EKSTERNA EVALUACIJA. NUŽNOST ILI ZABLUDA.

LAVOSLAV ČAKLOVIĆ

SAŽETAK. U članku se raspravlja o ciljevima, principima i metodama ocjenjivanja, posebno na javnim državnim ispitima (maturama). U engleskoj obrazovnoj literaturi te su rasprave oživjele, posebno nakon razvoja kompjutorske tehnologije. Iako pod jakim utjecajem psihometrijskih metoda i teorije, današnje mišljenje o tome što se mjeri ispitnom procedurom poprima drugačije dimenzije.

Na kvalitetu ispitanika sve više se gleda kao na mentalni konstrukt promatrača (evaluatora) u procesu 'mjerenja'. Ona ne postoji izvan tog procesa. Cilj kvantitativnih procedura vrednovanja može i treba jedino služiti ocjenjivaču kako bi formirao svoju percepciju i prosudbu o ispitanikovim rezultatima tako da vrednovanje bude pravedno i konzistentno za sve ispitanike.

Donošenje odluke koja utječe na svakog pojedinca je mentalni proces koji se jako razlikuje od običnog ispitivanja podataka — to je proces *normativnog mjerenja*. Pri tome posebnu pažnju treba posvetiti profilima koji su na granici dviju ocjena. Testovi tih ispitanika uspoređuju se u parovima i rangiraju, a zatim nagrade ocjenom. Metode višeatributnog odlučivanja osiguravaju pravedan i konzistentan postupak.

1. UVOD

Nije pitanje misle li strojevi, pitanje je čine li to ljudi.

B. F. Skinner (bihejviorist)

Ovaj tekst je inspiriran iskustvima provođenja državnih ispita u Engleskoj, specijalno GCSE[1]. Navedene procedure u vrednovanju ispitnih rezultata odnose se na glavne koordinatore i njihovu tehničku podršku. U procesu vrednovanja ispitanika, a u svrhu donošenja opće ocjene, ispitivačke komisije trebaju apsorbirati, analizirati i usporediti miljune ocjena. Sadašnje procedure su mješavina lokalno razvijenih statističkih procedura po uzoru na psihometrijske procedure. Neke su smislene i adekvatne, a neke su nejasne i dvojbene.

Kao prvo, glavni koordinatori trebaju analizirati podatke i provjeriti jesu li instrumenti vrednovanja korišteni kako je zamišljeno i jesu li ocjenjivači poštovali zamišljenu proceduru. U tu svrhu, mnogi autori sugeriraju korištenje alata deskriptivne statistike. Drugo, glavni ispitivači trebaju prosuditi o sposobnostima svakog ispitanika i na kraju ga ocijeniti. Za ovu fazu edukativna literatura sugerira

Key words and phrases. vrednovanje, državna matura, ocjenjivanje.

[1]General Certificate of Secondary Education

normativne mjerne tehnike, to su tehnike višekriterijskog odlučivanja, koje vode nepristranom i konzistentnom vrednovanju.

U poglavlju 2 (*Ciljevi vrednovanja znanja*) raspravlja se o ciljevima vrednovanja znanja u obrazovnoj literaturi, vrstama testova i pojmu ocjene. Poglavlje 3 (*Principi, snaga i ograničenost ljudskog prosuđivanja*) govori o principima ljudskog prosuđivanja i njihovoj primjeni kod individualnog i grupnog ocjenjivanja. Poglavlje 4 (*Procedura zaključivanja ocjena*) govori o procedurama kod donošenja ocjena i kvantitativnoj smislenosti nekih operacija. Principi višeatributnog i grupnog odlučivanja i njihova primjena kod ocjenjivanja objašnjeni su u poglavlju 5 (*Višeatributno odlučivanje i ocjenjivanje*). Naglasak je dan na teoriju korisnosti i metodu potencijala. Poglavlje 6 (*Kako (o)smisliti ocjenjivanje?*) se pita postoji li alternativa sadašnjoj praksi i daje neke sugestije. Poglavlje 7 (*Državna matura u Hrvatskoj*) se pita kuda ide državna matura u Hrvatskoj.

Postavlja se pitanje da li se teorija primjenjuje u praksi. U strogom smislu ne. Ipak, navedena pitanja i sugestije imaju za posljedicu da se rasprava o metodi fokusira oko budućeg praktičnog i teorijskog istraživanja posebno prema razvoju alata za potporu ispitivačima i njihovom prosuđivanju.

2. Ciljevi vrednovanja znanja

Što je cilj vrednovanja znanja? Ocijeniti trenutačno znanje, vještine, postignuća, urođene vještine, neke buduće sposobnosti, da služi u pedagoške svrhe ili kombinacija svega toga. Naivno gledajući, cilj vrednovanja znanja je 'utvrditi postignuća' ispitanika iz jednog predmeta ili grupe predmeta. Trenutačni stavovi o vrednovanju u obrazovnoj literaturi su pod utjecajem psihometrijskih koncepata i teorija korištenih u analizama psihometrijskih testova, kao IQ test na primjer. Te teorije zasnovane su na vjerovanju da:

- unutar svakog pojedinca postoji nešto što se može zvati 'sposobnošću' ili 'postignućem' u izvršavanju postavljenih zadataka,
- taj entitet je mjerljiv na objektivnoj, jednodimenzionalnoj skali,
- cilj testa je sakupiti podatke na temelju kojih ga je moguće procijeniti.

French (1989) [4] smatra da cilj javnih ispita (u Engleskoj i Walesu) nije objektivno mjerenje nečega o kandidatu i da ne postoji neki entitet unutar kandidata koji može biti mjeren na taj način. Prema Frenchu, cilj takvog ispita je iznijeti prosudbu donešenu od strane ispitivačke komisije. Bodovi i ocjene ne predstavljaju mjeru entiteta već ispitivačevu prosudbu kvalitete ipitanikove 'predstave', kako je oni vide, u ispitanikovom testu ili nekoj drugoj formi pokazivanja sposobnosti. Dalje, French iznosi da...

> ...ta distinkcija nije nevažan i ezoterijski akademski stav; ona ima značajne implikacije za formu mnogih numeričkih procedura korištenih kod manipulacije u postupku vrednovanja. Te procedure nisu tu da procjenjuju kandidatove sposobnostima u prisutnosti 'mjernih grešaka' niti imaju za cilj procjenu ispitivačevih sudova. Njhova je uloga suptilnija. One pomažu ispitivačima da bi formirali pravedne i konzistentne procjene.

2.1. **Standardizacija vrednovanja znanja.** Na koji način psihometrijski principi utječu na evaluaciju i razvoj ispitivanja znanja u školi kao što je državna matura na primjer? Bez obzira jesu li rezultati ispita prosudbe ili mjerenja, one trebaju biti *pouzdane, valjane, standardizirane* i *oslobođene pristranosti*. I na samu psihometriku danas se sve više gleda kao na nauku psihološkog vrednovanja a ne samo mjerenja.

Što se standardizacije procesa vrednovanja u obrazovanju tiče tu ima nekoliko pristupa. U literaturi se najčešće spominju dvije forme standardizacije: *normativno* orijenirana i *kriterijski* orijentirana. U praksi se gotovo uvijek radi o mješavini jedne i druge.

2.1.1. *Normativni test.* Normativno orijentirani test (eng. norm-referencing) daje rang ispitanika u odnosu na rezultate testa na nekoj predefiniranoj grupi ispitanika koji su bili ispitivani u istim uvjetima (normativna grupa). Očigledan nedostatak normativno orijentiranog testa je taj da ne može mjeriti napredak ispitanika. Prednost takvog testa je što i studenti i profesori znaju što se očekuje od testa i kako će test biti ocjenjivan. Uspjeh na takvom testu izražava se najčešće u *percentilima*.

Tipični primjer takvog testa je test inteligencije. Izborom normativne grupe već je intuitivno jasno što se tim testom želi mjeriti. Mišljenje autora je da škole ne bi trebale koristiti takve testove kao mjeru znanja jer zaobilaze zacrtani kurikulum i ne ukazuju na to što bi studenti trebali znati. Njihova je glavna namjena da razlikuju ispitanike i rang postignut na takvom testu ne govori o tome što student zna a što ne zna.

Neke institucije za normativnu grupu uzimaju grupu ispitanika[2]. Defekt takvog pristupa je taj da ocjena dobivena na ispitu danas i nakon godinu dana ne odražava jednaku kvalitetu ispitanika.

2.1.2. *Kriterijski test.* Kriterijski orijentirani testovi (eng. criterion-referencing) namijenjeni su da provjere u kojoj je mjeri osoba naučila specifično područje ili ovladala zacrtanim vještinama. Najčešće se sastoji od pitanja s više ponuđenih odgovora ili dozvoljava kratke rečenice kao odgovor. Pojam 'kriterij' ovdje ne treba miješati s prolaznim pragom kojim se zahtijeva od studenta da sakupi minimalan broj brodova kako bi zadovoljio na testu, već se misli na vještine koje se provjeravaju. Mnogi državni testovi u raznim zemljama koriste tu vrstu testova.

Čak i ako testiraju bliska znanja i vještine, test koji je dizajniran za provjeru sposobnosti koristit će drugačija pitanja nego test koji je namijenjen diferencijaciji ispitanika. To je stoga jer neka pitanja bolje odražavaju aktualna postignuća ispitanika, a neka druga pitanja su bolja za diferencijaciju između 'dobrih' i 'loših'. Mnoga pitanja čine i jedno i drugo. Kriterijski orijentirani test će koristiti pitanja na koja će korektno odgovoriti studenti koji poznaju određenu materiju, a normativno orijentirani test će koristiti pitanja na koja korektno odgovaraju 'najbolji', a nekorektno 'najlošiji' studenti.

Zanimljiva je sljedeća misao, Linn (1993) [12]:

[2]eng. cohort referencing, *cohort*—drug, ortak

> Općenito se priznaje da termini 'kriterijsko' i 'normativno', primijenjeni na instrumente vrednovanja, vode u zabludu jer se interpretacija daje ocjenama, a ne samim instrumentima.

2.1.3. *Još neki tipovi referenciranja.* U posljednje vrijeme sve više se govori o *limen referencing* u kojem se profil kandidata uspoređuje s tipičnim zamišljenim predstavnikom klase ispitanika za svaku ocjenu. Taj tip referenciranja postaje zanimljiviji utoliko što je porasla svijest o važnosti kurikuluma i sličnih standarda u obrazovanju. Važnost takvog referenciranja u ocjenjivanju dolazi do izražaja u fazi ocjenjivanja tzv. *graničnih testova*, v. odjeljak 5.1.

Slabo kriterijsko referenciranje je oslabljeno kriterijsko referenciranje koje nastoji zadržati kvalitetu i nivo izvođenja ispita. Pri tome se nastoji zadržati težina samog ispita ali se ne zahtijeva potvrda specifičnih znanja i vještina.

Construct referencing je način dodjeljivanja ocjena koji je fleksibilniji od kriterijskog, a uvažava i projektne zadatke i subjektivne dojmove o njima.

Niti jedan od navedenih tipova nije referentni sistem koji daje značenje ocjeni, to su metode koje pomažu u određivanju granica među ocjenama.

2.2. **Ocjena, što je to?** Ljudi su skloni vjerovati da je ocjena kvantitativna mjera sposobnosti ispitanika. Sama po sebi ocjena ne govori ama baš ništa o ispitaniku. Ako ispitanika i njegovu ocjenu stavimo u neki kontekst (školu, profesora, fakultet, sustav ...) tada o njemu stičemo neki dojam. Zapravo, mijenjamo sliku o sebi jer sebe stavljamo u kontekst iz kojeg promatramo osobu kroz prizmu njegove ocjene.

Ocjena dobivena na ispitu/testu, pretpostavljam, zamišljena je kao mjera koja odražava stupanj savladanosti zacrtanih obrazovnih ciljeva od strane kandidata. Način ocjenjivanja dio je kulturnog nasljeđa institucije, sustava, države. Ocjena ***kao pedagoški pokazatelj i motivacija*** u procesu savladavanja gradiva (interno ocjenjivanje) dio je obrazovnog sustava i nema drugih korisnika osim učenika, njegovih staratelja i nastavnika.

Ocjena na državnom ispitu/maturi (vanjsko ocjenjivanje) i opći uspjeh u srednjoj školi, ***kao informacija koja prati studenta*** i markira njegov budući život, koristi se u mnogim situacijama i korisnici te ocjene manipuliraju s njom na način kako njima to odgovara. Tako na primjer, fakulteti mogu i ne moraju koristiti uspjeh u srednjoj školi i/ili maturi prilikom upisa na fakultet. Ovdje se neka pitanja i sugestije otvaraju sami po sebi:

- Podupire li državna matura dinamiku razvoja društva?
- Razne institucije imaju potrebu za različitim profilima zaposlenika. Nije li bolje da država osigura intrumente evaluacije koje će institucije unajmljivati?
- Ti instrumenti evaluacije mogu se lakše internacionalizirati i modularno nadograđivati.

3. Principi, snaga i ograničenost ljudskog prosuđivanja

Postoje neki opći principi kojima je podložno ljudsko prosuđivanje i koji imaju utjecaj kako na organizaciju javnog testiranja tako i na organizaciju ispravljanja

testova sve do trenutka utvrđivanja liste profila, odnosno prije faze evaluacije i donošenja sveukupne ocjene, Grateaux (2007) [9].

1. princip. U procesu procjene, ocjene se donose uspoređivanjem u parovima. Apsolutne procjene se ne traže ni u kom slučaju. Laming (2004) [11] zaključuje da su sve procjene koje ljudi čine usporedbe jednog objekta s drugim i da su te procjene *nešto bolje od ordinalnih*. Osim toga, ljudski um nije u stanju držati precizan i stabilan referentni okvir u memoriji, a može razlikovati najviše pet kategorija danog kontinuuma ako nema neku dodatnu podršku. Stoga, kad ljudi donose niz uzastopnih procjena, njihove nedavne procjene su pod utjecajem prijašnjih procjena pa postaju zbunjeni i nije začuđujuće da to vodi greškama u procjenjivanju. Procjene postaju preciznije kada se objekti uspoređuju u parovima umjesto da se iznose apsolutne procjene koristeći prihvaćene standarde.

2. princip. Ocjenjivači donose odluke nezavisno jedan od drugoga. Procjene pojedinca u grupi podliježu grupnoj dinamici pa osciliraju. Ljudi često donose pogrešne procjene samo da bi se uklopili u gomilu.

3. princip. Ako je ikako moguće, smanjiti prisutnost nevažnih informacija. Na primjer, procjena kvalitete likovnog djela je pod velikim utjecajem (mogućeg) autorstva tog djela.

4. princip. Zahtijeva se adekvatno iskustvo ocjenjivača. Neposredno iskustvo koristi se kao referentna točka, a prošla iskustva čine procjenu pouzdanom i ponovljivom što je važno kod utvrđivanja granice među ocjenama.

5. princip. Ocjenjivači procjenjuju jedan zadatak (komponentu) testa. Za donošenje ukupne ocjene postoji softverska podrška. Ljudi donose preciznije i pouzdanije procjene ako se koncentriraju na komponentu umjesto na cjelinu, Dawes & Corrigan (1974) [3]. Na primjer, statistička kombinacija znakova bolesti koje dijagnosticira liječnik je preciznija nego njegovo sveobuhvatno mišljenje, Laming (2004) [11]. Jednako tako, eksperti su dobri u pronalaženju onoga što traže ali nisu dobri u povezivanju parcijalnih informacija u cjelokupnu procjenu kvalitete ispitanika.

6. princip. Odgovori ispitanika (prikazani ispitivaču) trebaju biti u obliku koji olakšava razumijevanje ispitanikovih odgovora. Na primjer, čitanje testova na monitoru otežava njihovo razumijevanje zbog nemogućnosti prelistavanja i istovremenog gledanja svih listova (testova).

7. princip. Ocjena koju donese ocjenjivač mora se moći rekonstruirati. Drugim riječima, da bude tako formirana da ju drugi ocjenjivač može ponoviti.

Jedan od zaključaka temeljen na tim principima je da ocjenjivač ne bi smio znati kojoj školi ili grupi ispitanika pripada test kojeg ocjenjuje. Prema principu 3. to je suvišna i nevažna informacija.

4. PROCEDURA ZAKLJUČIVANJA OCJENA

Rezultat vrednovanja svakog predmeta (komponente) je određen skup bodova — ***profil***, v. tablicu 1. Nakon toga, daljnje vrednovanje se isključivo bazira na tom profilu. Bodovi su brojevi i tu leži opasnost u manipulaciji s njima. Oni mogu

biti zbrajani, može se računati srednja vrijednost, mogu biti podvrgnuti statističkoj analizi čak i ako je to posve besmisleno.

Sintezu (dodjelu ocjena na temelju profila) radi glavni koordinator sa suradnicima. To je dvofazna procedura. U prvoj fazi daje se ocjena svakom kandidatu po nekoj uhodanoj proceduri, a nakon toga se analiziraju i pregledavaju testovi onih ispitanika čiji bodovi su blizu zacrtane granice za tu ocjenu, v. Forrest (1981.) [2]. Pri tome se koriste *deskriptivne* i *normativne* procedure.

Deskriptivne procedure služe za utvrđivanje činjenica i njihovu analizu. Jedna takva deskriptivna procedura je računanje zajedničke distribucije i kovarijance među zadacima i među ispitanicima u svrhu naknadne procjene težine zadataka ili mjerenje mogućeg odstupanja ocjenjivača od dogovorenih standarda u ocjenjivanju.

Normativne procedure su posve drugačije. One nastoje mijenjati i činiti konzistentnijim vjerovanja i preferencije pojedinca. Pojedinca se pita za njegove subjektivne stavove i njegovi odgovori se analiziraju i ispituju jesu li u skladu s unaprijed utvrđenim pravilima 'racionalnog' ponašanja. Njegova vlastita inkonzistentnost i nepoštivanje pravila mu se predočuju i on sebe shvaća, sagledava i evoluira u svojim vjerovanjima i preferencijama i teži sve većoj konzistenciji.

Uzmimo za primjer, ocjenjivača koji ocjenjuje određen broj testova ispitanika. Prije svega, on utvrđuje neku grubu shemu bodovanja koja bi trebala reflektirati kvalitetu samih testova. Pretpostavimo da je ocijenio određen broj testova i, kako napreduje, on profinjuje utvrđenu shemu ocjenjivanja jer postaje svjestan finesa u proceduri ocjenjivanja. U nekom trenutku, ocjenjivač postaje nezadovoljan rezultatom svog ocjenjivanja jer smatra da je određen broj ispitanika dobio više bodova nego je 'zaslužio', a neki drugi ispitanici su dobili manje od 'zasluženog' broja bodova. On radi pauzu u ocjenjivanju i razmišlja što da učini. Moguća su dva zaključka: (1) njegova je shema ocjena neprikladna; (2) njegova holistička procjena precjenjuje neke aspekte ispitanikovog uratka koje je već nagradio ili ih je predvidio ali ih je zaboravio nagraditi. U prvom slučaju on revidira shemu ocjenjivanja, a u drugom modificira mentalne procese koje koristi u holističkoj procjeni testova. U oba slučaja prisutna je inkonzistentnost između *numeričkih reprezentacija* njegovih procjena i samih *procjena* (mentalnih procesa). Kvaliteta njegovog rada evoluira.

Jednako tako, normativna je i procedura vrednovanja ispitanikovog 'općeg uspjeha' koja u prvom redu pomaže ocjenjivaču formirati procjenu na pošten i konzistentan način što je više moguće, v. odjeljak 5.

4.1. **Kvantitativna smislenost operacija.** U deksriptivnim i normativnim analizama koristimo brojeve kako bi reprezentirali razne entitete i relacije među njima. Manipulacije s brojevima mogu biti manje ili više smislene ovisno o tome održavaju li ti brojevi kvalitativne ili kvantitativne odnose među entitetima. Kvantitativni model zamjenuje kvalitativne relacije, kao na primjer:

i) ovaj je štap *dulji* nego onaj, ili
ii) ja *preferiram* objekt A u odnosu na objekt B,

s kvantitativnim. U prvom slučaju model pridružuje štapovima numeričku vrijednost koja predstavlja *duljinu* štapa na nekoj skali. U drugom slučaju objektu A pridajemo veću *korisnost* nego objektu B na nekoj skali. Sofisticiraniji modeli preferenciji u ii) pridružuju i *intenzitet preferencije* i na temelju tih podataka rekonstruiraju korisnost objekata. Razlozi za zamjenu riječi brojem su višestruki. Kvantitativni modeli su sažetiji, omogućavaju suptilnije izražavanje i omogućavaju razne kvantitativne analize.

4.2. **Skale u ocjenjivanju.** Evo primjera za početak. Pretpostavimo da je ocjenjivač bodovao testove tri ispitanika A, B, C na temelju profila u tablici 1. Najveći mogući broj bodova po predmetu je 100. Kvantitavni odnos između 66 i 63 pokazuje da je kandidat B postigao bolji uspjeh od kandidata A na testu ako se gleda samo predmet I. U posljednjem stupcu tabele dana je suma bodova svakog kandidata koja predstavlja *opći uspjeh* kandidata za sva tri predmeta. Srednja vrijednost (ili suma bodova) je načešće korištena mjera kvalitete općeg uspjeha.

Pretpostavimo da je neki drugi ispitivač također ocijenio iste testove i rezultat njegovog ocjenjivanja je skup profila predstavljen u tablici 2. Numeričke vrijednosti pridijeljene uspjehu ispitanika za svaki predmet predstavljaju isti kvalitativni poredak za oba ocjenjivača. Drugi ispitivač je bio nešto blaži u ocjenjivanju za drugi predmet i stroži za preostala dva. Suma bodova međutim daje inverzni poredak za opći uspjeh.

Gornji primjer ukazuje na ***besmislenost računanja prosjeka*** ocjena po profilu kao mjeru općeg uspjeha. Drugim riječima, ako bodovi za svaki predmet predstavljaju *poredak* kandidata za taj predmet, onda je računanje prosjeka *kvantitativno besmislena* operacija. Uzrok toj besmislenosti leži u izboru skale na kojoj mjerimo kandidate. Ovdje je to *ordinalna* skala jer reflektira samo poredak kandidata i ne uvažava njihove međusobne razlike na toj skali kao dodatnu mjeru kvalitete. Skala koja uvažava te razlike je *intervalna skala*. Što se ocjenjivača tiče, nije sasvim sigurno da su u stanju napraviti procjene koje uvažavaju razliku vrijednosti na skali, tj. nisu u stanju 'mjeriti' ispitanike na intervalnoj skali, French i Vassiloglou (1986) [6].

Ordinalna skala dozvoljava strogo rastuću promjenu skale dok intervalna skala dozvoljava pozitive afine transformacije oblika $x \mapsto \alpha x + \beta,\ \alpha > 0$. Preračunavanje stupnjeva ${}^oC$ u Fahrenheighte je primjer transformacije intervalne skale.

	Predmet			
	I	II	III	
A	63	59	66	188
B	66	57	69	192
C	69	56	68	193

TABLICA 1. Bodovanje ispitanika.

	Predmet			
	I	II	III	
A	59	65	46	170
B	60	61	48	169
C	61	58	47	166

TABLICA 2. Drugo moguće bodovanje ispitanika.

Za intervalnu skalu je prosjek kvantitativno smislena operacija jer je odnos između prosjeka dva ispitanika neovisan o dozvoljenoj transformaciji intervalne skale[3]. Asby i Townsend (1984) [1] argumentiraju da ako su podaci dani na ordinalnoj skali onda su sredine, varijance kvantitativno besmislene i bilo kakvi statistički zaključci na njima bazirani nisu prikladni. Samo neparametarski testovi bazirani na rangovima mogu biti primjereni. Nužni uvjet da bi parametarski testovi mogli biti korišteni je da su podaci mjereni na intervalnoj skali.

Postoje i drugačija mišljenja, Gaito (1980) [7], pa čak i uzrečica: *brojevi ne znaju odakle dolaze*. Ipak, opreznost je majka mudrosti. U članku French [4] dana su dva primjera koja pokazuju da statistički testovi nad ordinalnim podacima imaju smisla ovisno o pitanjima koja zanimaju analitičara.

Da bi numerički odnosi među brojevima[4] bili kvalitativno interpretabilni oni svakako moraju biti kvantitativno smisleni. Međutim, još jedan uvjet treba biti ispunjen. Ti odnosi trebaju biti *semantički smisleni* tj. trebaju biti interpretabilni u kvalitativnom smislu u granicama korisnikove percepcije. Na primjer, ako veliki broj profila podvrgnemo faktorskoj analizi, onda dobiveni faktori i njihove 'težine' trebaju biti kvalitativno percipirani od strane korisnika, v. Čaklović-Radas (u pripremi) [19].

Zanimljivo je da metoda potencijala (odjeljak 5.4) za profile u tablici 1 i 2 daje isto rangiranje: To je posljedica činjenice što metoda potencijala uzima razli-

	Tablica 1	Tablica 2
B	0.359	0.369
C	0.345	0.334
A	0.296	0.297

TABLICA 3. Rangiranje profila iz tablica 1 i 2 metodom potencijala.

[3]Ostavljamo čitaocu da sam dokaže navedenu tvrdnju.

[4]Prosjek ocjena na primjer.

ke vrijednosti u svakom stupcu kao input i još te vektore normira tako da najveća razlika po apsolutnoj vrijednosti ima vrijednost 1. Taj postupak normiranja odgovara 'trade-off' proceduri u višekriterijskom odlučivanju. Ovakvo rangiranje u oba primjera moglo bi se okarakterizirati više kao slučaj nego pravilo, ali se naslućuju neke 'privlačne' osobine metode potencijala. Na WEB adresi `http://decision.math.hr/programs/` » `Group decision` (*EduProfil I* i *EduProfil II*) čitalac može provjeriti gornji račun. Na istom mjestu čitatelj može i sam kreirati vlastiti profil i eksperimentirati po želji.

5. Višeatributno odlučivanje i ocjenjivanje

Glavni koordinator (s ekipom) koji ocjenjuje opći uspjeh na temelju profila svakog ispitanika može krenuti u fazu određivanja granica između dvije ocejen ako je siguran da su pojedine komponente ispita korektno bodovane. To još uvijek ne znači da pojedini testovi neće biti ponovno pregledani ako se nalaze na granici između dvije ocjene.

5.1. **Određivanje granica među ocjenama.** Pretpostavimo da je svakom ispitaniku dodijeljen njegov profil s ocjenama po komponentama kao u tablici 1 ili tablici 2 i da želimo svakom ispitaniku dati opću ocjenu na ispitu u skupu $\{1, 2, 3, 4, 5\}$. Opća ocjena nastaje *agregacijom* ili *akumulacijom* parcijalnih bodova po nekom principu. Kao što smo vidjeli u tablicama 1 i 2 prosjek nije dobar postupak agregacije bodova jer nije kvantitativno smislen. Problem je u tome što ne postoji dobar i smislen algoritam koji bi profilima automatski dodijelio ocjene. Mi ćemo, radi jednostavnijeg razumijevanja cijele procedure opisati rudimentarnu proceduru koja lako može biti implementirana. Vidi također Greatorex (2009) [10] za druge metode određivanja granice.

Metoda se provodi u dvije faze. U prvoj fazi se svakom profilu dodijeli ocjena ili dvije susjedne ocjene ako postoji nedoumica. Tu klasifikaciju radi ocjenjivač ekspert ili neki jednostavan algoritam koji još uvijek nije dovoljno inteligentan da razgraniči spada li pojedini test ispitanika u kategoriju (klasu) 'čiste ocjene'. Testovi kojima pripadaju dvije ocjene spadaju u tzv. *granične* testove. U drugoj fazi se granični testovi rangiraju na temelju uspoređivanja u parovima nekom od metoda. Proceduru uspoređivanja po parovima može učiniti jedan ekspert ili grupa njih. U slučaju grupne odluke rangiranje je dano konsenzusom grupe, v. Čaklović (2004) [18]. Nakon toga se skup graničnih testova podijeli u gornju i donju klasu koje se pridruže gornjoj i donjoj 'čistoj ocjeni'.

U predloženoj proceduri, kriterij prema kojem se testovi uspoređuju u parovima je *opći dojam*, dakle jedan kriterij. Tehnički nije tešto zakomplicirati proceduru na način da se pojedinim komponentama daju težine i rangiranje provede u odnosu na komponente ispita kao kriterije. Pitanje je koliko je to smisleno, no to je stvar odluke glavnog koordinatora i njegovih pomagača.

5.2. **Teorija korisnosti.** Klasičan pristup višeatributnom odlučivanju koji se razvija posljednjih 70tak godina koristi funkciju korisnosti, neki je još nazivaju i

funkcija vrijednosti. Za razliku od rangiranja, funkcija korisnosti pridružuje objektima koje 'mjerimo' vrijednosti na intervalnoj skali. Začetnici takvog pristupa su Von Neumann i Morgenstern (1944) [20], a za aksiomatski pristup zaslužan je Savage (1954) [17].

5.2.1. *Uspoređivanje u parovima.* Uspoređivanje u parovima prisutno je u gotovo svim metodama odlučivanja. Rezultate tog uspoređivanja Saatyjeva AHP metoda [16] zapisuje u pozitivnu recipročnu matricu, a metoda potencijala, Čaklović [18], ih pamti kao usmjeren graf $\mathcal{G} = (V, \mathcal{A})$ gdje je V skup vrhova (alternativa) a $\mathcal{A}$ skup lukova. Ako je vrh $a \in V$ više preferiran nego vrh b tada kreiramo luk $\alpha = (a, b) \in \mathcal{A}$ koji izlazi iz b i ulazi u a. Skup svih lukova predstavlja relaciju preferencije na kartezijevom skupu $V \times V$ što zapisujemo ovako

$$a \succcurlyeq b \iff \alpha \in \mathcal{A}.$$

Problem je kako iz zadane relacije preferencije $\succcurlyeq$ konstruirati funkcija $X : V \to \mathbb{R}$ na skupu alternativa V tako da su relacija $\succcurlyeq$ i funkcija X usklađene tj. da vrijedi:

$$a \succcurlyeq b \iff X(a) \geq X(b). \tag{1}$$

Ako je usklađenost prisutna kažemo da je X ***ordinalna funkcija*** vrijednosti generirana relacijom $\succcurlyeq$. Drugim riječima, preferencija među objektima može se izraziti brojem. Ordinalna funkcija vrijednosti X uvijek postoji ako je relacija $\succcurlyeq$ potpuna i tranzitivna i dana je formulom (Savage)

$$X(a) = \#\{b \in V \mid a \succcurlyeq b\} \tag{2}$$

gdje $\#$ označava kardinalni broj skupa.

Ordinalna funkcija vrijednosti nije jedinstvena. Svaka druga funkcija koja je nastala kompozicijom funkcije X i neke rastuće funkcije također zadovoljava zahtjev (1).

Teorija korisnosti ***pretpostavlja*** da je relacija $\succcurlyeq$ ***potpuna*** (svaka dva vrha su povezana lukom) i ***tranzitivna*** (graf nema nenegativnih ciklusa). To znači da donosilac odluke mora biti u stanju usporediti svake dvije alternative i pri tome ne smije narušiti tranzitivnost preferencije. To su dvije nezaobilazne pretpostavke na kojima se bazira čovjekova 'racionalnost' u prosuđivanju.

Kod većine problema odlučivanja gornje dvije pretpostavke teško se mogu zadovoljiti. Tako na primjer, neke usporedbe zahtijevaju puno vremena ili puno novaca ili jedno i drugo. Netranzitivnost često može biti posljedica loše procjene, površnosti u prosuđivanju, zamora zbog previše razmišljanja ili pak nepoznavanja problema od strane donositelja odluke. U daljnjem tekstu mi pretpostavljamo samo da graf preferencije bude povezan.[5]

[5]Iskusni procjenitelji obično zapišu redoslijed alternativa koji intuitivno očekuju i nakon toga prelaze na uspoređivanje u parovima. Uspoređivanje u parovima prvi puta se primijenilo u testovima 1927. g. i mnoga psihološka istraživanja sugeriraju da je čovjekov um zbunjen uvođenjem treće alternative u razmatranje. To je jedan od razloga zašto većina metoda za donošenje odluka koristi baš uspoređivanje u parovima.

Drugi zahtjev, koji teorija korisnosti postavlja donositelju odluke, jest taj da mora biti u stanju uspoređivati ***zamjene*** (eng. exchange). Zamjenu možemo poistovijetiti s lukom u grafu preferencije pa se uspoređivanje zamjena svodi na uspoređivanje lukova. Klasična oznaka za zamjenu je $(a \leftarrow b)$ i uspoređivanje zamjena definira relaciju na skupu svih zamjena $(V \leftarrow V)$ koju označavamo s $\succcurlyeq_e$. Od relacije $\succcurlyeq_e$ također se zahtijeva da bude potpuna i tranzitivna. Osim toga relacije $\succcurlyeq_e$ i $\succcurlyeq$ moraju biti usklađene na neki način. Ta se usklađenost izražava sljedećim odnosom

$$(a \leftarrow b) \succcurlyeq_e (c \leftarrow d) \iff X(a) - X(b) \geq X(c) - X(d), \tag{3}$$

gdje X zadovoljava (1).

Formula (1) ističe važnost same numeričke vrijednosti $X(a)$ objekta a, dok formula (3) daje važnost i razlici $X(a) - X(b)$. Drugim riječima, funkcija X koja zadovoljava (1) predstavlja *ordinalnu*, a ona koja zadovoljava (3) predstavlja *intervalnu* skalu[6]. U literaturi se intervalna skala još naziva i *izmjeriva* funkcija vrijednosti. Konstrukcija intervalne skale na temelju dviju relacija $\succcurlyeq$ i $\succcurlyeq_e$ je prilično kompliciran i zahtijeva još neke dodatne zahtjeve, v. French [5].

5.3. **Višeatributno odlučivanje.** Višeatributno odlučivanje koristi skup profila kao polazište za rangiranje alternativa. Posebno je zanimljiva situacija, u ocjenjivanju pogotovo, kad je funkcija korisnosti aditivna po atributima. Jedan od aksioma 'zaslužan' za aditivnost je aksiom *nezavisnosti atributa*.

Pretpostavimo da je rezultat ocjenjivanja nekog broja ispitanika dan skupom profila prema nekim kriterijima (atributima) $\mathcal{A}$ i da je $\succcurlyeq$ relacija preferencije na skupu profila izvedena iz parcijalnih bodova v_i gdje je $i \in \mathcal{A}$.

Aksiom nezavisnosti atributa.

> Neka je $\mathcal{A}$ skup atributa, $\mathcal{J} \subset \mathcal{A}$ i $\mathcal{J}' = \mathcal{A} \setminus \mathcal{J}$ skup komplementarnih atributa. Reći ćemo da je $\mathcal{J}$ preferencijalno nezavisan skup atributa u $\mathcal{A}$ ako za bilo koja 4 ispitanika a, b, c, d za koja vrijedi
>
> $$\begin{aligned} v_i(a) &= v_i(b), \quad \forall i \in \mathcal{J}' \\ v_i(c) &= v_i(d), \quad \forall i \in \mathcal{J}' \\ v_i(a) &= v_i(c), \quad \forall i \in \mathcal{J} \\ v_i(b) &= v_i(d), \quad \forall i \in \mathcal{J} \end{aligned}$$
>
> zaključujemo da je $a \succcurlyeq b$ ako i samo ako je $c \succcurlyeq d$.

Drugim riječima, $\mathcal{J}$ je preferencijalno nezavisan skup atributa ako preferencija u kvaliteti između dva ispitanika ne ovisi o vrijednostima atributa izvan $\mathcal{J}$. Tako na primjer, u tablici 4 atributi I i II su nezavisni ako $(a \succcurlyeq b \iff c \succcurlyeq d)$. Kad bi pravilo agregacije bilo takvo da iz svakog profila zbrojimo tri najbolje ocjene, i rangiramo ih na temelju te sume, takvo bi pravilo razbilo nezavisnost atributa.

[6]Preciznije: formula (3) je nužni uvjet za intervalnu skalu.

	I	II	III	IV
a	49	60	88	71
b	53	58	88	71
c	49	60	33	46
d	53	58	33	46

TABLICA 4. Nezavisnost atributa I i II.

U većini slučajeva (ispita) zahtijeva se nezavisnost atributa. Posljedica toga je da funkcija agregacije ima jednostavan oblik:

$$\sum_{i=1}^{n} v_i(x_i),$$

gdje je n broj komponenti koje agregiramo i x_i ocjena dobivena iz i-te komponente. Kako osigurati da nezavisnost atributa bude prisutna u ocjenjivanju je priča za sebe, v. French (1998) [5], Roberts (1979) [15].

Funkcije v_i određuju se tako da ispitivač procijeni koliko bodova jedne komponente vrijedi jedan (ili više bodova) druge komponente. Taj 'trade-off' (trgovina) ne mora biti konstantan, može ovisiti i o samim vrijednostima već dobivenih bodova. Drugim riječima v_i može biti nelinearna funkcija. Na primjer, u matematici ili fizici, praktični zadaci i teorijska pitanja mogu imati jednu trade-off vrijednost ako je ispitanik nešto iznad praga, a drugu vrijednost ako je ispitanik kandidat za visoku ocjenu.

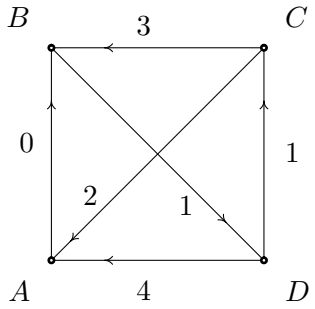
5.4. **Metoda potencijala.** Metoda potencijala, razvijena od strane autora, bazira donošenje odluke na grafu preferencija i uspoređivanju u parovima, a u uskoj je vezi sa Saatyjevom AHP metodom (Analytic Hierarchy Process) [16].

Kod uspoređivanja parova alternativa, osim orijentacije, usporedbi se obično pridružuje i intenzitet preferencije na nekoj skali. To znači da svakom luku α usmjerenog grafa pridružujemo nenegativan broj, označimo ga s $\mathcal{F}_\alpha$. U slučaju da obje alternative jednako vrednujemo onda je vrijednost intenziteta jednaka 0 (nula), a orijentacija luka je nevažna. Tako određenu funkciju $\mathcal{F}$ nazivamo ***tokom preferencije***. Na slici 1 dan je primjer jednog toka preferencije na skupu alternativa $V = \{A, B, C, D\}$. Primijetimo da su A i B jednako vrednovane ali zato A ima prednost ispred C i to intenziteta 2. U ovome je primjeru prisutna i nedosljednost donositelja odluke u ciklusu $D \to C \to A \to D$ duž kojeg zbroj intenziteta nije jednak nuli.

Konzistentan donosilac odluke trebao bi dati graf preferencije u kojem je

(4) zbroj *svih* intenziteta duž *svakog* ciklusa jednak nuli.

Praksa pokazuje da je u subjektivnim preferencijama donosilac odluke rijetko kada konzistentan. Usprkos tome, metoda potencijala je sposobna rangirati alternative i čak izmjeriti inkonzistentnost donosioca odluke, što je posebno korisno za naknadnu analizu odluke. Problem u donošenju odluke je sljedeći:



SLIKA 1. Tok preferencije, primjer.

Iz zadanog toka $\mathcal{F}$ konstruirati funkciju $X : V \to \mathbb{R}$ na skupu V tako da tok $\mathcal{F}$ i funkcija X budu usklađeni na sljedeći način:

$$\mathcal{F}_\alpha \geq 0 \iff X(a) \geq X(b), \tag{5}$$

za svaki luk $\alpha = (a, b) \in \mathcal{A}$ i

$$\mathcal{F}_\alpha \geq \mathcal{F}_\beta \iff X(a) - X(b) \geq X(c) - X(d), \tag{6}$$

gdje je $\alpha = (a, b)$ i $\beta = (c, d)$.

Uvjet (5) zahtijeva usklađenost relacije preferencije s X, a (6) zahtijeva, grubo govoreći, 'intervalnost' skale (eng. Condition of Order Preservation (COP)). Nužni zahtjev na tok $\mathcal{F}$, da bi uvjet (5) uopće mogao biti zadovoljen, je da relacija na skupu alternativa, koju inducira tok $\mathcal{F}$, bude tranzitivna. To je zahtjev na ordinalnu konzistentnost donositelja odluke. Uvjet konzistentnosti (4) je stroži zahtjev od ordinalne konzistentnosti jer je ordinalna konzistentnost njegova posljedica.

6. KAKO (O)SMISLITI OCJENJIVANJE?

Ne postoje razvijene teorijske osnove (principi) na kojima bi se proces ocjenjivanja zasnivao. Kad bi se takvo nešto poduzimalo onda to ne bi imalo smisla raditi neovisno o samom dizajnu ispitnog sustava. Pojmovi 'limen referencing' i 'slabo kriterijsko referenciranje' su nastali zato da bi opisali trenutačnu praksu a ne neke teoriske koncepte na kojima bi se te procedure trebale zasnivati.

6.1. **Neke nove/stare sugestije.** Postoji li uopće alternativa sadašnjoj praksi? Pollitt i Elliott (2003) [13], [14] sugeriraju da ocjenjivanje, ovakvo kakvo je opisano u ovom članku, može biti zamijenjeno 'širokopojasnim' uspoređivanjem u parovima. Takva razmišljanja postoje već neko vrijeme u literaturi ali je nejasno kako to sprovesti. Uspoređivanje u parovima je zamorno i dugotrajno. Takve bi procedure trebalo moderirati i standardizirati (za interno ocjenjivanje) i standardizirati (za vanjsko ocjenjivanje). Oni sugeriraju da se u takvo uspoređivanje u parovima uključe i testovi ranijih generacija radi održavanja nivoa i kvalitete ocjenjivanja umjesto da se koristi slabo kriterijsko referenciranje.

Kod elektroničkog bodovanja, statističke procedure mogu biti korištene za naknadno određivanje granica među ocjenama. Na primjer, za svako zasebno pitanje potrebno je odrediti granicu za svaku ocjenu i zatim sve te granice uklopiti u granicu za ispitanika. Druga mogućnost je da se pitanja sortiraju po težini i da se odredi koje je najteže pitanje, za svaku ocjenu, na koje ispitanik treba korektno odgovoriti. Prednost takvog pristupa je da ocjenjivači eksperti ne trebaju procjenjivati težinu pitanja, što inače (kako pokazuju istraživanja) rade loše. Mana takvog pristupa je da ne dozvoljava kompenzaciju.

6.2. **Održavanje standarda ili razvijanje sposobnosti?** Greatorex (2003) [8] zaključuje da sva nastojanja u praksi i teoriji, pod nazivnikom 'pravednijeg' ocjenjivanja, imaju u pozadini očuvanje standarda radije nego interpretaciju postignuća (rezultata).

Možda bi naglasak trebao biti na poboljšanju kvalitete svih sudionika u obrazovnom sustavu. U tom smislu bilo bi poželjno osmisliti i kreirati moderno i kvalitetno vrednovanje koje će pokazivati ispitanikove sposobnosti i znanje radije nego ih referencirati prema standardima. To opet kreira nove probleme jer je diskutabilno što znači termin 'sposobnosti'. S druge strane, promjene u obrazovnom sustavu zahtijevaju dodatni napor i za učenike i za nastavnike i čine ga teško razumljivim, ako ni zbog čega onda zbog pomanjkanja kontinuiteta, na primjer. U promjene treba ulaziti s jasnim zahtjevima od zainteresiranih (autoriteta) što je to obrazovni standard, a od profesionalaca ocjenjivača da razmotre valjane i poštene metode vrednovanja i ocjenjivanja. Ono što se pri tome najčešće zaboravlja je tko su sve korisnici ocjena i kakvi su njihovi interesi. Jedna od zabluda, ne samo u hrvatskom obrazovnom horizontu, je da javni ispiti (matura) mogu mjeriti i kvalitetu škola i nastavničkog kadra u njima. Čak se ta 'komponenta' mature ističe kao najvažnija. U svrhu mjerenja kvalitete škola treba razvijati drugačije 'mjerne instrumente'.

7. Državna matura u Hrvatskoj

Čini se da državna matura u Hrvatskoj proživljava 'dječje bolesti' državnih matura zemalja za koje se može reći da imaju više iskustava. Osnovni problem u svakoj zemlji izgleda da je nesporazum između zakonodavca i obrazovne struke. Zakonodavac nema jasne i transparentne namjere a struka nema jasne definicije i stavove. Postoji samo ogromno iskustvo. Dva su otvorena pitanja:

(1) ***Što se maturom mjeri?***
(2) ***Tko, zašto i kako koristi te informacije?***

Pretraživanjem hrvatskog web prostora naišao sam na svega par javnih informacija o državnoj maturi. Jedna je *Pravilnik o polaganju državne mature*, a druga je komentar jednog člana *Povjerenstva za provedbu mature* u javnim glasilima, koji potvrđuje ne samo neslaganje između države i struke nego i nerazumijevanje i/ili nepoznavanje sadašnjeg trenutka u znanstvenoj i obrazovnoj literaturi u svijetu.

7.1. **Iz pravilnika o polaganju državne mature, NN 87/08.**

Članak 2. *Cilj je državne mature provjera i vrjednovanje postignutih znanja i sposobnosti učenika, stečenih obrazovanjem prema propisanim općeobrazovnim nastavnim planovima i programima.*

Članak 10. *Ispitni sadržaji te način provjere i ocjenjivanja znanja i sposobnosti na ispitima uređuju se ispitnim katalozima prema nastavnim planovima i programima iz općeobrazovnih predmeta.*

Koliko je meni poznato katalozi (iz matematike) daju primjere i način bodovanja pojedinih zadataka. Nigdje se ne govori o zaključivanju ocjene na temelju profila.

7.2. **Iz javnih glasila.**

Pitanje: Što će se promijeniti uvođenjem državne mature?

> - Pa prvo više neće biti paušalnih dijeljenja ocjena jer je matura zapravo postupak vanjskog vrednovanja, možemo reći nekakav mjerni uređaj koji će objektivno procijeniti nečije znanje, ali i školu koju je netko pohađao te nastavnike. Rezultati mature zapravo će biti zlatni rudnik informacija.
>
> Vedran Mornar, član *Povjerenstva za provedbu mature* i bivši dekan zagrebačkoga FER-a. (Večernji list, 17. 01. 2010.)

8. THE FINAL TOUCH

Naslov članka je polemički i nakon svega napisanog mijenjam ga u ***Eksterna evaluacija, nužnost i zabluda.*** Riječ 'nužnost' zadržavam jer smatram da eksterna evaluacija na dužu stazu vodi ka kvalitetnijem obrazovnom sustavi, a akcentiram 'zabludu' jer zakonodavac, prema riječima spomenutog člana povjerenstva, pokušava riješiti jednadžbu

$$\text{evaluacija rezultata maturalnog ispita} = \text{evaluacija školstva}.$$

U zabludi su i profesori u srednjim školama koji školske aktivnosti usmjeravaju ka 'uspjehu na maturi' umjesto sticanju znanja i sposobnosti.

LITERATURA

[1] F. G. Ashby and J. T. Townsend. Measurement scales and statistics: the misconception misconceived. *Psych. Bull.*, 96:394–401, 1984.

[2] T. Christie and G. M. Forrest. *Defining Public Examination Standards*. Schools Council Research Studies, MacMillan, London, 1981.

[3] R. Dawes and B. Corrigan. Linear models in decision making. *Psychological Bulletin*, 81:95–101, 1974.

[4] S. French. Statistical and Decision Theoretic Aspects of Examination Assessment. *Trabajos de estadistica*, 4(1):33–66, 1989.

[5] S. French. *Decision theory - An introduction to the mathematics of rationality*. Ellis Horwood, Chichester, 1998.

[6] S. French and M. Vassiloglou. Strength of performane and and examination assessment. *Brit. J. Math. Statist. Psych.*, 39:1–14, 1986.
[7] J. Gaito. Measurement scales and statistics: resurgence of an old misconception. *Psych. Bull.*, 87:564–567, 1980.
[8] J. Greatorex. What happened to limen referencing? an exploration of how the awarding of public examinations has been and might be conceptualised. pages 1–15, Edinburgh, 2003. A paper presented at the BERA 2003.
[9] J. Greatorex. Contemporary GCSE and A-level Awarding: A psychological perspective on the decision-making process used to judge the quality of candidates' work., 2007. A paper presented at BERA 2007.
[10] J. Greatorex. How are archive scripts used in judgements about maintaining grading standards?, 2009. A paper presented at BERA 2009.
[11] D. Laming. *Human judgement The Eye of the Beholder*. Cambridge University Press, Cambridge, 2004.
[12] R. L. Linn, editor. *Educational Measurement*. Oryx Press, Phoenix, USA, 1993.
[13] A. Pollitt and G. Elliott. Finding a proper role for human judgement in the examination system. Technical report, Research and Evaluation Division, University of Cambridge Local Examinations Syndicate, 1 Hills Road, Cambridge, 2003.
[14] A. Pollitt and G. Elliott. Monitoring and investigating comparability: a proper role for human judgement. Technical report, Research and Evaluation Division, University of Cambridge Local Examinations Syndicate, 1 Hills Road, Cambridge, 2003.
[15] F. S. Roberts. *Measurement Theory*. Addison Wesley, Reading, Ma, 1979.
[16] T. L. Saaty. *Fundamentals of the Analytic Hierarchy Process*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15413, 2000.
[17] L. J. Savage. *The Foundations of Statistics*. John Wiley, New York, 1954.
[18] L. Čaklović. Stochastic preference and group decision. *Metodološki zvezki (Advances in Methodology and Statistics)*, 2(1):205–212, 2005.
[19] L. Čaklović and S. Radas. Incentives for industry-science collaboration. U pripremi.
[20] J. von Neumann and O. Morgenstern. *Theory of games and economic behavior. 1st paperback ed. (Repr. of 3rd ed. H/C)*. Princeton University Press, Princeton, 1944. Second edition in 1947, third in 1954.

PMF MATEMATIČKI ODJEL, BIJENIČKA 30, 10 000 ZAGREB, CROATIA
E-mail address: caklovic@math.hr