: Scores' changes under different faking instructions.* Paper presented at the scientific conference Current Trends in Psychology 2019 (STuP 2019), Novi Sad, Serbia.

Dinić, B., Mitrović, D. i Smederevac, S. (2014). Upitnik BODH (bes, osvetoljubivost, dominacija, hostilnost): novi upitnik za procenu agresivnosti. *Primenjena psihologija, 7-1*(Dodatak), 297–324.

Dilchert, S., & Ones, D. S. (2012). Application of preventive strategies. In M. Zeigler, C. MacCann, & R. D. Roberts (Eds.), *New perspective on faking in personality assessment* (pp. 177–200). New York: Oxford University Press.

DiStefano, C., & Motl, R. W. (2009). Personality correlates of method effects due to negatively worded items on the Rosenberg Self–Esteem scale. *Personality and Individual Differences, 46*, 309–313.

Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction (2nd ed.).* New York: Cambridge University Press.

Drobnjaković, E., Cmiljanović, M., Dinić, B. i Janičić, B. (2016). *Socijalno poželjno odgovaranje i HEXACO model ličnosti*. Rad prezentovan na 64. Naučno–stručnom skupu Društva psihologa Srbije, Zlatibor, Srbija.

Dwight, S. A. & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1–23.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: The Dryden Press.

Ehlers, C., Greene–Shortridge, T. M., Weekley, J. A., & Zajack, M. D. (2009). *The exploration of statistical methods in detecting random responding.* Paper presented at the Annual meeting of the Society for Industrial/Organizational Psychology, Atlanta, GA.

European Federation of Psychologists' Associations – EFPA (2013). *FPA review model for the description and evaluation of psychological tests: Test review form & notes (v.4.2.6).* Preuzeto sa http://www.efpa.eu/download/650d0d4ecd407a51139ca44ee704fda4

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Kent, UK: Hodder & Stoughton.

Fajgelj, S. (2013). *Psihometrija: metod i teorija psihološkog merenja (4. dopunjeno izdanje)*. Beograd, Srbija: Centar za primenjenu psihologiju.

Falk, C. F., & Cai, L. (2016). A flexibile full-information approach to the modeling of response styles. *Psychological Methods, 21*(3), 328–347.

Furr, M. R., & Bacharach, V. R. (2014). *Psychometrics: An introduction*. Thousand Oaks, Califormia: SAGE Publication Inc.

Galić, Z., Jerneić, Ž., & Prevendar, T. (2008). Socijalno poželjno odgovaranje, ličnost i inteligencija u selekcijskoj situaciji. *Psihologijske teme, 17*(1), 93–110.

Galić, Z., Jerneić, Ž. i Belavić, M. (2009). O svecima i superherojima: provjera Pauhusova modela socijalne poželjnosti. *Društvena istraživanja, 6*(104), 977–997.

Gough, H., & Bradley, P. (2005). *CPI 260 manual*. Mountain View: Consulting Psychologists Press.

Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. *Cross-Cultural Research, 33*(4), 369–381.

Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology, 1*, 308–311.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341–355.

Grimm, S. D., & Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality, 33*(4), 415–441.

Gudjonsson, G. H., & Young, S. J. (2001). Personality and deception. Are suggestibility, compliance and acquiescence related to socially

desirable responding? *Personality and Individual Differences, 50*, 192–195.

Hartley, J., & Betts, L. (2010). Four layouts and a finding: The effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology, 13,* 17–27.

Harzing, A. W. (2006). Response styles in cross–national survey research: A 26–country study. *International Journal of Cross Cultural Management, 6*(2), 243–266.

He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology, 45*(7), 1028–1045.

Holden, R. R., & Passey, J. (2009). Social desirability. In M. R. Leary, & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 441–454). New York, NY, US: The Guilford Press.

Holden, R. R., Book, A. S., Edwards, M. J, Wasylkiw, L., & Starzyk, K. B. (2003). Experimental faking in self-reported psychopathology: Unidimensional or multidimensional? *Personality and Individual Differences, 35*, 107–1117.

Holden, R. R., Kroner, D. G., Fakken, G. C., Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63*, 272–279.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114.

Jackson, D. N. (1984). *Personality Research Form manual (3rd ed.).* Port Huron, MI: Research Psychologists Press.

Jacobson, L. I., Kellogg, R. W., Cauce,A. M., & Slavin, R. S. (1977). A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society, 9*(2), 109–110.

Jerneić, Ž., Galić, Z. i Parmač, M. (2007). *Prijevod i adaptacija upitnika Sveobuhvatni upitnik socijalno poželjnog odgovaranja (Comprehensive Inventory of Desirable Responding).* Neobjavljeni materjal. Zagreb: Odsjek za psihologiju Filozofskog fakulteta.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103–129.

Kam, C., Risavy, S. D., & Perunovic, W. Q. E. (2015). Using over-claiming technique to probe social desirability ratings of personality items: A validity examination. *Personality and Individual Differences, 74,* 177–181.

Kašiković, B., Jerotijević, I., Novaković, S., Mijatović, K. i Žeželj, I. (2013). Kako izmeriti socijalno poželjno odgovaranje: Adaptacija Paulusovog testa "Precenjivanja znanja". Rad prezentovan na 19. Empirijskim istraživanjima u psihologiji, Beograd, Srbija.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *Public Opinion Research, 22*(3), 320–342.

Knežević, G. (2014). *Profesionalna selekcija: elementi, izazovi i jedno praktično rešenje. Knjiga II: Ličnost.* Beograd, Srbija: Centar za primenjenu psihologiju.

Knežević, G., Radović, B., & Peruničić, I. (2008) Can amorality be measured. Paper presented at the European Conference on Personality (XIV), Tartu, Estonia.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*(2), 379–386.

Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern? *Journal of Research in Personality, 31*(2), 293–301.

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In J. D. Wright, & P. V. Marsden (Eds.), *Handbook of Survey Research* (2nd Edition, pp. 263–313). San Diego, CA: Elsevier.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.

Kuncel, N. R., & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology, 62*(2), 201–228.

Kurtz, J. E., & Parish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO–PI–R. *Journal of Personality Assessment, 76*(2), 315–332.

Lanning, K. (1989). Detection of invalid response patterns on the California Psychological Inventory. *Applied Psychological Measurement, 13*, 45–56.

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*(2), 131–141.

MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing, 88*(4), 542–555.

Malhotra, N., Krosnick, J. A., & Thomas, R. K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quartely, 73*, 304–324.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology, 23*(4), 498–509.

Marsh, H. (1996). Positive and negative global self-esteem: A substantively

meaningful distinction or artifactors? *Journal of Personality and Social*

Psychology, 70, 810–819.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*(6), 882–888.

McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: effects on faking behavior and test measurement properties. *Journal of Personality Assessment, 78*(2), 348–369.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.

Mentus, T. (2014). *Značenje vremena odgovaranja u ispitivanju ličnosti* (neobjavljena doktorska disertacija). Filozofski fakultet, Univerzitet u Beogradu, Beograd, Srbija.

Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of

gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality & Quantity, 37*(3), 277–302.

Morey, L. C. (1991). The Personality Assessment Invenoryprofessional manual. Odessa, FL: Psychological Assessment Resources.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology, 142*(2), 71–89.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, *63*, 1–11.

Okanović, P. (2013). *Lažiranje testova u profesionalnoj selekciji: model baziran na Teoriji očekivanja* (neobjavljena doktorska disertacija). Filozofski fakultet, Univerzitet u Novom Sadu, Novi Sad, Srbija. Preuzeto sa http://www.ff.uns.ac.rs/studije/doktorske/uvid_javnosti/Predrag_Okanovic_disertacija.pdf

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assesment for personnel selection. *Human Performance, 11*(2–3), 245–269.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660–679.

Ong, A. D., & Weiss, D. J. (2000). The impact of anonymity of responses to sensitive questions. *Journal of Applied Social Psychology, 30*(8), 1691–1708.

Pace, V. L., & Borman, W. (2006). The use of warnings to discourage faking on noncognitive inventories. In M. H. Peterson, & R. L. Griffith (Eds.), *A closer examination of applicant faking behavior*

(pp. 281–302). Greenwich, Connecticut: Information Age Publishing.

Paulhus, D. L., & Bruce, N. (1990). *Validation of the OCQ: A preliminary study*. Paper presented at the annual convention of the Canadian Psychological Association, Ottawa, Ontario, Canada.

Paulhus, D. L., & Dubois, P. J. (2014). Application of the overclaiming technique to scholastic assessment. *Measurement, 74*(6), 975–990.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic bias in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025–1060.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in social desirable responding. *Journal of Personality and Social Psychology, 60,* 307–317.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychology attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Paulhus, D. L. (2002). Social desirable responding: The evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for the BIDR Version 6.* Unpublished manuscript. Department of Psychology, University of British Columbia at Vancouver.

Paulhus, D. L. (2006). *A Comprehensive Inventory of Desirable Responding (CIDR).* Poster presented at the Annual Meeting "The Association for Research in Personality", New Orleans.

Paulhus, D. L. (2018). The big two dimensions of desirability. In A. Abele, & B. Wojiscke (Eds.), *Agency and communion in social psychology* (pp. 79–89). New York: Taylor & Francis.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21,* 100–108.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*(4), 890–904.

Perinelli, E., & Gremigni, P. (2016). Use of social desirability scales in clinical psychology: A systematic review. *Journal of Clinical Psychology, 72*(6), 534–551.

Peterson, R. A., Rhi-Perez, P., & Albaum, G. (2014). A cross-national comparison of extreme response style measures. *International Jorinal of Market Research, 56*(1), 89–110.

Phillips, D. L., & Clancy, K. J. (1972). Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology, 77*(5), 921–940.

Plutchik, R., & Kellerman, H. (1964). *Theories of emotion* (1st edition). London: Academic Press.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 62,* 539–569.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review

of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Parmač Kovačić, M., Galić, Z., & Jerneić, Ž. (2014). Social desirability scales as indicators of self-enhancement and impression management. *Journal of Personality Assessment, 96*(5), 532–543.

Parmač, M., Galić, Z. i Jerneić, Ž. (2009). Vrijeme latencije kao indikator iskrivljavanja odgovora na upitnicima ličnosti. *Suvremena psihologija, 12*(1), 13–60.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding, *Personality and Individual Differences, 35*, 263–275.

Robie, C., Curtin, P. J., Foster, T. C., Phillips, H. L., Zbylut, M., & Tetrick, L. E. (2000). The effects of coaching on the utility of response la tencies in detecting fakers on a personality measure. *Canadian Journal of Behavioural Science, 32*, 226–233.

Rogers, R., Bagby, M., & Gillis, J. R. (1992). Improvements in the M test as a screening measure for malingering. *Bulletin of the American Academy Psychiatry and Law, 20*(1), 101–104.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*(2), 155–180.

Sedikides, C., & Alicke, M. D. (2018, in press). The five pillars of self-enhancement and self-protection. In R. M. Ryan (Ed.), *The Oxford Handbook of Motivation (2nd ed.).* New York: Oxford University Press.

Snell, A. F., Sydell, E. J. & Lueke, S. B. (1999). Towards a theory of applicant faking: integrating studies of deception. *Human Resource Management Review, 9*, 219–242.

Sturgis, P., Roberts, C., & Smith, P. (2012). Middle aleternatives revisited: How the neither/nor response acts a way of saying "I don't know"? *Sociological Methods & Research, 43*, 15–38.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.

Tellegen, A. (1982). *Brief manual for the Multidimensional Personality Questionnaire.* Unpublished manuscript. Minneapolis: University of Minnesota.

Tonković, M., Galić, Z., & Jerneić, Ž. (2011). The sonstruct validity of over-claiming as a measure of egoistic enhancement. *Review of Psychology, 18*(1), 11–19.

Vaerenbergh, Y. V., & Thomas, T. D. (2013). Response style in survey research: A literature review of antecedents, consequences, and remedies. International *Journal of Public Opinion Research, 25*(2), 195–217.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling, 13*, 99–117.

Velez, P., & Ashworth, S. D. (2007). The impact of item readability on the nedorsement of the midpoint response in surveys. *Survey Research Method, 1*, 69–74.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measure of Positive and Negative Affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.

Weathers, D., Sharma, S., & Niedrich, R. W. (2005). The impact of the number of scale points, dispositional factors, and the status quo

decision heuristic on scale reliability and response accuracy. *Journal of Business Research*, 58(11), 1516–1524.

Weems, G. H., Onwuegbuzie, A. J., Schreiber, J. B., & Eggers, S. J. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. *Assessment and Evaluation in Higher Education, 28* (6), 587–607.

Weijters, B. (2006). *Response styles in consumer research.* Doctoral dissertation. Ghent University. Faculty of Economics and Business Administration. Retrieved from https://biblio.ugent.be/ publication/4100284

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods, 15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science, 36*, 409–422.

Wiggins, J. S. (1966). Social desirability estimation and "faking good" well. *Educational and Psychological Measurement, 26*(2), 329–341.

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti, & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl, Vol. 1. Matters of public interest; Vol. 2. Personality and psychopathology* (pp. 89–113). Minneapolis, MN, US: University of Minnesota Press.

Ziegler, M., MacCann, C., & Roberts, R. D. (Eds.). (2012). *New perspectives on faking in personality assessment*. New York, NY, US: Oxford University Press.

IZAZOVI PRIMENE TESTA

Već od samog početka korišćenja, za testove se vezuju različite kontroverze, posebno u slučaju testova visokog uloga. S kraja 19. veka pa sve do polovine 20. veka u Sjedinjene Američke Države (SAD) su dolazili imigranti iz različitih evropskih zemalja, te se razvila potreba za selekcijom imigranata, a selekcija se obavljala, između ostalog, na osnovu procene inteligencije. Na osnovu rezultata na testovima donosila se odluka o tome da li će nekome biti odobren ulazak u SAD ili ne, što ima veliki uticaj na život datog imigranta. Mnogi imigranti su odbijeni zbog toga što su ostvarivali niske skorove na testovima inteligencije. U tadašnje vreme se smatralo da je inteligencija urođena, te su pravljene rang-liste inteligencije imigranata po rasi i nacionalnosti. Tako su neke nacionalnosti bile proglašene "slaboumnima" s obzirom na to da su njeni pripadnici ostvarivali dosledno niže rezultate na testovima inteligencije (više u Hedrih, 2018). Tek se kasnije uvidelo da postoji problem prilikom komparacije rezultata između pripadnika različitih kultura, tj. jezika, te da testovi moraju biti nezavisni od kulture da bi rezultati na njima bili uporedivi. Današnje stanovište je da je inteligencija od 40 do 60% nasledna (npr. Plomin, DeFries, Craig, & McGuffin, 2003) i da na nju deluju genski i različiti sredinski činioci. Pored toga, danas se zna da ne postoji test koji je nezavistan od kulture, te da se u svakoj kulturi test mora standardizovati i da svaka kultura mora imati svoje norme (Hedrih, 2018).

Današnja praksa adekvatne primene testova regulisana je etičkim principima i standardima koji su najbolje uobličeni u okviru Etičkih principa psihologa i kodeksa ponašanja koje je izdala Američka

psihološka asocijacija (APA, 2017) i Standarda za pedagoško i psihološko testiranje (AERA, APA, & NCME, 2006). Pored ovih opštih standarda, renomirana profesionalna udruženja takođe imaju propisane etičke principe, npr. Američka asocijacija savetnika (American Counseling Association – ACA)[34]. U poslednjih nekoliko decenija sve je više regulativa o pravilnoj upotrebi testova i to za specifične potrebe i uslove. Na primer, APA je publikovala smernice za odgovorno i kompetetno korišćenje testova (Turner, DeMers, Fox, & Reed, 2001), a Međunarodna komisija za testove (International Test Commision – ITC)[35] nudi široki opseg smernica za različite primene testa, npr. za prevod i adaptaciju testova, za računarski podržano i onlajn testiranje itd.

Kod nas je upotreba testova regulisana Kodeksom etike psihologa Srbije[36] (Društvo psihologa Srbije, 2000) u okviru kojeg su propisana pravila psihološkog testiranja i dužnosti psihologa, od pripreme za psihološko testiranje do interpretacije i prezentacije rezultata testiranja, uključujući i čuvanje psiholoških mernih instrumenata i dokumentacije o ispitanicima.

[34] https://www.counseling.org/docs/default-source/ethics/2014-code-of-ethics.pdf?sfvrsn=2d58522c_4
[35] https://www.intestcom.org/page/5
[36] http://dps.org.rs/images/kodeksEtike.pdf

Standardi za pedagoško i psihološko testiranje

Već su na nekoliko mesta u tekstu pomenuti Standardi za pedagoško i psihološko testiranje. Oni predstavljaju najvažniji izvor kriterijuma za evaluaciju testova, praktičnu primenu testova i utvrđivanje efekata njihove upotrebe. Standardi su nastali kao rezultat saradnje tri udruženja: 1. Američke asocijacije za istraživanje u obrazovanju (American Educational Research Association – AERA), 2. Američke psihološke asocijacije (American Psychological Association – APA) i 3. Nacionalnog veća za merenje u obrazovanju (National Council on Measurement in Education – NCME). Cilj Standarda je promovisanje kvalitetne i etične upotrebe testova i drugih standardizovanih postupaka, kao i uspostavljanje kriterijuma za ocenu testova, kvaliteta postupaka testiranja i utvrđivanja posledica upotrebe testova (AERA, APA, & NCME, 2006). U sadašnjem obliku, Standardi obuhvataju tri dela:

1. izrada, evaluacija i dokumentacija testova, u kojem se opisuju osnovne psihometrijske karakteristike – validnost i pouzdanost, kao i greška merenja, te način izrade i preuređivanja testova, normiranje i uporedivost rezultata, primena testa, ocenjivanje izveštavanja o rezultatima, oprema i dokumentacija testova;
2. nepristrasnost u testiranju, u kojem se opisuju uslovi postizanja nepristrasnosti, prava i odgovornosti ispitanika, testiranje ispitanika različitog jezičkog porekla, testiranje osoba s invaliditetom;
3. primena testiranja, u kojem se opisuju odgovornosti korisnika testova, te psihološko testiranje i procena u različitim kontekstima kao što su obrazovanje, zapošljavanje i provera kvalifikacija, evaluacija programa i socijalna politika.

Standardi ne podrazumevaju obavezujuće postupke prilikom evaluacije testova, psihometrijske instrukcije, niti su zamena za zakonske propise, već su to preporuke za unapređenje testovne delatnosti. Svaki praktičar bi trebao biti upoznat s ovim Standardima i primenjivati ih u svojoj testovnoj praksi.

Prvi Standardi su objavljeni 1966. godine, pod nazivom Standardi za pedagoške i psihološke testove i priručnike, potom su 1974. objavljeni pod sadašnjim nazivom, a u međuvremenu su revidirani i prošireni izdanjima tokom 1985. i 1999. godine, dok je aktuelna verzija iz 2014. Iako su Standardi prvi put objavljeni 1966, prethodila su joj dva izdanja, jedno koje je objavila APA 1954. i drugo koje su objavili AERA i NCME 1955, a u vezi sa tehničkim preporukama testiranja. Više o Standardima može se videti na: http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition

Da li je psihološki test potreban?

Test predstavlja najkorišćeniji alat psihologa koji pomaže u donošenju različitih odluka. Iako je najkorišćeniji alat psihologa, postavlja se pitanje da li je neizbežan i uvek neophodan. Prilikom definisanja testa, istaknuto je da je test metod prikupljanja podataka. Pored testa, podaci se mogu prikupiti i na druge načine, putem intervjua, posmatranja, anamneze itd. Još je Goldman (Goldman, 1971) napravio analizu sličnu listama "za" i "protiv", u okviru koje je postavio korisnicima testa sledeća pitanja:

1. Koju vrstu informacija želim da dobijem testom?
2. Kako će te informacije biti iskorišćene?
3. Koliko je tih informacija dostupno iz drugih izvora?
4. Koje druge metode mogu iskoristiti da bih dobio informacije koje tražim?
5. Koje su prednosti upotrebe testa u odnosu na druge izvore informacija?
6. Koji su nedostaci upotrebe testa u pogledu vremena, resursa i materijalnih troškova u odnosu na druge izvore informacija?

Ukoliko već na samom početku primene testa ne postoje eksplicitni razlozi zašto bi se test primenio, onda ni rezultati testa neće biti od koristi. Oni, čak, mogu da odmognu u donošenju odluke, ukoliko se ne upotrebe na adekvatan način. Takođe, ukoliko do informacija možemo doći uz pomoć neke druge metode, koja je ujedno i jeftinija ili vremenski manje zahtevna, onda nam test nije potreban. Na primer, ne treba nam test sklonosti ka fizičkoj agresiji da bismo procenili da li neko pripada rizičnoj grupi za ispoljavanje nasilja, već tu informaciju možemo dobiti na osnovu istorije događaja ispitanika (da li je inicirao ili učestvovao u tučama, da li je privođen za prekršaje ili krivična dela koja

uključuju elemente nasilja i sl.). Izuzetak je situacija kada rezultatima testa želimo nešto da potvrdimo, kao npr. u slučaju dobijanja sertifikata o poznavanju jezika. Takođe, u cilju psihološke procene, test može biti sastavni deo metoda prikupljanja podataka, ali svakako ne jedini. Kada se vrši psihološka procena, na primer u dijagnostičke svhe ili svrhe određivanja tretmana, test nikada ne sme biti jedini pokazatelj funkcionisanja ispitanika, već treba koristiti višestruke izvore informacija.

Osnovni razlozi za primenu testa

Uvažavajući specifične potrebe za primenom testa, Urbina (Urbina, 2004) navodi dva osnovna razloga za njegovu primenu. Prvi se odnosi na **efikasnost**. Naime, većina, ako ne i sve informacije potrebne za donošenje odluka, mogu se naći u drugim izvorima koji ne uključuju testove. Na primer, informaciju o tome da li je neko obučen za rad na određenim mašinama možemo dobiti od nadležnog rukovodioca sektora, informaciju o tome kakva je neko ličnost možemo dobiti od njegovih bliskih osoba i slično. Međutim, test štedi vreme i resurse, te umesto da ispitamo nekoliko bliskih osoba ispitanika u vezi s tim kakva je on ličnost, jednostavnije i brže je ispitaniku dati test ličnosti na osnovu kojeg ćemo te informacije dobiti. Postoje dve situacije kada je test jedina opcija za dobijanje informacija o osobi, a to je kada onaj ko treba da donese odluku nije u mogućnosti longitudinalno da prati ispitanika ili nije u mogućnosti da ima interakciju s njim.

Drugi osnovni razlog za primenu testa je **objektivnost**. Čak i kada postoji mogućnost longitudinalnog posmatranja ispitanika i interakcije s njim, informacije dobijene ovim putem mogu biti podložne subjektivnoj proceni i impresiji, ili bi njihova generalizacija bila

ograničena na uži skup situacija. Subjektivnost predstavlja poseban problem u posmatranju s učešćem (u kojem ispitivač aktivno učestvuje u životu ispitanika i stupa s njim u direktan kontakt i interakciju) i kada informacije dobijamo od osoba bliskih ispitaniku. Pored toga, rezultati najbolje organizovanog posmatranja ne mogu se adekvatno interpretirati i evaluirati bez referetnog okvira. Rezultati testova se uvek interpretiraju u kontekstu nekog referentog okvira, bio to normativni, kriterijumski ili neki drugi. Tumačenje rezultata spram referentnog okvira čini interpretaciju testa nepristrasnom, što je jedna od osnovnih prednosti testova.

Naravno, tumačenje je nepristrasno ako se test primenjuje na adekvatan i standardizovani način, u adekvatnim uslovima, na populaciji kojoj je namenjen i slično. Ukoliko se ne vodi računa o relevantnim karakteristikama ispitanika koje mogu uticati na testovne rezultate, rezultati testova verovatno neće odražavati realnu izraženost svojsta koje je predmet merenja. Na početku ovog poglavlja već je spomenuto da su testovi inteligencije nastali u jednoj kulturi korišćeni na drugoj, bez prethodne adekvatne adaptacije (testovi su samo prevedeni na drugi jezik, a nije utvrđeno da li su psihometrijski ekvivalentni). Stoga je tumačenje rezultata testova bilo pristrasno u smislu da je favorizovalo ispitanike kulture u kojoj je originalno nastao test. Međutim, postoje neke situacije kada se mora odstupiti od standardizovane procedure testiranje, npr. prilikom testiranja osoba s invaliditetom. Kasnije će biti reči o ovim situacijama, ali je ovde bitno istaći da je, iako testovi mogu biti pristrasni, lakše proceniti njihovu pristrasnost, nego drugih metoda prikupljanja podataka.

Na kraju, dolazimo do zlatnog pravila za primenu testova koje glasi:

Neki test se može primeniti samo ako se unapred zna da će dati bolje i nepristrasnije rezultate, nego neke druge metode prikupljanja podataka.

Top 10 razloga zašto NE koristiti test

1. Namena je nepoznata ili nejasna korisniku testa.
2. Korisnik testa nije u potpunosti upoznat sa svim aspektima primene testa.
3. Korisnik testa ne zna šta se dešava sa rezultatima testa.
4. Informacija koja bise dobila testom je već dostupna ili se može dobiti na efikasniji način.
5. Ispitanik ne želi da pristupi testiranju ili nije u mogućnosti.
6. Postoji opasnost da ispitanik naruši proces testiranja.
7. Uslovi testiranja su neadekvatni.
8. Test nije primeren ispitanikovim karakteristikama (starost, pol, kulturološki ili lingvistički kontekst, zdravstveni status...).
9. Norme su zastarele ili neprimenljive na ispitaniku.
10. Dokumentacija o psihometrijskim karakteristikama testa je neadekvatna (preuzeto iz Urbina, 2004).

Nepristrasnost testiranja

Apsolutno pravedan tretman ispitanika je nemoguće postići, u krajnjoj liniji zbog toga što nijedan test nema savršenu pouzdanost i validnost, tj. obe ove psihometrijske karakteristike su stvar stepena. Bez obzira na to, pravičnost testiranja u svakom kontekstu se prosuđuje u odnosu na alternativne postupke prikupljanja podataka, te i nesavršen testovni alat može dati pouzdanije i validnije rezultate od alternativnog postupka procene. Već je spomenuto da je objektivnost jedan od dva osnovna razloga za primenu testa. Objektivnost testa bi trebalo razmotriti sa stanovišta tehničkih karakteristika testa, načina na koji se tumače i izveštavaju rezultati testa, i faktora za koje se valjano ili pogrešno

pretpostavlja da mogu uticati na rezultate testa. Pored toga, trebalo bi razmotriti pravilnike, zakone i druge pravne odrednice na kojima se bazira upotreba testova i sankcije za nepropisanu upotrebu ili zloupotrebu testova.

Nepristrasnost u testiranju može se shvatiti na različite načine. U Standardima (AERA, APA, & NCME, 2006) navode se četiri značenja nepristrasnosti u testiranju. U prvom, tehničkom značenju, nepristrasnost se odnosi na **jednako funkcionisanje testa i njegovih delova**. Ovo određenje se odnosi na sledeće: značenje rezultata treba da bude isto i nepristrasno za pripadnike različitih grupa, a ukoliko se dokaže drugačije, to se treba jasno naznačiti, kao i način pravilne interpretacije. Razlike u funkcionisanju testa utvrđuju se statističkim i psihometrijskim postupcima (npr. preko razlika u diskriminativnosti ajtema ili preko utvrđivanja tzv. diferencijalnog funkcionisanja ajtema ili testa). Nejednako funkcionisanje ajtema ili testa postoji kada ispitanici koji imaju isti nivo crte, ali pripadaju različitim grupama, ostvaruju različite skorove. Na primer, moguće je da neki ajtem drugačije shvataju muškarci i žene, te da razlike u skorovima između muškaraca i žena na tom ajtemu ne potiču od stepena prisustva merenog svojstva, već od toga što je taj ajtem pristrasan u odnosu na pol. Diferencijalno funkcionisanje testa ima važne posledice po validnost testa, a samim tim i na interpretaciju rezultata.

Prema drugom značenju, nepristrasnost se odosi na **jednak tretman ispitanika** u procesu testiranja. Ovo određenje se odnosi na jednake uslove i tretman svih ispitanika u procesu testiranja, kao što su izloženost istim ili komparabilnim uslovima testiranja (uključujući, na primer, jednake mogućnosti za upoznavanje s testovnim materijalom kroz primere i vežbu), jednaka upotreba testovnih rezultata, primena adekvatnih normi i interpretacija rezultata koja podrazumeva preciznost,

potpunu informativnost i korišćenje odgovarajućeg rečnika i formulacija koje će smanjiti pogrešnu interpretaciju. Da bi test i testiranje bili nepristrasni, svi ispitanici moraju imati isti i pravičan tretman, odnosno jednake mogućnosti da ispolje i izraze ono što je predmet merenja testa u skladu sa stepenom izraženosti datog svojstva i načinom njegove manifestacije. U ovom slučaju se nepristrasnost posmatra iz ugla primene testa i kaže se da test nije sam po sebi nepristrasan ili pristrasan, već je nepristrasna ili pristrasna upotreba testa u određenim okolnostima ili nad određenim ispitanicima (AERA, APA, & NCME, 2006).

Prema trećem određenju, nepristrasnost predstavlja **jednakost ishoda testiranja**. Ovo značenje se, pre svega, odnosi na jednakost ishoda testiranja za ispitanike koji potiču iz različitih grupa u odnosu na rasnu, etničku, polnu, starosnu i drugu pripadnost ili pripadnost osobama s invaliditetom i drugim senzitivnim grupama. Međutim, ideja da nepristrasnost podrazumeva približno podjednake proporcije uspešnih u različitim grupama (kao kod npr. pozitivne diskriminacije) generalno nije prihvaćena u stručnoj literaturi i navodi se da međugrupne razlike u ishodima testiranja same po sebi ne upućuju na pristrasnost ili nefer tretman tokom testiranja. To se opravdava na sledeći način: kada bi se radila, na primer, profesionalna selekcija, osobe koje bi dobro obavljale posao bi trebalo da imaju podjednak izgled za izbor, bez obzira na to kojoj grupi pripadaju. Ukoliko već nekim zakonskim i drugim propisima nije drugačije regulisano, međugrupne razlike u ishodima testiranja nemaju neposredne veze s pristrasnošću testiranja. Drugim rečima, vladajuće stanovište je da ispitanici podjednakog statusa, s obzirom na predmet merenja testa, treba u proseku da imaju isti rezultat, bez obzira na to kojoj grupi pripadaju.

Prema poslednjem, četvtom određenju, nepristrasnost podrazumeva **jednakost prilika za učenje**. Ovo shvatanje nepristrasnosti odnosi se na testove postignuća i jednakost mogućnosti za učenje sadržaja koji obuhvataju takvi testovi. Naime, ako neko nije imao prilike da nauči ono što je sadržaj testa postignuća, ili je imao priliku, ali je nije iskoristio, verovatno će imati nizak skor. Tako je upotreba testova postignuća u cilju selekcije učenika koji će dobiti stipendiju nepristrasna ukoliko svi kandidati nisu imali prilike da se upoznaju sa sadržajem koji će biti pokriven testom. Međutim, postoji nekoliko poteškoća u vezi s ovakvim shvatanjem nepristrasnosti, a dve se mogu smatrati najvažnijima. Prva je u definsanju prilika za učenje, što je više stvar stepena nego crno-belog ishoda da li je neko imao ili nije imao prilike za učenje. Na primer, sve škole jedne države imaju propisan kurikulum i za svaki predmet postoji spisak nastavnih jedinica koje se moraju obraditi. Međutim, nisu sve škole uspešne u obradi obaveznih i propisanih nastavnih jedinica, te tako imamo škole u kojima učenici generalno ostvaruju dobre rezultate i one u kojima učenici ostavaruju generalno loše rezultate. Drugim rečima, iako su svi učenici prošli obavezno nastavno gradivo, nisu svi to prošli na isti način i s istim kvalitetom, uz pomoć adekvatne opreme i slično. Pored toga, nekada se namerno traži od ispitanika da radi s materijalom s kojim prethodno nije imao iskustva kako bi se procenili snalaženje s novim materijalom, primena znanja, način učenja i slično. U ovoj situaciji je uslov da prethodnog upoznavanja sa istim sadržajem nije bilo, ali ne znači da nije bilo upoznavanja sa sličnim sadržajima. Druga poteškoća je u tome što ispitanici mogu da znaju sadržaj testa, odnosno teme, ali ne i pokrivenost određenih tema pitanjima u testu. Neke nastavne jedinice su važnije od drugih, te im se posvećuje veća pažnja u testu, što učenici ne mogu unapred da znaju.

Takođe, nekad i od samog predmetnog nastavnika zavisi kojoj nastavnoj jedinici će posvetiti veću pažnju.

Vrste pristrasnosti testiranja

Opozit nepristrasnosti testiranja je pristrasnost, te se ona odnosi na komponente koje su irelevantne za procenu predmeta merenja testa, ali koje čine da se dobijaju sistematski niži ili viši skorovi kod pripadnika određenih grupa. Kako se prvo značenje nepristrasnosti odnosi na tehničke kvalitete testa, tj. jednako funkcionisanje testa i svih njegovih delova, u psihometrijskoj literaturi je najviše pažnje posvećeno upravo ovom aspektu i proveri postojanja pristrasnosti u funkcionisanju testa. U literaturi se najčešće razlikuju dve vrste pristrasnosti testa spram funkcionisanja testa, a to su konstruktna i prediktivna pristrasnost (npr. Furr & Bacharach, 2014; Nunnally & Bernstein, 1994).

Konstruktna pristrasnost

Konstruktna pristrasnost se naziva još i interna pristrasnost ili pristrasnost merenja (eng. *construct bias, internal bias, measurement bias*) i odnosi se na pristrasnost u **značenju testa**. Drugim rečima, kada je test konstruktno pristrasan, onda skor na testu ima različito značenje za pripadnike različitih grupa. Konstruktna pristrasnost obuhvata relacije između pravih i opaženih skorova i ukoliko ove relacije nisu iste u različitim grupama, možemo zaključiti da je test pristrasan, odnosno da odražava mereno svojstvo s različitom preciznošću. To bi značilo da u jednoj grupi opaženi skor reflektuju pravi skor, a u drugoj grupi ne. Dakle, konstruktna pristrasnost se odnosi na situaciju da pripadnici dve grupe imaju podjednake prave skorove, ali ne i opažene. Ukoliko postoji konstruktna pristrasnost, skorovi pripadnika različitih grupa ne mogu se

porediti – njihovo poređenje nema psihološkog značenja i opravdanja, jer može voditi u pogrešne zaključke.

Važno je naglasiti da postojanje međugrupnih razlika ne znači samo po sebi postojanje konstruktne pristrasnosti. Za konstruktnu pristrasnost je važno da su ispunjena dva uslova: 1. da postoje značajne razlike između grupa i 2. da te razlike *ne potiču* od onoga što je mereno svojstvo. Uzmimo na primer polne razlike u agresivnom ponašanju. Istraživanja dosledno pokazuju da postoje polne razlike u agresiji, pri čemu izraženost tih razlika zavisi od vrste agresije, te su razlike najizraženije u slučaju fizičke agresije (Archer, 2004). Međutim, može se postaviti pitanje da li ove razlike zaista potiču od razlike u agresiji, ili su ajtemi u upitnicima za procenu agresije pristrasni u smislu da se muškarci više slažu s njima bez obzira na to što su i muškarci i žene podjednako agresivni. Prilikom analize ajtema Upitnika reaktivne i proaktivne agresije dobijeno je da postoji pristrasnost u odgovaranju na ajtem "Koliko često ste udarili drugoga da biste se odbranili?" (Dinić & Raine, 2019). Očekivano je bilo da će i muškarci i žene koji pokazuju nisku sklonost ka agresiji izveštavati o manjoj učestalosti ovog ponašanja, da će muškarci i žene koji pokazuju prosečnu sklonost ka agresiji izveštavati o umerenoj učestalosti ovog ponašanja i da će muškarci i žene koji pokazuju visoku sklonost ka agresiji izveštavati o većoj učestalosti ovog ponašanja. Međutim, dobijeni rezultati pokazuju da muškarci izveštavaju o većoj učestalosti ovog ponašanja, bez obzira na to kolika je njihova sklonost ka agresiji, dok se odgovori žena distribuiraju u skladu s očekivanjem. Dobijeni rezultat upućuje na činjenicu da je ovaj ajtem pristrasan, te da razlike koje se dobijaju između muškaraca i žena na ovom ajtemu ne potiču od razlika u agresiji, već od nečeg drugog, što nije predmet merenja upitnika. Moguće je da je odbrambena fizička agresija nešto što je

karakteristično za sve muškarce, što se smatra prihvatljivim, pa čak i očekivanim i poželjnim ponašanjem za ulogu muškarca.

Detekcija konstruktne pristrasnosti

Postoji nekoliko načina detekcije konstruktne pristrasnosti koji se razlikuju po svojoj kompleksnosti i zahtevima. Uvid u pristrasnosti je nekada dovoljan na osnovu **kvalitativne analize** sadržaja ajtema. Dobra praksa je da se ajtemi testa ocene od strane nezavisnih procenjivača u odnosu na različite aspekte, kao što su jezik i upotreba termina koji su razumljivi svim ispitanicima, zastupljenost sadržaja spram relevantnosti za određene grupe i slično.

U drugim slučajevima, uvid u pristrasnosti se može dobiti na osnovu kvantitativnih, odnosno psihometrijskih pokazatelja. Prvi od njih podrazumeva izračunavanje **indeksa diskriminativnosti ajtema** na različitim grupama. Diskriminativnost je svojstvo ajtema da razlikuje ispitanika s niskom i visokom crtom, a izračunava se kao korelacija ajtema s ukupnim skorom na testu. To bi značilo da ispitanici koji odgovore potvrdno ili tačno na ajtem ili izraze veći stepen slaganja, imaju ujedno i viši skor na testu od onih ispitanika koji odgovore odrično ili netačno, ili izraze manji stepen slaganja sa tvrdnjom. Visoka diskriminativnost ajtema znači da je taj ajtem konceptualno sličan većini ostalih ajtema u testu. Ukoliko je indeks diskriminativnosti za ajtem međusobno sličan u različitim grupama, može se zaključiti da ajtem na isti način odražava konstrukt u tim grupama, odnosno da je nepristrasan. S druge strane, ukoliko se indeksi diskriminativnosti ajtema razlikuju po grupama, onda je ajtem pristrasan i ne reflektuje isti konstrukt u ovim grupama. U ovom slučaju, ukoliko takav ajtem zadržimo u testu, skorovi na testu neće biti komparabilni po grupama. Da bi test bio nepristrasan, ajteme za koje se

utvrdi da su pristrasni treba izbaciti iz testa ili ih revidirati na način da ne budu pristrasni. Na ovom mestu je važno istaći da indeks diskriminativnosti ne zavisi od broja ispitanika po grupama koje se porede, a koji su odgovorili tačno ili netačno. Na primer, ukoliko u jednoj grupi imamo 30% ispitanika koji su potvrdno odgovorili na ajtem, a u drugoj 60%, indeks diskriminativnosti može biti isti za taj ajtem u obe grupe. Ono što možemo konstatovati u slučaju istog indeksa diskriminativnosti, a različitih ajtemskih skorova je da razlike između grupa zaista postoje na tom ajtemu, te da one potiču od predmeta merenja, a ne od pristrasnosti ajtema.

Drugi način obuhvata **rang težina ajtema** testa u okviru grupa. Naime, ukoliko je rang težina ajtema (aritmetičkih sredina) u testu isti za različite grupe, možemo konstatovati da je test nepristrasan. Jednakonst ranga težina se lako utvrđuje na sledeći način: najpre se napravi rang ajtema po težini za svaku grupu posebno, te se izračunaju Spirmanovi ρ koeficijenti korelacije. Ukoliko je korelacija > ,90, može se smatrati da je test nepristrasan (Furr & Bacharach, 2014). Korelacija između rangova težina može biti visoka, čak iako težine istog ajtema nisu slične između grupa. Moguće je da, na primer, u jednoj grupi težine tri ajtema redom budu 0,6, 0,4 i 0,2, a u drugoj tih istih ajtema redom 0,9, 0,8 i 0,4, te da korelacija između rangova bude visoka bez obzira na to što težine pojedinačnih ajtema nisu podudarne. Dakle, bitno je samo da je rang težina ajtema nepromenjen u grupama.

Treći način određivanja pristrasnosti testa je preko uvida u **faktorsku strukturu** testa u različitim grupama. Pritom, ukoliko se u različitim grupama izdvoji različiti broj faktora na testu, to je jasan znak da postoji konstruktna pristrasnost. Međutim, ukoliko se ispostavi da se isti broj faktora može zadržati u različitim grupama, to i dalje ne znači da

je test konstruktno nepristrasan. Da bi test bio konstruktno nepristrasan, potrebno je da struktura faktora bude ista u različitim grupama. Ista struktura faktora se procenjuje na dva nivoa:

1. **osnovni – konfiguralni nivo** podrazumeva da ajtemi pripadaju istim faktorima na različitim grupama. Ukoliko je ovaj nivo ostvaren, to znači da je konstrukt koji je predmet merenja na isti način konceptualizovan u grupama;
2. **metrični nivo** podrazumeva da su i opterećenja ajtema na faktorima ista u različitim grupama. Metrični nivo podrazumeva da konkretan test meri ekvivalentan konstrukt u grupama. Kako je indeks diskriminativnosti ništa drugo do opterećenje na faktoru, provera sličnosti faktorske strukture po grupama se zapravo svodi na proveru sličnosti indeksa diskriminativnosti ajtema, i obrnuto. Sličnost faktorske strukutre može se proveriti preko Takerovih koeficijenata kongruencije ili slaganja, čija vrednost od ,85 do ,94 ukazuje na zadovoljavajuću sličnost, a ,95 i više na dobru sličnost, tj. faktori se mogu smatrati ekvivalentim (Lorenzo-Seva & Berge, 2006).

Razvijeni su i sofisticiraniji načini provere preko konfirmatorne faktorske analize, a u okviru postupaka provere tzv. invarijatnosti merenja (eng. *measurement invariance*). I u ovom slučaju se proveravaju nivoi sličnosti, te ostvarivanje sličnosti na konfiguralnom i metričnom nivou pripada slaboj invarijatnosti, odnosno zadovoljenje ta dva nivoa je minimum da bismo mogli da kažemo da je test konstruktno nepristrasan u odnosu na primenu na datim grupama.

Međutim, to što je struktura faktora ista u grupama i dalje ne znači da se skorovi pripadnika tih grupa mogu porediti. Da bi to bilo moguće, potrebno je dokazati postojanje **skalarnog nivoa**, što se smatra jakom

invarijatnošću. Skalarni nivo podrazumeva da, pored opterećenja ajtema na faktorima, i odsečci (intercepti) ili pragovi ajtema budu jednaki u grupama. Kada je skalarni nivo ostvaren, to znači da skorovi predstavljaju iste nivoe konstrukta između grupa i da se onda grupe međusobno mogu porediti. Tako, na primer, prilikom utvrđivanja invarijatnosti jednog upitnika narcizma na uzorcima iz Srbije i Sjedinjenih Američkih Država, dobijeno je da je upitnik invarijantan u odnosu na sva tri nivoa, što omogućava poređenje skorova između pripadnika ovih kultura. Pokazano je da ispitanici iz Sjedinjenih Američkih Država imaju više skorove na nekim aspektima narcizma u odnosu na ispitanike iz Srbije, čime se odražavaju razlike u individualistički i kolektivistički orijentisanim kuluturama ova dva uzorka (Dinić & Vujić, 2019). S druge strane, u drugom istraživanju nije ostvarena skalarna invarijantost jednog aspekta narcizma između engleske, poljske i japanske adaptacije upitnika, te nije bilo moguće donositi zaključke o razlikama između tih zemalja na tom konkretnom aspektu narcizma (Żemojtel-Piotrowska et al., 2018). Detalji u vezi s ovim tehnikama prevazilaze cilj ovog udžbenika, ali se zainteresovani čitaoci upućuju npr. na Brauna (Brown, 2006).

Prediktivna pristrasnost

Prediktivna pristrasnost naziva se još i eksterna pristrasnost ili pristrasnost diferencijalne validnosti (eng. *predictive bias, external bias, diferential validity bias*). Prediktivna pristrasnost odnosi se na pristrasnost prilikom **upotrebe testa**, odnosno kada test ima drugačije implikacije za pripadnike različitih grupa. Prediktivna pristrasnost odnosi se na relacije između skorova na dva testa, odnosno testa i kriterijuma, a koji se razlikuju u različitim grupama (npr. razlikuje se značajnost povezanosti ili intenzitet, pa čak i smer povezanosti). To bi značilo da se

relacije između pravih skorova na testu i opaženih skorova na kriterijumu razlikuju po grupama. U Standardima se prediktivna validnost definiše kao „sistematsko precenjivanje ili podcenjivanje predviđenog učinka na kriterijumu za osobe koje pripadaju grupama koje se razlikuju po svojstvima koja su irelevantna za kriterijumski učinak" (AERA, APA, & NCME, 2006, str. 269).

Uzmimo, na primer, predikciju sklonosti ka agresivnom ponašanju na osnovu besa kao crte ličnosti. Možemo pretpostaviti da će osobe višeg besa biti sklonije agresiji. Međutim, sumnjamo u to da ovaj odnos važi i za muški i za ženski pol. Ukoliko se ustanovi da postoji značajna i pozitivna povezanost skora na upitniku besa sa skorom na upitnku agresivnog ponašanja i to samo na muškom polu (ali ne i na ženskom), može se sumnjati u to da je prediktor, tj. upitnik besa prediktivno pristrasan. Takođe, može se sumnjati i na to da je kriterijum, tj. upitnik agresivnog ponašanja prediktivno pritrasan, a moguće da su oba upitnika pristrasna.

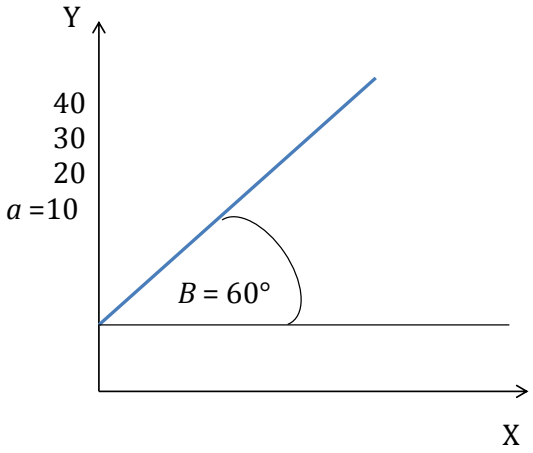
Detekcija prediktivne pristrasnosti

Prediktivna pristrasnost se, takođe, može utvrditi na različite načine, ali u osnovi svih načina stoji **regresiona analiza**. U regresionoj jednačini možemo razlikovati sledeće elemente:

$$Y = a + BX$$

gde je Y kriterijum, a je odsečak ili intercept, B je nagib ili regresioni koeficijent i X je prediktor. Odsečak je vrednost na kriterijumu kada je prediktor jednak nuli, dok je nagib vrednost koja govori o tome za koliko jedinica se menja kriterijum kada se prediktor poveća za jednu jedinicu merenja. Na Slici 34 dat je primer regresione linije na kojoj se može videti

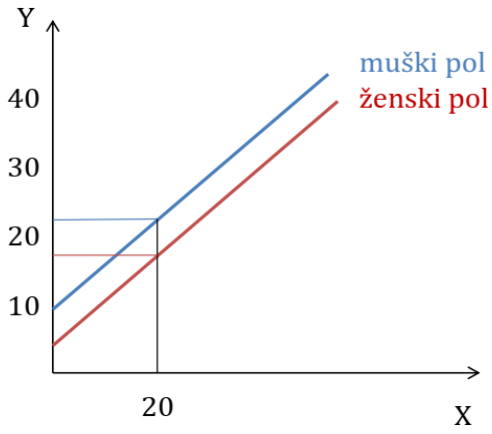
da odsečak iznosi 10 (to je odsečak na Y-osi u tački u kojoj je seče regresiona linija), dok je izračunati nagib ,60.



Slika 34. Primer regresione krive.

Prediktivna pristrasnost se može manifestovati u vidu različitih odsečaka za različite grupe, u vidu različitih nagiba, ili u vidu različitih i odsečaka i nagiba. Ukoliko postoje razlike u odsečcima, to se naziva **pristrasnošću odsečka**. Pristrasnost odsečka ukazuje na to da će grupa koja ima viši odsečak imati sistematski više skorove na kriterijumu na svim nivoima prediktora, u odnosu na grupu koja ima niži odsečak. U pomenutom primeru predikcije agresije na osnovu besa, ukoliko je na muškom polu odsečak regresione linije veći nego na ženskom, to znači da predikcija neće biti ista za muškarca i ženu koji ostvaruju iste skorove na upitniku besa. Tako, ako je odsečak npr. 10 za muškarce i 5 za žene, to znači da će muškarac koji ima skor na upitniku besa 20 imati predviđeni skor na upitniku agresije 22 (kad se zamene vrednosti u regresionoj jednačini, to je 10 + 0,60 * 20), a žena 17 (5 + 0,60 * 20). Ukoliko je nagib isti i za muškarce i za žene, razlika u predikovanim skorovima će biti

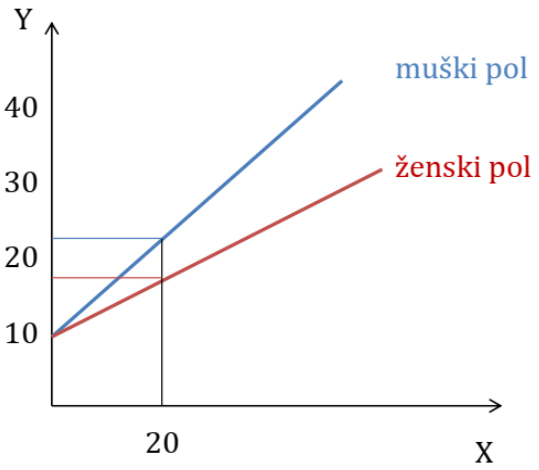
jednaka duž celog kontinuuma skorova, tj. žene će uvek imati za 5 jedinica niži predikovani skor na agresiji od muškaraca (pod pretpostavkom da upitnik agresije nije pristrasan, videti Sliku 35).



Slika 35. Primer pristrasnosti odsečka za dve grupe.

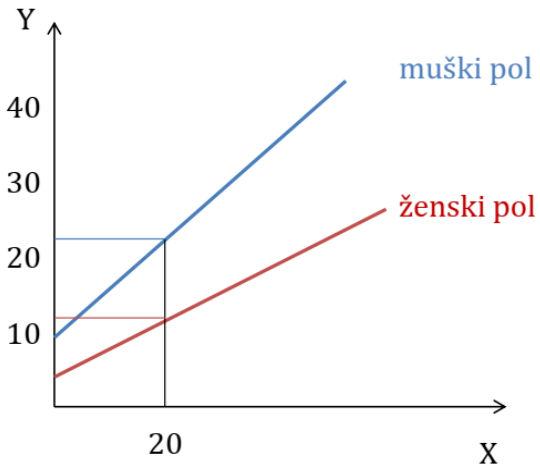
Druga vrsta pristrasnosti je **pristrasnost nagiba**. Kada ova vrsta pristrasnosti postoji, to znači da pristrasnost nije konstantna duž celog kontinuuma skorova prediktora. Moguće je da pristrasnosti ne postoje ili su manje u slučaju niskih skorova, a da postoje ili su više u slučaju visokih, i obrnuto. U našem primeru, ukoliko je recimo nagib regresione linije kod muškaraca ,60, a kod žena ,40, pod pretpostavkom da je odsečak isti kod oba pola (10), muškarac koji ima skor 20 na upitniku samopoštovanja imaće predikovani skor na upitniku agresije 22 (10 + 0,60 * 20), a žena 18 (10 + 0,40 * 20). Razlika će biti 4 jedinice. Međutim, u slučaju viših skorova, npr. 40, muškarčev prediktivni skor će biti 34 (10 + 0,60 * 40), a ženin s istim skorom na upitniku besa će biti 26 (10 + 0,40 * 40). Može se primetiti da je razlika sada čak 8 jedinica (Slika 36). Dakle, pristrasnosti su manje u slučaju niskih skorova, ali izraženije u slučaju visokih skorova na prediktoru. Diskrepanca između predikovanih skorova muškaraca i

žena će biti sve veća kako rastu skorovi na prediktoru, tj. upitniku samopoštovanja.



Slika 36. Primer pristrasnosti nagiba.

Postoji i varijanta pristrasnosti u kojoj su prisutne obe vrste pristrasnosti – **pristrasnosti odsečka i nagiba**, i ona je mnogo češća u praksi. Postoje različite kombinacije varijacija odsečka i nagiba po grupama, npr. da su odsečak i nagib veći u jednoj grupi u odnosu na drugu, ili da je odsečak veći a nagib manji u jednoj grupi u odnosu na drugu i sl. Navešćemo samo jedan primer ilustracije radi. Ukoliko imamo situaciju da su na muškom polu i nagib (10) i odsečak (0,60) viši nego na ženskom polu (nagib 5 i odsečak 0,40), za muškarca koji ima skor 20 bismo dobili prediktivni skor na agresiji 22 (10 + 0,60 * 20), a za ženu 13 (5 + 0,40 * 20). Razlika bi bila 9 jedinica. U slučaju viših skorova, za muškarca koji ima skor 40 na upitniku besa bismo dobili predikovani skor na agresiji 34 (10 + 0,60 * 40), a za ženu 21 (5 + 0,40 * 40), tj. razlika bi bila čak 13 jedinica (Slika 37).



Slika 37. Pristrasnost odsečka i nagiba.

Na kraju treba istaći da su konstruktivna i prediktivna pristrasnost međusobno nezavisne, te da test može biti konstruktno pristrasan, ali ne i prediktivno, i obrnuto. Na primer, test znanja prilikom upisa na fakultet može biti konstruktno nepristrasan u odnosu na pol, tj. da ima isto značenje i za mladiće i za devojke, ali da ne predviđa uspešnost studiranja na isti način kod mladića i devojaka, odnosno da nije prediktivno nepristrasan.

Pravedna upotreba testa

Prema Standardima (AERA, APA, & NCME, 2006) samo se u jednom značenju termina nepristrasnosti ona dovodi u vezu s metrijskim karakteristikama, a to je prvo značenje opisano kao jednako funkcionisanje testa i njegovih delova. Dakle, u ovom značenju nepristrasnost testa se dokazuje preko statističkih i psihometrijskih metoda. U svim ostalim značenjima nepristrasnost je nešto što nema veze sa psihometrijskim kvalitetom testa, već sa ocenom pravičnosti upotrebe

testovnih rezultata (eng. *test fairness*) sa stanovišta socijalnih, moralnih, filozofskih, pa i legalnih aspekata (Jensen, 1980). Sud o fer i pravednoj upotrebi testa, kao i o odlukama baziranim na testovnim rezultatima, u ovom slučaju je u očima posmatrača. Naime, test može imati odlične metrijske karakteristike i na osnovu prethodno opisanih statističkih metoda može se pokazati da nije pristrasan, ali neki ljudi i dalje procenjuju da test nije upotrebljen na pravedan i pošten način, i pri tome mogu biti u pravu (Furr & Bacharach, 2014). Na primer, standardizovanom testu SAT (koji je ekvivalent standardizovanim testovima znanja, kao test na završnom maturskom ispitu), koji se koristi za procenu podobnosti za upis na fakultet u Sjedinjenim Američkim Državama, često se zamera da favorizuje određene grupe, npr. pripadnike bele rase i one sa višim socio-ekonomskim statusom (npr. Rattani, 2016). Kritičari SAT testa i načelno upotrebe SAT skorova pri selekciji kandidata za upis na fakultet ističu, između ostalog, da razlike u postignuću na ovom testu potiču od razlika u socio-ekonomskom statusu. Tako, imućniji učenici imaju prilike da idu u bolje škole i dobiju bolje obrazovanje koje im omogućava bolje rezultate na SAT testu. Kako se polaganje SAT testa plaća, imućniji onda mogu lakše da priušte ponovno polaganje testa u slučaju da ostvare nezadovoljavajući rezultat. Dakle, SAT ima dobre metrijske karakteristike i test dobro funkcioniše u različitim grupama, tj. to što neka grupa ima sistematično niže skorove potiče zaista od izraženosti predmeta merenja SAT-a. Ali razlog zašto neka grupa ima sistematično niže skorove je izvor nepristrasnosti (loši socio-ekonomski uslovi koji povlače i lošije obrazovanje).

Izvori pristrasnosti

Izvori pristrasnosti testa mogu biti različiti, ali među njima su dva najvažnija. To su: sadržajni izvori pristrasnosti i izvori pristrasnosti povezani s odgovorima.

Sadržajni izvori pristrasnosti testa

Sadržajni izvori pristrasnosti mogu se odnositi na neadekvatno uzorkovanje testovnog sadržaja, kao što je neuravnoteženost sadržaja za sve grupe s obzirom na familijarnost, zanimljivost, relevatnost itd. Na primer, ako se ispituju učenici različitih srednjih škola, a skup pitanja u datom testu je direktno povezan s medicinskom strukom, učenici medicinskih škola bi bili favorizovani. Sadržajni izvori pristrasnosti mogu se odnositi i na nejasno uputstvo, te neki ispitanici mogu odgovarati u skladu s tim šta misle o sebi ili šta misle da se od njih očekuje. Takođe, i sam izbor jezičkih termina u testu može biti izvor pristrasnosti, npr. ukoliko se koristi rečnik koji je uvredljiv za neku grupu, koji izaziva uznemirenost kod pripadnika neke grupe i slično.

Izvori pristrasnosti povezani s odgovorima

Ova grupa pristrasnosti odnosi se na to da ajtemi, osim predviđenih ponašanja, misli i osećanja, izazivaju i druge odgovore ili da se pitanja mogu rešiti i na drugačiji način od predviđenog. Na primer, učinak na testu znanja ne sme biti povezan s brzinom pisanja, ali je moguće da bolji rezultat na testu znanja imaju oni učenici koij brže pišu. Samim tim, skor na ovom testu je pristrasan. Potom, ukoliko je protokol skorovanja takav da se više vrednuju neka rešenja od drugih, a da su ta rešenja zastupljenija u nekoj grupi i tada, takođe, postoji pristrasnost.

Problemi u procesu testiranja

Neadekvatna upotreba testa ili zloupotreba može se javiti u svakom koraku psihološkog testiranja, od izbora cija testiranja do saopštavanja rezultata testa. Ipak, kao najčešći problem u upotrebi testa ističu se nedovoljne kompetencije korisnika testa (DeMers et al., 2000), te će se ovom problemu posvetiti posebna pažnja. U narednim odeljcima će se predstaviti primeri dobre prakse i potencijalni problemi u procesu testiranja po fazama.

Problemi koji potiču od korisnika testa

Već je bilo rečeno da korisnik testa ima najveću odgovornost za selekciju, pravilnu upotrebu, primenu i tumačenje rezultata testa, kao i donošenje odluke na bazi rezultata testa (AERA, APA, & NCME, 1999; DeMers et al., 2000). Ključna uloga korisnika testa je u tome da izvrši validnu interpretaciju skorova i dobijenih podataka, često iz različitih izvora (DeMers et al., 2000; Turner et al., 2001). Čak i najbolje proučen i opremljen test može biti beskoristan, zloupotrebljen ili ići na štetu ako je neadekvatno i nestručno upotrebljen (Krković, 1974). Stoga se kao jedan od osnovnih razloga niskog kvaliteta upotrebe testa, neadekvatne primene testa, kao i njegove zloupotrebe, navodi nedovoljno znanje i kompetencije korisnika testa (Turner et al., 2001). Postoje dva osnovna domena kvalifikacije korisnika:

1. **opšta psihometrijska znanja i veštine** koje predstavljaju osnovu za razumevanje primene i upotrebe testa; ona obuhvataju sržna znanja koja se odnose na poznavanje statističkih (normalna raspodela, deskriptivni pokazatelji, pokazatelji asocijacije između mera, transformacije) i psihometrijskih pojmova i procedura (teorije merenja, vrste

skorova, psihometrijske karakteristike, selekcija testova, procedura zadavanja testa), kao i opšta znanja i veštine vezane za testiranje u određenom kontekstu i uslovima (npr. profesionalna selekcija) ili za testiranje određenih subpopulacija (npr. testiranje osoba različitog jezičkog porekla, kulture, dece, osoba s invaliditetom). Među ovim opštim znanjima posebno treba istaći poznavanje greške merenja i izvora varijabilnosti greške merenja (za detaljan prikaz kompetencija korisnika testa videti DeMers et al., 2000). Da bi korisnik testa mogao da pruži validnu interpretaciju testovnih rezultata, potrebno je da može da integriše znanja iz psihometrijskih i metodoloških principa, teorije koja stoji u osnovi merenog konstrukta, znanja o psihometrijskim karakteristikama konkretnog testa i o odnosu između konkretnog testa i svrhe testiranja u određenom kontekstu.

2. **specifične kvalifikacije za primenu testa u odnosu na cilj i kontekst primene**; test se može primeniti u cilju klasifikacije, evaluacije tretmana, samospoznaje i ličnog razvoja ili u istraživačke svrhe (o čemu je bilo reči u prvom poglavlju), a kada je reč o određenom kontekstu, najčešće se razlikuju pet konteksta: organizacijski kontekst (testiranje u zapošljavanju i proveri kvalifikacija), obrazovni kontekst (uključujući i individualnu i grupnu, populacijsku procenu), profesionalno i karijerno savetovanje, zdravstvena nega i forenzički kontekst.

Ova znanja i veštine obezbeđuju se edukacijom, prvenstveno u okviru studijskih programa psihologije i to kako na obaveznim tako i na izbornim kursevima, a dodatno posebnim obukama i seminarima. Obuke obezbeđuju različite organizacije i udruženja, kao što je Društvo psihologa

Srbije kod nas, ali takođe i izdavačke kuće. U slučaju kupovine nekih testova, obuka za njegovu primenu je obavezna, te se test ne može kupiti, a da prethodno nije završena obuka. Obuke i radionice se mogu organizovati i u okviru seminara i naučno-stručnih skupova i konferencija. Korisnici testova se, takođe, mogu edukovati i praćenjem savremene literature, naučnih i drugih publikacija, iako je ovo neformalni vid obuke. Nažalost, ne postoji neko udruženje korisnika gde bi korisnici mogli razmeniti iskustva i unaprediti svoju testovnu praksu, niti supervizirali razne obuke koje se nude na tržištu.

Međutim, ono što se često zanemaruje, a što je važno za obezbeđivanje kvalifikacija korisnika, je i adekvatna **supervizija** svih relevantnih aspekata znanja i veština za primenu konkretnih testova, kao i preduzetih akcija i odluka uključenih u proces testiranja. Supervizijom se obezbeđuju razvoj i usavršavanje primene testa u datom kontekstu i za specifične namene. Ona se sprovodi još na studijama, posebno u okviru studijskih praksi, ali i kasnije prilikom pohađanja edukacija. Ipak, ona ne bi trebala da se završi sa dobijanjem diplome ili serfitikata, već bi iskusni i akreditovani profesionalaci u oblasti primene testa trebalo da vode kontinuiranu superviziju. Supervizija kompetencija i adekvatne primene testa treba biti regulisana od strane psihološkog udruženja ili komore. Nažalost, kod nas još uvek ne postoji psihološka komora koja bi se bavila ovim pitanjem, te ne postoji ni adekvatna supervizija upotrebe testova. Takođe, radna inspekcija na lokalnom ili državnom nivou se bavi nekim drugim aspektima rada, te ne pokriva proveru kompetencija korisnika testova. Na ovom mestu treba pomenuti da superviziju takođe može sprovoditi i izdavačka kuća. U praksi, izdavačka kuća utvrđuje da li potencijalni korisnik testa ispunjava uslove za primenu određenog testa (da li ima diplomu, završenu edukaciju za primenu određenog testa...), a

kada se korisniku proda test, izdavačku kuću više interesuje da li je test nelegalno umnožavan i kako se test čuva, nego da li je adekvatno primenjen.

Opšta znanja potrebna za adekvatnu selekciju testova

- svrha testiranja i upotrebe skorova;
- znanje o predmetu merenja testa i upotrebi testa u namenjenom kontekstu;
- tip i procedura skorovanja (npr. da li postoji korekcija za pogađanje ili za odgovarački stil...);
- referentni okvir za interpretaciju skorova (normativni, kriterijumski...);
- psihometrijske karakteristike (kakva je pouzdanost i da li je adekvatan pokazatelj pouzdanosti odabran, koji su dokazi o validnosti, da li su testirane sve vrste validnosti, da li se mogu izvući adekvatni zaključci o generalizaciji validnosti na osnovu izbora i veličine uzorka, pouzdanosti uključenih mera za ocenu validnosti, restrikcije ranga skorova i sl.)
- format ajtema (npr. pitanje višestrukog izbora više angažuje receptivne veštine, a otvorenog tipa ekspresivne veštine, te ukoliko ispitanici imaju različito razvijene ove veštine, to može uticati na njihove skorove);
- težina testa (da li je prilagođena ispitanicima);
- pristrasnost testa u odnosu na primenu na specifičnim grupama (npr. na manjinskim grupama čiji maternji jezik nije jezik na kojem je nastao test);
- medijum testa (iskustvo s testovima različitog medija, npr. računarski podržanih, može uticati na skorove, čak i ako se obezbede primer i vežba);
- opis i veličina uzorka (da li obuhvata sve relevantne poduzorke za dati predmet merenja, npr. pol, starost, obrazovanje, rasa, socio-ekonomski status...);
- aktuelnost normi;
- procedura zadavanja testa (npr. da li postoje primer i vežba, da li postoji vremensko ograničenje...);
- upućenost ispitanika u cilj testiranja i sam test, i karakteristike ispitanika koje mogu uticati na validnost i interpretaciju skorova (motivacija, mentalni status, iskustvo sa istim ili sličnim testom...);
- slaganje karakteristika testa i cilja procene u odnosu na predmet merenja, interpretacije rezultata, težinu testa, sličnost ispitanika sa referentnom grupom itd.;
- da li ima posebnih zahteva za primenu testa i koja su ograničenja testa.

Opšta znanja potrebna za adekvatnu administraciju testova

- prava i odgovornosti ispitanika;
- standardizovanost procedure zadavanja testa;
- procedura skorovanja;
- poverljivost podataka i testovnog materijala;
- sigurnost testa i sprečavanje nelegalnog umnožavanja, neovlašćene upotrebe i sl.;
- opšta etička načela u procesu testiranja;
- saopštavanje rezultata ispitaniku, roditeljima ili starateljima ili drugim licima koji su naručioci testiranja (količina informacija, način prikazivanja rezultata, način interpretacije i sl. (adaptirano prema DeMers et al., 2000; Turner et al., 2001; Urbina, 2004).

Sistem kvalifikacije korisnika testa

Da bi se predupredila neadekvatna upotreba testova, razvijen je sistem kvalifikacije korisnika testova s jasno naznačenim minimalnim veštinama i kompetencijama koje korisnici treba da imaju. Ovaj sistem je naveden u okviru pređašnjih verzija Standarda, međutim u sadašnjoj verziji ovaj sistem kvalifikacije se više ne navodi. Bez obzira na to što kvalifikacije korisnika nije više deo Standarda, mnogi izdavači praktikuju ovaj sistem prilikom odluke o davanju dozvole za korišćenje određenih testova.

Sistem kvalifikacije obuhvata tri nivoa koji su označeni kao A, B i C nivo. Ovaj sistem je usvojilo i Društvo psihologa Srbije i može se naći u okviru Pravilnika o standardima i procedurama izrade i upotrebe psiholoških mernih instumenata[37]. Detalji u vezi sa potrebnim kompetencijama za svaki od nivoa dati su u Tabeli 27.

[37] http://www.dps.org.rs/pocetna/594

Tabela 27

Kategorizacija testova i kvalifikacija korisnika

	A	B	C
Namena testa	za istraživačke svrhe/projekte	psihodijagnostika i prognoza	proceduralno i interpretativno složena psihodijagnostika i prognoza
Potrebne kvalifikacije	bazična akademska znanja iz psihometrije i statistike; min. 180 ESPB	bez dodatne obuke, dovoljno je fakultetski stečeno znanje i uputstvo iz priručnika; min. 240 ESPB	min. 240 ESPB i dodatna teorijska i praktična edukcija

Napomena: Potebne kvalifikacije preuzete iz su Pravilnika o standardima i procedurama izrade i upotrebe psiholoških mernih instumenata Društva psihologa Srbije.

U praksi, kada neki korisnik želi da naruči određeni test, on najpre mora ispuniti obrazac za nabavku. U okviru tog obrasca navode se i kompetencije korisnika, te na osnovu toga izdavač procenjuje da li korisnik ima potrebne kvalifikacije za primenu naručenog testa. Na primer, u prethodnim verzijama Standarda je navedeno da za primenu C testova, pored master diplome iz oblasti psihologije, korisnik treba da ima i minimum jednu godinu superviziranog iskustva u oblasti primene testa (APA, 1950). Tako, ukoliko neko potražuje test C nivoa, potrebno je da u prijavi navede broj diplome i od strane koje ustanove je diploma izdata, završene edukacije i slično. Izdavač zadržava prava da zatraži formalni dokaz o navedenim kvalifikacijama korisnika, kao što su kopija diplome, sertifikat o završenoj edukaciji itd. Međutim, nekada nije tako lako proveriti kvalifikacije korisnika, posebno kada je reč o zahtevima za upotrebu C klase testova. Na primer, neko može raditi u kliničkom centru, ali nema iskustva u radu sa pacijentima sa neurološkim problemima a želi da koristi neki neurološki test. U tim situacijama

izdavačka kuća treba da obezbedi edukaciju koja ide s kupovinom testa, ali se ona, naravno, dodatno plaća. Pored toga, nevolja je i u tome što ne postoji saglasnost među izdavačkim kućama u vezi sa određivanjem nivoa testa, i to, prema nekim nalazima, čak u dve trećine testova (Simmer, 1994).

Najmanji zahtevi se postavljaju pred testove koji se koriste u istraživačke svrhe zato što se na osnovu njih ne donosi odluka o pojedincu, te takvi testovi nikada nisu testovi s visokim ulogom. Za njihovu primenu i adekvatno tumačenje rezultata dovoljna su osnovna znanja iz statistike i psihometrije, te bi podjednako uspešno trebalo da ih primeni i neko s ovim osnovnim znanjima i neko ko ima više kompetencije. Primer takvog komercijalnog testa je Test opšte informisanosti – Sinapsa ediciije (TOI-SE: Janičić i Dinić, 2016), koji je namenjen proceni sposobnosti prikupljanja informacija i brzog učenja, te se može koristiti kao dopuna proceni intelektualnih kapaciteta osobe, ali ne i kao zamena za procenu istih. Za primenu ovog testa je dovoljno bazično znanje iz psihometrije i statistike i priručnik, te ga mogu primeniti i nastavnici iz društvenih nauka, sociolozi, pedagozi i sl.

Valja napomenuti da sam predmet merenja testa nije nešto što određuje njegovu kategoriju, već cilj primene. Tako je moguće je da postoji test inteligencije koji se zadaje grupno i za čiju je primenu i tumačenje rezultata dovoljno uputstvo iz priručnika, ali isto tako postoji i test inteligencije namenjen individualnoj primeni, koji obuhvata i manipulativne testove, praćenje postupka rešavanja problema kod ispitanika i sl., te zahteva dodatnu edukaciju za njegovu primenu i tumačenje profila mentalnih sposobnosti.

Odgovornosti korisnika testa

Na nekoliko mesta je rečeno da korisnik testa ima najveću odgovornost za pravilnu selekciju, upotrebu i interpretaciju testovnih rezultata. Odgovornost kvalifikovanog stručnjaka odnosi se na sve faze procesa testiranja. Korisnik testa najpre treba da poseduje određena znanja, kompetencije i veštine za primenu konkretnog testa, a koja su predviđena kategorizacijom testa. Ukoliko korisnik testa nije kompetentan da zadaje, na primer test kategorije C, ne treba ni da ga zadaje. Studijski programi psihologije su uglavnom organizovani po modulima, barem kod nas, te se može desiti da neki student završi master studije psihologije, a da tokom nije studija video MMPI jer nije završio klinički modul, niti birao kliničke predmete na kojima je MMPI obrađivan. U slučaju da takav master psiholog dobije posao na klinici, iako pravno gledano on ima kvalifikacije da zada kliničke testove kategorije B jer ima master diplomu psihologa, njegova je odgovornost da potraži pomoć u tome jer je činjenica da on za to nije obučen tokom studija. Iako poslodavac može tražiti spisak položenih ispita kako bi utvrdio kompetencije kandidata, u praksi se ovo retko dešava.

Potom, korisnik testa treba da poštuje autorska prava u vezi s testom. Upotreba komercijalnih testova uključuje ograničenja koja određuje vlasnik autorskih prava (najčešće izdavačka kuća ili udruženje) o tome ko može koristiti testovni materijal, da se testovni materijal može koristiti samo u originalom obliku (da je zabranjeno nelegalno umnožavanje testova), kao i kako će se obezbediti sigurnost i čuvanje testovnog materijala i tesovnih podataka. Jedno od ograničenja koje vlasnik autorskih prava može dati je da se test ne može slobodno, tj. besplatno koristiti za istraživačke svrhe, te da se i za te svrhe mora otkupiti potreban broj primeraka testa. Ukoliko se test koristi za

istraživačke svrhe (bilo uz dozvolu da se koristi sa ili bez novčane nadoknade), vlasnik autorskih prava ne sme da određuje prirodu i cilj istraživanja (ITC, 2014). Takođe, vlasnik autorskih prava može odbiti zahtev da se test prevede na drugi jezik ili da se uradi bilo kakva modifikacija testa. Svaka modifikacija testa mora prethodno da prođe pismeno odobrenje vlasnika autorskih prava.

Za ostale testove koji se nalaze u javnom domenu ili otvorenom pristupu, vlasnik autorskih prava je obično autor testa. Iako se testovi koji su u javnom domenu mogu koristiti bez odobrenja autora (osim ukoliko nije drugačije naznačeno), uvek se preporučuje da se autor kontaktira pre primene testa i prevođenja, ukoliko test nije na srpskom jeziku. Pored toga što je kontaktiranje autora lep akademski gest kojim se iskazuje poštovanje prema autoru i uspostavlja saradnja s njim, na taj način se može dobiti i informacija da li je test već preveden na srpski. Tako se ne bi radio dupli posao i da ne bi postojale različite verzije testova.

Korisniku testa mora biti potpuno jasna svrha testiranja, mora da zna zašto neki test primenjuje i mora da bude u potpunosti upoznat sa procedurom testiranja konkretnog testa. Dakle, korisnik testa treba da argumentuje selekciju testa. Izbor testa se argumentuje predmetom merenja testa, a koji je u skladu sa ciljem procene, potom dokazima o psihometrijskim kvalitetima testa, validnosti zaključaka koji se donose na osnovu rezultata testa, opravdanosti upotrebe testa na konkretnom ispitaniku ili uzorku ispitanika, aktuelnošću normi i slično, te da se rezultati testiranja koriste zajedno s ostalim podacima. Ukoliko se odstupa od navedenih standarda, onda to treba dodatno argumentovati (npr. zašto se koristi test sa zastarelim normama, ili ako se test koristi za svrhe koje autor testa nije predvideo, onda treba dostaviti dokaze za

korišćenje testa za svrhu koju je korisnik odredio). Potom, korisnik treba u potpunosti da prouči priručnik i testovni materijal, te provežba zadavanje testa. U slučaju testova B i C kategorije, ovo vežbanje zadavanja testova se obično radi pod supervizijom (npr. predmetnog nastavnika na fakultetu u slučaju B testova ili edukatora u slučaju C testova).

Pored toga, korisnik ima odgovornost da obezbedi nepristrasno testiranje. Korisnik testa unapred treba da pretpostavi moguće negativne posledice i da ih predupredi. Na primer, ukoliko zna da će mu na ispitivanje doći osoba s invaliditetom, treba da pripremi uslove i sam se pripremi za testiranje. Svakako da se ne mogu svi problemi prilikom primene testa predvideti, ali bi korisnik testa trebalo da bude upoznat sa opštim standardima šta treba da radi u specifičnim okolnostima ili ukoliko dođe do nepredviđenih problema. U situacijama kada se šteta ne može sanirati, npr. testiranje je moralo biti prekinuto, korisnik treba o tome napisati izveštaj koji će priložiti zajedno s rezultatima testa čiji je rad prekinut ili rezultatima ponovnog testiranja.

Korisnik testa, takođe, ima odgovornost da ispitanicima i svim zainteresovanim stranama i onima koji imaju prava na to, adekvatno objasni i interpretira rezultate. Nekada treća lica pojednostavljeno gledaju na rezultate testova, traže instant odgovore samo na osnovu testovnih rezultata, ili traže objašnjenja zašto su testovi korišćeni. U skladu sa ciljem testiranja i time kome je namenjen izveštaj o rezultatima, korisnik testa treba na razumljiv način da nekome ko nije psiholog intepretira rezultate testa, i osvrne se na ograničenja rezultata testa – koji se zaključak može, a koji ne može izvesti na osnovu rezultata testa. Korisnik testa ima jedini odgovornost u vezi s interpretacijom testovnih rezultata kada testove koristi za procenu svojih klijenata,

pacijenata i slično, ali i tada treba na adekvatan način da saopšti rezultate klijentima i odluke donete u skladu s psihološkom procenom. Kada je ispitanik dete, obično se rezultati saopštavaju roditeljima i/ili vaspitačima, učiteljima i nastavnicima. Ukoliko je testiranje naručeno, korisnik ima ulogu konsultanta, te je odgovornost korisnika podeljena, ali da bi treća strana mogla adekvatno da donese odluku, potrebno je da izveštaj o rezultatima bude jasan, razumljiv i pažljivo napisan da ne bi bilo zloupotrebe ili pogrešno donetih odluka. Jedina situacija kada se povratna informacija o rezultatima obično ne daje je kada se testovi primenjuju u svrhe selekcije ili unapređivanja kadrova, dok se u svim ostalim situacijama očekuje da se dâ povratna informacija o rezultatima, bilo u pisanoj ili usmenoj formi.

Korisnik testa takođe ima odgovornost i za testovnu dokumentaciju, njeno uredno čuvanje i obezbeđivanje njene sigurnosti i zaštite od zloupotrebe. Ukoliko korisnik isti test aktivno koristi godinama, posebno u grupnim testiranjima na lokalu, onda bi trebalo da obezbedi da ponovna primena testa ne ugrozi validnost rezultata (npr. ako se pročuje koji se test koristi i kakava su pitanja u njemu).

Na kraju, treba podvući da se od korisnika testa očekuje da bude profesionalan i etičan, da iskaže kompetentnost u primeni testova i interpretaciji rezultata, da preuzima odgovornost za sve akcije i odluke donete u procesu testiranja, da osigura zaštitu testovnog materijala, kao i poverljivost testovnih rezultata (AERA, APA, & NCME, 2006).

Problemi prilikom selekcije testova

Problem prilikom selekcije testova povezan je najpre sa njihovom dostupnošću. Kao što je rečeno, kod nas postoje dva izvora od kojih se mogu nabaviti testovi koji su standardizovani na teritoriji

Republike Srbije, a to su Društvo psihologa Srbije i izdavačka kuća Sinapsa edicije. Do skoro je postojao samo jedan izvor, te je testovna ponuda bila priilčno ograničena. Međutim, u ponudi testova postoji balanas između testova koji su nastali od strane domaćih autora, i postojećih testova koji su adaptirani na srpski jezik. Drugi problem prilikom selekcije testova povezan je sa neinformisanošću korisnika testa o postojećim testovima na tržištu. Naime, korisnici testova koriste one testove s kojima su imali iskustvo na fakultetu, bez obzira na to da li je njihova primena za konkretni cilj procene adekvatna (npr. korišćenje MMPI-ja u cilju profesionalne selekcije). Česta je praksa i da se koriste zastarele verzije testa (kao što se kod nas u zdravstvu, pa i na klinikama često može sresti korišćenje pređašnje verzija MMPI-ja).

Takođe, korisnik testa treba da selektuje test u skladu sa njegovom namenom na način kako ju je autor odredio. Neki autori kao zlatno pravilo prilikom selekcije testa ističu upravo to da test mora biti prikladan za svrhu za koju se koristi (Cronbach, 1990; Gronlund & Linn, 1990). U tom smislu, korisnik testa treba da razume način konstrukcije, administracije, skorovanja i interpretacije konkretnog testa i da odredi da li se ti aspekti podudaraju sa aktuelnim potrebama za testiranjem. Ukoliko se test koristi za neku drugu namenu u odnosu na to kako ju je autor testa odredio, onda korisnik testa to treba da argumentuje. U skladu s tim, korisnik testa treba da odabere test koji je adekvatan za upotrebu na konkretnom ispitaniku, odnosno koji je razvijen na uzorku koji se poklapa sa važnim karakteristikama ispitanika (pol, starost, obrazovanje, opšta ili klinička populacija, pripadnost jezičkoj ili drugoj manjini). Na primer, ukoliko ne postoje norme za stare za neki test, ni tumačenje normi za starijeg ispitanika neće biti adekvatno, te je upitno zašto je takav test odabran. Kada se razmatraju karakteristike

normativnog uzorka, treba uzeti u obzir i veličinu normativnog uzorka, kao i aktuelnost normi. Na ovom mestu treba pomenuti i sve vezano za obezbeđivanje nepristranosti testiranja, posebno kada je reč o testiranju osoba s invaliditetom, osoba iz manjinskih grupa i slično. Ukoliko test već nije prilagođen testiranju ispitanika koji pripada nekoj od navedenih grupa, onda korisnik testa treba da obrazloži upotrebu konkretnog testa.

Načelno, ne bi trebalo koristiti test ukoliko važne karakteristike ispitanika nisu obuhvaćene referentnom grupom na kojoj je test razvijen ili normativnim uzorkom, ukoliko procedura administracije ili skorovanja ne sledi onu koja je primenjena prilikom standardizacije testa, ukoliko karakteristike testa mogu uticati na upotrebljivost rezultata (npr. dobijanje efekta poda ili plafona), ukoliko test sadrži kulturološki specifične zadatke i pitanja, i ukoliko dokazi o validnosti testa ne podržavaju odluke koje bi se donele na osnovu skorova na testu (DeMers et al., 2000).

Kada je reč o selekciji testova na osnovu psihometrijskih karakteristika i tu se mogu javiti problemi, te da se za psihološku procenu odabere test koji nije dovoljno pouzdan ili validan. Naime, to što je neki test standardizovan i za koji postoji priručnik ne znači nužno i da ispunjava sve visoke standarde za primenu. Korisnik testa treba kritički da pristupi oceni testa na osnovu dostupnih informacija iz priručnika (na kom uzorku je standardizovan test, kakve su mu psihometrijske karakteristike, koji podaci o validnosti su prikazani...). U profesionalnoj praksi gde se testovi upotrebljavaju u cilju psihološke procene, testovi moraju biti visoko pouzdani i validni kako bi omogućili preciznu procenu ispitanika. Zahtevi za psihometrijskim karakteristikama su fleksibilniji kad se testovi koriste u cilju trijaže, tj.

grube procene ili u istraživačke svhe. Postoji nekoliko razloga zašto su zahtevi fleksibilniji kada se testovi koriste u istraživačke svrhe. Prvi je što se testovi mogu konstruisati u cilju provere nekih inovativnih hipoteza ili ispitivanja nedovoljno istraženih konstrukata, kako bi se proverila stabilnost pretpostavljenog modela ili teorije. Drugi razlog je taj što testovi koji se koriste u istraživačke svrhe često nisu standardizovani, niti su im populacijske vrednosti i psihometrijske karakteristike uvek poznate pre sprovođenja istraživanja (npr. ako se test prvi put koristi na određenom jeziku). Treći razlog je što se testovni podaci najčešće koriste na uprosečenom ili grupnom nivou, bez izvođenja zaključaka o pojedinačnim ispitanicima i ovo su, po pravilu, testovi niskog uloga (ITC, 2014).

Problemi prilikom administracije testova

Adekvatno zadavanje testova podrazumeva pažljivu pripremu uslova testiranja, ispitanika i ispitivača.

Priprema uslova testiranja

Ovaj aspekt pripreme odnosi se na anticipaciju i uklanjanje svih potencijalnih izvora distrakcije u prostoru u kojem se odvija testiranje. U to spadaju fizički uslovi – prostorija u kojoj se radi treba da je osvetljena i provetrena, da je zvučno izolovana, da ispitanik ima dovoljno mesta za rad i da ispitivač može da posmatra rad ispitanika tako da ne ometa samog ispitanika, da u prostoriji nema vizuelnih distrakcija, niti pića i hrane (osim vode ako je potrebna ispitaniku tokom rada), da se mobilni telefoni utišaju ili isključe i ne drže na stolu, da se alarmi na satovima isključe i sl. Takođe, prostorija ne sme biti prometna u smislu da druge osobe mogu da ulaze u nju ili da se

šetaju po njoj tokom testiranja. U prostoriji tokom rada treba da budu prisutni samo ispitanik i ispitivač, ukoliko je reč o individualnom testiranju, odnosno ispitanici i ispitivač i njegovi pomoćnici ukoliko je reč o grupnom testiranju. Kod grupnog testiranja treba povesti računa o tome da ispitanici budu na dovoljnoj udaljenosti, da ne prepisuju i ne komuniciraju međusobno rokom rada i da ne ometaju jedni druge u radu. Priprema uslova testiranja podrazumeva strogo pridržavanje standardizovane procedure zadavanja testa, prateći instrukcije iz priručnika za test. Kada se sledi standardizovana procedura bitno je da u onome što ispitivač kaže ispitanicima nema ni viška, ni manjka. Na primer, čak iako je svima jasno uputstvo za rešavanje nekog testa i svi su uspešno rešili primer, ukoliko su testom predviđeni zadaci za vežbu, oni se moraju proći sa ispitanicima. S druge strane, ukoliko nekom ispitaniku nije jasan princip rada zadataka čak i posle primera i vežbe, ne može mu se dati novi primer koji nije predviđen uputstvom, već se jedino mogu ponoviti primeri i zadaci za vežbu. Takođe, ne može mu se dati ni dodatno objašnjenje koje nije predviđeno uputstvom, niti nikakva naznaka načina rada koja nije predviđena uputstvom, već mu se samo može ponoviti ono što već stoji u datom uputstvu.

Ukoliko ispitaniku nisu obezbeđeni adekvatni uslovi za rad, ili se odstupa od strandardizovane procedure zadavanja testa, postoji opasnost da dobijen rezultat neće odražavati pravi stepen merenog svojstva ispitanika. Nekada su ove situacije nepredvidive, npr. uključi se požarni alarm ili počnu građevinski radovi u blizini prostorije testiranja. Ponekad je prisustvo trećih lica neophodno, npr. prevodilaca, asistenata za ispitanike s invaliditetom ili roditelja, što može narušiti standardizovanost procedure. U toj situaciji, kao i u bilo

kojoj drugoj gde se odstupa od standardizovane procedure, to se treba ubeležiti u izveštaj o rezultatima.

Priprema ispitanika

Priprema ispitanika odnosi se na dva aspekta. Prvi aspekt zavisi od ispitivača i podrazumeva uspostavljanje prijatne i profesionalne atmosfere. O tome je već bilo reči prilikom isticanja važne činjenice o procesu testiranja, a to je da on uvek podrazumeva **međuljudski odnos** (AERA, APA, & NCME, 2006). Interakcija ispitivača i ispitanika je nekada veoma intenzivna, kao u slučaju individualnih testova, a nekada svedena na minimum, u slučaju masovnih testiranja, ali je uvek prisutna. Važna karakteristika ovog odnosa je u *disbalansu*, jer ispitivač i ispitanik nisu u istom položaju – ispitanik je uvek u podređenom položaju u odnosu na ispitivača. Ovaj disbalans je posebno uočljiv prilikom testiranja testovima s visokim ulogom kada izveštaj o rezultatima značajno može uticati na život ispitanika. Stoga, odnos između ispitivača i ispitanika mora biti profesionalan, ljubazan, bižan i s uzajamnim poštovanjem. Cilj uspostavljanja adekvatne atmosfere testiranja je u redukciji disbalansa između ispitivača i ispitanika i pružanja mogućnosti da ispitanik maksimalno izrazi svoje kapacitete na testu. Iako se razlika u poziciji moći između ispitanika i ispitivača može relaksirati prijatnom i profesionalnom atmosferom, ovaj disbalans u pozicijama je neizbežan i uvek prisutan (Drenth & Sijtsima, 2006).

Ostvarivanje adekvatne atmosfere pruža se jasnim i ljubaznim informisanjem ispitanika o svrsi testiranja, predmetu merenja testa (sposobnosti, stavovi, osobine ličnosti...), kriterijumu selekcije, pa potom načinu rada, načinu korekcije odgovora i slično. Isticanjem

međuljudskog odnosa u procesu testiranja ukazuje se na to da se pažnja ovom aspektu testiranja takođe mora posvetiti, pored pažnje koja se posvećuje tehničkim kvalitetima testa. Naime, odnos između ispitanika i ispitivača može uticati na objektivnost testiranja i validnost interpretacije testovnih rezultata. U Standardima (AERA, APA, & NCME, 2006) se posebno ističe odnos ispitivača i ispitanika kada je ispitanik osoba s invaliditetom, pripadnik druge etničke, rasne, religiozne grupe, drugog pola ili polne orijentacije, socioekonomskog statusa, starosti i drugih atributa koja mogu uticati na interakciju ispitivača i ispitanika. Problemi u ovom aspektu procedure testiranja najčešće potiču od neiskustva ispitivača, kada ispitivači možda nisu dovoljno osetljivi na signale zbunjenosti ili uplašenosti kod ispitanika, kada brzaju u davanju uputstva i slično.

Drugi aspekt pripreme ispitanika zavisi od samog ispitanika. U ovom kontekstu tri konstrukta se najčešće pominju u literaturi, a to su: testovna anksioznost, test i testovna mudrost i testovna praksa. **Testovna anksioznost**[38] (eng. *test anxiety*) je anksioznost koja postoji pre i tokom testiranja. Ona se definiše kao skup fizioloških, kognitivnih i bihejvioralnih reakcija koje prate zabrinutost zbog mogućih negativnih posledica neuspeha na testu, ispitu ili u nekoj drugoj evaluativnoj situaciji koja uključuje testiranje (Sieber, O'Neil, & Tobias, 1977). Rezultati dosledno pokazuju da je testovna anksioznost negativno povezana s postignućem na testovima sposobnosti, a takođe ostavlja negativne posledice i po mentalno zdravlje (više u Genc, 2014). Testovna anksioznost čini da ispitanik nije u stanju da isprati i

[38] U literaturi se nekada naziva i ispitna anksioznost kada se koristi u širem značenju, i obuhvata sve ispitne provere koje ne moraju nužno uključiti test.

razume instrukcije za rešavanje zadataka, da razborito razmišlja i da ima teškoća u prisećanju naučenog materijala. Stoga je ishod prisustva testovne anksioznosti takav da dve osobe koje imaju podjednako znanje ili sposobnosti, imaju različite skorove na testu. Osoba koja ima testovnu anksioznost ostvaruje sistematski niže skorove. Iako su opšta anksioznost i testovna anksioznost povezane, pokazano je da su u pitanju distinktivni konstrukti (više u Genc, 2014). Razlozi pojave testovne anksioznosti mogu biti povezani sa karakteristikama testa, npr. ako je u pitanju test brzine ili numerički test, sa karakteristikama ispitanika, npr. ako ima negativna očekivanja od ishoda testiranja, strah od negativne evaluacije, kao i sa namenom testiranja, poput selekcije, ili mogu poticati od kombinacije navedenih karakteristika (Urbina, 2014).

Testovna mudrost (eng. *test wiseness, test sophistication, test-taking skills*) je „bilo koja vrsta sposobnosti koja omogućuje ispitaniku da poboljša svoj rezultat koristeći različite dopunske informacije kao što su tekst pitanja, format pitanja, tekst stavki i informacije iz opšteg konteksta testiranja" (Fajgelj, 2013, str. 252). Testovna mudrost je povezana sa znanjem i sposobnostima koje su predmet merenja testa, i zapravo predstavlja spoj poluznanja i snalažljivosti na testu. Testovna mudrost predstavlja korišćenje poluznanja i samih informacija iz testa koje pomažu u prepoznavanju tačnih odgovora ili eliminaciji netačnih. Ishod prisustva testovne mudrosti je da dve osobe koje imaju podjednako znanje ili sposobnosti ostvaruju različite skorove u zavisnosti od testovne mudrosti, te da osoba koja je testovno mudrija ima viši skor. Kako Fajgelj (2013) navodi, testovna mudrost ne meri znanje nego uspešnu reakciju na loše konstruisane testove. Iako je testovna mudrost nešto što zavisi od ispitanika, odgovornost za

ugožavanje validnosti testa od strane ispitanika koji ima izraženu testovnu mudrost je na autoru testa. Dakle, ispitanik koga karakteriše testovna mudrost samo uspešno iskorišćava propuste koji su napravljeni prilikom konstrukcije testa.

Najičigledniji primer je kada jedno pitanje u testu sugeriše odgovor na neko drugo pitanje u testu. Takođe, kada se neki učenik navikne na kognitivni stil nastavnika, može da pretpostavi način na koji nastavnik pravi pitanja, pa onda i lakše može da pretpostavi šta je tačan odgovor, da li je nešto trik pitanje i slično.

Testovna praksa (eng. *test practice, practice effect*) ili uvežbavanje podrazumeva iskustvo sa testiranjem uopšte, uključujući sve korake testiranja kroz koje prolazi ispitanik, kao i iskustvo sa konkretnim testovima i tipovima pitanja. Ispitanik ne mora imati iskustvo sa istim testom, već je dovoljno da ima iskustvo sa sličnim testom po pitanju vrste testa i formata ajtema, sa alternativnom formom testa ili sa medijumom zadavanja testa (npr. papir-olovka ili računarski podržan). U jednom slučaju testovna praksa poboljšava učinak na testu, te se u tom kontekstu može shvatiti kao jednostavniji oblik testovne mudrosti. Poboljšanje učinka na testu usled iskustva s testovima može biti zbog pamćenja konkretnih rešenja ili načina rešavanja problema u zadacima (na koji način se mogu menjati direkcije u rešavanju zadataka), usled toga što sadržaj "slegne" ili redukcije testovne anksioznosti. Na primer, ukoliko imamo dva kandidata s podjednako izraženim sposobnostima, ali je jedan već prethodno bio testiran testovima sposobnosti, a drugi nije, prvi ispitanik će biti u prednosti.

Na ovom mestu možemo se podsetiti zahteva nepristrasnosti testiranja a to je obezbeđivanje jednakih uslova za sve i prilika za

učenje. U ovakvoj situaciji, da bi testiranje bilo pošteno, kandidatu koji nema iskustva sa testovima treba obezbediti vežbu sa sličnim testovima, ili istim onim koje je prethodno prošao drugi ispitanik. Svrha ove vežbe nije samo u iskustvu s testovnim materijalom, već i zarad redukovanja testovne anksioznosti i usmerenosti na rad. Nekada se ti testovi za vežbu mogu naći u sklopu testovnog materijala kao odvojeni testovi, a nekada su ti testovi sastavni deo glavnih testova i zadaju se svima, bez obzira na testovno iskustvo, ali se ne skoruju. Istraživanja pokazuju da je efekat testovne prakse veći kod primene testova brzine, nešto veći kod grupih i kod neverbalnih testova, kada je interval između zadavanja testova kraći (do par meseci), a skoro dvostruko veći kada se zadaje isti test ponovo (test-retest), nego kada se zadaje alternativna forma (Benedict & Zgaljardic, 1998). Otud, kada se radi test-retest i testiranje alternativnom formom, ukoliko drugačije nije navedeno u priručniku, preporuka je da prođe minimum 6 meseci između dva testiranja (Urbina, 2014). Na ovom mestu ću napraviti samo jednu digresiju u vezi s omogućavanjem vežbe ili obuke za rad testa. Još je Kronbah govorio o tome da „ako je obuka za neki test opšta pojava, nešto nije u redu s tim testom" (Cronbach, 1990, str. 86). Naime, ukoliko i oni iskusni s testovnim materijalom imaju dobiti od obuke na radu konkretnog testa, taj test se onda čini lošim jer očigledno meri još nešto, pored glavnog predmeta merenja. Kada su u pitanju karakteristike ispitanika, od testovne prakse najviše koristi imaju ispitanici s nižim obrazovanjem i ispitanici visokih sposobnosti (Benedict & Zgaljardic, 1998).

Iako je korist od testovne prakse uočljivija kod kognitivnih testova, ona je prisutna i kod testova ličnosti. U njima se ispoljava kao doslednije odgovaranje na naredna pitanja i testove, tj. odgovaranje u

skladu sa zapamćenim odgovorima na prethodna, slična pitanja. Ova pojava je već bila opisana kao efekat dužine testa na odgovaranje. U drugom slučaju, kada se zadaju velika baterija testova i dugački testovi, može doći do zamora i zasićenja. Bez obzira na to da li je efekat pozitivan ili negativan po krajnji učinak, u oba slučaja govorimo o uticaju prethodnog merenja ili testiranja na naredno.

Priprema ispitanika podrazumeva i podizanje motivacije za rad, kao i uspostavljanje atmosfere koja bi redukovala odgovaračke pristrasnosti i, pre svega, socijalnu poželjnost ili nepoželjnost, ali i pogađanje, o čemu je već bilo reči.

Saveti za ispitanike kako da rešavaju test

Pažljivo pročitaj uputstvo.
Radi brzo.
Ukoliko ti neki zadatak oduzima previše vremena, preskoči ga.
Odgovori čak i kad nisi sasvim siguran, ukoliko nema kaznenih poena.
Ukoliko nema kaznenih poena, pogađaj ako si nesiguran.
Ukoliko ti ostane vremena, proveri odgovore (preuzeto iz Jackson, 2003).

Priprema ispitivača

Ispitivač najpre mora biti upoznat sa svrhom testiranja, testovnim materijalom i biti u stanju da odgovori na sva ispitanikova pitanja u vezi s testiranjem, kao i da adekvatno reaguje u slučaju narušavanja standardizovane procedure testiranja. Grupno testiranje ne zahteva dodatnu pripremu, ali individualno zahteva. Najjednostavniji primer je kada pitanja u testu moramo čitati ispitaniku. Čitanje pitanja mora biti doslovno i ne sme se pitanje prepričavati ili postaviti na drugačiji način, a takođe ispitaniku se treba dati dovoljno vremena za formulaciju odgovora. Odgovori ispitanika se

moraju beležiti doslovno i bez prekidanja toka misli ispitanika. Ukoliko se zadaje manipulativni test, materijal mora na isti način biti prezentovan svim ispitanicima, u istoj poziciji i orijentaciji i slično. Ovo je nešto što ispitivač mora da uvežba (brzo skupljanje i slaganje materijala, praćenje i beleženje odgovora ispitanika bez prekidanja, posebno kada je reč o vremenski ograničenim testovima ili testovima u kojima se beleži vreme rada i sl.).

Primena testova na osobama s invaliditetom

Primena testova na osobama s invaliditetom najčešće podrazumeva modifikaciju postojeće standardizovane primene testa. Međutim, najpre treba odrediti da li je modifikacija testa potrebna. Naime, mnogi ljudi imaju posebne potrebe koje ne utiču na učinak na testu. Kada je odeđena posebna potreba povezana s merenim konstruktom, i verovatno je da će uticati na učinak ispitanika, tada treba razmisliti o modifikaciji postupka testiranja.

Svaka modifikacija testa se posmatra kao nov test kojem je potrebno utvrditi psihometrijske karakteristike. S obzirom na to da je uzorak za ovakve modifikacije često mali i raznovrstan po pitanju tipa invaliditeta, u literaturi se ne nalaze poređenja različitih aspekata procedure testiranja u strandardizovanim i modifikovanim uslovima, što otežava donošenje odluke o adekvatnom načinu modifikacije procedure testiranja. Ipak, osnovno pravilo je da modifikacija mora biti takva da ne podrazumeva promenu merenog konstrukta, ali može da podrazumeva izmenu oblika prezentovanja ajtema (medijum, tip pitanja...), načina odgovaranja, uslova primene testa (individualno umesto grupno, promena mesta testiranja ukoliko mesto nije prilagođeno osobama s invaliditetom...), produženo vreme za rad, zadavanje samo dela testa ili selektovanih ajtema ukoliko je to adekvatno (prilikom sticanja sertifikata i zapošljavanja nije, ali u nekim drugim situacijama procene jeste). Najbolja modifikacija je upotreba zamenskih testova, posebno konstruisanih za primenu na osobama s invaliditetom, ali su ovakvi testovi u praksi retki.

Prilikom izveštavanja o rezultatima testova trećim licima, kada se radi o individualnom testiranju, izveštaj treba da sadrži i podatke o modifikaciji. Kada je reč o grupnom testiranju, ukoliko postoje dokazi da modifikovana i standardna primena daju uporedive rezultate, ne treba posebno označavati da je u pitanju modifikovana primena. Ukoliko nedostaju takvi dokazi, onda nikako nije poželjno samo označiti protokole koji su modifikovani (jer se time, na neki

način, vrši stigmatizacija osoba s invaliditetom), već treba dati specifične informacije o prirodi modifikacije, ukoliko je zakonski dopuštena.
Više o primeni testova na osobama s invaliditetom može se naći u Standardima (AERA, APA, & NCME, 2006) i u Međunarodnim smernicama za upotrebu testova (ITC, 2000).

Procedura testiranja mora biti transparentna ispitaniku. To znači da ispitanik treba da bude informisan o cilju i svrsi testiranja, o tome šta se testovima procenjuje (sposobnosti, stavovi, osobine ličnosti), o kriterijumu u slučaju klasifikacije (npr. najbolja tri kandidata će biti primljena) i o svojim rezultatima. Ukoliko je neki od ovih aspekata informisanja ispitanika zanemaren, postoji problem s nepravilnom upotrebom testa. U skladu sa etičkim principima koje je propisala APA (APA, 2017), ispitanik pre učešća u testiranju mora da dâ svesnu saglasnost, tzv. **saglasnost uz informisanje** ili **informisanu saglasnost** (eng. *informed consent*). Svesna saglasnost proističe iz ispitanikovog prava na slobodu i samoopredeljenje (Frabkfort-Nachmisas & Nachmias, 1992). Ovu saglasnost treba ispitivač da pripremi, a ona treba da sadrži sledeće informacije: koji je cilj testiranja, ko sve ima uvid u podatke i kako će podaci biti zaštićeni i čuvani, kako će se saznati rezultati na testu i gde i kako će se izveštavati, gde i za šta će se rezultati upotrebiti, šta ako ispitanik želi da odustane u toku testiranja, da li snosi neke posledice. Ukoliko postoje neki specifični uslovi testiranja, npr. novčana nadoknada, dobijanje bodova, uključivanje trećih lica i slično, to sve treba navesti u informisanoj saglasnosti. Takođe, ukoliko se očekuje da testiranje izazove neku nelagodu, uznemirenost i sl., na to treba upozoriti ispitanike i opisati kakvi se rizici očekuju. Ukoliko treba da se testiraju maloletna lica, saglasnost za testiranje treba da daju roditelji, a u

zavisnoti od uzrasta ispitanika i sami ispitanici. Pre potpisivanja informisane saglasnosti, ispitanik treba da razreši sve dileme koje ima u vezi s testiranjem i rezultatima. Ispitaniku se treba omogućiti prilika da postavi pitanja i dobije odgovore.

Saglasnost za učestvovanje u testiranju

Naziv istraživačkog projekta: Standardizacija i provera psihometrijskih karakteristika Testa opšte kulture.

Cilj ovog projekta je standardizacija i provera psihometrijskih karakteristika Testa opšte kulture. Ovaj projekat će pomoći da se dobiju norme za Republiku Srbiju i da se ujednači rad psihologa.

Svi podaci dobijeni u ovom istraživanju biće poverljivi. Samo će istraživački tim imati uvid u podatke. Učesnici u istraživanju neće na samim testovima ostavljati nikakve lične podatke i nikada neće biti imenom i prezimenom navedeni ni u jednom izveštaju ili publikaciji o rezultatima projekta. Pisani izveštaji ili publikcije o rezultatima istraživanja sadržaće informacije samo o grupi, a ne individualne rezultate ili rezultate na nivou odeljenja, škola i drugih institucija. Učestvovanje u ispitivanju je dobrovoljno i od njega se može odustati u bilo kojem trenutku bez ikakvih posledica.

Za sve dodatne informacije o testiranju možete se obratiti koordinatoru projekta, Milici Milićević sa Odseka za psihologiju Filozofskog fakulteta u Novom Sadu (kontakt: milica.milicevic@ff.uns.ac.rs).

Pročitao/la sam i razumeo/la date informacije.

a. Pristajem na učestovanje b. Ne pristajem na učestovanje

U nekim specijalnim okolnostima informisana saglasnost nije potrebna, tj. podrazumeva se (npr. prilikom zapošljavanja ukoliko se rezultati koriste isključivo za selekciju kandidata, a ne i za proveru validnosti testa i slično) ili regulisana pravilinkom ustanove (npr. svi pacijenti koji leže na psihijatrijskom odeljenju moraju proći psihološku procenu) ili je naložena od strane pravnog lica ili institucije (prilikom sudske odluke o upućivanju na psihološku procenu), a nekada se namerno ne kaže pravi cilj merenja, npr. prilikom prikrivenog posmatranja ili neke eksperimentalne manipulacije. Na primer, ukoliko

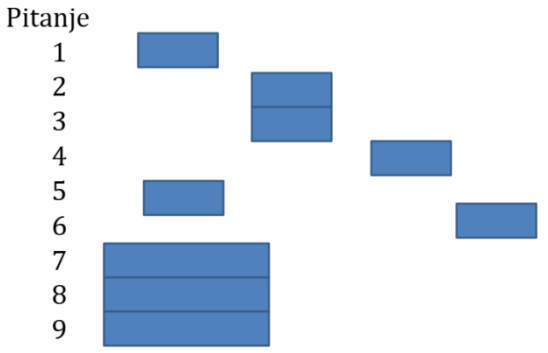
nas zanima agresivno ponašanje ispitanika koji se međusobno takmiče u paru, pri čemu je reakcija protivnika unapred podešena tako da odražava različite nivoe provokacije, da bi ispitanik dao validne odgovore svakako ne sme znati da mu protivnik nije stvaran. U toj situaciji cilj testiranja se treba zamaskirati, ali na kraju procedure se sa ispitanicima mora obaviti debrifing u kojem mu se moraju dati informcije o pravom cilju testiranja. Za detalje u vezi sa specijalnim uslovima, mogu se videti Etički principi psihologa i kodeks ponašanja koje je propisala APA (APA, 2017).

U pripremu ispitivača spada i obezbeđivanje poverljivosti testovnog materijala i testovnih protokola i podataka. Ispitivač mora biti upoznat sa legalnim pravima u vezi s distribucijom i čuvanjem testovnog materijala i čuvanjem testovnih podataka, i ne sme dozvoliti umnožavanje testovnog materijala ili korišćenje testovnog materijala od strane nekvalifikovanih korisnika. Takođe, ispitivač mora biti upoznat sa pravima ispitanika i drugim specifičnim pravnim aspektima koji obuhvataju konkretnu vrstu posla ispitivača.

Problemi prilikom skorovanja testova

Greške u skorovanju se mogu javiti usled nepažnje onoga ko skoruje test. Ukoliko se skorovanje radi ručno, olakšica je ukoliko postoje šabloni za skorovanje (Slika 38). Pored toga, savetuje se i da se sabira strana po strana testa, pa onda da se dobije ukupni skor (Mužić, 1968). Poželjna je, takođe, dvostruka provera, posebno ukoliko je reč o testiranju visokog uloga.

Pitanje				
1	T	N		
2	T	N		
3	T	N		
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	_____			
8	_____			
9	_____			



Slika 38. Primer liste za odgovore i šablona za skorovanje.

Svakako da računarski podržano zadavanje i skorovanje čini da nema grešaka prilikom skorovanja, a ovi softveri ujedno mogu pružiti i prevođenje skorova u norme. Skorovanje olakšavaju i optički skeneri, posebni uređaji koji skeniraju test i kao izlaz daju tabelu sa podacima ispitanika i bodovanjem za svako pitanje, kao i ukupni skor. Kod primene skorovanja preko skenera bitno je list za odgovore unapred pripremiti tako da skener lako može očitati odgovor (npr. u kućicama, ili svako slovo da bude u posebnom kvadratiću i slično). Izdavačke kuće često imaju razvijene softvere za unos i skorovanje testova, a nije retka praksa i da se sam test nudi u računarski podržanom okruženju.

Skorovanje zatvorenih pitanja je jednostavno, ali otvorena pitanja zahtevaju dodatni angažman. Naime, nekada je teško predvideti varijacije odgovora a koji upućuju na tačno rešenje, te ukoliko unapred nisu definisane sve varijacije tačnog odgovora, računarski podržano skorovanje neće biti ispravno. Kod skrovanja otvorenih pitanja kao dodatna provera tačnosti rešenja savetuje se da se odgovori podvrgnu proveri intersubjektivne saglasnosti eksperata.

Problemi prilikom tumačenja rezultata testova

Donošenje bilo kakve odluke na osnovu testovnih rezultata mora biti argumentovano i opravdano. To znači da onaj ko donosi odluke, a to je najčešće korisnik testa, mora da se pozove na jasan empirijski dokaz koji bi potkrepio njegovu odluku, npr. o tome zašto se očekuje da će ispitanici koji su prošli selekciju pokazati bolji radni učinak u odnosu na one koji nisu prošli selekciju, potom o tome da li je upotrebljen granični skor najoptimalniji za konkretan cilj testiranja, šta sve ispitanik može ili ne može na osnovu rezultata testa i slično. Iako ovaj aspekt primene testova podrazumeva i obrazloženje selekcije testova, više je u vezi sa interpretacijom skora jer se odnosi na utemeljnost donete odluke na osnovu rezultata testa. Na ovom mestu je samo važno podsetiti se razlika između psihološkog testiranja i procene, a osnovna razlika je u tome što psihološka procena zahteva i ekspertsko znanje iz konkretne oblasti procene, a ne samo znanje iz psihometrije i statistike. U kontekstu problema prilikom interpretacije rezultata ovde ćemo se ograničiti na psihološko testiranje, a ne procenu, i to na komunikaciju između psihologa i drugih lica koja koriste testovne rezultate.

U nekim situacijama ispitanici ni ne dobijaju svoje rezultate, npr. kada se testovi koriste u istraživačke svrhe ili za potrebe selekcije i unapređenje kadra. Međutim, kada je reč o primeni testova u istraživačke svrhe, ispitanicima se mogu saopštiti njihovi rezultati pod uslovom da je anonimnost ispitanika garantovana (u smislu da uvid u rezultate neće imati treća lica, niti da će javno biti istaknuti pojedinačni rezultati) i da se pojedinačni ispitanici mogu na neki način identifikovati (preko ostavljenog imejla, šifre, imena i prezimena, ali uz garanciju sigurnosti podataka). Međunarodna komisija za testove naglašava da su istraživači, kad god je to moguće, dužni da se pobrinu oko davanja povratne informacije ispitanicima kao sastavnog dela istraživačkih procedura (*ITC*, 2014*).* Naravno, uvek je bolje da se fidbek dobije odmah, ali to zavisi od mogućnosti istraživača. Izveštaji o rezultatima učesnika u istraživanju uglavnom sadrže sirove skorove ili norme i kratke opise visokih i niskih skorova na primenjenim skalama (ali ne i opis profila). Na primer, na sajtu http://hexaco.org/hexaco-online možete popuniti inventar HEXACO-PI-R i dobiti podatke o vašim skorovima, medijani i rasponu od 80% skorova na referentnom uzorku, objašnjenje o tome kako tumačiti skorove, kao i o značenju visokih i niskih skorova na skalama inventara. Kada se rezultati saopštavaju ispitanicima na ovaj način, potrebno je dati i neko opšte uputstvo kako se skorovi tumače, šta su visoke, a šta niske vrednosti i slično. Ukoliko je reč o istraživanju u kojem je bilo obmanjivanja ili neke druge eksperimentalne manipulacije, obavezno je sa ispitanicima uraditi debrifing po završetku istraživanja, te je to prilika i da im se saopšte rezultati ukoliko je moguće da se automatski odmah izračunaju.

Direktno saopštavanje rezultata ispitaniku ima mnoge prednosti, a osnovne su dobijanje informacija o ispitanikovoj reakciji na testiranje i o kontekstu testiranja. Na ovaj način možemo dobiti informaciju o tome da li je test kod ispitanika izazvao anksioznost i uznemirenost, što možda ne bismo primetili u slučaju grupnog zadavanja testova, potom kako je pristupio testovnom materijalu, pa samim tim i kako reaguje na rezultate (da li se slaže s njima, ili ne, da kritikuje uslove testiranja ili sebe za dobijeni niski rezultat, da li je iznenađen rezultatima), te ukoliko je bilo nekih propusta u zadavanju testa, možemo dobiti informaciju o tome da li je to uticalo na ispitanikov rezultat ili ne.

Interpretacija rezultata nije samo izveštavanje o skorovima i normama i svrstavanje ispitanika u kategorije ili tipove. Ispitaniku ništa ne znači što na skali Ekstraverzije ostvaruje skor 30 ukoliko ne zna referentnu vrednost spram koje može svoj skor da poredi. Takođe, ispitaniku ništa ne znači da na skali Ekstraverzije ostvaruje 80. percentil ukoliko ne zna značenje percentila i u odnosu na koju grupu on zauzima 80. percentil. Ukoliko se ispitaniku pripiše neka dijagnostička kategorija, nije dovoljno navesti naziv te kategorije, već je potrebno i objasniti je u kontekstu skorova ispitanika.

Ono što interpretacija rezultata treba da sadrži na osnovnom nivou, uključuje sledeće elemente:

1. šta je predmet merenja testa;
2. koje je značenje konkretnih skorova ispitanika, da li su niski, prosečni, visoki itd.;
3. koja su ograničenja preciznosti merenja, a koja potiču od greške merenja, u kom opsegu varira pravi skor ispitanika;

4. treba predupredititi neke uobičajene pogrešne interpretacije, npr. u vezi s IQ skorom;
5. način na koji će rezultati testova biti iskorišćeni.

Izveštaj treba biti smislen, upotrebljiv i razumljiv onome kome je namenjen, i da sadrži dovoljno informacija, dakle ne premalo, ali ne i previše. Na višem nivou, ukoliko je cilj testiranja složeniji, te obuhvata i uvod u profile i integraciju podataka iz različtih izvora, interpretacija treba da sadrži izvođenje korisnih i ekološki validnih zaključaka i preporuka, npr. za dalji tretman, usavršavanje itd. Ono što je sigurno je da u takvim interpretacijama treba izbegavati opšte fraze koje važe za sve ili za većinu ljudi (tzv. Barnum efekat kojim se opisuje neupotrebljivost ovakvih generalnih opisa). Prilikom saopštavanja rezultata na testu jasno treba naglasiti da su oni samo jedan od izvora podataka i da ih uvek treba sagledati s ostalim informacijama o ispitaniku, uslovima testiranja i slično (ITC, 2000).

Primer pisane interpretacije rezultata na Testu opšte informisanosti

Ovaj test meri opštu sposobnost izraženu kao uočavanje odnosa među elementima i jasno mišljenje.

Vaš rezultat na testu iznosi 26 od mogućih 36, što znači da ste tačno odgovorili na 72% zadataka. Vaš rezultat je bolji od 70% ispitanika s istim karakteristikama kao Vašim, kao što su pol i starosna kategorija. Rezultati na testu mogu biti promenljivi, odnosno Vi ste mogli postići nešto bolji ili gori rezultat u nekoj drugoj situaciji. Granica greške rezultata na ovom testu je +/-3 boda, i postoji verovatnoća od 95% da je Vaš rezultat u rasponu od 23–29.

Rezultati ovog testa koristiće se za donošenje odluke o stipendiranju, pri čemu će se uzeti u obzir i drugi podaci kao što su prosek i dužina studiranja.

U slučaju da imate pitanja u vezi sa testom, načinom korišćenja podataka ili pojašnjenjem rezultata na testu, molimo Vas da se obratite Petru Petroviću telefonom (021/552-882) ili imejlom (petar.petrovic@vojvodina.gov.rs).

Prava i odgovornosti ispitanika

U okviru Međunarodnih smernica za upotrebu testova (ITC, 2000) navode se, između ostalog, i prava i odgovornosti ispitanika, a one su detaljno izlistane i u okviru Prava i odgovornosti ispitanika koje je predložila APA (APA, 1998; DeMers et al., 2000), zajedno sa odgovornostima korisnika testa. Ispitanici imaju pravo da budu informisani o svojim pravima i odgovornostima vezanim za upotrebu njihovih testovnih rezultata i pristup tim rezultatima; da se prema njima ophodi pristojno, s poštovanjem i pravedno s obzirom na njihovu, rasu, pol, starost, invaliditet itd.; da budu informisani o svrsi testiranja, prirodi testa, ko će imati uvid u rezultate i šta se planira s rezultatima (sve ovo treba da stoji u informisanoj saglasnosti za ispitanika); da unapred budu obavešteni o terminu testiranja, da li i kada će rezultati biti dostupni itd.; da traže posebne uslove testiranja ukoliko za to imaju osnova; da testiranje i interpretaciju rezultata obavlja kompetntna osoba; da znaju da li je testiranje obavezno i koje su posledice ako ne pristanu ili u toku testiranja odustanu od njega, i slično.

Obaveze ispitanika su da pažljivo pročitaju ili slušaju koja su njihova prava i obaveze; da se prema drugima odnose pristojno i s poštovanjem tokom procedure testiranja; da postave pitanja u vezi sa svim što im nije jasno pre početka testiranja; da pažljivo prate instrukcije koje im daje ispitivač i da se na testu predstave iskreno; da budu svesni posledica neučestvovanja u testiranju ili odustajanja od njega ukoliko ih ima, i spremni da ih prihvate; informišu ispitivača ili odgovarajuću osobu u vezi s tim ako veruju da uslovi testiranja mogu uticati na rezultat, tj. smatraju da njihovi rezultati nisu adekvatan pokazatelj njihovih svojstava, i slično. U Standardima (AERA, APA, & NCME, 2006) se još dodaje da ispitanici imaju odgovornost u očuvanju

sigurnosti testa, tj. da ne treba da zloupotrebljavaju testovni materijal koji im se daje, odaju drugima pitanja iz testa i sl.

Pitanja za razmišljanje

1. Koja su dva osnovna razloga za primenu testa? Objasni.
2. Kako glasi zlatno pravilo za primenu testa. Navedite jednu situaciju u kojoj primena testa nije potrebna.
3. Šta podrazumeva nepristrasnost testiranja?
4. Opiši probleme prilikom primene testa koji potiču od korisnika testa.
5. Kako treba pripremiti ispitanika za testiranje?

Reference

American Psychological Association (2017). *Ethical principles of psychologists and code of conduct* (2002, Amended June 1, 2010 and January 1, 2017). Preuzeto sa: https://www.apa.org/ethics/code/ethics-code-2017.pdf

American Psychological Association (1998). *Test taker rights and responsibilities working group of the Joint Committee on testing practices*. Washington, DC: American Psychological Association. Preuzeto sa: https://www.apa.org/science/programs/testing/rights

American Psychological Association (1950). Ethical standards for the distribution of psychological tests and diagnostic aids. *American Psychologist, 5,* 620–626.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education

(2006). *Standardi za pedagoško i psihološko testiranje*. Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of Gender Psychology, 8*(4), 291–322.

Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternative forms. *Journal of Clinical and Experimental Neuropsychology, 20*, 339–352.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York, NY: Harper Collins Publisher.

DeMers, S. T., Turner, S. M., Andberg, M., Foote, W., Hough, L.,... Rey-Casserly, C. M. (2000). *Report of the Task Force on test user qualifications.* Preuzeto sa https://www.apa.org/science/programs/testing/qualifications.pdf

Dinić, B. M., & Raine, A. (2019). An Item Response Theory Analysis and Further Validation of the Reactive–Proactive Aggression Questionnaire (RPQ): The Serbian Adaptation of the RPQ. *Journal of Personality Assessment*. Manuscript accepted for publication. https://doi.org/10.1080/00223891.2019.1573430

Dinić, B. M., & Vujić, A. (2019). The Pathological Narcissism Inventory: Measurement invariance across Serbian and USA samples and further validation. *European Journal of Psychological Assessment*. Manuscript accepted for publication.

Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen.* Houten, the Netherlands: Bohn Stafleu van Loghum.

Društvo psihologa Srbije (2000). *Kodeks etike psihologa Srbije.* Preuzeto sa: http://dps.org.rs/images/kodeksEtike.pdf

Fajgelj, S. (2013). *Psihometrija. Metod i teorija psihološkog merenja* (IV dopunjeno izdanje). Beograd, Srbija: Centar za primenjenu psihologiju.

Frabkfort-Nachmisas, C., & Nachmias, D. (1992). Research methods in the social sciences. London, UK: Edward Arnold.

Furr, M. R., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Thousand Oaks, California: SAGE Publications Ltd.

Genc, A. (2014). *Relacije između stres-procesa i ispitne anksioznosti – distorzije u sećanjima na emocije iz prošlih stresnih transakcija* (neobjavljena doktorska disertacija). Filozofski fakultet, Univerzitet u Novom Sadu, Novi Sad, Srbija.

Goldman, L. (1971). *Using tests in counseling* (2nd ed.). New York, NY: Appleton-Century-Crofts.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York, NY: Macmillan.

Hedrih, V. (2018). *Kroskulturalna adaptacija psiholoških mernih instrumenata.* Niš, Srbija: Filozofski fakultet.

International Test Commission (2000). *International guidelines for test-use: Version 2000.* Preuzeto sa: https://www.intestcom.org/page/5

Jackson, C. (2003). *Psihologijsko testiranje* (2. izdanje). Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Janičić, B. i Dinić, B. (2016). *Test opšte informisanosti – Sinapsa edicije (TOI-SE).* Beograd, Srbija: Sinapsa edicije.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: The Free Press.

Krković, A. (1974). *Klasifikacija ispitanika*. Zagreb: Zavod za produktivnost rada.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2*(2), 57–64.

Mužić, V. (1968). *Testovi znanja*. Zagreb, Hrvatska: Školska knjiga.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.

International Test Commission (2014). *Očitovanje Međunarodnog povjerenstva za testove o upotrebi testova i drugih mjernih instrumenata u istraživačke svrhe*. Zagreb, Hrvatska: Hrvatsko psihološko društvo, za delo prevedeno na hrvatski jezik. Preuzeto sa: https://www.intestcom.org/files/statement_ using_ tests_for_research_croatian.pdf

Plomin, R., DeFries, J. C., Craig, I. W., & McGuffin, P. (Eds.). (2003). *Behavioral genetics in the postgenomic era*. Washington, DC, USA: American Psychological Association.

Rattani, S. A. (2016). SAT: Does racial bias exist? *Creative Education, 7*(15), 2151–2162.

Sieber, J. E., O'Neil, H. F., & Tobias, S. (1977). *Anxiety, learning, and instruction.* Oxford, England: Lawrence Erlbaum Associates.

Simner, M. L. (1994). *Draft of final report of the Professional Affairs Committee*

Working Group on Test Publishing Industry Safeguards. Ottawa, Ontario, Canada: Canadian Psychological Association.

Turner, S. M., DeMers, S. T., Fox, H. R., & Reed, G. M. (2001). APA's guidelines for test user qualifications: An executive summary. *American Psychologist, 56*(12), 1099–1113.

Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons Inc.

Żemojtel-Piotrowska, M., Piotrowski, J., Rogoza, R., Baran, T., Hitokoto, H., & Maltby, J. (2018). Cross-cultural invariance of NPI-13: Entitlement as culturally specific, leadership and grandiosity as culturally universal. *International Journal of Psychology*. Advance online publication. https://doi.org/10.1002/ijop.12487

Ajtem-analiza

Preduslov

Osnovni preduslov za sprovođenje ajtem-analize je homogenost testa. Ukoliko se neki test sastoji od nekoliko skala, i ukoliko se one zasebno skoruju, onda ajtem-analizu treba raditi na svakoj skali posebno. Na primer, inventari ličnosti se obično sastoje od nekoliko skala, te bi za svaku skalu trebalo raditi posebnu ajtem-analizu. S druge strane, neki test se može sastojati od suptestova ili supskala, pa test možemo skorovati tako da dobijemo pojedinačne skorove za svaku supskalu, ali i skor na celom testu. Na primer, skala Neuroticizam iz inventara Velikih pet plus dva (VP+2: Smederevac, Mitrović i Čolović, 2010) sastoji se od tri supskale – Anksioznost, Depresivnost i Negativni afekat. Tako, možemo dobiti zasebne skorove za ove tri supskale, ali možemo dobiti i ukupni skor na skali Neuroticizam. U ovom slučaju, ajtem-analizu možemo raditi za svaku supskalu posebno, u cilju evaluacije pojedinačnih supskala, ali i na svim ajtemima skale Neuroticizma zajedno, u cilju evaluacije te skale u celini.

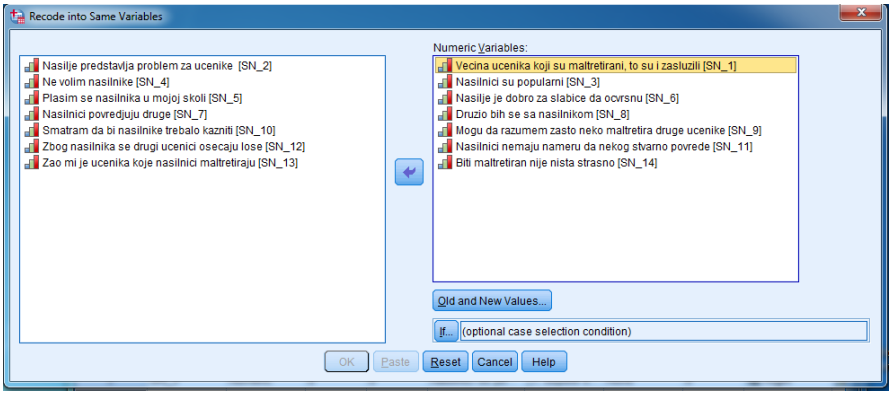
Matrica za vežbu

Matrica se nalazi na https://osf.io/n5hx6/ (Bullying Attitudinal Scale: Swearer & Carry, 2003). Viši skor treba da odražava pozitivan stav prema nasilju.

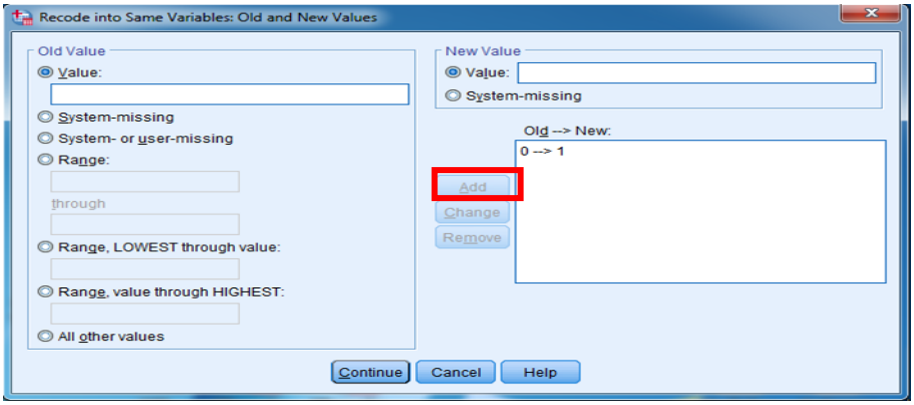
Rekodiranje

Pre ajtem-analize obavezno je obrnuto formulisane ajteme rekodirati. U suprotnom, dobiće se negativne diskriminativnosti greškom, i dobijeni pokazatelji neće biti tačni. Da bi se proverilo da li su ajtemi dobro rekodirani, treba proveriti koeficijente diskriminativnosti (Corrected Item-Total Correlation u okviru tabele Item-Total Statistics). Ukoliko postoji ajtem s negativnom diskriminativnošću, najverovatnije taj ajtem nije dobro rekodiran. U suprotnom, ukoliko ajtem jeste dobro kodiran, ali i dalje ima negativnu diskriminativnost, to znači da taj ajtem meri nešto sasvim drugo što nije predmet merenja testa, te treba razmotriti razlog negativne diskriminativnosti i ako se pokaže da je zaista ajtem loš, izbaciti ga.

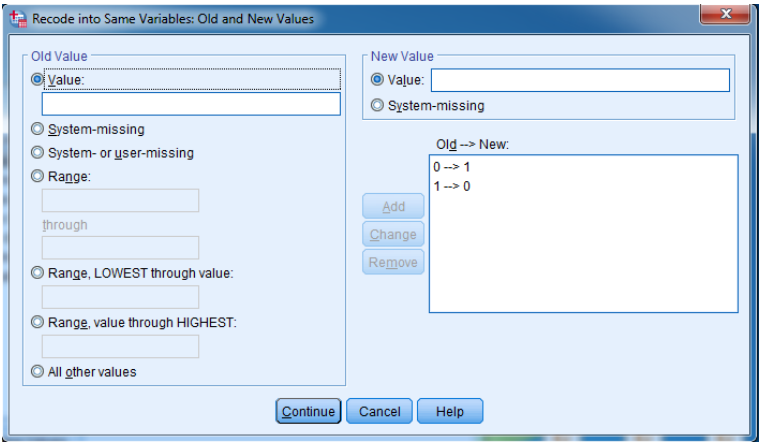
Rekodiranje ajtema može se vršiti u okviru date varijable, ili se može napraviti nova varijabla koja bi sadržala rekodirane vrednosti, a originalne, nerekodirane varijable bi bile sačuvane u matrici. U ovom primeru će se objasniti prva varijanta. Najpre, iz padajućeg menija Tranform treba odabrati Recode into Same Variables. Potom, iz levog prozora treba odabrati varijable koje treba rekodirati i prebaciti ih u prozor Numeric Variables, kao na sledećoj slici:



Ukoliko je reč o binarnim varijablama, onda 0 želimo da rekodiramo u 1, ali moramo ujedno uraditi i inverzno rekodiranje 1 u 0. To radimo tako što najpre kliknemo na dugme Old and New Values... i u dobijenom novom prozoru, u polje Old Value, ukucamo vrednost iz koje rekodiramo, a u polje New Value ukucamo vrednost u koju želimo da je rekodiramo (npr. ako 1 treba da se rekodira u 0, onda u prvi prozor – Old Value treba ukucati 1, a u drugi – New Value vrednost 0). Nakon ovoga treba kliknuti dugme Add i u prozoru Old->New bi trebalo da se pojavi 1->0, kao na slici:



Potom, treba ubaciti i obrnutu varijantu, tj. 0 treba rekodirati u 1. To znači da treba u polje Old Value uneti vrednost 0, a u New Value vrednost 1, pa kliknuti Add i u prozoru Old->New bi trebalo da se pojavi i 0->1. Dakle, mora se rekodirati i 1 u 0 i 0 u 1:



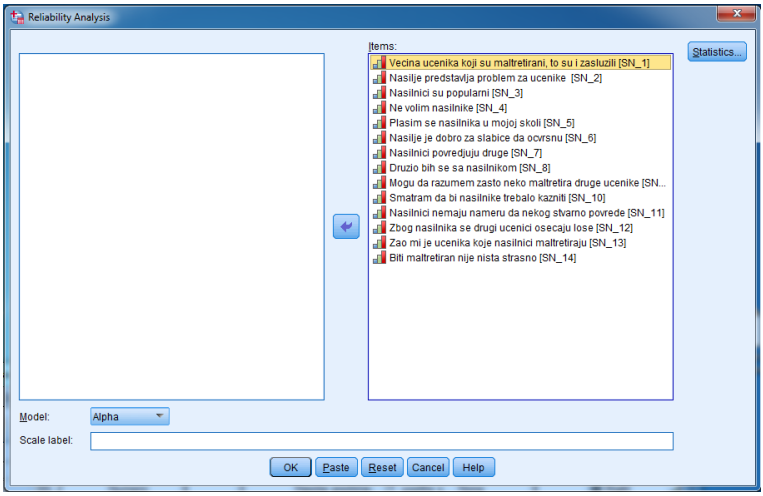
Na kraju treba kliknuti Continue, pa OK i u okviru istih varijabli će vrednosti u selektovanim ajtemima biti rekodirane.

U slučaju skala procene, npr. petostepene skale Likertovog tipa, treba izvršiti sledeće rekodiranje: 1->5, 2->4, 4->2, 5->1 (kategorija odgovora 3 ostaje 3, tako da se ona ne menja). U slučaju četvostepene skale Liketovog tipa, rekodiranje bi trebalo da bude: 1->4, 2->3, 3->2, 4->1.

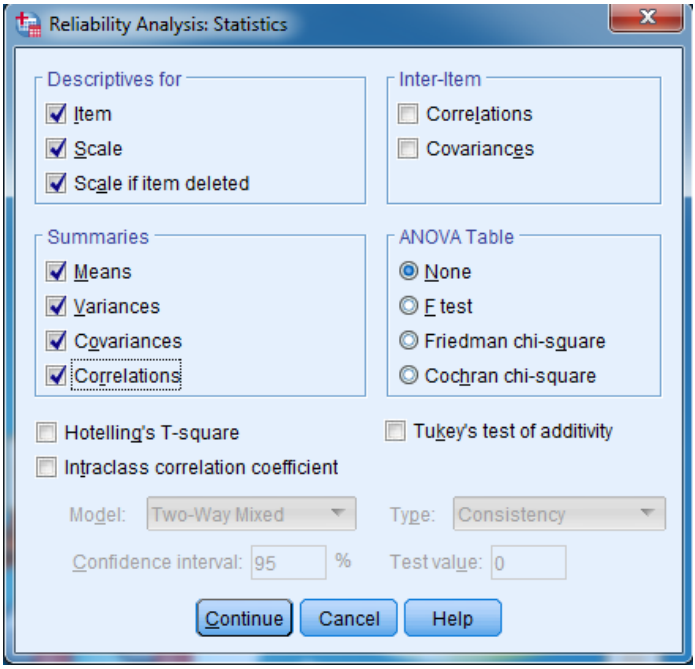
U našem primeru, ajteme koje treba rekodirati su 1, 3, 6, 8, 9, 11 i 14. Nakon rekodiranja, svi ajtemi su usmereni tako da viši skor ukazuje na pozitivniji stav prema nasilju.

Koje opcije treba odabrati za ajtem-analizu?

Iz padajućeg menija odabrati Analyze/Scale/Reliability Analysis... Potom, prebaciti sve ajteme za koje želite da radite ajtem-analzu u polje za varijable, kao na slici:



Potom, kliknuti na dugme Statistics i čekirati sledeće opcije:



Nakon toga, kliknuti Continue, pa OK.

Tumačenje autputa iz SPSS-a

Pouzdanost interne konzistencije - Kronbahov alfa koeficijent (α)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,825	,834	14

Pouzdanost interne konzistencije

Pouzdanost i validnost su osnovne metrijske karakteristike testa. Uopšteno, pouzdanost predstavlja odsustvo greške merenja. U Standardima za pedagoško i psihološko testiranje, pouzdanost se definiše kao „stepen u kojem su testovni rezultati za grupu ispitanika dosledni u ponovljenim primenama mernog postupka i time verodostojni i ponovoljivi za pojedinog ispitanika, tj. stepen u kojem su rezultati nezavisni od delovanja greške merenja za datu grupu" (AERA, APA, & NCME, 2006, str. 268). Postoji nekoliko načina određivanja pouzdanosti, a u ovom autputu dat je Kronbahov alfa koeficijent (α) koji predstavlja pouzdanost interne konzistencije ili unutrašnje saglasnosti testa, odn. stepen srodnosti ajtema od kojih se test sastoji. Preporuke za tumačenje ovog koeficijenta su da on ne bi trebalo da bude niži od ,70 (DeVellis, 2003), dok se koeficijent ,80 i više smatra dobrom pouzdanošću, a ,90 i više odličnom (George & Mallery, 2003).

Alfa koeficijent zavisi od nekoliko činilaca, a jedan od njih je broj ajtema u testu. Naime, alfa koeficijent ima tendenciju da bude viši što je broj ajtema u testu veći. Tako, ukoliko neki test od 5 ajtema ima alfu ,70,

smatraćemo da je ta vrednost adekvatna. Međutim, ukoliko neki test od 30 ajtema ima alfu ,70, iako je to prihvatljiva vrednost alfe, nije i dovoljno dobra, tj. mogla bi biti viša. Iz tog razloga se prilikom tumačenja alfe uzima u obzir i broj ajtema na koji se ona odnosi i taj broj ajtema se prikazuje uz alfu.

Pouzdanost nekog testa uvek je pouzdanost konkretnih podataka tj. skorova. Prema savremenim preporukama, prilikom tumačenja pouzdanosti uvek treba jasno istaći da se rezultati odnose na pouzdanost skorova (na konkretnom uzorku ispitanika), a ne testa (Appelbaum et al., 2018). Na ovaj način se ističe to da metrijske karakteristike nisu karakteristike testa koje su nezavisne od uzorka ili interpretacije skorova, tj. da se dobijene karakteristike na jednom uzorku u jednoj prilici ne mogu preslikati na buduće i ostale primene. U našem primeru, tumačenje bi glasilo: Pouzdanost skorova iznosi α = ,825. Ili, ako želimo detaljnije: Za dati broj ajtema u skali (14), pouzdanost skorova je dobra i iznosi α = ,825.

Ajtem statistici - osnovni deksriptivni pokazatelji

Item Statistics

	Mean	Std. Deviation	N
SN_1 Vecina ucenika koji su maltretirani, to su i zasluzili	2,04	1,058	599
SN_2 Nasilje predstavlja problem za ucenike	1,82	1,120	599
SN_3 Nasilnici su popularni	2,94	1,323	599
SN_4 Ne volim nasilnike	1,66	1,148	599
SN_5 Plasim se nasilnika u mojoj skoli	3,92	1,241	599
SN_6 Nasilje je dobro za slabice da ocvrsnu	2,03	1,280	599
SN_7 Nasilnici povredjuju druge	1,71	1,047	599
SN_8 Druzio bih se sa nasilnikom	2,06	1,129	599
SN_9 Mogu da razumem zasto neko maltretira druge ucenike	2,22	1,257	599
SN_10 Smatram da bi nasilnike trebalo kazniti	1,68	,972	599
SN_11 Nasilnici nemaju nameru da nekog stvarno povrede	2,26	1,172	599
SN_12 Zbog nasilnika se drugi ucenici osecaju lose	1,95	1,037	599
SN_13 Zao mi je ucenika koje nasilnici maltretiraju	1,83	1,051	599
SN_14 Biti maltretiran nije nista strasno	1,64	,986	599

Težina ajtema

Osnovne metrijske karakteristike ajtema su težina i diskriminativnost. Aritmetička sredina ajtema (Mean – *M*) u psihometrijskom smislu predstavlja težinu ajtema. Kod binarnih stavki ("da/ne", "tačno/netačno"), težina se kreće u rasponu od 0 do 1 (ukoliko sve netačne odgovore ili odgovore "ne" kodiramo kao 0, a tačne ili odgovore "da" kao 1, što je uobičajeno). Kod skala procene, težina se kreće u rasponu koliko ima kategorija odgovora. Na primer, za petostepenu Likertovu skalu procene, težina bi se kretala u rasponu od 1 do 5.

Stavke možemo grubo interpretirati kao lake, prosečne težine ili teške u odnosu na teorijski prosek. U slučaju binarnih stavki, teorijski prosek je 0,5, te se stavka čija je težina oko nule može se protumačiti kao teška (ili čak, veoma teška), stavka čija je tećina oko teorijskog proseka tj. 0,5 može se protumačiti kao prosečne težine, a stavka čija je težina oko 1 kao laka (ili čak, veoma laka). U slučaju testova ličnosti (ili testova koji pripadaju drugoj paradigmi merenja), umesto termina "težina" primerenije je upotrebiti izraz "slaganje sa stavkom". U slučaju petostepene Likertove skale procene, teorijski prosek je 3, tako da stavku čija je "težina" oko 1 možemo protumačiti kao veoma tešku, odn. takvu da se ispitanici češće *nisu* slagali s tvrdnjom, potom, stavku čija je težina oko 3 kao prosečnu, i stavku čija je težina oko 5 možemo protumačiti kao laku, odn. takvu da su se ispitanici češče slagali sa tvrdnjom. Svakako, prilikom tumačenja treba uzeti u obzir nijanse određivanja težine, te se u tom slučaju stavke mogu okarakterisati i kao umereno lake ili umereno teške, a ne samo lake, prosečne ili teške.

Iako poređenje sa teorijskim prosekom može biti informativno, ipak treba obratiti pažnju na to kakve su težine svi ajtemi u testu (da li su generalno laki ili teški), i da li postoji neki ajtem (ili skup ajtema) koji se

upadljivo razlikuje od ostalih (upadljivo je lakši ili teži u odosu na ostale ajteme u testu).

Prema Klasičnoj testnoj teoriji (KTT) favorizuju se stavke koje imaju približno jednaku težinu. Drugim rečima, poželjno je ako je varijabilnost težina stavki mala. Pored toga, favorizuju se stavke koje imaju prosečnu težinu, ali i dovoljnu varijansu. Na primer, stavka s petostepenom skalom procene za odgovaranje koja ima prosečnu težinu 3, ali suženu varijansu (u smislu da je većina ispitanika zaokružila kategoriju odgovora 3), nije dobra stavka. Međutim, stavka koja ima prosečnu težinu 3, ali dovoljnu varijansu, u smislu npr. da je 20% ispitanika zaokružilo kategoriju 1, 20% kategoriju 2 i tako redom, primer je dobre stavke. S druge strane, ekstremno teške ili lake stavke obično imaju loše metrijske karakteristike i njih treba izbegavati u testu. Takve stavke obično imaju suženu varijansu, te samim tim i nižu korelaciju s ostalim stavkama u testu, što posledično dovodi i do nižih metrijskih karakteristika koje su bazirane na interkorelacijama ajtema (npr. diskriminativnost).

Postoje izuzeci kada u testu treba zadržati ekstremno lake ili ekstremno teške stavke. U testovima koji pripadaju prvoj paradigmi merenja tj. testovima sposobnosti postoji potreba da se zadrže ekstremno laki zadaci ukoliko su to npr. zadaci u testovima brzine na početku testa koje svi ispitanici, ili barem velika većina, uspešno reši, ili u testovima znanja ukoliko su to pitanja koja se odnose na neka kritična znanja. U testovima koji pripadaju drugoj paradigmi merenja tj. testovima ličnosti, to mogu biti tzv. marker-ajtemi koji jesu "teški", ali veoma diskriminativni (npr. ajtem "Bio sam uhapšen" u upitniku procene sklonosti ka antisocijalnim ponašanjima). Takođe, to mogu biti ajtemi koji pripadaju skalama validnosti za proveru davanja socijalno poželjnih odgovora ili za

proveru stila odgovaranja, poput slaganja. Takve stavke obično sadrže tvrdnju s kojom se ne bi složila velika većina ispitanika (npr. "Trebalo bi odbaciti sve zakone.").

U našem primeru, sve stavke su "teže" (samo jedna prelazi teorijsku prosečnu vrednost 3), tj. ispitanici su se manje slagali s tvrdnjama. Budući da viši skor odražava podržavajući stav prema nasilju, to je i očekivano, tj. većina ispitanika ima negativan stav prema nasilju. Stavke su uglavnom ujednačene po težini, tj. varijabilnost težina je mala. "Najteže" stavke su stavke br. 4 (M = 1,64), 14 (M = 1,66) i 10 (M = 1,68), ali u ovom kontekstu gde su sve stavke "teže", ove stavke se ne ističu po tome da su upadljivo "teže" u odnosu na ostale. "Najlakša" stavka je br. 5 (M = 3,92) i ona se time ističe u odnosu na ostale. Primetićete da iako je težina stavke 5 nešto iznad teorijskog proseka, u kontekstu težina ostalih stavki, ova stavka je upadljivo "lakša". Kao što je rečeno, sa stanovišta KTT poželjno je da stavke međusobno ne variraju previše po težini.

Standardna devijacija ajtema

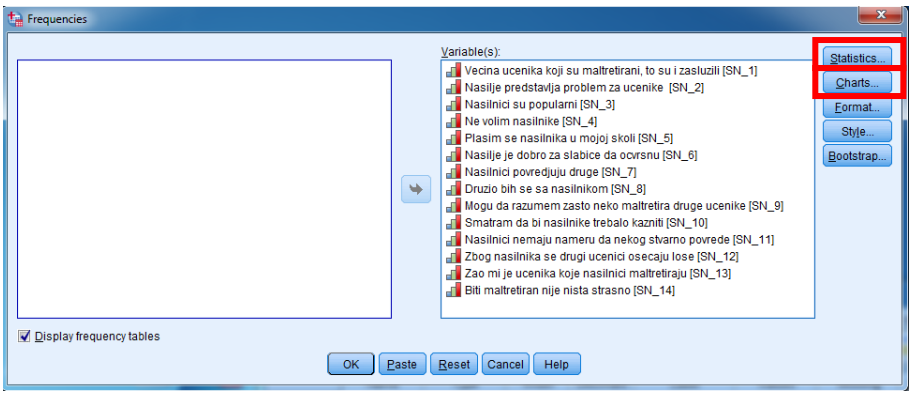
U autputu, pored aritmetičke sredine prikazan je i drugi pokazatelj distribucije skorova, a to je standardna devijacija (Std. Deviation – *SD*). Standardna devijacija nema neku interpretativnu vrednost prilikom selekcije stavki, ali može dati dodatni uvid u karakteristike stavki. Stavke koje su loše su one koje imaju malu varijansu tj. *SD*. Obično ekstremno laki ili ekstremno teški ajtemi imaju nisku varijansu, jer je kod njih manja varijabilnost u odgovaranju tj. većina ispitanika isto ili slično odgovara na ajtem (npr. ili većina odgovori tačno, ili se većina slaže sa tvrdnjom, pa zaokružuje kategoriju odgovora 4 ili 5 na skali procene). Ne postoje granične vrednosti za određivanje niske i visoke *SD*, već se o tome zaključuje na osnovu *SD* vrednosti svih stavki u

testu. U našem primeru, stavke koje imaju nižu *SD* su stavke br. 14 (*SD* = ,986) i 10 (*SD* = ,972) koje su već okarakterisane kao "teže" stavke.

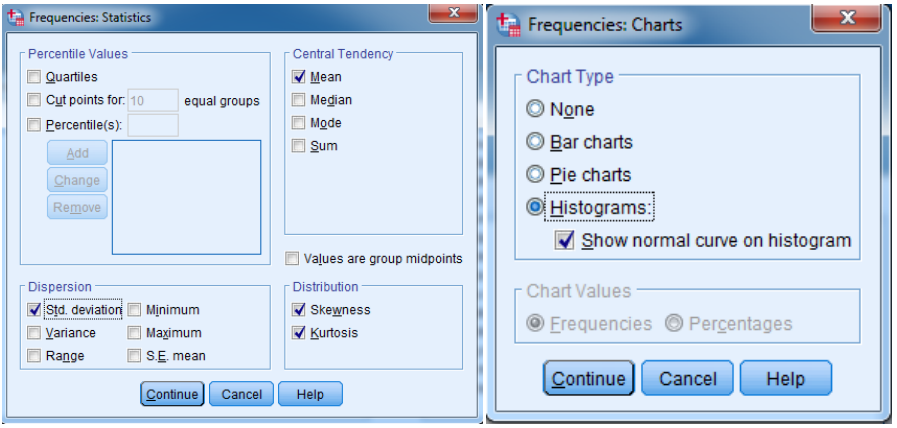
Ono što je još uočljivo na osnovu *SD* pokazatelja je da stavka br. 3 ima najvišu *SD* (*SD* = 1,323). Iako visina *SD* za ovu stavku nije upadljivo različita od većine ostalih stavki, svakako ćemo obratiti pažnju na ovu stavku u pogledu drugih metrijskih karakteristika. Kao što je rečeno, poželjno je da stavka ima dovoljnu varijansu, ali ne i preveliku.

Oblik distribucije u kontekstu određenja težine ajtema

Oblik distribucije ajtema presudno utiče na metrijske karakteristike. Ajtemi koji imaju sličniju distribuciju međusobno više koreliraju. Samim tim, bolje karakteristike će imati ajtemi koji imaju sličniju distribuciju u setu svih ajtema testa, nego ajtemi koji se razlikuju po distribuciji od ostalih. S obzirom na to, korisno može biti pogledati distribuciju pojedinačnih ajtema. Iako u autputu imamo *M* i *SD*, različitim distribucijama se može doći do istih ovih vrednosti, te zato treba pogledati i skjunis – *Sk* (asimetričnost ili zakošenost) i kurtozis – *Ku* (spoljštenost ili izduženost) ajtema, kao i samu distribuciju ajtemskih skorova. Uvid u oblik distribucije ajtema možemo dobiti tako što odaberemo Analyze/Descriptive Statistics/Frequencies... pa u polje Variable(s) prebacimo ajteme, kao na slici:



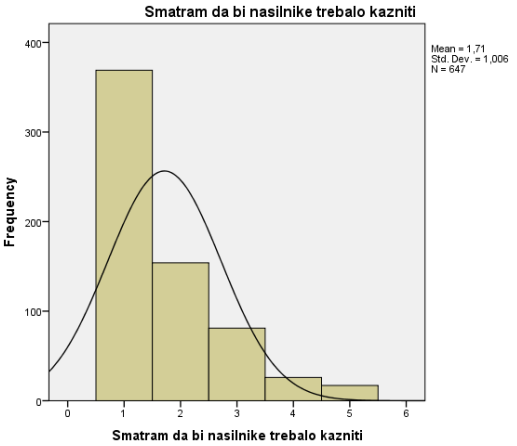
Potom, u polju Statistics... čekiramo Mean, Std. deviation, Skewness i Kurtosis, a u polju Charts čekiramo Histograms i Show normal curve on histogram, pa Continue i OK:



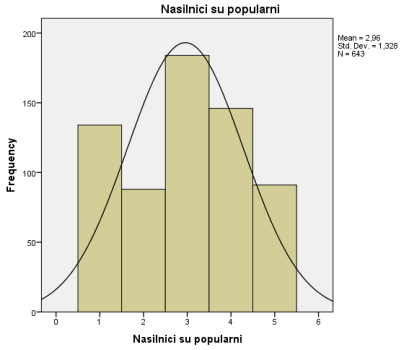
U autputu se dobijaju vrednosti pokazatelja oblika distribucije. Postoje različite preporuke u vezi s tim koliki mogu biti skjunis i kurtozis da bi se distribucija i dalje smatrala normalnom, i u ovom priručniku će se usvojiti fleksibilniji kriterijum prema kojem bi skjunis i kurtozis trebali biti u rasponu od –2 do +2 da bi se distribucija smatrala normalnom (Gravetter & Wallnau, 2014). Međutim, nas ovde više interesuje da li se oblik distribucije nekog ajtema razlikuje u odnosu na ostale ajteme, a ne toliko da li je distribucija normalna.

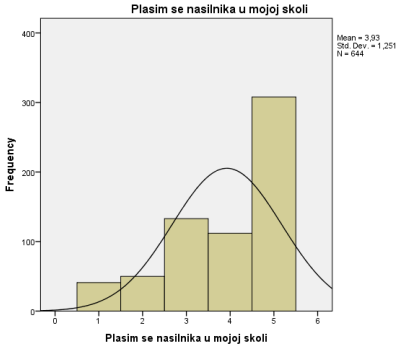
Uvidom u skjunis i kurtozis možemo primetiti da su oni nešto viši u slučaju ajtema br. 4 i 14, ali da su oko 2 (za ajtem br. 4 *Sk* = 1,61, *Ku* = 1,48, a za ajtem br. 14 *Sk* = 1,63, *Ku* = 2,13). Podsetićemo, ajtemi br. 4 i 14 su okarakterisani kao "teži", te nije iznenađujuće što im je distribucija zakošena (i to pozitivno) i leptokurtična tj. sužena. Ono što je primetno je da su oblici distribucije ajtema br. 3 i 5 drugačiji u odnosu na ostale ajteme. Naime, većina ajtema ima uglavnom pozitivno

zakošenu i leptokurtičnu distribuciju (voditi računa o tome da je u pitanju rekodirani ajtem, tj. da na originalnom ajtemu većina ispitanika bira odgovor 5 = *uopšte se ne slažem*), kao na slici:



Međutim, ajtemi br. 3 i 5 imaju negativan skjunis tj. negativno su zakošeni:





Iako je kurtozis, takođe, negativan za ajteme br. 3 i 5, on je negativan kod još nekih ajtema, te to nije osnovna razlika između ovih ajtema i ostalih, već upravo njihov skjunis.

Sumirane vrednosti za ajtem statistike

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	2,126	1,641	3,925	2,284	2,392	,384	14
Item Variances	1,289	,945	1,749	,805	1,852	,066	14
Inter-Item Covariances	,324	-,131	,628	,758	-4,803	,029	14
Inter-Item Correlations	,264	-,088	,473	,562	-5,369	,020	14

Težina testa

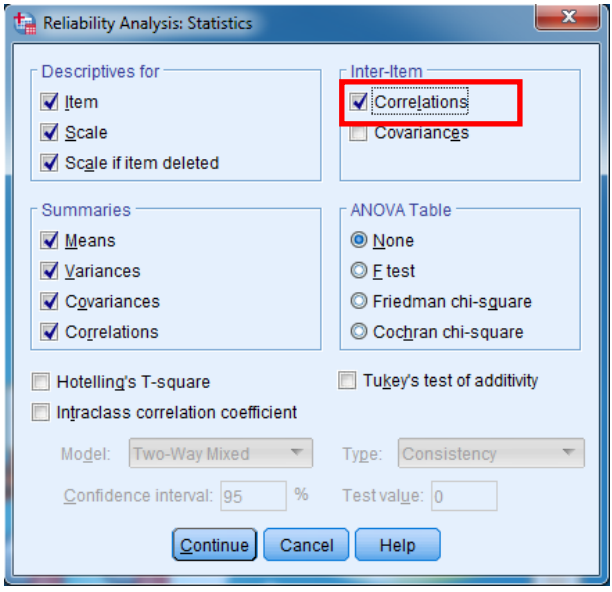
Težina testa može se izraziti na dva načina. Prvi način je prosečna težina ajtema u testu (Item Means u koloni Mean), koja se dobija uprosečavanjem svih artmetičkih sredina ajtema u testu iz kolone Mean u tabeli Item Statistics. Uvidom u pojedinačne težine ajtema zaključili smo da su stavke ovog upitnika generalno "teške", što se odražava i na prosečnu "težinu" svih stavki u upitniku koja iznosi 2,126. Prosečna težina svih stavki u upitniku tumači se kao i težina pojedinačnih stavki, te bismo ovaj upitnik okarakterisali kao umereno "težak".

Drugi način ordeđivanja težine testa je preko proseka sumacionih skorova na uzorku ispitanika na testu, i biće objašnjen kasnije, uz tabelu Scale Statistics.

Homogenost

Iz tabele Summary Item Statistic nam treba još jedan pozatelj, a to je pokazatelj homogenosti. Homogenost predstavlja stepen u kojem

svi ajtemi testa mere istu osobinu. Postoji nekoliko načina određivanja homogenosti, a u ovom autputu data je prosečna interajtemska korelacija (Mean Inter-Item Correlations - MIC, poznat i kao h_1 koeficijent). Drugim rečima, to je prosek međusobnih korelacija svih ajtema u testu. Uvid u pojedinačne korelacije ajtema može se dobiti ako se u polju Inter-Item čekira Correlations:



Preporuka za ovaj koeficijent je da on treba biti u rasponu od ,20 do ,50 da bi se test smatrao homogenim (Clark & Watson, 1995). Ukoliko je ovaj koeficijent ispod ,20 – test nije homogen tj. nema jedan predmet merenja. Valja napomenuti da se u slučaju mera širih konstrukata, koji obuhvataju skup različitih indikatora (npr. dimenzije višeg reda poput Ekstraverzije i drugih bazičnih osobina ličnosti), može tolerisati niži MIC koeficijent (npr. ,15, ili čak, poželjnije je da MIC koeficijent bude u rasponu od ,15 do ,20, više u Clark & Watson, 1995). U slučaju mera užih

konstrukata, kao šti su npr. supskale Ekstraverzije – Pozitivan afekat, Aktivitet i slično, za očekivati je da imaju viši MIC koeficijent budući da im je i predmet merenja uži. Ipak, ukoliko je MIC koeficijent preko ,50 to znači da je test "previše" homogen tj. da stavke ili grupa stavki u testu suviše visoko koreliraju na način da su međusobno redundantne. U ovom slučaju najverovatnije imamo vrlo slično formulisane stavke ili tautologije, ili smo stavkama obuhvatili samo manji skup indikatora, pa na sličan način pitamo ispitanike o istom indikatoru (npr. stavke "Često sam nervozan" i "Češće se unervozim u odnosu na većinu drugih ljudi"). U praksi se često dešava da autori kratkih skala pribegavaju sličnim formulacijama ajtema, jer na taj način povećavaju korelaciju između ajtema, pa samim tim i alfu, ali je zapravo reč o veštački povećanoj alfi, te se ovakva praksa nikako ne preporučuje. Drugim rečima, ako imate kraću skalu (npr. od 4 ajtema) i visoku alfu (npr. ,90), to može biti signal da postoji previsoka korelacija između stavki, te treba pogledati koliko iznosi MIC koeficijent. U našem primeru koeficijent homogenosti iznosi ,264 i skalu bismo okarakterisali kao zadovoljavajuće homogenu.

Valja napomenuti da je MIC koeficijent dosta kritikovan kao pokazatelj homogenosti, budući da *nije* osetljiv na to koji deo varijanse ajtema kovarira sa ostalim ajtemima, da li prava varijansa ili varijansa greške. Tako, moguće je da neki skup ajtema visoko međusobno korelira jer deli zajedničku grešku (npr. svi ajtemi koji počinju formulacijom "Kad sam... "), te je MIC koeficijent u tom slučaju veštački visok. Iz tog razloga o homogenosti se zaključuje najčešće preko faktorske analize i to na više načina:

1. uvidom u broj faktora koji se preporučuje kao optimalan u odnosu na različite kriterijume za određenje broja faktora (npr. Scree dijagram, paralelna analiza i sl., više u Subotić, 2013);

2. uvidom u procenat varijanse koju objašnjava prvi faktor u inicijalnoj soluciji, a koji treba da bude minimum 20% (Reckase, 1979);
3. odnosom između procenta objašnjenje varijanse koji objašnjava prvi i drugi izolovani faktor, koji treba da bude minimum 4:1 (Reeve et al., 2007).

Ajtem-total statistici

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
SN_1 Vecina ucenika koji su maltretirani, to su i zasluzili	27,72	67,261	,497	,293	,811
SN_2 Nasilje predstavlja problem za ucenike	27,94	66,916	,482	,331	,811
SN_3 Nasilnici su popularni	26,82	74,053	,053	,057	,845
SN_4 Ne volim nasilnike	28,10	65,495	,548	,371	,807
SN_5 Plasim se nasilnika u mojoj skoli	25,83	71,653	,181	,065	,834
SN_6 Nasilje je dobro za slabice da ocvrsnu	27,73	63,879	,561	,363	,805

SN_7 Nasilnici povredjuju druge	28,05	67,979	,459	,288	,813
SN_8 Druzio bih se sa nasilnikom	27,70	65,134	,581	,368	,804
SN_9 Mogu da razumem zasto neko maltretira druge ucenike	27,54	65,510	,487	,284	,811
SN_10 Smatram da bi nasilnike trebalo kazniti	28,08	65,804	,650	,481	,802
SN_11 Nasilnici nemaju nameru da nekog stvarno povrede	27,50	66,314	,488	,287	,811
SN_12 Zbog nasilnika se drugi ucenici osecaju lose	27,81	68,886	,409	,292	,816
SN_13 Zao mi je ucenika koje nasilnici maltretiraju	27,93	66,356	,557	,374	,807
SN_14 Biti maltretiran nije nista strasno	28,12	66,834	,570	,365	,807

Diskriminativnost ajtema

Diskriminativnost (ili osetljivost) ajtema predstavlja karakteristiku ajtema da razlikuje ispitanike s niskom crtom od ispitanika s visokom crtom. Ona se računa kao korigovana korelacija između ajtema i ukupnog (total) skora, što se uobičajeno naziva korigovana ajtem-total korelacija (Corrected Item-Total Correlation) i predstavlja korelaciju ajtema sa ukupnim ili total skorom. Drugim rečima, ona predstavlja stepen u kojem su razlike u ajtemskih skorovima između ispitanika podudarne s razlikama u njihovim ukupnim skorovima. Međutim, ovde je prikazana korigovana korelacija, a korekcija se sastoji u tome što je iz ukupnog skora izbačena stavka za koju se računa diskriminativnost. Kada ove korekcije ne bi bilo, korelacije bi bile više jer imamo situaciju da stavka korelira sa ukupnim skorom u kojem se već i ona sama nalazi. Drugim rečima, veštački bi korelacija bila viša, te se otud tumači samo korigovana korelacija. Ovo je posebno uočljivo kod kraćih testova gde je učešće svakog pojedinačnog ajtema u ukupnom skoru veliko.

Preporučene vrednosti za diskriminativnosti ajtema su od ,30 do ,80. Ukoliko je diskriminativnost nekog ajtema ispod ,30 to znači da ajtem nije dovoljno diskriminativan, te ga treba izbaciti. Ukoliko je diskriminativnost nekog ajtema preko ,80 to znači da ajtem suviše visoko korelira sa ostalim ajtemima u testu, te ne doprinosti informativnosti testa. Visoka diskriminativnost je indikator toga da najverovatnije ima redundantnih ajtema. Na osnovu matrice korelacija između ajtema treba detektovati s kojim sve ajtemima taj ajtem s diskriminativnošću preko ,80 ostvaruje visoku korelaciju. Potom, treba videti koji od tih ajtema ima gore metrijske karakteristike, i eliminisati ajtem s gorim metrijskim karakteristikama. Kao što je pomenuto, uvid u

pojedinačne korelacije ajtema može se dobiti ako se u polju Inter-Item čekira opcija Correlations.

U našem primeru dve stavke imaju diskriminativnost nižu od preporučene, stavka br. 3 (,053) i br. 5 (,181). Diskriminativnosti ostalih stavki se kreću u rasponu od ,409 do ,650, što je odražava dobre diskriminativnosti ajtema.

Treba napomenuti da se ove preporuke o poželjnim vrednostima diskriminativnosti shvate fleksibilno. Naime, ukoliko je test takav da retko koji ajtem ima višu diskriminativnost (npr. preko ,40 ili čak preko ,30), onda je u redu spustiti donju granicu prihvatljivosti i kao diskriminativne prihvatiti ajteme s diskriminativnošću ,20 i više, ali ne niže od toga. Niske diskriminativnosti su posebno čest slučaj kod testova znanja koji obuhvataju gradivo iz različitih oblasti, te testova opšte informisanosti, kao i testova inteligencije generalno, a ređe se mogu naći kod konativnih testova. S druge strane, moguće je da neki ajtem, ili par ajtema, ima upadljivo visoku diskriminativnost, koja iako ne prelazi ,80 ostvaruje vrednost oko ,80 (npr. oko ,70). Za takav ajtem, s upadljivo visokom diskriminativnošću u odnosu na ostale, bez obzira na to što je ta diskriminativnost ispod ,80, treba proveriti kolike korelacije ostvaruje s drugim ajtemima u testu, i da li je u pitanju redundantan ajtem.

Već je spomenuto kako su težina i standardna devijacija povezane, a na ovom mestu će samo biti istaknuto to kako su težina i diskriminatvinost povezane. Naime, ajtemi koji imaju ekstremnu težinu u kontekstu svih ajtema u testu (ekstremno laki ili ekstremno teški ajtemi), imaće suženu varijansu, pa samim tim i nižu korelaciju s ostalim ajtemima. To se odražava i na njihovu nižu diskriminativnost, pošto će i korelacija sa ukupnim skorom takvog ajtema biti niža. Kako sličnost

oblika distribucije ajtema utiče na visinu korelacije između njih, ajtemi koji imaju drugačiju distribuciju od većine ostalih će imati niže korelacije sa ostalim ajtemima u testu, pa samim tim i nižu diskriminativnost. Ajtemi s drugačijom distribucijom su obično ti ajtemi ekstremne težine. Na ovom mestu valja napomenuti da ekstremnost težine nekog ajtema uvek treba sagledati u kontekstu težina svih ajtema u testu. U našem primeru, gde su svi ajtemi "teži", ajtem koji se po svojoj težini upadljivo razlikuje od ostralih je ajtem br. 3 koji iako ima težinu oko teorijskog proseka, u ovom kontekstu je prelak, pa ima drugačiju distribuciju od ostalih ajtema.

Koeficijent determinacije

Kvadrirana multipla korelacija (Squared Multiple Correlation - SMC) predstavlja koeficijent regresije, tačnije koeficijent determinacije (R^2) u kojem je kriterijum ajtem za koji se računa SMC, a prediktori su svi ostali ajtemi. Kada ovaj koeficijent pomnožimo sa 100, dobijamo procenat objašnjene varijanse koji taj ajtem deli sa ostalim ajtemima. Na primer, ajtem br. 1 ima SMC = ,293, što znači da taj ajtem deli oko 29% varijanse sa ostalim ajtemima. Što je viša vrednost ovog koeficijenta, to znači da dati ajtem meri iste ili slične karakteristike kao i drugi ajtemi.

SMC koeficijenti nisu ništa drugo do inicijalni komunaliteti u okviru faktorske analize u užem smislu (komunaliteti se obeležavaju sa h^2). Radi provere, ukoliko nad istim ajtemima Upitnika stavova prema nasilju primenimo neki od metoda faktorske u užem smislu, npr. Maximum Likelihood - ML, dobićemo sledeći ispis komunaliteta:

Communalities

	Initial	Extraction
SN_1 Vecina ucenika koji su maltretirani, to su i zasluzili	,293	,271
SN_2 Nasilje predstavlja problem za ucenike	,331	,313
SN_3 Nasilnici su popularni	,057	,003
SN_4 Ne volim nasilnike	,371	,376
SN_5 Plasim se nasilnika u mojoj skoli	,065	,036
SN_6 Nasilje je dobro za slabice da ocvrsnu	,363	,366
SN_7 Nasilnici povredjuju druge	,288	,290
SN_8 Druzio bih se sa nasilnikom	,368	,369
SN_9 Mogu da razumem zasto neko maltretira druge ucenike	,284	,270
SN_10 Smatram da bi nasilnike trebalo kazniti	,481	,544
SN_11 Nasilnici nemaju nameru da nekog stvarno povrede	,287	,302
SN_12 Zbog nasilnika se drugi ucenici osecaju lose	,292	,226
SN_13 Zao mi je ucenika koje nasilnici maltretiraju	,374	,368
SN_14 Biti maltretiran nije nista strasno	,365	,408

Extraction Method: Maximum Likelihood.

Kao što se može primetiti, inicijalni komunaliteti su isti kao i SMC koeficijenti. Podsetićemo, komunalitet ajtema odnosi se na proporciju varijanse tog ajtema koja je objašnjena svim zajedničkim faktorima, tj. proporciju tog ajtema u okviru zajedničke varijanse. Komunalitet se izračunava kao suma kvadriranih faktorskih opterećenja datog ajtema na svim zajedničkim faktorima. Kako je ovde u pitanju inicijalni komunalitet, on se odnosi na varijansu ajtema obuhvaćenu *svim* faktorima, dok se komunalitet nakon ekstrakcije odnosi na varijansu ajtema obuhvaćenu *zadržanim* faktorima. Drugim rečima, inicijalni komunalitet se ne menja s menjanjem broja zadržanih faktora, dok se komunalitet nakon ekstrakcije menja s menjanjem broja zadržanih faktora. Tako, u datom primeru ajtem br. 1 ima inicijalni komunalitet ,293, a komunalitet nakon ekstrakcije (u ovom slučaju, kada je zadržan jedan faktor) ,271. Ukoliko bismo na istom setu varijabli zadržali dva faktora, komunalitet nakon ekstrakcije za ovaj ajtem bi bio ,333.

U literaturi se mogu naći različite preporuke u vezi s minimalnim prihvatljivim vrednostima komunaliteta. Neki autori ističu da u okviru primene faktorske analize u društvenim naukama komunaliteti obično budu niski do umereni, u rasponu od ,40 do ,70, što se smatra preporučenim vrednostima (Costello & Osborne, 2005). Drugi autori zastupaju fleksibilniji stav i kao minimalnu preporučenu vrednost ističu ,20 (Child, 2006). Ipak, savetuje se da se istraživači ne oslanjaju toliko na ove preporuke, jer one zavise od veličine uzorka i broja varijabli, već da imaju u vidu visine komunaliteta svih varijabli, te mapiraju one varijable s upadljivo nižim komunalitetima. Isti princip se može primeniti i pri tumačenju SMC koeficijenata, tj. treba sagledati SMC koeficijente za sve ajteme i uočiti da li neki ajtemi imaju upadljivo niži SMC koeficijent u

odnosu na ostale. Takvi ajtemi su nam kandidati za izbacivanje jer dele malo zajedničke varijanse sa ostalim ajtemima. U našem primeru, takvi su ajtemi br. 3 i 5.

Treba napomenuti da se nekada dešava da se SMC koeficijenti ne mogu izračunati, posebno kod binarnih stavki.

Promena alfe ukoliko se ajtem ukloni

U poslednjoj koloni tabele Item-Total Statistics (Cronbach's Alpha if Item Deleted) nalazi se podatak o tome kolika bi bila vrednost Kronbahove alfe ukoliko se dati ajtem ukloni iz testa. Pritom, treba voditi računa o tome da se taj pokazatelj odnosi na uklanjanje jednog, konkretnog ajtema, te nam ne govori o tome koliko bi se alfa promenila ako uklonimo dva ili više ajtema. U toj koloni treba da tražimo vrednost koja je veća od dobijene alfe s početka autputa. U našem slučaju α = ,825 i možemo primetiti da ukoliko se izbaci samo ajtem br. 3, alfa bi porasla na ,845. Ukoliko bismo izbacili samo ajtem br. 5, alfa bi porasla na ,834.

Ne postoje jasne preporuke do koje razlike u originalnoj i novoj alfi treba ići, ali je bitno da se izvuče maksimalna alfa koja je moguća, na način da se ne remete ostale važne karakteristike testa, pre svega validnost. U našem primeru, izbacivanjem ajtema br. 3 bismo povećali alfu, te taj ajtem svakako treba izbaciti. Primetićete, taj ajtem je bio loš i po drugim metrijskim karakteristikama (npr. niska diskriminativnost). U slučaju ajtema br. 5 situacija nije tako jednostavna, ukoliko se samo posmatra promena alfe. Naime, izbacivanjem ovog ajtema alfa bi i dalje ostala relativno ista kao originalna, tj. u domenu iste kategorije (,80 i više, što upućuje na dobru pouzdanost). Međutim, iako izbacivanjem ovog ajtema ne poboljšavamo pouzdanost, skraćivanje testa je takođe poželjno, te bismo svakako probali i rešenje kada izbacimo i ovaj ajtem.

Ono što doprinosi odluci da se ovaj ajtem izbaci su ostale loše metrijske karakteristike (npr. diskriminativnost), te je ovaj ajtem svakako još jedan kandidat za izbacivanje.

Kako je revizija testa iterativni postupak tj. radi se korak po korak, odnosno ajtem po ajtem, ukoliko bismo se oslonili na pokazatelj promene alfe, najpre bismo izbacili ajtem br. 3, pa proverili rešenje svih metrijskih karakteristika. Ukoliko bi se ajtem br. 5 i dalje ponašao isto, njega bismo izbacili u drugom koraku. Uzevši u obzir sve pokazatelje, a pre svega težinu i diskriminativnost ajtema, svakako su ajtemi br. 3 i 5 kandidati za izbacivanje, a nakon izbacivanja ovih ajtema treba proveriti da li se izdvajaju još neki ajtemi.

Mogući su i drugačiji scenariji. Na primer, moguće je da će izbacivanjem ajtema br. 5 doći do drugačijih metrijskih karakteristika, te da će se kao loši izdvojiti neki drugi ajtemi. Takođe, moguće je da u nekoj iteraciji pogoršamo karakteristike, pa moramo da vratimo izbačen ajtem. U svakom slučaju, treba proveriti karakteristike izbacivanjem jednog po jednog ajtema i pratiti promene. Kao finalno rešenje treba zadržati ono koje je najoptimalnije, po kojem su novodobijene metrijske karakteristike bolje od originalnih, ili barem nisu promenjene (tj. nisu narušene), ali je prednost novog rešenja u tome što je test skraćen.

Statistici za skalu – deskriptivni podaci za skalu

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
29,76	77,002	8,775	14

Težina testa

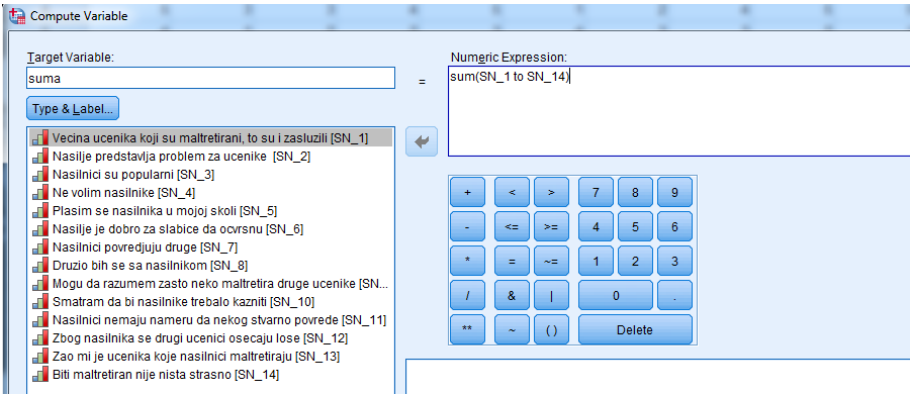
Kao što je rečeno, drugi način određivanja težine testa je preko proseka sumacionih skorova na testu (Mean u tabeli Scale Statistics). Ovaj prosek može se uporediti sa teorijskim prosekom. Teorijski prosek se dobija tako što se minimalni mogući skor na testu sabere sa maksimalnim mogućim skorom na testu, i to podeli sa 2 po formuli: (*Min* + *Max*)/2. U našem primeru imamo 14 ajtema s priloženom petostepenom skalom, dakle minimalni mogući skor je 14 (kada se na svaki ajtem zaokruži 1), a maksimalni skor je 14 * 5 = 70, tako da je teorijska težina testa (14 + 70) / 2 = 42. Empirijski prosek sumacionih skorova je 29,76. Kako je empirijski prosek manji od teorijskog, zaključujemo da je ovaj upitnik "težak", tj. da su se ispitanici generalno slabo slagali s tvrdnjama. Ukoliko bi test imao npr. binarne stavke koje su kodirane kao 0 = *ne* i 1 = *da*, i 14 stavki, teorijska aritmetička sredina bi bila (0 + 14) / 2 = 7.

Oblik distribucije u kontekstu određenja težine testa

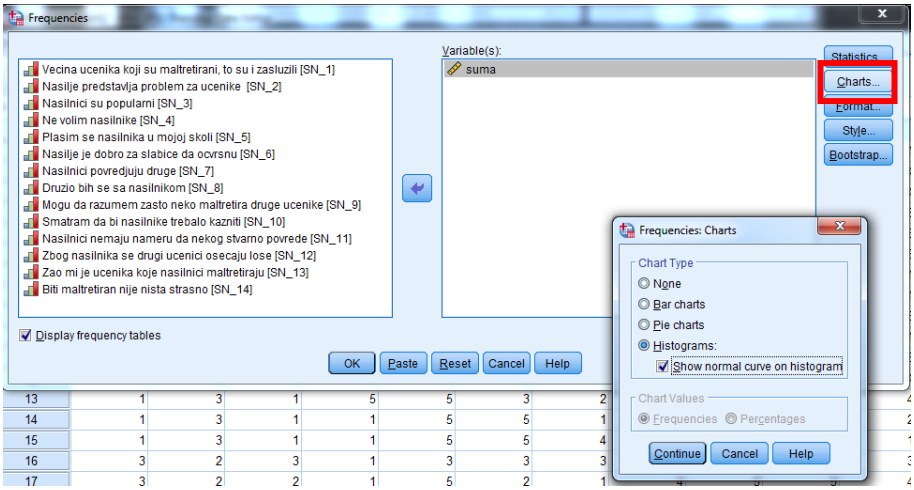
Bolji uvid u težinu testa ćemo dobiti ako pogledamo oblik distribucije skorova na testu. Možemo pogledati distribuciju sumacionih ili prosečnih skorova na testu, svejedno je. U oba slučaja najpre treba napraviti novu varijablu koja će sadržati ukupne skorove. U slučaju da želimo sumacione skorove, treba odabrati opciju Transform/Compute Variable... pa u polje Target Variable treba uneti naziv nove varijable, npr. suma (naziv ne sme da sadrži spejs i specijalne znakove poput , ! i slično). Potom, u polje Numeric Expression treba uneti izraz za izračunavanje sumacionog skora. Taj izraz se sastoj od komande sum(V1 to Vn), pri čemu umesto V1 i Vn treba da stoje nazivi prve i poslednje varijable za koju se računa sumacioni skor, onako kako su one

nazvane u matrici. U našem slučaju su to varijable SN_1 i SN_14, pa bi komanda glasila sum(SN_1 to SN_14). Valja napomenuti da je ovakvo računanje moguće samo ukoliko svi ajtemi u matrici za koje želimo da računamo skor idu redom, od V1 do Vn. Ukoliko su ajtemi u testu randomizirani po skalama, npr. da prvu skalu čine ajtemi br. 1, 4 i 7, a drugu ajtemi br. 2, 3 i 5, onda ne možemo na ovaj način računati sumacioni skor, već u Numeric Expression moramo zadati pun izraz, tj. SN_1 + SN_4 + SN_7.

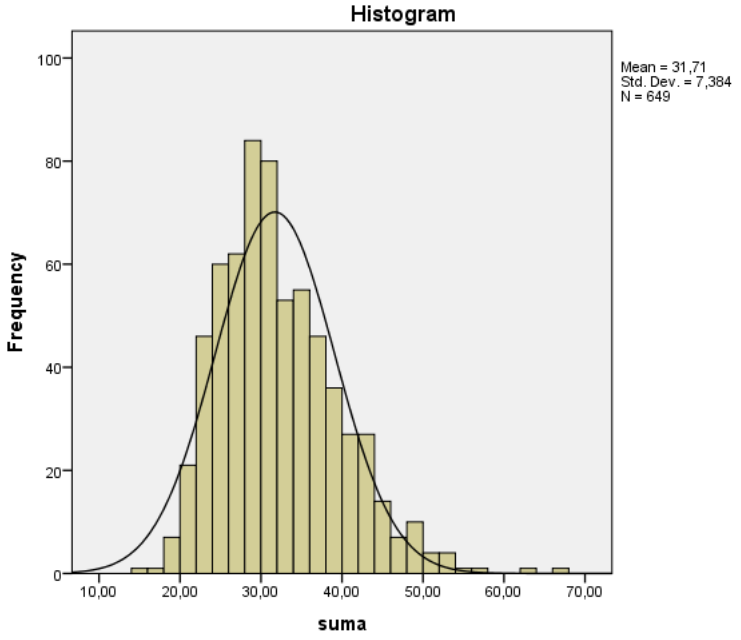
Ukoliko želimo da dobijemo varijablu s prosečnim skorovima, onda bismo primenili komandu mean(V1 to Vn).



Uvid u oblik distribucije možemo dobiti tako što odaberemo Analyze/Descriptive Statistics/Frequencies... pa u polje Variable(s) prebacimo novu varijablu suma, i u polju Charts čekiramo Histograms i Show normal curve on histogram, pa Continue i OK.



U autputu dobijamo sledeći oblik distribucije:



Iako su vrednosti skjunisa i kurtozisa u opsegu prihvatljivih za normalnu distribuciju, možemo primetiti da je distribucija asimetrična, i to pozitivno zakošena, ka nižim vrednostima (*Sk* = 0,83), a ujedno je i

leptokurtična tj. izdužena, šiljasta (*Ku* = 0,89). Na osnovu oblika distribucije zaključujemo da je test "teži" ispitanicima, tj. da su se ispitanici manje slagali s tvrdnjama. U kontekstu predmeta merenja, to znači da ispitanici generalno imaju negativniji stav prema nasilju, što se i očekuje.

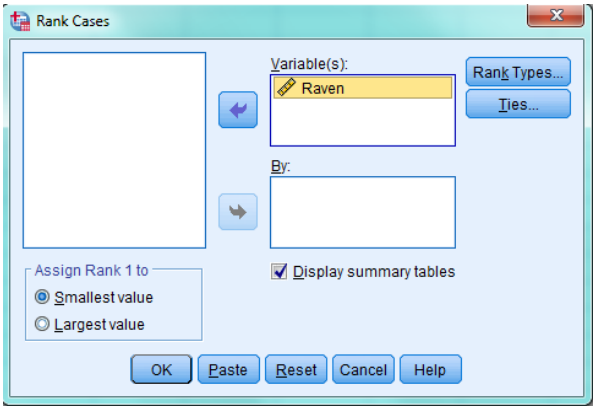
Primer izveštaja

Pouzdanost interne konzistencije skorova Skale stavova prema nasilju je dobra (α = ,83). Ispitanici se uglavnom nisu slagali sa tvrdnjama, te se raspon "težina" stavki kreće od 1,64 do 3,92 (na petostepenoj skali), pri čemu prosečna "težina" stavki iznosi 2,13. Ujedno, oblik distribucije sumacionih skorova je pozitivno zakošen, te je skala "teška" ispitanicima, odnosno ispitanici generalno imaju negativniji stav prema nasilju. Dve stavke imaju diskriminativnost ispod preporučene vrednosti ,30 (stavke br. 3 i 5) i ujedno niži SMC koeficijent u odnosu na ostale stavke i njihovim uklanjanjem bi se povećala pouzdanost skorova. S obzirom na to, preporuka je da se ove stavke izbace iz skale u budućim primenama. Diskriminativnosti ostalih ajtema su dobre i kreću se u rasponu od ,41 do ,65.

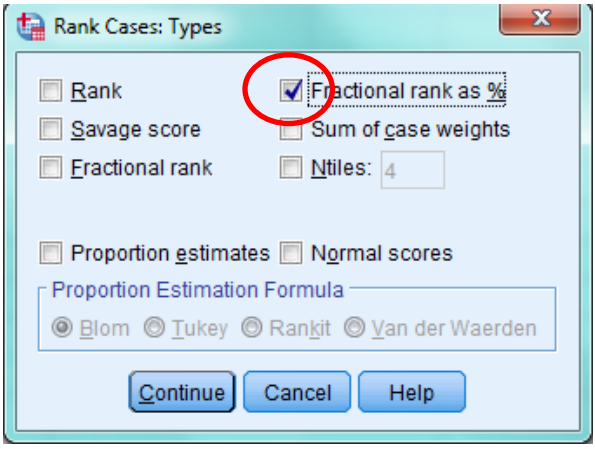
Računanje normi

Percentili

Iz padajućeg menija odabrati Transform/Rank Cases, te u polje za varijable prebaciti varijablu sa sirovim skorovima za test za koji želite da računate percentilne rangove:



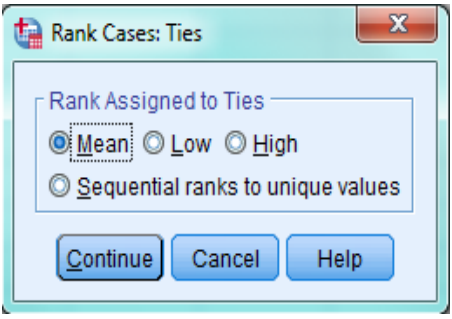
Odaberite opciju Rank Types sa desne strane i čekirajte Factional rank as %, kliknite Continue i OK:



Na kraju matrice će se formirati nova varijabla koja će u svom nazivu imati slovo P ispred imena varijable i ona sadrži percentilne rangove. Sada možete očitati kom sirovom skoru odgovara koji percentil.

Posebno se može odabrati način tretiranja združenih skorova, tako što ćete odaberati dugme Ties. Opcije koje se nude su da se svim

istim združenim skorovima dodeli prosečan percentilni rang (Mean), najmanji (Low), najviši (High) ili sekvencijalni, bez preskakanja dobijenih rangova (Sequential ranks to unique values):



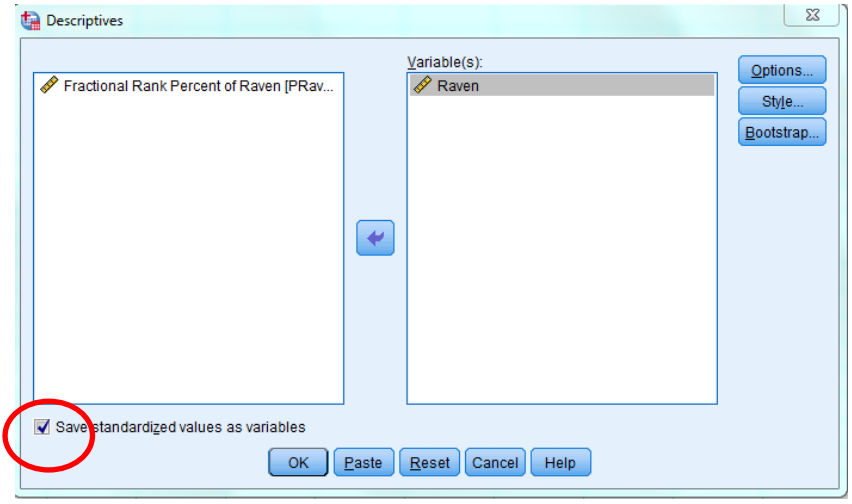
Najčešće se zadržava opcija koja združenim rangovima dodeljuje prosečan rang (Mean), ali ilustracije radi, možete uporediti percentilne rangove dobijene različitim tretmanima:

Raven	Mean	Low	High	Sequential
27,00	73,29	67,12	79,45	20,55
28,00	80,82	80,82	80,82	21,92
29,00	83,56	82,19	84,93	23,29
29,00	83,56	82,19	84,93	23,29
29,00	83,56	82,19	84,93	23,29
30,00	86,99	86,30	87,67	24,66
30,00	86,99	86,30	87,67	24,66
31,00	91,10	89,04	93,15	26,03
31,00	91,10	89,04	93,15	26,03
31,00	91,10	89,04	93,15	26,03
31,00	91,10	89,04	93,15	26,03
32,00	96,58	94,52	98,63	27,40
32,00	96,58	94,52	98,63	27,40
32,00	96,58	94,52	98,63	27,40
32,00	96,58	94,52	98,63	27,40
33,00	100,00	100,00	100,00	28,77

Kao što se može videti, za sirovi skor 27, najniži rang je 67,12, a najviši je 79,45, što znači da je prosečni rang (67,12 + 79,45) / 2 = 73,29. Sekvencijalni rang pošto izlistava redom rangove, bez preskakanja, od najnižih sirovih skorova nije još ni stigao do najnižeg ranga za skor 27 (67,12), već je najviši sekvencijalni rang koji je ostvaren na ovom uzorku tek 28,77.

Standardni skorovi

Iz padajućeg menija odabrati Analyze/Descriptive Statistics/Descriptives, te u polje za varijable prebaciti varijablu sa sirovim skorovima. Čekirajte opciju Save standardized values as variables koja se nalazi na dnu prozora za dijalog:



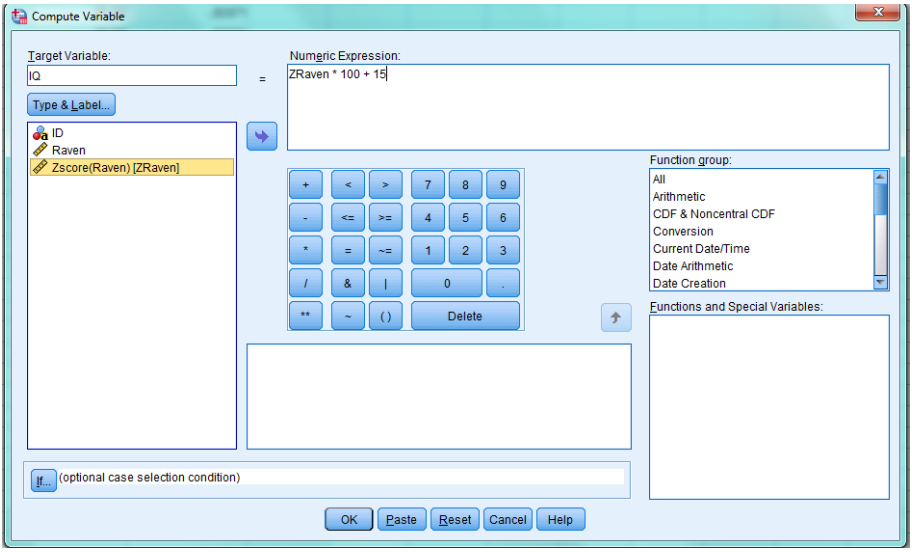
Na kraju matrice će se formirati nova varijabla koja će u svom nazivu imati slovo Z ispred imena varijable i ona sadrži standardne skorove.

Izvedeni standardni skorovi

Za izvedene standardne skorove vam treba varijabla sa standardnim skorovima, na kojoj potom primenjujete transformaciju. Iz padajućeg menija treba odabrati Transform/Compute Variable. U polje Target Variable treba uneti naziv nove varijable koja će sadržati izvedene standardne skorove, npr. IQ, a u polje Numeric Extression treba uneti formulu za izračunavanje izvedenog standardnog skora. U slučaju IQ skorova, ona bi glasila:

ZRaven * 100 + 15

pri čemu je ZRaven varijabla sa standardnim skorovima.

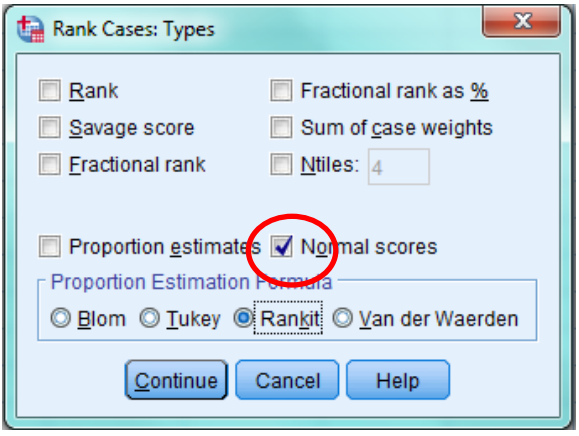


Na kraju matrice će se formirati nova varijabla IQ koja će sadržati IQ skorove. Istim postupkom se mogu izračunati i bilo koji drugi izvedeni standardni skorovi.

U okviru istog prozora za dijalog može se računati i ukupni *z*-skor – dovoljno je samo uneti izraz za izračunavanje sume nekoliko *z*-skorova. Ukoliko nam se varijable sa *z*-skorovima na suptestovima nazivaju, na primer, ZAL4, ZIT1, ZS1, onda bi izraz glasio: ZAL4 + ZIT1 + ZS1.

Normalizovani standardni skorovi

Prilikom računanja normalizovanih standardnih skorova, najpre treba normalizovati skorove. To se radi u okviru istog prozora za dijalog gde se računaju i percentilni rangovi, samo treba odabrati opciju Normal scores. Kada se navedena opcija odabere, aktiviraju se moguće transformacije, a ona koja se preporučuje je Rankit (Solomon & Sawilowsky, 2009). Mogu se, naravno, probati i ostale opcije:



Na kraju matrice će se formirati nova varijabla koja će u svom nazivu imati slovo N ispred imena varijable i ona sadrži normalizovane skorove. Da bi se dobili standardni skorovi, potrebno je primeniti iste korake kao u postupku računanja standardnih skorova, s tim što će

varijabla za koju se računa standardni skor biti varijabla s normalizovanim skorovima.

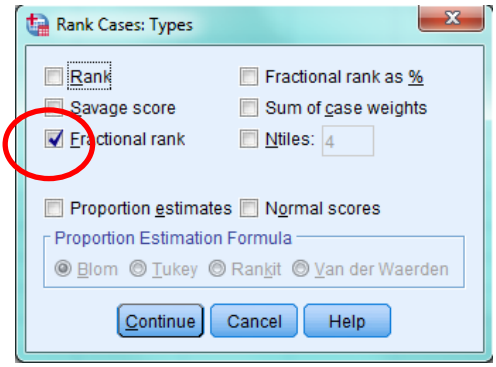
Ovaj način se uobičajeno koristi kada se računaju normalizovani standardni skorovi. Na str. 119 opisan je i postupak dobijanja normalizovanih standardnih skorova preko kumulativnih procenata (Urbina, 2004). Kumulativni procenti mogu se dobiti na različite načine, a u SPSS-u ih je moguće dobiti preko Analize/Descriptive Statistics/Frequencies. Dovoljno je samo u polje za varijable prebaciti varijablu sa sirovim skorovima i kliknuti OK. U autputu se dobija tabela sa sirovim skorovima i pripadajućim procentima odgovora i kumulativnim procentima:

Raven

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	7,00	1	1,4	1,4	1,4
	11,00	1	1,4	1,4	2,7
	14,00	3	4,1	4,1	6,8
	16,00	1	1,4	1,4	8,2
	17,00	1	1,4	1,4	9,6
	18,00	3	4,1	4,1	13,7
	19,00	3	4,1	4,1	17,8
	20,00	4	5,5	5,5	23,3
	21,00	7	9,6	9,6	32,9
	22,00	4	5,5	5,5	38,4
	23,00	6	8,2	8,2	46,6
	24,00	7	9,6	9,6	56,2
	25,00	2	2,7	2,7	58,9
	26,00	5	6,8	6,8	65,8
	27,00	10	13,7	13,7	79,5
	28,00	1	1,4	1,4	80,8
	29,00	3	4,1	4,1	84,9
	30,00	2	2,7	2,7	87,7
	31,00	4	5,5	5,5	93,2
	32,00	4	5,5	5,5	98,6
	33,00	1	1,4	1,4	100,0
	Total	73	100,0	100,0	

Potrebno je formirati novu matricu u SPSS-u koja će sadržati dve kolone iz autputa – kolonu sa sirovim skorovima i sa kumulativnim

procentima. Potom treba otići na Transform/Rank Cases, te u polje za varijable prebaciti varijablu sa kumulativnim procentima, a u okviru polja Rank Cases odabrati Fractional rank (ne Fractional rank as %) koji služi za izračunavanje proporcija:



Na kraju matrice ćete dobiti novu varijablu sa proporcijma koja će sadržati slovo R ispred naziva varijable, npr. RRaven. Sada treba preko funkcije *idf.normal* dobiti normalizovane z-skorove. Ova komanda olakšava proceduru jer se u istom koraku računa i normalizacija i standardizacija. Treba otići na Transform/Compute Variable, te u polje Target Variable dati naziv novoj varijabli, npr. NZskor, a zatim u polju Numeric Expression zadati funkciju:

idf.normal(RRaven, 0, 1)

pri čemu RRaven označava varijablu s kumulativnim proporcijama. Na kraju matrice će se formirati nova varijabla, NZskor koja će sadržati normalizovane standarne skorove, te se ona može podvrgnuti bilo kojoj transformaciji kako bi se dobili izvedeni standardni skorovi, npr. IQ.

Način dobijanja normalizovanih standardnih skorova preko percentila (Furr & Bacharach, 2014) opisan je na strani 119. Prema ovom postupku najpre se izračunavaju percentili, pa se oni podele sa 100 kako bi se dobile proporcije, ali u SPSS-u proporcije mogu direktno da se izračunaju preko opcije Fractional rank. Potom se novodobijena varijabla s proporcijama standardizuje (Analyse/Descriptive Statistics/Descriptives). Iz ovako dobijene varijable mogu se računati izvedeni *z*-skorovi. Razlika u odnosu na primer s kumulativnim procentima je u tome što ne moramo praviti posebnu matricu s kumulativnim procentima, već se direktno iz matrice sa sirovim skorovima mogu dobiti norme. Valja napomenuti da, iako skorovi dobijeni različitim postupcima normalizacije visoko koreliraju, ipak postoje razlike te uvek treba jasno naglasiti koji postupak normalizacije je korišćen.

Sten i stenajn

Ukoliko je distribucija normalna, sten, stenajn i C-skorovi mogu se izračunati kao i bilo koji drugi izvedeni standardni skorovi, preko Transform/Compute Variable.

Ukoliko distribucija nije normalna, skorovi se mogu najpre normalizovati, pa standardizovati, pa primeniti izraz za izračunavanje sten, stenajn ili C-skorova iz z-skorova. Moguće je, takođe, iz percentila računati sten i stenajn skorove, a sintakse za prebacivanje percentila u sten ili stenajn skorove se mogu naći na platformi Open Science Framework: https://osf.io/twyxb/.

Reference

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2006). *Standardi za pedagoško i psihološko testiranje*. Jastrebarsko, Hrvatska: Naklada Slap, za delo prevedeno na hrvatski jezik.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25.

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continuum International Publishing Group.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 20*(2), 1–7. Retrieved from: http://pareonline.net/getvn.asp?v=20&n=2

DeVellis, R. F. (2003). *Scale development: Theory and applications (2nd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Furr, M. R., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Thousand Oaks, California: SAGE Publications Ltd.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Gravetter, F. J., & Wallnau, L. B. (2014). *Statistics for the behavioral sciences* (9th ed.). Wadsworth, Belmon, CA: Jon-David Hague.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207–230.

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., ... Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*, S22–S31.

Smederevac, S., Mitrović, D. i Čolović, P. (2010). *Velikih pet plus dva, primena i interpretacija.* Beograd, Srbija: Centar za primenjenu psihologiju.

Solomon, S. R., & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods, 8*(2), 448–462.

Subotić, S. (2013). Pregled metoda za utvrđivanje broja faktora i komponenti (u EFA i PCA). *Primenjena psihologija, 6*(3), 203–229.

Swearer, S. M. & Cary, P. T. (2003). Perceptions and attitudes toward bullying in middle school youth: A developmental examination across the bully/victim continuum. *Journal of Applied School Psychology*, *19*(2), 63–79.

Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons Inc.

UNIVERZITET U NOVOM SADU
FILOZOFSKI FAKULTET NOVI SAD
21000 Novi Sad
Dr Zorana Đinđića 2
www.ff.uns.ac.rs

Elektronsko izdanje
http://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-540-2

CIP - Каталогизација у публикацији
Библиотеке Матице српске, Нови Сад

159.9.072

ДИНИЋ, Бојана, 1983-
Principi psihološkog testiranja [Elektronski izvor] / Bojana Dinić. - Novi Sad : Univerzitet, Filozofski fakultet, 2019

Način pristupa (URL): http://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-540-2. - Nasl. sa naslovnog ekrana. - Opis zasnovan na stanju na dan: 17.9.2019.

ISBN 978-86-6065-540-2

a) Психолошки тестови

COBISS.SR-ID 330688775