

Predrag Živković
Fakultet pedagoških nauka
Univerziteta u Kragujevcu
Jagodina
pedjazvk@gmail.com

UDK 373.211.24
159.923.5-053.4
373.23

MEĐUPOSMATRAČKA POUZDANOST I SLAGANJE PROCENJIVAČA – PREDŠKOLSKIH VASPITAČA U PROCENI SOCIJALNIH ODNOSA U GRUPI

Apstrakt. U strukturisanom posmatranju procenjivači izveštavaju o ponašanju koje je vidljivo, ali uvek procenjuju iz sopstvene i jedinstvene perspektive. Subjektivna komponenta procene (za razliku od objektivne i kontekstualne) izražena je idiosinkrazijom procenjivača; svaki procenjivač učestvuje sopstvenim predrasudama i interpretacijama u proceni. U ovom radu, koji je deo šireg istraživanja problema multiperspektivnosti etnografskih istraživača i anamorfičkog prigovora relijabilnosti naturalističkih istraživanja, prikazani su rezultati ispitivanja međuposmatračke pouzdanosti i saglasnosti procenjivača – predškolskih vaspitača u proceni socijalnih odnosa u vršnjačkoj grupi. Cilj istraživanja je bio da se utvrdi stepen slaganja procenjivača u proceni socijalnih odnosa u grupi. Na uzorku 28 vaspitača i 138 dece predškolskog uzrasta primenjen je modifikovan Protokol za posmatranje socijalnih odnosa (OIM02) od 15 deskriptora (sa 2 i 5 stepeni procene; za izračunavanje *Cohen's* kappa koeficijenta i interklasnog koeficijenta korelacije ICC, respektivno). Četrnaest parova vaspitača procenilo je po 10 istih članova svojih grupa. Dobijene su vrednosti dobre interne konzistencije (*Cronbach* $\alpha = 0.83$), zadovoljavajuće međuposmatračke pouzdanosti (*Cohen's* kappa = 0.59) i odlične saglasnosti između procenjivača izražene interklasnim koeficijentom korelacije ICC = 0.80).

Ključne reči: međuposmatračka pouzdanost, slaganje procenjivača, kappa statistika, predškolski vaspitači.

UVOD

Indeks *pouzdanosti* pokazuje preciznost instrumenta, on se odnosi na psihometrijski kapacitet instrumenta da proizvodi konstantno slične ili iste rezultate. Postoje različite mogućnosti da se pouzdanost meri: između procenjivača koji procenjuju iste učesnike (međuposmatračka pouzdanost) ili kroz različite vremenske tačke procene istih učesnika (test-retest pouzdanost) (Borsboom et al., 2004).

Procena pouzdanosti često je ograničena samo na rezultat linearne korelacije, pa nedostaje jasno razumevanje metodoloških pristupa, što limitira interpretativnost i komparativni potencijal izveštaja o rezultatima.

Za razliku od valjanosti procene socijalnih odnosa i veština (kompetencija) dece predškolskog uzrasta, o *pouzdanosti* (pouzdanosti procene) procenjivača se retko izveštava. Autori se uglavnom oslanjaju na prikaze i interpretaciju linearne korelacije između sirovih skorova dobijenih od različitih procenjivača (Houwer et al., 2005; Vagh et al., 2009). U tako sačinjenim izveštajima prikazuju se rezultati istraživanja u kojima su dobijene umerene vrednosti korelacija, koje variraju između $r = 0.30$ i $r = 0.60$. Ove korelacije su slične onima dobijenim za roditelji - vaspitači i roditelj - roditelj procenjivačke parove (Janus, 2001; Norbury et al., 2004; Bishop et al., 2006; Massa et al., 2008; Gudmundsson and Gretarsson, 2009; Kosh et al., 2011).

U istraživanjima i proceni socijalnih kompetencija i ponašanja dece predškolskog uzrasta (McCabe and Altamura, 2011) *meta i komparativne analize* pokazuju:

Tabela 1. Pregled korišćenih instrumenata u proceni socijalnog ponašanja i postignute vrednosti međuposmatračke pouzdanosti.

INSTRUMENT	KARAKTERISTIKE INSTRUMENTA	VREDNOSTI MEĐUPOSMATRAČKE POUZDANOSTI
Achenbach System of Empirical Based Assessment – Preschool Module (Achenbach & Rescorla, 2000, Achenbach, Dumenci & Rescorla, 2003).	CBCL/1.5–5 (The Child Behavior Checklist) i C-TRF (Caregiver – Teacher Report Form) za procenu adaptivnog funkcionisanja na 99 stavki, opis tri šira sindroma: internalizacija, eksternalizacija i totalitet problema	Dobijene su, za CBCL/1.5–5, vrednosti međuposmatračke pouzdanosti procenjivačkih parova roditelji-roditelji (.51–.67) i vaspitači-vaspitači (.38–.50). Za C-TRF dobijene vrednosti međuposmatračke pouzdanosti vaspitača .21–.79.
Behavior Assessment System for Children (Reynolds & Kamphaus, 2004)	Korišćene su TRS (Teacher Rating Scales) i PRS (Parent Rating Scales) za procenu problema adaptacije i ponašanja u grupi dece predškolskog uzrasta (2 do 5 godina). Koristi procenu na 139 stavki i dve skale: funkcionalna komunikacija i socijalne veštine.	Dobijene su umerene vrednosti test-retest pouzdanosti i međuposmatračke pouzdanosti.
Devereux Early Childhood Assessment Program – DECA i Clinical Form DECA-C) (LeBuffe & Naglieri, 2003)	Korišćeni za procenu zdravog socijalnog i emocionalnog rasta. Na 37 stavki petostepene Likertove skale meri: inicijativu, samokontrolu i socijalno povezivanje.	Za DECA međuposmatračka povezanost .59–.77; za DECA-C međuposmatračka pouzdanost procene parova vaspitača .32–.77.

Early Coping Inventory (Zeitlin, Williamson & Szczepanski, 2007)	Na 48 stavki meri ponašanje u tri kopirg klastera: senzomotorna organizacija, reaktivno ponašanje i samo-inicirano ponašanje.	Dobijene su umerene vrednosti test-retest pouzdanosti i međuposmatračke pouzdanosti na uzorku obučanih vaspitača.
Early Screening Project (Walker, Severson & Feil, 1995)	Identifikuje decu sa rizikom problema prilagođavanja i obrascem povlačenja.	Za uzorak vaspitača-procenjivača dobijene su prihvatljive vrednosti test-retest pouzdanosti i međuposmatračke pouzdanosti.
Eyberg Child Behavior Inventory i njegova revidirana verzije SESBI-R (Eyberg & Pincus, 1999)	Na 38 ajtema skale Likertovog tipa i procene da/ne za svaki ajtem	Za osnovnu verziju vrednosti međuposmatračke pouzdanosti od .61–.69, i za revidiranu verziju (vaspitač-vaspitač) vrednosti međuposmatračke pouzdanosti od .43–.86.
Infant/Toddler Social Emotional Assessment ITSEA + BITSEA (Briggs-Gowan & Carter, 2006)	Na 166 i 42 stavki meri socijalno-emocionalne probleme i identifikuje potrebu za planiranjem intervencija.	Na osnovnoj verziji, sa trostepenom skalom Likertovog tipa, dobijene su vrednosti međuposmatračke pouzdanosti (vaspitač-vaspitač) od .24–.66. Na BITSEA verziji (vaspitač-vaspitač) dobijene su vrednosti međuposmatračke pouzdanost od .28–.59.
Preschool and Kindergarten Behavior Scale (2nd ed., PBKS-2 Merrell, 1994)	Skala se sastoji od 34 stavke kojima meri socijalne veštine i 42 stavke za skalu problema u ponašanju dece predškolskog uzrasta.	Dobijene su umerene vrednosti međuposmatračke pouzdanosti (vaspitač-vaspitač).
Competence and Behavior Evaluation (SCBE) Scale (LaFraniere & Dumas, 1995)	Na 80 stavki, sa 6-stepenom skalom Likertovog tipa, meri socijalne kompetencije, afektivnu ekspresivnost i adaptaciju.	Unutrašnja konzistentnost: .79–.91. Vrednosti međuposmatračke pouzdanosti od .72–.89.
Social Skills Improvement System-Rating Scale SSIS-RS (Gresham & Elliott, 2008)	Na 4-stepenoj skali Likertovog tipa meri tri oblasti: socijalne veštine, probleme u ponašanju i akademsku kompetentnost.	Unutrašnja konzistentnost i vrednosti test-retest pouzdanosti (vaspitač-vaspitač) su dobre.

Iako ova korelaciona analiza (uglavnom Personova) daje informacije o veličini relacija između dve grupe vrednosti, ona nikako ne objašnjava detaljnije slaganje između procenjivača (Blind and Altman, 2003; Kottner et al., 2011). Pored toga, rezultati o međuposmatračkoj *pouzdanosti* (inter-rater reliability) i *slaganju* (inter-rater agreement) često su isključeni iz korelacione analize (Bishop and Baird, 2001; Van Noord and Prevatt, 2002). Ovo je zbog toga što savršena linearna korelacija može biti postignuta i ukoliko se jedna grupa procenjivača sistematski isto razlikuje (skoro konzistentnim iznosom) od druge grupe, čak i kada nijedno apsolutno slaganje ne postoji.

Slaganje se postiže samo kada tačke poklapanja leže na liniji ili unutar oblasti jednakosti obe procene (Blind and Altman, 1986; Liao et al., 2010). Tako, analize koje se oslanjaju samo i jedino na linearnim korelacijama i koje ne obezbeđuju meru i merenje saglasnosti procenjivača nisu dovoljne za potpunu procenu međuposmatračke pouzdanosti. Kao što je Stemler naglasio (Stemler, 2004), pouzdanost nije pojedinačan i poseban, unitaran koncept i ne može biti dosegnuta i obezbeđena samo dokazima i evidencijom o lineranoj korelaciji.

Cilj ovog rada je pokazati kako se ova tri koncepta međuposmatračke pouzdanosti: koeficijent interklasne korelacije (ICC) (Liao et al., 2010; Kottner et al., 2011), slaganje procenjivača i linearna korelacija (Personovog koeficijenta korelacije), dopunjuju kod *redundantnosti* procenjivača (Stemler, 2004).

Zaključci dobijeni na osnovu procena različitih procenjivača ili u procenama u različitim vremenskim tačkama su relevantni za veliki broj naučnih (sub)disciplina u kojima se sposobnosti, ponašanje i simptomičesto procenjuju, vrednuju i upoređuju. U nameri da se utvrdi stepen slaganja između procenjivača, važno je da ovde razmotrimo tri različita aspekta:

1. *Međuposmatračka pouzdanost* (inter-rater reliability), kojom procenjujemo do kog nivoa su korišćene mere i merenja u stanju da diferenciraju učesnike sa različitim nivoima sposobnosti u situaciji kada ih ocenjuju različiti procenjivači. Merenja međuposmatračke pouzdanosti mogu, takođe, da posluže da se odredi najmanji obim divergencije dva skora koja su neophodna da bi se uspostavila pouzdana razlika.
2. *Međuposmatračka saglasnost* (inter-rater agreement) uključuje proporciju apsolutnog slaganja procenjivača, koji je primenljiv za utvrđivanje veličine i smera razlike.
3. *Veličina (stepen) povezanosti* između procena, merena linearnom korelacijom.

Autori sa polja istraživanja obrazovanja (Brown et al., 2004; Stemler, 2004) i biheviornalne psihologije (Mitchell, 1979) takođe su naglasili potrebu i neophodnost da se jasno istaknu razlike posebnih aspekata u merenju odnosa procene i njihovog doprinosa proceni slaganja i pouzdanosti. Precizne definicije i distinkcije konceptata štite nas od mogućih stranputica u interpretaciji podataka. Kako se različiti ali komplementarni koncepti, slaganje, korelacija i međuposmatračka pouzdanost često poistovećuju (mešaju) i koriste naizmenično (sinonimno) (Van Noord and Prevatt, 2002; Massa et al., 2008), pokušaćemo da ponudimo precizne definicije i metodološku osnovu (zasnovanost).

Slaganje se odnosi na stepen i nivo identičnosti procena (Vet et al., 2006; Shoukri, 2010), često opisano i kao proporcija identičnosti kod različitih parova procenjivača (Kottner et al., 2011). U nameri da se utvrdi kada se dve procene statistički razlikuju (jedna od druge, međusobno), psihometrijski aspekti upotrebljenih instrumenata, kao što je pouzdanost (npr. test-retest pouzdanost ili interklasni koeficijent korelacije kao mera međuposmatračke pouzdanosti), moraju se ozbiljno razmotriti. Opšte karakteristike instrumenta, npr. valjanost kategorija skorovanja (Jonsson and Svingby, 2007) ili broj pojedinačnih ajtema (stavki) koji obuhvata skor, utiće direktno na verovatnoću apsolutnog slaganja. Na primer, veći broj stavki koje pokriva sirovi skor može značiti manju verovatnoću da se dostigne apsolutno slaganje skorova. Dakle, dva sirova skora ili dve standarizovane vrednosti (kao što je t-skor) koje se razlikuju u apsolutnom broju nisu nužno i statistički različita. Razlika (apsolutna) može biti suviše mala da bi mogla da izrazi sistematsku razliku u odnosu na distribuciju skorova. Tako, veličina nesistematske greške mora biti uzeta u razmatranje pre donošenja suda o proporciji slaganja.¹

U nameri da pokušamo da obezbedimo adekvatnu populaciono-specifičnu procenu pouzdanosti u ovom istraživanju, računacemo međuposmatračku pouzdanost izraženu kao interklasni koeficijent korelacije (ICC). Interklasni koeficijent korelacije pokazuje stepen do kojeg su mere u stanju da naprave razliku između participanata sa različitim skorovima, izraženu od dva ili više procenjivača koji postižu sličnosti pri korišćenju istog instrumenta (Liao et al., 2010; Kottner et al., 2011). Interklasni koeficijent korelacije uzima u obzir varijansu procena za jedno dete koje procenjuju dva procenjivača, kao i varijansu između kompletnih grupa dece. Može tako poslužiti da se uporedi pouzdanost procena između dve grupe procenjivača i da se proceni pouzdanost instrumenta korišćenog u istraživanju (Bockmann and Kiese-Himmel, 2006).

Kao što smo naglasili, izračunati koeficijenti korelacije su najčešće korišćeni statistici u utvrđivanju međuposmatračke pouzdanosti. Međutim, linearna korelacija nam ne daje odgovarajuće informacije o slaganju procena. Ona daje korisne informacije o *odnosu dve varijable*. U specifičnom slučaju, kada koristimo koeficijente korelacije kao indirektnu meru konzistentnosti procene, te se linearna povezanost može očekivati – Personova korelacija je odgovarajući statistički pristup. Ne bi trebalo da služi samo kao jedina mera međuposmatračke pouzdanosti, već i kao mera veličine linearne asocijacije. Ima dodatnih prednosti jer omogućuje upoređivanje,

¹ Nažalost, u mnogim istraživanjima u kojima se meri slaganje procenjivača, potpuno se ignoriše distinkcija *apsolutne* razlike i *statistički pouzdane* razlike i ne koriste se standardizovane vrednosti (Bishop and Baird, 2001; Bishop et al., 2006; Gudmundsson and Gretarsson, 2009).

koje je korisno kada ispitujemo razlike između grupa u vezi sa stepenom povezanosti procena.

METOD

Ispitanici na kojima je primenjena Skala za procenu socijalnih odnosa u grupi su vaspitači iz predškolskih ustanova sa teritorije šumadijsko-pomoravskog regiona (N = 28) i deca predškolskog uzrasta (N = 138). U studiju je uključeno 48% vaspitača sa više od 10 godina radnog staža, 39,5% vaspitača sa više od 5 a manje od 10 godina radnog staža i 12,5% njih sa manje od 5 godina radnog iskustva.

Skala procene socijalnih odnosa u grupi dece predškolskog uzrasta modifikovana je verzija Protokola za posmatranje aktivnosti u vrtiću: praćenje društvenih relacija iz *Vodiča za samoo vrednovanje u predškolskim ustanovama* (Observation Items / High Scope Child Observation Record for Infant and Toddlers, 2002; prema Bennett, 2013: 68). Protokol je pretočen u skalu sa 13 stavki, na kojima su parovi vaspitača koji rade sa istom grupom dece procenjivali po 10 istih članova grupe. Skala je Likertovog tipa sa dvostepenim i petostepenim procenama stavki. Dvostepena (kategorijalna) procena je uvrštena zbog neophodnog uslova za izračunavanje Koenovog kapa koeficijenta (*Cohen kappa*). Rezultati sa petostepene skale korišćeni su zbog izračunavanja interklasnog koeficijenta korelacije (ICC).

Pouzdanost Skale procene socijalnih odnosa u grupi ispitali smo na dva načina: ocenom interne konzistencije, izračunavanjem alfa-koeficijenta (*Cronbach alpha*) i ocenom međuposmatračke pouzdanosti, izračunavanjem *kappa* koeficijenta. Konstrukcionu validnost ispitali smo faktorskom analizom da bismo dobili podatke o dimenzijama i strukturi skale.

REZULTATI

Uopšteno govoreći, instrument pokazuje zadovoljavajuće merne karakteristike. Pouzdanost je određivana Krombahovim alfa koeficijentom. Unutrašnja pouzdanost korišćenog upitnika je preko .80 (Krombahov alfa koeficijent $\alpha = .837$, sa standardizacijom $\alpha = .853$).

Tabela 2. *Interna konzistencija skale*

Krombahova alfa	Krombah alfa – stand. stavki	Spirman-Braun	Gutman	N
.837	.857	.838	.821	13

Faktorska analiza izdvaja 4 faktora koji objašnjavaju 74.58% varijanse u procenama (za društvene nauke zadovoljavajuće iznad 60%), odnosno odgovori na svim stavkama su (manje ili više) povezani. KMO (Kaiser-Meyer-Olkin) test (faktorabilnosti) pokazao je vrlo dobru vrednost (KMO = 0.78). Bartleov test sferičnosti dostigao je statističku značajnost na nivou $r < 0.001$ ($r = 0.000$). Podaci ukazuju na *faktorabilnost* matrice i *opravdanost* faktorske analize.

Podaci pokazuju da je zadovoljen opšti Terstonov kriterijum komunaliteta, jer su sve vrednosti komunaliteta *veće od preporučenih* $h = 0.30$. Faktorskom analizom (primenom metod glavnih komponenti uz Varimax rotaciju osa) izdvojena je četvorofaktorska interpretabilna struktura kojom je objašnjeno 74.58% varijanse. Izbor četiri faktora izvršen je na osnovu Kajzer-Gutmanovog kriterijuma i scree plota (eigenvrednosti > 1).

Tabela 3. Procenat objašnjene varijanse

Komponente	Ejgenvrednosti	% varijanse	Kumulativni %
1.	5.177	39.822	39.882
2.	1.945	14.963	54.785
3.	1.346	10.382	65.137
4.	1.228	9.446	74.583

Rezultati po stavkama pokazuju umerene vrednosti svih korišćenih statistika. Vrednosti *kappa* koeficijenata međuposmatračke pouzdanosti se kreću od 0.40 do 0.97. Ukupna vrednost *kappa* koeficijenta za celu skalu iznosi $\kappa = 0.597$ i statistički je značajna.

Tabela 4. Analiza stavki sa *kappa* vrednostima

Stavka	AS	SD	Kappa	Ekstr.
Da li dete inicira, uživa ili postaje uplašeno u igrama jurenja, skrivanja, šale...?	3.15	1.40	0.73	.791
Da li dete traži vaspitača i da li komunicira o svojim potrebama i željama?	3.98	0.98	0.96	.722
Da li se dete oseća dovoljno sigurno da spontano potraži novu odraslu osobu i ponudi mu knjigu, igračku ili predmet?	3.44	1.04	0.91	.765
Da li dete spontano donosi objekat i pokazuje ga drugom detetu?	4.19	0.87	0.95	.695
Da li grli, potapšava, nudi pomoć drugom detetu?	3.41	1.31	0.40	.792
Da li dete koristi ime drugog deteta?	3.81	1.45	0.92	.842
Da li dete priča direktno sa drugim detetom, poziva ga da se igraju ...?	4.04	1.04	0.61	.748

Da li dete pokazuje zadovoljstvo kada završi aktivnost – pokazuje zadovoljstvo tako što nešto kaže, pljeska rukama, skače, zove nekog da dođe i pogleda...?	3.91	0.90	0.93	.517
Da li dete etiketira osećanje rečju ili frazom?	3.35	0.98	0.97	.725
Da li dete donosi utešan predmet onome ko je potresen?	2.95	1.11	0.58	.827
Da li dete koristi reči ili fraze da priča o emocijama koje je pokazalo drugo dete?	3.01	0.89	0.41	.713
Da li dete počinje da inicira i uživa u igrama sa drugima (krije predmet od druge osobe kako bi ga ona našla ili beži od druge osobe kako bi ga uhvatila)?	3.73	0.77	0.95	.815
Da li dete posmatra i isprobava nešto od onoga što radi dete koje se blizu njega igra?	3.75	0.77	0.92	.743
Total Kappa: 0.597				

Tabela 5. Ukupne vrednosti kappa

	Vrednost	Standardna greška	Aproks. T	Značajnost
Kappa ukupno	.597	.110	6.814	.000

Vrednosti interklasnog koeficijenta korelacije su niže (od očekivanih) za pojedinačno merenje ICC = 0.243, .178 za nižu i .328 za višu granicu) i zadovoljavajuće za prosečno merenje (ICC = 0.807, .738 za nižu i .864 za višu granicu). Na oba merenja su dobijene statistički značajne vrednosti korelacije. Dobijena vrednost Personovog koeficijenta korelacije je $r = .598$.

Tabela 6. Interklasni koeficijent korelacije

	Interklasna korelacije	Niža granica	Viša granica	F	Sig
Pojedinačno merenje	0.243	.178	.328	5.173	.000
Prosečno merenje	0.807	.738	.864	5.173	.000

DISKUSIJA

U ispitivanju pouzdanosti inicijalnog konstrukta dobili smo zadovoljavajuću vrednost alfa koeficijenta pouzdanosti, a komunaliteti pokazuju vrednosti koje ukazuju na faktorabilnost matrice i opravdanost faktorske analize. Dobijena četvorfaktorska struktura objašnjava skoro 75% varijanse. Ovde nije interpretirana struktura faktora, ali dobijene vrednosti ukazuju na interpretabilnost.

Dobijene vrednosti koeficijena međuposmatračke pouzdanosti slažu se sa rezultatima istraživanja u kojima su ispitivani socijalni odnosi u preškolskoj grupi dece, na sličnim uzorcima ispitanika i stavki (Tabela 1): Achenbach System of Empirical Based Assessment – Preschool Module (Achenbach & Rescorla, 2000, Achenbach, Dumenci & Rescorla, 2003) dobijene su, za CBCL/1.5–5, vrednosti međuposmatračke pouzdanosti procenjivačkih parova roditelji – roditelji (.51-.67) i vaspitači – vaspitači (.38-.50) za C-TRF dobijene vrednosti međuposmatračke pouzdanosti vaspitača .21-.79; Devereux Early Childhood Assessment Program – DECA i Clinical Form DECA-C (LeBuffe & Naglieri, 2003) za DECA međuposmatračka povezanost .59-.77.; za DECA-C međuposmatračka pouzdanost procene parova vaspitača .32-.77; Infant/Toddler Social Emotional Assessment ITSEA + BITSEA (Briggs-Gowan & Carter, 2006) na osnovnoj verziji, sa trostepenom skalom Likertovog tipa, dobijene su vrednosti međuposmatračke pouzdanosti (vaspitač – vaspitač) od .24-.66. na BITSEA verziji (vaspitač – vaspitač) dobijene su vrednosti međuposmatračke pouzdanosti od .28-.59.

Iako standardi za procenu vrednosti međuposmatračke pouzdanosti nisu u potpunosti usaglašeni, pa ih različiti autori različito interpretiraju, možemo primetiti da se naši rezultati mogu svrstati u grupu onih u kojima je dobijena vrednost umerene međuposmatračke pouzdanosti.

Mali broj stavki mogao bi ukazati na nedostatak pri izračunavanju odgovarajućih koeficijena pouzdanosti i korelacije, ali taj nedostatak se ne pokazuje kao odlučujući (što pokazuje faktorska analiza skale).

ZAKLJUČAK

Iako se procene posmatrača suštinski oslanjaju na profesionalni sud i ocenu, tu je i upitanost o tome koliko procena zavisi od procenjivača, pre nego od posmatranih performansi. Praktičari i istraživači često navode da je merenje konzistentnosti procena i sudova procenjivača i evaluatora o osobama ili objektima – međuposmatračka pouzdanost. Ovo široko podrazumevanje i korišćenje može „maskirati“ važnu tehničku distinkciju između međuposmatračkog slaganja i međuposmatračke pouzdanosti.

Komentari o subjektivnoj i idiosinkratičnoj prirodi istraživanja koja se zasnivaju na opažanju učesnika, u naturalističkom istraživačkom kontekstu, javljaju se uz pitanje: Kako znamo da su rezultati tog istraživanja primenjivi na druge situacije? Neki autori kao odgovor na pitanje predlažu triangulaciju izvora podataka i metodologije, kao i da pojam „verodostojnosti“ zameni uobičajeno gledanje na pouzdanost, i da se taj

pojam prenese na pitanja *mogućnosti verovanjama, potvrđivanja, transfera i pouzdanja*.

Ako je opservacijsko istraživanje po prirodi strukturisanije i daje kvantitativne podatke, tada se primenjuju pravila pouzdanosti tipa slaganja između višestrukih procena jednog procenjivača i pouzdanost tipa slaganja između različitih procenjivača. Multiperspektivnost u naturalističkim istraživanjima, sa procenjivačima koji svako iz svog ugla opisuje i vrednuje situaciju, može biti obezbeđena verodostojno samo kada se istraživanja dopune procenom međuposmatračke pouzdanosti.

LITERATURA

- Bennett, M., Madigan, I., Radulović, L., Miškeljin, L. (2013). *Vodič za samovrednovanje u predškolskim ustanovama*. Beograd: MNTRP.
- Bishop, D. V., Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: use of the children's communication checklist in a clinical setting. *Dev. Med. Child Neurol.*, 43(809–818).
- Bishop, D. V. M., Laws, G., Adams, C., Norbury, C. F. (2006). High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behav. Genet.*, 36(173–184).
- Blind, J. M., Altman, D. G. (2003). Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet. Gynecol.*, 22(85–93).
- Bland, M. J., Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327(307–310).
- Bockmann, A. K., Kiese-Himmel, C. (2006). *ELAN – Eltern Antworten: Elternfragebogen zur Wortschatzentwicklung im frühen Kindesalter*, 1st Edn. Göttingen: Hogrefe.
- Borsboom, D., Mellenbergh, G. J., Van Heerden J. (2004). The concept of validity. *Psychol. Rev.*, 111(1061–1074).
- Brown, G. T., Glasswell, K., Harland, D. (2004). Accuracy in the scoring of writing: studies of reliability and validity using a new Zealand writing assessment system. *Assess. Writ.*, 9(105–121).
- Houwer, A., Bornstein, M. H., Leach, D. B. (2005). Assessing early communicative ability: a cross-reporter cumulative score for the macarthur cdi. *J. Child Lang.*, 32(735–758).
- Graham, M. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Rating*. Center for Educator Compensation Reform. CECR: University of Wisconsin.

- Gudmundsson, E., Gretarsson, S. J. (2009). Comparison of mothers' and fathers' ratings of their children's verbal and motor development. *Nordic Psychol.*, 61(14–25).
- Janus, M. (2001). *Validation of a Teacher Measure of School Readiness with Parent and Child-Care Provider Reports*. Department of Psychiatry Research Day, Canadian Centre for Studies of Children at Risk.
- Jonsson, A., Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Edu. Res. Rev.*, 2(130–144).
- Koch, H., Kastner-Koller, U., Deimann, P., Kossmeier, C., Koitz, C., Steiner, M. (2011). The development of kindergarten children as evaluated by their kindergarten teachers and mothers. *Psychol. Test Assess.*, Model. 53 (241–257).
- Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartssonm A. et al. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int. J. Nurs.*, Stud. 48(661–671).
- Liao, S. C., Hunt, E. A., Chen, W. (2010). Comparison between inter-rater reliability and inter-rater agreement in performance assessment. *Annal. Acad. Med.*, 39 (613–625). Singapore.
- Maassen, G. H. (2004). The standard error in the Jacobson and Truax reliable change index: the classical approach to the assessment of reliable change. *J. Int. Neuropsychol. Soc.*, 10 (888–893).
- Massa, J., Gomes, H., Tartter, V., Wolfson, V., Halperin, J. M. (2008). Concordance rates between parent and teacher clinical evaluation of language fundamentals observational rating scale. *Int. J. Lang. Commun. Disord.*, 43 (99–110).
- McCabe, P., Altamura, M. (2011). Empirically valid strategies to improve social and emotional competence of preschool children. *Psychology in the School*, Vol.48 (5) (513–539).
- Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychol. Bull.*, 86(376–390).
- Norbury, C. F., Nash, M., Baird, G., Bishop, D. V. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist – 2. *Int. J. Lang. Commun. Disord.*, 39(345–364).
- Shoukri, M. M. (2010). *Measures of Interobserver Agreement and Reliability*, 2nd Edn. Boca Raton, FL: CRC Press.
- Shrout, P. E., Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.*, 86(420–428).
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Pract. Assess. Res. Eval.*, 9 (66–78).
- Stolarova, M., Wolf, C., Rinker, T., Briemann, A. (2014). How to assess and compare inter-rater reliability, agreement and correlation of ratings: an exem-

- plary analysis of mother-father and parent-teacher expressive vocabulary rating pairs. *Front Psychology*, June, Vol.4 (481–509).
- Vagh, S. B., Pan, B. A., Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: the utility of combining parent and teacher report. *Child Dev.*, 80(1545–1563).
- Van Noord, R. G., Prevatt, F. F. (2002). Rater agreement on iq and achievement tests: effect on evaluations of learning disabilities. *J. School Psychol.*, 40(167–176).
- Zahra, D., Hedge, C. (2010). The reliable change index: Why isn't it more popular in academic psychology. *Psychol. Postgrad. Aff. Group Q.*, 76 (14–19).

Predrag Živković

INTER-RATER RELIABILITY AND AGREEMENT BETWEEN RATERS – PRESCHOOL EDUCATORS IN ASSESSMENT OF SOCIAL RELATIONS IN THE GROUP OF CHILDREN

Abstract. In the structured observation raters report on behavior that is visible, but always valued through their own and unique perspective. The subjective component of assessment (as opposed to objective and contextual) is expressed by idiosyncrasies of evaluators; each rater participating with own biases and interpretations in the judgment. In this study, which is part of a wider investigation of the problem of multiperspectivity of ethnographic researchers and anamorphic complaints of reliability in naturalistic research, author present results of testing of inter-rater reliability and agreement between raters - preschool teachers in the assessment of social relations in the group of preschool children. The aim of this study was to determine the degree of agreement in the assessment of social relationships in the group. In a sample of 28 teachers and 138 preschool children, a modified protocol for observation of social relations (OIM02) was applied on 15 descriptors (with 2 and 5 degrees assessment; for calculating Cohen's kappa coefficient and interclass correlation coefficient ICC, respectively). Fourteen pairs of teachers judged 10 members of their group. The values of good internal consistency (Cronbach $\alpha = 0.83$), satisfactory inter-rater reliability (Cohens kappa = 0.59) and excellent agreement between raters expressed by interclass correlation coefficient (ICC = 0.80) was obtained.

Key words: inter-rater reliability, raters' agreement, kappa statistic, preschool teachers.