

Efekti metode - analiza odabranih karakteristika grupe fenomena koji utječu na rezultate anketnih istraživanja

Mandić, Luka

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:131:116865>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom](#).

Download date / Datum preuzimanja: **2024-07-14**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)



Diplomski rad

Luka Mandić

Mentorica: doc. dr. sc. Ksenija Klasnić

**Efekti metode - analiza odabranih karakteristika grupe fenomena koji utječu
na rezultate anketnih istraživanja**

Diplomski rad

Zagreb, svibanj 2021.

Sadržaj

1. Uvod	3
2. Formulacija iskaza	6
2.1. Mjerenje utjecaja formulacije iskaza u istraživanjima	8
3. Nemarno odgovaranje	14
3.1. Detektiranje nemarnog odgovaranja.....	15
4. Priistranost slaganja	17
4.1.Mjerenje pristranosti slaganja.....	18
5. Nacrt istraživanja	20
6. Ciljevi rada	20
7. Studija 1	21
7.1. Metode	21
7.2.Rezultati i diskusija	23
7.3.Kontroliranje utjecaja nemarnog odgovaranja	26
7.4.Zaključak i prijedlozi.....	28
8. Studija 2	31
8.1.Metode	31
8.2.Rezultati.....	36
8.3.Razlike u nemarnom odgovaranju s obzirom na sociodemografske karakteristike	39
8.4.Dimenzionalnost skale suvremenog seksizma	41
8.5.Kontroliranje utjecaja nemarnog odgovaranja i pristranosti slaganja	41
9. Sveukupni zaključak i prijedlozi	52
10. Literatura	54
Sažetak	62

1. Uvod

Klasična teorija testova je tradicionalni pristup unutar kvantitativne metodologije koji se koristi prilikom testiranja pouzdanosti i validnosti neke skale koja je sastavljena od određenog broja čestica (Cappelleri, Lundy i Hays, 2014). Ona svoju definiciju uvelike duguje Novickom (1966), a njene opise moguće pronaći i u Lord i Novick (1968) te u Allen i Yen (1979). Prema klasičnoj teoriji testova, rezultat dobiven mjerenjem u sebi inherentno sadrži određenu komponentu pogreške. Sukladno tome, rezultat koji je dobiven nakon što ispitanik popuni test odnosno ispuni upitnik naziva se bruto rezultat. Bruto rezultat se sastoji od dva djela: pravog rezultata i komponente pogreške.

$$X_b = X_t + X_e$$

Pravi rezultat (engl. *true score*), kao što mu i ime govori, predstavlja stvarni rezultat na nekom testu ili skali. Kada bi bila riječ o savršenim uvjetima i kada bi bilo moguće eliminirati pogrešku mjerenja, osigurali bismo pravi rezultat te bismo upravo njega koristili umjesto bruto rezultata (Husremović, 2016). No, zbog prisutnosti pogreške mjerenja, pravom rezultatu potrebno je pripisati upravo i pogrešku mjerenja. Pogreška mjerenja dijeli se na dva tipa: sustavnu i slučajnu pogrešku. Sustavna pogreška označuje efekte koji utječu na rezultate mjerenja varijabli, a njihov izvor se većinom nalazi u samome mjernom instrumentu ili načinu mjerenja. Na primjer, do sustavne pogreške prilikom mjerenja tjelesne temperature može doći na minimalno dva načina:

1. Ako tjelesnu temperaturu mjerimo uređajem koji je iz nekog razloga neispravan (npr. neispravnost u toplinskom senzoru)
2. Ako tjelesnu temperaturu mjerimo na neispravan način (ako termostat držimo u zraku, a ne ispod pazuha)

Ako slijedimo logiku iz ova dva primjera, zaključujemo da sustavna pogreška uvijek mijenja vrijednost rezultata u istome smjeru. Na primjer, toplinski senzor može biti neispravan te on može sustavno podcjenjivati ili precjenjivati tjelesnu temperaturu svakog ispitanika, pod uvjetom da neispravnost ne mijenja svoju prirodu iz mjerenja u mjerenje. Samim time, one će svakim mjerenjem pokazivati višu/nižu tjelesnu temperaturu od one prave ili ispravne. Isto tako, držanje

termostata u zraku većinom će rezultirati registracijom nižeg rezultata od onog koji bi se dobio ispravnim mjerenjem tjelesne temperature, osim ako se ne nalazimo u iznimno toploj klimi.

Nasuprot sustavne pogreške, izvor slučajne pogreške nepoznat je i nepredvidiv, a sama slučajna pogreška varira od mjerenja do mjerenja. Takvi nesustavni faktori predstavljaju varijable koje slučajno djeluju te rezultate mijenjaju u oba smjera s različitim intenzitetom. Slučajne pogreške često slijede Gaussoliku normalnu distribuciju. Kada bismo mjerenje ponovili veliki broj puta, slučajna pogreška precjenjivanja i podcjenjivanja međusobno bi se gotovo poništile te bi se bruto rezultat približio pravom rezultatu (Husremović, 2016). Zbog ovoga, moguće je definirati još jednu bitnu pretpostavku klasične teorije testova: predmet mjerenja stabilan je kroz vrijeme. Ako je predmet mjerenja zaista stabilan kroz vrijeme, ako eliminiramo sustavnu pogrešku i ako mjerenje izvršimo dovoljno veliki broj puta, dobiveni rezultat trebao bi biti legitiman i validan. Upravo se na ovim pretpostavkama temelji kvantitativna analiza i sukcesivna interpretacija rezultata koji su prikupljeni koristeći metodu ankete. Uspoređujući opis slučajne i sustavne pogreške, postaje jasno da bi se prilikom konceptualizacije i operacionalizacije znanstvenih istraživanja istraživači trebali fokusirati na eliminiranje sustavnih pogrešaka. U teoriji, sustavne pogreške trebale bi biti jasno uočljive, ali pod pretpostavkom da je priroda efekta sama po sebi jednostavna (Harris i Smith, 2009). No, postoji dodatna grupa efekata koji neminovno utječu na rezultate anketnih istraživanja, koji po svojoj prirodi/definiciji ne spadaju u sustavne pogreške, ali im po svojim utjecajima slične, a to su efekti metode (engl. *method effect*). Efekti metode su skup pojava koji predstavljaju neželjenu varijancu u prikupljenim podacima, nezavisnu od mjernog instrumenta (Maul, 2013). Ovakvi efekti često se javlja u skalama samoprocjene, koje su poznate po tome da su podložne mjernim pristranostima (Podsakoff, Mackenzie i Podsakoff, 2012).

Kako bismo ilustrirali utjecaje efekata metode koji su po svojoj prirodi slični sustavnim pogreškama, iskoristit ćemo primjer mjerenja seksizma. Sljedeći logiku klasične teorije mjerenja, rezultati neke skale seksizma sastojat će se od pravog rezultata i komponente pogreške. U pravilu ćemo biti sigurniji u rezultate što je uzorak veći i heterogeniji. Suprotni predznaci slučajne pogreške se međusobno gotovo poništavaju u većim uzorcima, a izbjegavanjem sustavne pogreške osigurava se dodatna validnost i preciznost rezultata, što zajedno s heterogenim i/ili reprezentativnim uzorkom olakšava generaliziranje rezultata na širu populaciju. No, ovakav

pristup sustavno zanemaruje moguće utjecaje efekta metode. Na primjer, način na koji vršimo mjerenje može utjecati na rezultate. Rezultati anketnog upitnika koji je distribuiran putem raznih *online* kanala mogli bi se razlikovati od rezultata prikupljenih CAPI (engl. *Computer-assisted personal interviewing*) metodom. Osobe koje ispunjavaju anketni upitnik putem *online* kanala, mogu paralelno nesmetano raditi druge stvari, čime im se kognitivni kapaciteti za ispunjavanje samog upitnika smanjuju, a posljedično i kvaliteta rezultata. Razlike mogu biti prisutne i u samoj CAPI metodi, jer možemo pretpostaviti da ispitanici na pitanja o seksizmu neće na isti način odgovoriti ako je anketar muškarac s jasno izraženim tradicionalnim vrijednostima i tradicionalnom percepcijom rodnih uloga ili žena s jasno izraženim egalitarnim svjetonazorom. Može se pretpostaviti da će rezultati anketnih upitnika prikupljeni od strane muškarca s jasno izraženim tradicionalnim vrijednostima i tradicionalnom percepcijom rodnih uloga pokazivati nešto veće razine seksizma nego rezultati prikupljeni od strane žene s jasno izraženim egalitarnim vrijednostima. No, unutar domene ovog rada, fokusirali smo se na efekte metode koji su vezani uz samu konstrukciju instrumenata koji su korišteni unutar nekog anketnog upitnika, nezavisno o mogućem utjecaju anketara, jer pretpostavljamo da se muškarac s jasno izraženim tradicionalnom vrijednostima i percepcijom rodnih uloga neće angažirati za mjerenje seksizma CAPI metodom.

Kako bismo ilustrirali utjecaj efekata metode u anketnim upitnicima, iskoristili smo primjer mjerenja zadovoljstva poslom (engl. *job satisfaction*). Naime, rezultati na skali zadovoljstva poslom ovisit će o načinu na koji su formulirani iskazi skale (engl. *item wording*) odnosno koliko iskaza u skali je pozitivno formulirano, a koliko negativno formulirano (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018). Nadalje, rezultati na skali *zadovoljstva zaposlenjem* ovisit će o tome je li prvi iskaz skale pozitivan ili negativan, odnosno javit će se potvrdna pristranost (engl. *confirmatory bias*). Smjer prve čestice može utjecati na sve sljedeće čestice u skali, jer se prilikom odgovaranja na prvu česticu aktiviraju uvjerenja koja su u skladu s njome te se ona prenose na čestice koje slijede (Kunda, Fong, Santioso, i Reber, 1993). Postoji jednostavan način kako se potvrdna pristranost može pretvoriti u slučajnu pogrešku, a to je randomizacija prikaza čestica u skali, što je opcija koja je dostupna u većini anketnih *software*-a. Jednakom broju sudionika bit će prikazana skala gdje je prva čestica pozitivna, kao i broju sudionika gdje je prva čestica skale negativna, čime će se na dovoljno velikom uzorku ti utjecaji međusobno gotovo poništiti i pretvoriti u slučajnu pogrešku.

Iz tih razloga, u ovome radu se nećemo baviti potvrdnom pristranosti. Više o utjecaju nekih efekata metode na skalu zadovoljstva zaposlenjem objasnili smo u sljedećem poglavlju.

Postoje brojni drugi efekti metode, koji na sebi svojstvene načine djeluju na rezultate anketnih upitnika, dovodeći do pristranih rezultata: pristranost ekstremnog stila odgovaranja (engl. *extreme response style*), pristranost slaganja (engl. *acquiescence bias*) i nemarno odgovaranje (engl. *careless responding*) su samo od nekih efekata metode čiji se utjecaji često zanemaruju u konstrukciji i interpretaciji rezultata anketnih istraživanja. Ove skupine efekata često se zanemaruju, a njihovo djelovanje podsjeća na tip pogreške koji se nastoji izbjegavati u svim znanstvenim istraživanjima: pojedinačni efekt metode sustavno iskrivljuje rezultat u istome smjeru, jednako kao i sustavna pogreška. Unutar ovog rada, bavili smo se utjecajima formulacije iskaza u prvoj studiji, a utjecajima nemarnog odgovaranja i pristranosti slaganja u drugoj studiji, jer smo koristili skalu sačinjenu od isključivo pozitivnih iskaza.

2. Formulacija iskaza

Izraz formulacija iskaza odnosi se na smjer u kojemu je iskaz formuliran odnosno na njegov odnos s mjernim konstruktom. S obzirom na smjer, postoje dvije vrste iskaza: pozitivni (engl. *regular item*) i negativni (engl. *reversed item*). Pozitivnim iskazima smatramo one iskaze čiji smjer odgovara smjeru mjerenog konstrukta, što znači da taj iskaz nije potrebno rekodirati prije obrade podataka. Na primjer, zamislimo unidimenzionalnu skalu pomoću koje mjerimo konstrukt *religioznosti* koja u sebi sadrži iskaz „Često odlazim u crkvu“. Navedeni iskaz trebamo smatrati pozitivnim, jer možemo tvrditi da veći rezultati na njemu u pravilu iskazuju veći rezultat na skali *religioznosti*. Iskaz koji u sebi sadrži jezične negacije se također može smatrati pozitivnim iskazom, čime postaje jasno kako prilikom kategorizacije pojedinog iskaza kao pozitivnog ili negativnog u obzir treba uzeti isključivo njegov odnos s konstruktom kojeg mjeri, bez obzira sadrži li on u sebi ikakav jezični oblik negacije. Na primjer, iskaz „Ateisti ne bi trebali raditi s djecom“ se može smatrati pozitivnim iskazom kada se mjeri konstrukt *religioznosti*, makar on u sebi sadrži negaciju. Pozitivan je iz razloga što veći rezultat na njemu generalno može značiti veći rezultat na skali *religioznosti*. S druge strane, negativni iskazi su oni iskazi koji su formulirani u smjeru suprotnom od mjernog konstrukta te ih je potrebno rekodirati, odnosno

potrebno je učiniti refleksiju skale, kako bi svi uključeni iskazi imali isti smjer s konstruktom mjerenja. U skladu s prethodno navedenim formulacijama, iskaz „*Podržavam ukidanje vjeronauka u školama*“ trebao bi se smatrati negativnim, jer veći rezultat na njemu u pravilu znači manji rezultat na skali *religioznosti*.

Skale korištene u anketnim upitnicima moguće je podijeliti s obzirom na broj pozitivnih i negativnih iskaza koji ih sačinjavaju. Uravnotežene skale (engl. *balanced scales*) naziv je za skale koje su sastavljene od jednakog broja pozitivnih i negativnih iskaza. Korištenje takvih skala prvobitno je uvedeno s ciljem povećanja validacije skala (Nunnally, 1978; Paulhus, 1991) te su one postale neizostavan dio društvenih istraživanja. Smatralo se da uravnotežene skale kontroliraju pristranost nemarnog slaganja s iskazima bez obzira na njihov sadržaj, jer se sastoje od pozitivnih i negativnih iskaza, ali i da smanjuju pristranost stila odgovaranja (engl. *Response style bias*). Pristranost stila odgovaranja označava sustavnu tendenciju da se na neku grupu iskaza odgovara u skladu s nekim vanjskim utjecajem koji nije vezan uz sam sadržaj iskaza (Paulhus, 1991). Dakle, smatralo se da ubacivanjem suprotnih polova neke dimenzije skala dobiva na validnosti jer se zahvaća širi spektar stavova, ali i da se implementacijom i pozitivnih i negativnih iskaza u skalu uklanjaju razne pristranosti iz krajnjih rezultata. Nasuprot tome, recentna istraživanja pokazuju kako implementacija negativnih iskaza u skale umanjuje njihovu validnost (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018). Naime, uravnotežene skale, u usporedbi s drugim skalama, u pravilu imaju slabiju diskriminacijsku snagu te često rezultiraju modelima koji slabije pristaju podatcima. Isto tako, smatra se da uravnotežene skale, u usporedbi s drugim skalama, češće dovode do problema u analizi, poput odbijanja jednodimenzionalnih modela u korist višedimenzionalne solucije (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018). Nadalje, kognitivno procesuiranje nije isto za pozitivne i negativne iskaze, pogotovo u slučaju negacija u formulaciji, čime se može zaključiti da na uspješnost procesuiranja negativnih iskaza utječe razvijenost vokabulara, čime se u varijancu rezultata uključuju i utjecaji koji nisu vezani uz samo značenje iskaza (Marsh, 1986; Mestre, 1988).

Iako recentna istraživanja pronalaze sve više mogućih nedostataka korištenja negativnih iskaza, njihovo korištenje u skalama se i dalje preporuča (Weijters i Baumgartner, 2012). Generalno, postoje dva argumenta za korištenje negativnih iskaza u mjernim skalama. Prema

prvom argumentu, kombiniranje pozitivnih i negativnih iskaza unaprjeđuje validnost mjernog konstrukta, jer takva kombinacija iskaza predstavlja širi raspon situacija i uvjerenja na kojima se odgovor sudionika može zasnivati (Weijters i Baumgartner, 2012). Takva logika nalaže da valja ispitati oba pola istog konstrukta, kako bi rezultati bili vjerodostojniji. Neka istraživanja su pokazala kako, u usporedbi s ostalim skalama (npr. skalama sačinjenim od samo pozitivnih ili samo negativnih iskaza), uravnotežene skale imaju najmanju varijancu, čime su temelji na kojima je prvi argument zasnovan poljuljani (Suárez-Alvarez et al., 2018). Prema drugom argumentu, kombiniranje pozitivnih i negativnih iskaza trebalo bi smanjiti utjecaje potvrdne pristranosti (engl. *confirmatory bias*) i utjecaje pristranosti slaganja (engl. *acquiescence bias*).

Bez obzira na brojne nedostatke, uravnotežene skale su i dalje korištene u anketnim upitnicima. I zaista, istraživanja su pokazala kako su skale Likertovog tipa, to jest najkorištenije skale u društvenim istraživanjima, sklone utjecajima potvrdne pristranosti i pristranosti slaganja s iskazima (McClendon, 1991), što je argument u korist implementacije negativnih iskaza u skale, makar ne postoje jasni dokazi da njihova implementacija smanjuje razine tih pristranosti. Empirijski je potvrđeno da (ne)korištenje negativnih iskaza u skalama ima posljedice na rezultate, što znači da, ako ih se koristi u skalama, njihove utjecaje treba moći na neki način kontrolirati i uzeti u obzir prilikom interpretacije rezultata. Bez obzira na desetljeća rada i brojne studije koje se bave tematikom utjecaja formulacije iskaza, jasni zaključci o ovom efektu metode i dalje nisu izneseni te oko njihovog (ne)korištenja ne postoji konsenzus (Kam, 2016).

2.1. Mjerenje utjecaja formulacije iskaza u istraživanjima

Formulacija iskaza može imati višestruke utjecaje na rezultate anketnih istraživanja. Prvo, ona može utjecati na aritmetičke sredine pojedinog iskaza. Zamislimo dva iskaza koja su polarna suprotnost jedan drugome. Za početak, to mora značiti da su iskazi međusobno u potpunosti isključivi. Na primjer, osoba koja tvrdi da je u potpunosti zadovoljna svojim zaposlenjem ne može u isto vrijeme biti u potpunosti nezadovoljna svojim zaposlenjem, jer jedno isključuje drugo. Pretpostavimo da postoji skala *zadovoljstva zaposlenjem* koja se sastoji od isključivo pozitivnih iskaza te skala *nezadovoljstva zaposlenjem*, koja sadrži iste iskaze kao skala *zadovoljstva zaposlenjem*, samo što su u ovome slučaju oni preformulirani tako da skalu sad tvore isključivo negativni iskazi. Zamislimo da skala *zadovoljstva zaposlenjem* sadrži iskaz „*Iznimno*

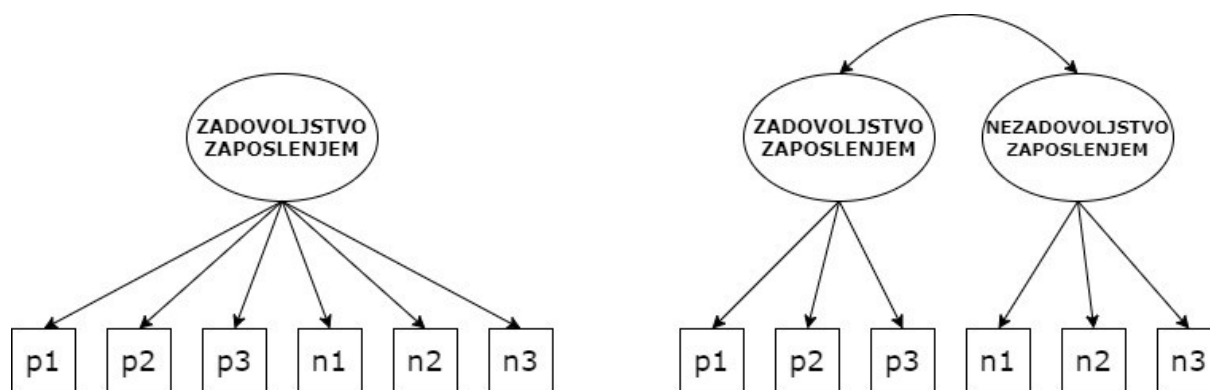
sam sretan/na svojim zaposlenjem“. U tom slučaju, skala *nezadovoljstva zaposlenjem* sadržavala bi polarno suprotni iskaz koji bi glasio „*Iznimno sam nesretan/na svojim zaposlenjem*“ ili „*Uopće nisam sretan/na svojim zaposlenjem*“. Nadalje, pretpostavimo da smo skale *zadovoljstva zaposlenjem* i *nezadovoljstva zaposlenjem* primijenili unutar istog istraživanja na nekom uzorku. Ako i dalje vrijedi konsenzus da je jedna skala potpuna refleksija druge, hipotetske aritmetičke sredine iskaza i skala trebale bi biti podjednake nakon rekodiranja iskaza jedne od tih dvaju skala. Ali, aritmetičke sredine međusobno obrnutih iskaza na istom uzorku mogu se statistički značajno razlikovati (Mandić i Klasnić, 2021). Ako ispitanici dva polarno obrnuta iskaza shvaćaju u potpunosti suprotnima, njihovi iznosi korelacija trebali bi se približiti iznosu -1. No, tomu ne mora biti slučaj (Mandić i Klasnić, 2021), jer kognitivni procesi za interpretaciju pozitivnih iskaza nisu isti kao oni za interpretaciju negativnih iskaza (Marsh, 1986). Aritmetičke sredine i korelacije mogu poslužiti za ilustraciju problema, no one ne pridonose njegovom rješenju odnosno kontroli utjecaja formulacije iskaza. Korištenje pozitivnih i negativnih iskaza počiva na pretpostavci da obje vrste iskaza mjere isti konstrukt (Marsh, 1986). Bez obzira na to, međukorelacije iskaza unutar balansirane skale slabije su nego u skali koja se sastoji od samo pozitivnih ili samo negativnih iskaza (DiStefano i Motl, 2006; Zeng, Wen i Zhang, 2020), a korištenje balansiranih skala umanjuje internu konzistentnost skale. Ovo je navelo jednu struju istraživača ovog efekta metode na hipotezu kako pozitivni i negativni iskazi mjere različite, ali međusobno povezane konstrukte (Chang, Maydeu-Olivares i D'Zurilla, 1997).

Kako bi omogućili istraživanje utjecaja efekta metode na validnost skale, istraživači često koriste neki oblik modela više svojstva-više metoda (engl. *Multitrait-Multimethod*; MTMM; Campbell i Fiske, 1959). MTMM model generalno ubraja dva podtipa specifičnih modela, model povezanih svojstva-povezanih metoda (engl. *Correlated Trait-Correlated Method*; CT-CM) i model povezanih svojstva-povezanih unikateta (engl. *Correlated Trait-Correlated Uniqueness*; CT-CU). CT-CM specificira efekt metode kao latentnu varijablu te istovremeno odvaja njegovu varijancu od varijance konstrukta (Zeng, Wen i Zhang, 2020). S druge strane, CT-CU model specificira efekt metode kao rezidualne, ali budući da ne pretpostavlja nezavisni latentni faktor efekta metode nije pogodan za mjerenje jačine samog efekta metode, zbog čega se CT-CM prevalentno koristi prilikom procjene utjecaja efekta metode (Tomás i Oliver, 1999). Postoji i treći model, gdje su svojstva korelirana, ali metode nisu, a istraživanja pokazuju da upravo takvi modeli rezultiraju modelima koji najbolje odgovaraju podacima (Reise et al., 2007). Sadržajni

konstrukti mogu biti povezani s latentnim faktorom metode, što sugerira da efekti metode nisu sustavna pogreška, već su produkt karakteristika skale mjernog konstrukta (DiStefano i Motl, 2009). Upravo je ta hipoteza, po kojoj se utjecaji formulacije iskaza ne mogu generalizirati među nezavisnim studijama jer ovise o skali mjerenja, potvrđena u nekim istraživanjima (Kam, 2016).

Ranije smo se referirali na istraživanja čiji rezultati pokazuju da uravnotežene skale mogu rezultirati odbijanjem jednofaktorskih modela u korist višedimenzionalnih solucija (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018). Za ilustriranje tog problema, poslužili smo se jednom od najkorištenijih multivarijantnih metoda u društvenim znanostima: eksploratornom faktorskom analizom. Zamislimo da smo dobili zadatak za konstrukciju uravnotežene *skale zadovoljstva zaposlenjem*. Makar je viđenje ovog konstrukta kao jednodimenzionalnog prevladavalo većinom prošlog stoljeća, neki noviji autori tvrde da se unutar tog koncepta radi o dva različita konstrukta odnosno *zadovoljstva zaposlenjem* i *nezadovoljstva zaposlenjem* (Crede, Oleksandr, Bagraim, i Sully, 2009). Rezultati istraživanja navedenih autora pokazali su da iskazi vezani uz zadovoljstvo zaposlenjem visoko saturiraju na jednom faktoru, a iskazi vezani uz nezadovoljstvo zaposlenjem na drugome faktoru. Nadalje, njihovoj tezi o odbacivanju jednodimenzionalnosti u prilog idu i međusobno različite korelacije koje skale *zadovoljstva zaposlenjem* i *nezadovoljstva zaposlenjem* imaju s drugim skalama. Rezultati njihovog istraživanja pokazali su da skala *zadovoljstva zaposlenjem* snažnije korelira s drugim pozitivnim konstruktima, a skala *nezadovoljstva zaposlenjem* s negativnim konstruktima, što može poslužiti kao razumna linija argumentacije za odbacivanje jednodimenzionalne koncepcije ovog konstrukta. Ali, može se reći da ovakvi rezultati idu u prilog tezi da pozitivni iskazi mjere jednu vrstu konstrukata, a negativni drugu vrstu (Chang, Maydeu-Olivares i D'Zurilla, 1997). Bez obzira na dosadašnja istraživanja o konstruktima *(ne)zadovoljstva zaposlenjem*, za potrebe vizualizacije utjecaja formulacije iskaza, pretpostavit ćemo da se zapravo radi o jednodimenzionalnom konstruktima. Zamislimo da smo dobili zadatak za konstrukciju uravnotežene *skale zadovoljstva zaposlenjem* koja se sastoji od šest iskaza (npr. Kam i Meyer, 2015). Budući da se radi o uravnoteženoj skali, morali smo implementirati tri pozitivna iskaza, koji su se direktno odnosili na *zadovoljstvo zaposlenjem*, te tri negativna iskaza, koji su se direktno odnosili na *nezadovoljstvo zaposlenjem*.

Nakon toga, pretpostavimo da smo proveli eksploratornu faktorsku analizu te smo primijetili da su se po nekom kriteriju umjesto očekivanog jednog pojavila dva faktora. Primjerice, sukladno Guttman-Kaiser kriteriju (Guttman, 1954; Kaiser, 1956), zadržane su dvije komponente/dva faktora čiji su karakteristični korijeni veći od 1. Ovo pravilo je vjerojatno najpopularnije i najčešće korišteno na osnovi uvida u objavljene radove koji koriste eksploratornu faktorsku analizu (Costello i Osborne, 2005; Fabrigar et al., 1999). Nakon što bismo analizirali koji iskazi visoko saturiraju na koji faktor, primijetili bismo da pozitivni iskazi visoko saturiraju na jedan faktor (*zadovoljstvo zaposlenjem*), a negativni iskazi na drugi faktor (*nezadovoljstvo zaposlenjem*). Ovakvi rezultati mogli bi nas navesti da odbijemo jednodimenzionalnu ideju konstrukta te prihvatimo i argumentiramo da se zapravo radi o dvije dimenzije.

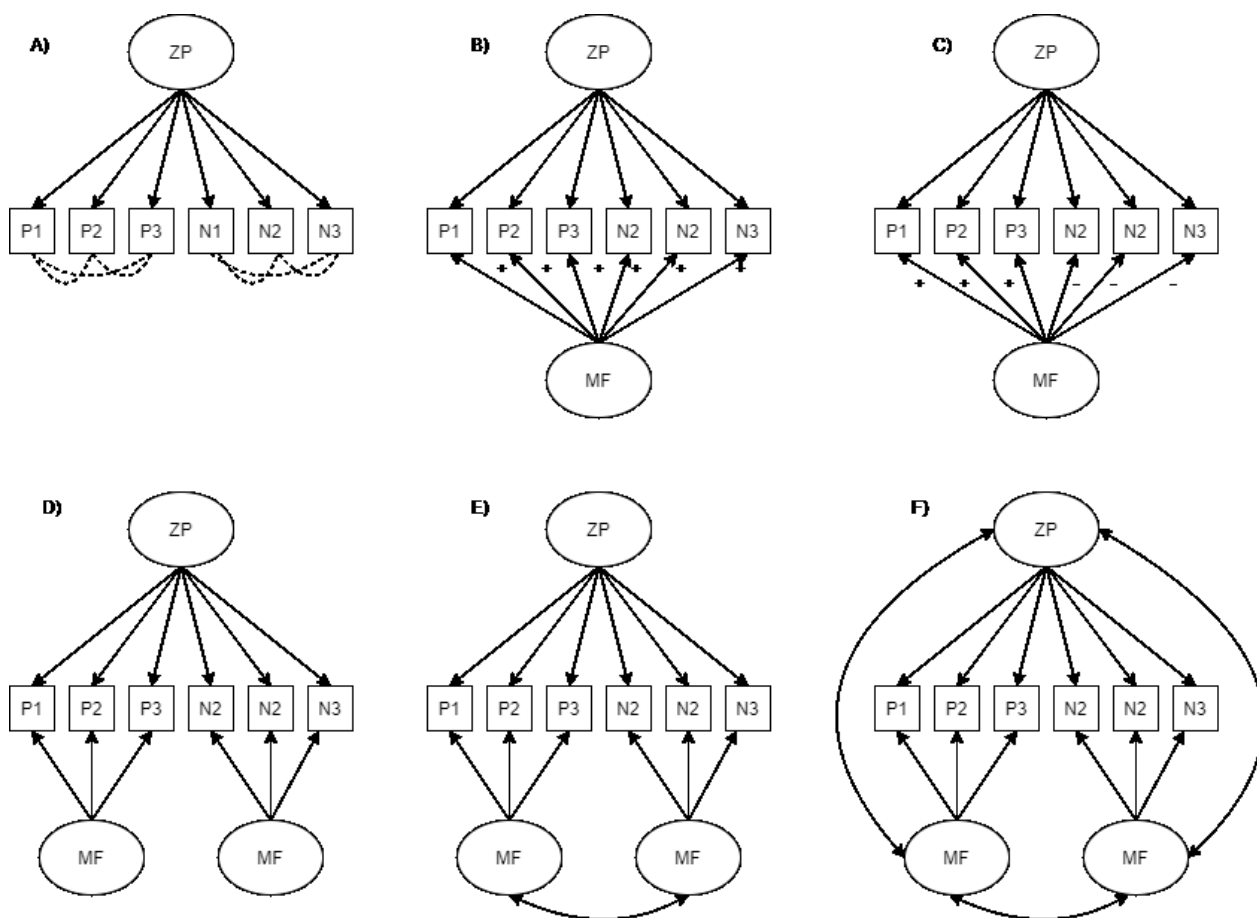


Prikaz 1: Grafički prikaz mogućeg utjecaja formulacije iskaza. Lijevi model prikazuje inicijalno zamišljeni jednodimenzionalni model. Desni model prikazuje mogući utjecaj formulacije iskaza, gdje se oko jednog faktora grupiraju pozitivni iskazi, a oko drugog negativni, čime može doći do odbacivanja inicijalno zamišljenog jednodimenzionalnog modela. Napomene: oba modela su pojednostavljena; iskazi p1-p3 označavaju pozitivne, a iskazi n1-n3 negativne iskaze.

Ovakvi zaključci mogli bi biti neispravni, budući da se prilikom interpretacije u obzir nije uzeo utjecaj formulacije iskaza. Upravo iz ovih razloga, potrebno je detaljnije proći kroz pristupe koji nastoje kontrolirati utjecaj ovog efekta metode. Makar uključivanje međusobno polarno suprotnih iskaza može poslužiti u detektiranju prisutnosti utjecaja formulacije iskaza, ovakav pristup neće nužno doprinijeti njegovoj kontroli. Naime, uključivanje takvih iskaza (npr. „U potpunosti sam zadovoljan/na svojim zaposlenjem“ i „U potpunosti sam nezadovoljan/na svojim zaposlenjem“) može poslužiti u testiranju razlike aritmetičkih sredina, koje će potvrditi prisutnost efekta, ali primjena ovoga u kontroli je limitirana. Jedan od mogućih pristupa kontroli ovog

efekta nalazi se u strukturalnom modeliranju (engl. *Structural equation modeling*; SEM). Prilikom specifikacije modela *zadovoljstva zaposlenjem*, potrebno je specificirati dvofaktorski model, gdje će jedan faktor biti sadržajni faktor (u ovome slučaju *zadovoljstvo zaposlenjem*), a drugi faktor latentna dimenzija efekta metode formulacije iskaza. Ovoj specifikaciji se može pristupiti na više načina:

- a) specificira se samo jedan latentni faktor efekta metode formulacije iskaza. Ovaj pristup podrazumijeva da pozitivni i negativni iskazi imaju isti pozadinski izvor. Ne preporučamo ovakav pristup, jer ako se smatra da formulacija iskaza utječe na rezultate anketnih istraživanja, logično je za očekivati da različite vrste iskaza imaju različite izvore. Ovaj argument može potkrijepiti i hipoteza određenog broj autora koji tvrde da pozitivni i negativni iskazi mjere različite, ali povezane konstrukte (Chang, Maydeu-Olivares i D'Zurilla, 1997).
- b) specificiraju se dva latentna faktora efekta metode formulacije iskaza, gdje se jedan odnosi na efekt metode pozitivnih iskaza, a drugi na efekt metode negativnih iskaza. Unutar ovakvog STMM (engl. *Single Trait Multiple Method*) modela, moguće je specificirati da su efekti metode (ne)korelirani. Budući da u ovom slučaju oba latentna faktora efekta metode predstavljaju utjecaje formulacije iskaza, preporučamo dopuštanje njihove korelacije.
- c) Nakon specifikacije latentnog/ih faktora efekta metode, moguće je napraviti dodatni korak. Sadržajni konstrukti mogu biti povezani s latentnim faktorom metode, što sugerira da su efekti metode produkt skale sadržajnog konstrukta (DiStefano i Motl, 2009). Preporučamo specifikaciju svih navedenih modela u svakom istraživanju koje se bavi utjecajima formulacije iskaza, kako bismo utvrdili najbolji pristup.



Prikaz 2: Specifikacija nekih mogućih modela kontrole efekata formulacije iskaza. Model A = model s povezanim unikvitetima; ST-CU model. Model B = model s jednim latentnim faktorom efekta metode, koji isto utječe i na pozitivne i na negativne iskaze; ST-SM model. Model C = model s jednim latentnim faktorom efekta metode, koji različito utječe na pozitivne i negativne iskaze; ST-SM model. Model D = model s dva latentna faktora efekta metode koji nisu korelirani; ST-MM model. Model E = model s dva latentna faktora efekta metode koji su korelirani; ST-MCM model. Model F = model s dva latentna faktora metode koji su korelirani, ali su i korelirani sa sadržajnim faktorom; SCT-MCM. ZP = Zadovoljstvo zaposlenjem; P1-P3 = pozitivno formulirani iskazi; N1-N3 = negativno formulirani iskazi; MF = latentni faktor efekta metode. *Napomena:* isprekidane linije predstavljaju povezane unikvitete, dok zaobljene dvostruko usmjerene pune linije označavaju korelirane latentne konstrukte.

Dodatni pristup koji se može izabrati unutar SEM pristupa kontroliranja utjecaja formulacije iskaza jest razlika rezultata između pozitivnih i negativnih iskaza. Na primjer, ukoliko se radi o testiranju utjecaja koju formulacija iskaza ima na neku skalu, od zbroja aritmetičkih sredina pozitivnih iskaza oduzet će se zbroj aritmetičkih sredina negativnih iskaza (ili obrnuto) te će dobiveni rezultat predstavljati utjecaj formulacije iskaza na rezultate skale. Dobiveni rezultat se u

formi indikatora može koristiti u strukturalnom modelu (npr. Weijters, Baumgartner i Schillewaert, 2013).

3. Nemarno odgovaranje

Nemarno odgovaranje (engl. *careless responding*, CR; *insufficient effort responding*, IER) je naziv za efekt metode gdje sudionici ne obraćaju dovoljno pozornosti na sadržaj iskaza te ne ulažu trud prilikom interpretacije i odgovora na njega (Huang, Curran, Keeney, Poposki i DeShon, 2012; Meade i Craig, 2012). Ne postoji konsenzus oko njegovog izvora, makar se nemarno odgovaranje kao efekt metode često pripisuje nedostatku pažnje i motivacije (Weijters et al., 2013). Bez obzira na nedostajanje konsenzusa oko izvora ove pojave, konsenzus postoji oko postojanja njegovih posljedica. Naime, nemarno odgovaranje može dovesti do kontradiktornih odgovora u anketnom upitniku, čime se mogu pojaviti nevaljani vektori odgovora koji imaju potencijal iskriviti i izmijeniti rezultate istraživanja (Maniaci i Rogge, 2014). Rezultati određenih simulacijskih studija potvrđuju ovu premisu, jer je dobiveno da ako samo 10% sudionika nemarno odgovara na pitanja, dolazi do pojave dodatnog faktora u hipotetski pretpostavljenom jednodimenzionalnom konstrukt (Schmitt i Stults, 1985; Woods, 2006; Huang, Liu i Bowling, 2015). Ako se dodatni faktor pojavi, rezultati će većinom biti sljedeći: jedan faktor sastojat će se od pozitivnih iskaza koji visoko saturiraju na upravo tu latentnu dimenziju, dok će se drugi faktor sačinjavati od negativnih iskaza (Arias, Garrido, Jenaro, Martínez-Molina i Arias, 2020). U ovome slučaju, odbacivanje jednodimenzionalnog konstrukta predstavlja resultantnu silu spurioznih¹ korelacija (ili lažnih korelacija) koje su nastale zbog nekonzistentnih i često kontradiktornih odgovora koje nemarni sudionici anketnih istraživanja daju. To znači da je jednodimenzionalni konstrukt zapravo ispravan, ali se zbog odsustva kontroliranja utjecaja efekta metode pokazao kao neispravan. Ako se nemarno odgovaranje ne kontrolira, njegovi utjecaji mogu imati neželjene posljedice na interpretaciju pa i replikaciju rezultata anketnih istraživanja (Curran, 2016).

¹ U statistici, spuriozne korelacije predstavljaju matematičku vezu između dvije varijable. Dobivena korelacija može biti utjecaj trećeg faktora, koji utječe na prve dvije varijable. Ako je korelacija spuriozna, nakon kontroliranja utjecaja treće varijable, korelacija između prve dvije varijable prestaje biti značajna.

Zapravo, većina istraživanja koja su se bavila ovom temom, pokazuju da je i niski postotak nemarnih odgovora dovoljan da se stvore značajne izmjene u korelacijama, statističkim značajnostima, dimenzionalnosti podataka i jačini efekata (Baumgartner i Steenkamp, 2001; Woods, 2006; Huang et al., 2012, Maniaci i Rogge, 2014; DeSimone i Harms, 2018). Osim spomenutih utjecaja, nemarno odgovaranje može dovesti i do veće šanse pojave pogreške tipa II, koja označava prihvatanje nulte hipoteze koja se trebala odbaciti (Crede, 2010; Meade i Craig, 2012). Jedan od najčešćih oblika nemarnog odgovaranja je odgovaranje po ravnoj liniji (engl. *Straightlining*; SL), gdje sudionik daje slične odgovore bez obzira na sadržaj ili smjer iskaza (DeSimone, DeSimone, Harms i Wood, 2017). Oblici u kojima se SL manifestira se međusobno razlikuju, od ispitanika koji daju identične odgovore na sve iskaze, do ispitanika koji namjerno koriste sve ponuđene opcije kako bi se doimalo da su pažljivo pročitali sve iskaze (DeSimone i Harms, 2018).

Slučajno odgovaranje (engl. *Random responding*) je bio jedan od prvih korištenih termina koji je opisivao obrasce odgovora koji su bili rezultat nemara ili manjka koncentracije (Eden i Leviatan, 1975). Takav naziv za ovaj efekt metode nije validan jer ljudi nemaju sposobnost generiranja slučajnih odgovora, obrazaca ili brojeva, čak i ako ih se traži da tako odgovaraju (Neuringer, 1986; Huang et al., 2012). Makar ne postoji puno radova koji se nemarnim odgovaranjem bave iz longitudinalne i individualne perspektive, postoje radovi unutar kojih je prisutnost nemarnog odgovaranja bila temporalno konzistentna, čime se pretpostavlja da je ono manifestacija neke osobine ličnosti (Bowling et al., 2016).

3.1. Detektiranje nemarnog odgovaranja

Mjerenju prisutnosti nemarnog odgovaranja generalno se može pristupiti na dva načina: *a priori* metodama i *a posteriori* metodama. *A priori* metode odnose se na posebne iskaze, skale ili druge oblike koji su implementirani unutar anketnog upitnika i koji služe isključivo pri distinkciji nemarnog od pažljivog odgovaranja među sudionicima. *A posteriori* metode odnose se na načine detektiranja nemarnog odgovaranja koje ne uključuju implementaciju posebnih iskaza i/ili skala, već primjenu nekih statističkih tehnika na postojećim opaženim podacima.

Provjera instruktivne manipulacije

Jedan od najčešće korištenih *a priori* metoda mjerenja nemarnog odgovaranja naziva se provjera instruktivne manipulacije (engl. *Instruction manipulation check*; IMC). IMC se sastoji od jednog ili više posebnih iskaza koji se zbog svojeg sadržaja mogu koristiti za potrebe klasifikacije nemarnih i pažljivih sudionika. Takav iskaz može glasiti: „*Molimo Vas da na ovom pitanju ne odaberete niti jedan odgovor i da odgovor na pitanje ostavite praznim*“ ili „*Molimo Vas da odaberete opciju Uopće se ne slažem za ovo pitanje*“, ako od sudionika direktno zahtijevamo odabir točno određenog odgovora, ili „*Trenutno se koristim elektroničkim uređajem*“ ako se radi o iskazu za koji je samo jedan odgovor ispravan. Poželjno je u svakom upitniku koristiti što je više moguće ovakvih iskaza, a da oni u maksimalnoj mjeri ostanu suptilni i što manje primjetni. Čak i u kraćim upitnicima, poželjno je koristiti minimalno dva ovakva iskaza, jer vjerojatnost da se na peterotomnoj skali slaganja slučajno odabere prava/trazena opcija iznosi 20%, dok je ta ista vjerojatnost 4% kada se radi o dva nezavisna IMC iskaza. Upravo iz tih razloga, potrebno je napomenuti kako se svi potencijalni *a priori* IMC iskazi koji služe za provjeravanje nemarnog odgovaranja moraju postaviti prije ključne skale od interesa, kako bi se moglo odrediti u kojoj su mjeri njezini rezultati pod utjecajem nemarnog odgovaranja. Ovakve situacije su dodatni argument za korištenje više IMC iskaza unutar anketnog upitnika, jer neki ispitanici mogu biti pažljivi, a u nekom trenutku postati nemarni, dok se kod nekih može dogoditi suprotni proces.

Vrijeme unutar kojeg je upitnik ispunjen

A posteriori metode mjerenja nemarnog odgovaranja predstavljaju razne oblike manipulacije na već dobivenim podacima. Vrijeme koje je bilo potrebno da se upitnik ispuni često služi kao upravo takav pristup mjerenja nemarnog odgovaranja. Generalna premisa je da ispunjavanje upitnika u mnogo kraćem vremenskom intervalu od prosjeka označava prisutnost nemarnog odgovaranja, jer sudionici nisu uzeli dovoljno vremena da pažljivo pročitaju i odgovore na svako pitanje (Beck, Albano i Smith, 2018). Ako se radi o anketnim upitnicima koji se ispunjavaju na računalu, ista se stvar može tvrditi i za one sudionike kojima je trebalo mnogo

više vremena od prosjeka da ispune anketni upitnik. Razlog tome je što se takvi upitnici mogu ispunjavati u anonimnim, izoliranim i nekontroliranim uvjetima, gdje sudionici hipotetski mogu raditi više stvari istovremeno. Isto tako, sudionici mogu započeti s ispunjavanjem upitnika, početi raditi nešto drugo usred njegova ispunjavanja te se nakon nekoliko minuta ili sati vratiti upitniku. Problem kod ovog pristupa je što je često teško odrediti vremenske granice koje razlikuju nemarne od pažljivih sudionika. No, ako se takva granica ispravno detektira, ovaj *a posteriori* pristup mjerenju nemarnog odgovaranja može biti koristan indikator mjerenja nemarnog odgovaranja (Wise i Kong, 2005).

Makar se sve veći broj *a posteriori* metoda predlaže i testira, a priori su većinom one koje su preporučene (Beck, Albano i Smith, 2018). Takvi iskazi su jednostavni za implementirati u anketni upitnik, a njihovo korištenje i implementiranje nije komplicirano te daje dodatne argumente za valjanost dobivenih rezultata. Njihova implementacija predstavlja potez niskog rizika (engl. *low-risk high-reward*) prilikom kontrole utjecaja nemarnog odgovaranja.

4. Pristranost slaganja

Pristranost slaganja s iskazima bez obzira na njihov sadržaj (engl. *Acquiescence bias*), definira se kao tendencija prema pozitivnoj strani skale, odgovarajući na iskaze uz zanemarivanje značenja samog iskaza (Bentler, Jackson, i Messick, 1971; Weijters, Baumgartner, i Schillewaert, 2013). Kao i prethodno spomenuti fenomeni, pristranost slaganja često uzrokuje spuriozne korelacije između varijabli, čime umanjuje valjanost dobivenih rezultata (Aichholzer, 2014; McCrae, Herbst i Costa, 2001). U skladu s definicijom, zbog sudionika koji odgovaraju na ovakav način, korelacije među konstruktima istog smjera bit će „napuhane“ (engl. *Inflated*), a korelacije među konstruktima suprotnog smjera smanjene (Kam i Mayer, 2015). Zbog sudionika koji odgovaraju na ovakav način, doći će do viših aritmetičkih sredina pojedinih varijabli, bez obzira na njihov sadržaj ili smjer. Za pristranost slaganja s iskazima bez obzira na njihov sadržaj se smatra da je stabilna kroz vrijeme (Billiet i Davidov, 2008; Weijters, Geuens i Schillewaert, 2010). Također, na nju se gleda kao na individualnu karakteristiku sudionika, koja je većinom van utjecaja konstrukta mjerenja (Weijters, Geuens i Schillewaert, 2010).

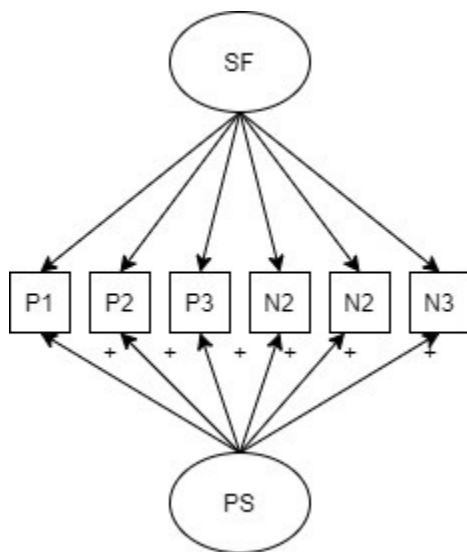
Na peterotomnoj skali slaganja, ovakva pristranost ne mora se odnositi isključivo na sudionike koji redovito odabiru opciju pet, već i ona one koji odabiru opciju četiri, jer ona također predstavlja određenu dozu slaganja. Isto tako, kod sudionika se može javiti suprotan, ali po prirodi sličan proces, a to je pristranost neslaganja bez obzira na sadržaj iskaza. U toj instanci, sudionici će najčešće odabirati opciju jedan ili opciju dva, ne uključujući dovoljan napor u proces interpretacije samog iskaza.

4.1.Mjerenje pristranosti slaganja

Jedan od klasičnih pristupa kontroliranja pristranosti slaganja jest korištenje balansiranih skala (Nunally, 1978; Paulhus, 1991). Osim navedenog, mogu se koristiti i antonimi (Primi, Santos, De Fruyt i John, 2019). Ako sudionik za vrijeme ispunjanja upitnika podliježe pristranosti slaganja, korelacija između antonima neće biti ni blizu iznosa -1. Ukoliko antonimi međusobno imaju istu težinu, od njih se može konstruirati indeks pristranosti slaganja (Primi et al., 2015). Prije konstruiranja ovog indeksa, potrebno je rekodirati sve antonime u istome smjeru, kako bismo vidjeli odstupanje rezultata od očekivanog iznosa 0. Ovakav pristup često se rabi u balansiranim skalama, dok je u nebalansiranim prisutan problem pristranosti koje su vezane uz predominantni smjer iskaza. Generalno veći broj pozitivnih iskaza rezultirat će pristranosti prema pozitivnoj strani, a veći broj negativnih iskaza prema negativnoj strani (Ferrando i Lorenzo-Seva, 2010).

Jedna od osnovnih razlika u modernom mjerenju pristranosti slaganja je u tome mjeri li se ona indirektno ili direktno (Aicholzer, 2015). Indirektan pristup označava korištenje nezavisne skale koja služi isključivo za mjerenje pristranosti slaganja (npr. Weijters, Baumgartner, i Schillewaert, 2013). Ovakve skale se većinom sastoje od međusobno nepovezanih i heterogenih iskaza koji u suštini ne mjere ništa. Upravo se na temelju nepovezanosti tih iskaza konstruira indeks koji predstavlja individualni rezultat na skali pristranosti slaganja (engl. *Net acquiescence response style*; NARS). Direktni pristup mjerenju pristranosti slaganja predstavlja pristup unutar strukturalnog modeliranja, gdje su iskazi koji predstavljaju sadržajnu latentnu dimenziju isti oni koji mjere i pristranost slaganja (npr. Savalei i Falk, 2014). Ovakav pristup zahtijeva balansiranu skalu koja se koristi prilikom mjerenja sadržajnog konstrukta, za razliku od indirektnog pristupa koji može biti korišten i za nebalansirane skale (Watson, 1992). Jedan od direktnih pristupa

mjerenja pristranosti slaganja naziva se slučajnim preprečnim faktorskim modelom (engl. *Random intercept factor model*; RIFA). Unutar ovakvog pristupa, modelu koji se sastoji od sadržajnog/sadržajnih konstrukata dodaje se latentna dimenzija te se sve saturacije manifestnih varijabli na dodanu latentnu dimenziju postave na 1 ili se ostave slobodnima (Maydeu, Olivares i Coffman, 2006). Isti modeli koriste se i prilikom kontrole utjecaja formulacije iskaza te stoga ovaj pristup smatramo problematičnim. Naime, ovakva statistička manipulacija specifikacijom modela većinom rezultira boljim pristajanjem podataka (de la Fuente i Abad, 2020), ali to primarno treba pripisati respecifikaciji modela, a ne kontroli nekog izvanjskog utjecaja. Problem je u tome što ne možemo sa sigurnošću znati što dodana latentna dimenzija predstavlja, već joj možemo samo pretpostaviti sadržaj. Iz teorijske perspektive, novu specificiranu latentnu dimenziju moguće je opisati i opravdati iz raznih polazišta, npr. ili kao latentnu dimenziju utjecaja formulacije iskaza ili kao latentnu dimenziju pristranosti slaganja ili kao nešto treće, ali niti jedno polazište ne možemo sa sigurnošću potvrditi. Bez obzira na ovo, većina znanstvenika koja proučava fenomen pristranosti slaganja redovito koristi ovaj pristup (Hidalgo-Rasmussen i González-Betanzos, 2019; Kam, Schermer, Harris i Vernon, 2013). No, smatramo da je umjesto ovog „direktnog“ pristupa mjerenju pristranosti slaganja, bolje koristiti „indirektan“ način gdje se koristi skala posebno konstruirana za potrebe mjerenja te pristranosti.



Prikaz 3: Prikaz RIFA modela. SF = sadržajni faktor odnosno konstrukt; PS = latentna dimenzija pristranosti slaganja.

Još jedan od mogućih pristupa kontroliranja pristranosti slaganja naziva se *ex post* standardizacija rezultata gdje se od svakog pojedinačnog rezultata sudionika oduzme njegov prosječni rezultat na svim iskazima (npr. Fischer, 2004). Ovakav pristup, ipsatizacija rezultata, često se koristio u kombinaciji s raznim faktorskim analizama te se smatra da on uklanja utjecaj pristranosti slaganja iz rezultata (Aicholzer, 2015).

5. Nacrt istraživanja

Analizirali smo podatke iz dvije studije. Unutar prve studije testirali smo utjecaj formulacije iskaza na skale ličnosti te na kraju dali prijedloge maksimalne kontrole tog utjecaja prilikom konstrukcije novi skala koje sadrže i pozitivne i negativne iskaze. U drugoj studiji testirali smo utjecaje nemarnog odgovaranja i pristranosti slaganja. Nismo gledali utjecaje formulacije iskaza jer je korištena skala u drugoj studiji bila sačinjena od isključivo pozitivnih iskaza.

6. Ciljevi rada

Glavni cilj rada je utvrditi utjecaje raznih efekata metode i pristranosti, kao i moguće načine njihove kontrole. Cilj prve studije je prikazivanje utjecaja koje formulacija iskaza može imati na rezultate anketnih istraživanja te utvrditi na kojim se sve razinama ti utjecaji mogu manifestirati. Dodatno, iznosimo prijedlog pomoću kojega bi se ti utjecaji trebali smanjiti u situacijama kada se koriste skale koje sačinjavaju i pozitivni i negativni iskazi.

Cilj druge studije je prikazati utjecaje nemarnog odgovaranja i pristranosti slaganja na rezultate anketnih istraživanja te konstruirati indeks pristranosti. Navedeni utjecaji kontrolirani su na dva načina: pomoću strukturalnih modela te pomoću otežavanja opaženih varijabli primjenom indeksa pristranosti.

7. Studija 1

Prilikom analize rezultata studije 1 fokusirali smo se isključivo na utjecaje formulacije iskaza na rezultate. Cilj ove studije bio je prikazati utjecaje formulacije iskaza na rezultate anketnih istraživanja iz više perspektiva, a određeni dijelovi studije se nalaze u postupku objavljivanja/sudijelovanja dio rada nagrađenog za rektorovu nagradu 2019./2020 (opis uzorka, opis načina prikupljanja podataka, dio analitičkog pristupa i dio rezultata o utjecajima na aritmetičke sredine).

7.1. Metode

Prikupljanje podataka

Podatci su prikupljeni putem dvije inačice *online* anketnog upitnika. Poveznica za pristup anketnom upitniku distribuirana je putem raznih društvenih mreža, čime je uzorak neprobabilistički i prigodan. Poveznica za upitnik je primarno dijeljena putem Facebook-a te pretpostavljamo da je većina sudionika pristupila upitniku preko ove platforme. Anketni upitnik napisan je na hrvatskom jeziku te su se podatci prikupljali tijekom ožujka 2020. Provedbu istraživanja odobrilo je nadležno Povjerenstvo Odsjeka za sociologiju za prosudbu etičnosti istraživanja, Filozofskog fakulteta Sveučilišta u Zagrebu. Prvo pitanje u anketnom upitniku služilo je kao filter varijabla pomoću koje je sudionicima dodijeljena jedna od inačica anketnog upitnika. Sudionici kojima je rođendan u parnom mjesecu dodijeljena je jedna inačica, a onima kojima je rođendan u neparnom mjesecu druga.

Uzorak

Istraživanje je provedeno na prigodnom uzorku korisnika društvenih mreža u Hrvatskoj. Uzorak sudionika koji su rođeni na parni mjesec sastojao je od $N=391$ sudionika, a uzorak onih koji su rođeni na neparni mjesec od $N=400$ sudionika. 82,3% od svih sudionika odgovorilo je na sva pitanja u anketnom upitniku. 97% od svih sudionika nije imalo više od 3,4% praznih odgovora. Iz analize su isključeni sudionici kojima je nedostajalo više od 10% odgovora. Nisu pronađene statističke značajne razlike među uzorcima s obzirom na varijable *spol*, *dob*, i *obrazovanje*. Prosječna dob svih sudionika bila je 30 godina [Min = 15; Max = 83; SD = 12,9).

71,5% sudionika bili su pripadnici muškog spola. 45,3% sudionika ostvarilo je neki oblik visokog obrazovanja, gdje visoko obrazovanje označava ono obrazovanje koje je iznad sekundarnog nivo (srednja stručna sprema). Ovaj postotak uključuje i studente preddiplomskog studija koji još nisu završili svoje obrazovanje.

Analitički pristup

Odlučili smo testirati utjecaj formulacije iskaza koristeći tri skale iz Mini-IPIP skala ličnosti (Donellan, Oswald, Baird, i Lucas, 2006). Navedene skale odabrane su zbog velikog broja empirijskih istraživanja u području ličnosti, ali i zbog njihove praktičnosti za upotrebu. Naime, skale su kratke te se svaka od pet skala sastoji od samo četiri iskaza. Svi iskazi bodovali su se na skali procjene slaganja od 1 do 5, kako bismo osigurali usporedive varijance.

Kako bismo testirali utjecaje formulacije iskaza odnosno utjecaj promjene pozitivnog iskaza u negativni (ili obrnuto), izabrali smo dvije facete iz Mini-IPIP-a, *savjesnost* i *ugodnost* (Mandić, 2020). Navedene facete izabrane su iz razloga što mislimo kako je preformuliranje njihovih iskaza rezultiralo iskazima koji zvuče prirodnije, nego da se ista stvar napravi s nekom od preostalih faceta. Na primjer, iskaz „*Empatičan sam*“ u drugoj će inačici postati „*Nisam empatičan*“. Jedna inačica skala *savjesnosti* i *ugodnosti* sastojala od četiri pozitivna iskaza po skali, a druga od četiri negativna iskaza po skali. Testirali smo razlike u aritmetičkim sredinama nakon što smo sve iskaze rekodirali u istom smjeru, jer pretpostavljamo da je formulacija iskaza imala statistički značajan utjecaj na aritmetičke sredine. Potom smo proveli testove invarijantnosti među dvjema inačicama istog konstrukta, u cilju dobivanja dodatnih saznanja o utjecaju formulacije iskaza. Za treću skalu smo koristili onu *intelekta*, koja nije izmijenjena u obje inačice i koja je služila kao kontrolni instrument. Svaka skala ličnosti nalazi se na posebnom ekranu, kako bismo maksimalno smanjili takozvani *spill-over* utjecaj jedne skale na drugu. Ekрани s pitanjima pojavljivali su se istim redoslijedom, bez obzira na inačicu, s time da su jedine razlike bile u formulaciji pojedinih iskaza (Mandić, 2020).

Analize su provedene koristeći R sustav za statističko računanje 4.0.2. (R Development Core Team, 2020). Svi univarijantni testovi provedeni su koristeći dodatni statistički paket *psych*, koji je razvio Revelle (2019). Svi strukturalni modeli i svi testovi metrijske invarijantnosti provedeni su koristeći dodatni statistički paket *lavaan*, koji je razvio Rosseel (2012).

7.2. Rezultati i diskusija

Rezultati su bili u skladu s očekivanjima. Među iskazima skale intelekta, koja nije adaptirana, nije pronađena statistički značajna razlika u aritmetičkim sredinama s obzirom na poduzorak, što argumentira valjanost sljedećih rezultata. Naime, budući da u neizmijenjenoj skali nije pronađena statistički značajna razlika u prosjecima između dva poduzorka, možemo pretpostaviti da je statistički značajna razlika u aritmetičkim sredinama koja je pronađena u drugim skalama rezultat utjecaja formulacije iskaza. Među sedam od osam parova iskaza skala *savjesnosti* i *ugodnosti* pronađene su statistički značajne razlike u prosjecima, što jasno pokazuje da obrtanje iskaza utječe na aritmetičke sredine odnosno na rezultate. Negativni iskazi većinom javljaju više rezultate na aritmetičkim sredinama. Nakon detaljnijeg uvida u distribucije odgovora, pronašli smo da je u šest od osam međusobno obrnutih parova iskaza, veći postotak sudionika odabrali opciju „*U potpunosti se slažem*“ za negativno formulirane iskaze u usporedbi s pozitivno formuliranim iskazima. 27,2% sudionika odabrali je opciju „*U potpunosti se slažem*“ na negativno formuliranim iskazima, a 15,6% na pozitivno formuliran iskazima. Ovo se može pripisati određenom obliku nemarnog odgovaranja ili pristranosti slaganja zbog sljedećih razloga: kognitivni procesi korišteni od strane sudionika prilikom odgovaranja na pozitivno formulirane iskaze nije jednak kao onaj korišten prilikom odgovaranja na negativno formulirane iskaze, a za jasno sadržajno razumijevanje negativnih iskaza potrebne su više razine lingvističkih vještina i truda (Suárez-Alvarez et al., 2018). Zbog ovoga, sudionici sa slabijim lingvističkim vještinama, ali i sudionici s manjim razinama motivacije, postaju podložni kontradiktornim odgovorima koji će utjecati na rezultate.

Tablica 1. Utjecaj formulacije iskaza na aritmetičke sredine (prethodno objavljeno u: Mandić, 2020:11)

UTJECAJ FORMULACIJE ISKAZA				
	Uzorak parnih mjeseci (N = 400)	Način promjene iskaza iz uzorka parnih mjeseci u uzorak neparnih mjeseci	Uzorak neparnih mjeseci (N = 391)	p
<i>Savjesnost, iskaz 1</i>	3,01	Negativni u pozitivni	3,22	,007
<i>Savjesnost, iskaz 2</i>	3,42	Negativni u pozitivni	3,76	,000
<i>Savjesnost, iskaz 3</i>	3,80	Negativni u pozitivni	3,65	,008
<i>Savjesnost, iskaz 4</i>	4,02	Negativni u pozitivni	4,02	,974
<i>Ugodnost, iskaz 1</i>	3,82	Pozitivni u negativni	4,13	,000
<i>Ugodnost, iskaz 2</i>	4,07	Pozitivni u negativni	4,28	,000
<i>Ugodnost, iskaz 3</i>	3,37	Pozitivni u negativni	3,95	,000
<i>Ugodnost, iskaz 4</i>	2,98	Pozitivni u negativni	3,69	,000
NEIZMJENJENA SKALA				
	Uzorak parnih mjeseci (N = 400)	Promjena između uzorka parnih mjeseci I uzorak neparnih mjeseci	Uzorak neparnih mjeseci (N = 391)	p
<i>Intelekt, iskaz 1</i>	3,66	Nikakva	3,73	,302
<i>Intelekt, iskaz 2</i>	3,51	Nikakva	3,53	,544
<i>Intelekt, iskaz 3</i>	3,69	Nikakva	3,74	,400
<i>Intelekt, iskaz 4</i>	3,64	Nikakva	3,70	,173

Napomene: svi iskazi su rekodirani tako da budu u istom smjeru u odnosu na konstrukt mjerenja. Interpretacija statističke značajnosti se nije promijenila primjenom Bonferonni korekcije.

Kako bismo detaljnije analizirali utjecaje formulacije iskaza na rezultate anketnih istraživanja, spojili smo rezultate dvaju poduzoraka te proveli testove mjerne invarijantnosti, gdje su grupe predstavljale inačicu anketnog upitnika na koji su sudionici odgovarali. Kako bismo proveli ove analize, bilo je potrebno specificirati jednostavne konfirmatorne mjerne modele, koji su se sastojali od jedne latentne dimenzije (skala ličnosti) te od četiri indikatora kojima se mjeri njima nadređena dimenzija. To znači da su se rezultati na iskazu jedan iz obje inačice anketnog upitnika spojili te su bili predstavljeni putem jednog indikatora. Prilikom specifikacije modela *ugodnosti*, paket *lavaan* je upozorio na prisutnost negativnih varijanci. Budući da su negativne varijance često rezultat neispravno specificiranih modela (Browne, 2015), možemo pretpostaviti da je utjecaj formulacije iskaza potencijalno u utjecao na rezultate, da se ovako jednostavan model ne može ispravno specificirati i ne može konvergirati. Naravno, ne možemo sa sigurnošću tvrditi da je pojava negativnih varijanci definitivna posljedica formulacije iskaza, pogotovo zbog veličine uzorka i asimetrične prirode uključenih i ordinalnih varijabli. Zbog ovoga, odlučili smo ne uključiti ga u testove mjerne invarijantnosti. Prilikom interpretacije rezultata mjerne

invarijantnosti, većina istraživača prelazi s interpretacije χ^2 statistika na alternative pokazatelje slaganja, poput ΔCFI jer je χ^2 previše osjetljiv na slabe i neznčajne razlike (Chen, 2007; Putnick i Bornstein, 2016). Slijedili smo kriterij promjene CFI parametra u iznosu -,01 za ugnježdene modele, koji predlažu Cheung i Rensvold (2002). Ako se između dva koraka mjerne invarijantnosti u CFI parametru dogodi promjena veća od ,01, veća razina mjerne invarijantnosti nije ostvarena.

Spojeni model skale *savjesnosti* nije ostvario konfiguralnu invarijantnost [$\chi^2 = 45,2$, $df = 4$, $RMSEA = ,194$, $CFI = ,672$], čime bi se moglo pretpostaviti da u ovome slučaju pozitivni iskazi predstavljaju različiti konstrukt od negativnih iskaza, što bi potvrdilo tezu nekih radova (Chang et al., 1997). Rezultati pristajanja modela savjesnosti u parnom poduzorku [$\chi^2 = 6,41$, $df = 2$, $p < ,001$, $CFI = ,966$, $TLI = ,898$, $RMSEA = ,082$, $90\%CI = ,015 - ,158$, $SRMR = ,033$], ali i u neparnom poduzorku [$\chi^2 = 5,74$, $df = 2$, $p = ,06$, $CFI = ,983$, $TLI = ,948$, $RMSEA = ,074$, $90\%CI = ,035 - ,148$, $SRMR = ,029$] su na granici prihvatljivosti, što je dodatan razlog da se neostvarivanje konfiguralne invarijantnosti jednim dijelom može pripisati utjecaju formulacije iskaza. S druge strane, spojeni model skale *intelekt* ostvario je konfiguralnu invarijantnost [$\chi^2 = 5,14$, $df = 4$, $RMSEA = ,027$, $CFI = ,998$] te je u skladu s ΔCFI kriterijem dosegao nije maskimalna razina maskimalnu razinu mjerne invarijantnosti, što dodatno validira argument da preformuliranje iskaza značajno utječe na rezultate.

Tablica 2. Testovi mjerne invarijantnosti

	χ^2 (df)	p	BIC	AIC	RMSEA	CFI	ΔCFI
SAVJESNOST							
Konfiguralna invarijantnost	19,08 (4)		8236,2	8124	,098	,760	
INTELEKT							
Konfiguralna invarijantnost	5,15 (4)		8726,6	8614,5	,027	,998	
Metrijska invarijantnost	7,17 (7)	,569	8708,6	8610,5	,008	1,000	,002
Skalarna invarijantnost	8,12 (10)	,813	8689,6	8605,5	,000	1,000	,000
Rezidualna invarijantnost	9,21 (14)	,895	8664	8598,6	,000	1,000	,000
Invarijantnost latentnih ar. sredina	10,65 (15)	,231	8658,7	8598	,000	1,000	,000

AIC= Akaike information criterion; BIC= Bayesian information criterion Napomene: ako je $\Delta CFI > ,01$, viša razina metrijske invarijantnosti nije ostvarena.

7.3. Kontroliranje utjecaja nemarnog odgovaranja

Kako bismo testirali pristajanje svih modela iz prikaza 2, trebali smo koristiti skalu koja se sastoji i od pozitivnih i od negativnih iskaza. Budući da su skale *ugodnosti* i *savjesnosti* bile sačinjene ili u potpunosti od pozitivnih ili od negativnih iskaza, bilo je potrebno koristiti druge skale. U istom istraživanju, koristili smo i skale *emocionalne stabilnosti* i *ekstraverzije*, koje se sastojale od dva pozitivna i od dva negativna iskaza. S obzirom da bi specifikacija određenih modela iz prikaza 2 rezultirala podidentificiranim ili tek-identificiranim modelom s nama dostupna četiri indikatora, odlučili smo testirati sve dostupne pristupe kontrole formulacije iskaza na dvofaktorskom modelu *ekstraverzije* i *emocionalne stabilnosti*. Jedini model koji nije ostvario dovoljan broj stupnjeva slobode je model A iz prikaza 2, tako da on nije uključen u analizu. Prilikom interpretacije pristajanja u konfirmatornoj faktorskoj analizi (CFA), vodili smo se često korištenim parametrima slaganja i njihovim preporučenim graničnim vrijednostima. Općenito, prilikom interpretacije komparativnih parametara slaganja (CFI; Bentler, 1990; TLI; Tucker i Lewis, 1973), vrijednosti TLI i CFI koeficijenata između ,90 - ,95 ukazuju na zadovoljavajuće pristajanje podacima (Bentler, 1990). Parametar RMSEA (engl. *Root mean square error of approximation*) jedan je od najkorištenijih parsimonično-korektivnih indeksa (Steiger i Lind, 1980). Hu i Bentler (1999) preporučaju da je pristajanje modela adekvatno ako je vrijednost parametra RMSEA ispod ,06 ili da je barem njezina donja vrijednosti 90% intervala pouzdanosti ispod te vrijednosti, makar neki autori tvrde da su ovi parametri prestrogi za modele koji se koriste u društvenim znanostima (Marsh, Hau i Wen, 2004). No, budući da je inicijalni model jako jednostavan, navodit ćemo se originalno navedenim vrijednostima pri interpretaciji rezultata.

Rezultati su u većoj mjeri bili u skladu s očekivanjima: što je model bio kompleksniji, to su se parametri slaganja popravljali po svojim apsolutnim vrijednostima. Testirali smo ugniježdene modele kako bismo provjerili statističku značajnost u pristajanju podacima. Originalni model ni u uzorku parnih ni neparnih mjeseci nije pokazao zadovoljavajuće pristajanje podacima, pogotovo ako u obzir uzmemo njegovu jednostavnost. Model B pokazuje statistički značajno bolje pristajanje podacima u odnosu na originalni model [$p = ,013$], ali se njegovo pristajanje isto može opovrgnuti kada se u obzir uzme jednostavnost modela. Model C i model B imaju isti broj stupnjeva slobode, tako da nisu međusobno alternativni niti ugniježdjeni modeli, što znači da

se njihova usporedba treba učiniti s obzirom na parametre slaganja, ali i s obzirom na AIC i BIC. Generalno, Model C [$\Delta\chi^2 = -12,278$, $\Delta CFI = ,013$, $\Delta TLI = ,021$, $\Delta RMSEA = -,009$, $\Delta SMRM = -,005$, $\Delta AIC = -13,1$, $\Delta BIC = -13,1$] pokazuje bolje parametre slaganja u odnosu na Model B, a također pokazuje statistički značajno bolje slaganje u odnosu na originalni model [$p = <,001$]. Makar parametri slaganja Modela D bolje pristaju podacima u odnosu na prijašnje modele, a razlika je čak statistički značajna u odnosu na Model C [$p = ,011$], varijanca jednog latentnog faktora utjecaja formulacije iskaze je negativna, čime smo prisiljeni odbaciti ga. Nadalje, Modeli E i F iz prikaza 2 nisu uspjeli konvergirati, što može biti rezultat premalog uzorka, ali i neispravne specifikacije modela, čime njihova uloga u kontroli formulacije iskaza postaje neznačajna.

Situacija je slična kod uzorka neparnih mjeseci. Originalni model ne pokazuje zadovoljavajuće parametre slaganja, pogotovo kad se u obzir uzme njegova jednostavnost. Model B, koji pretpostavlja latentnu dimenziju efekta metode, statistički bolje pristaje podacima u odnosu na originalni model [$p = ,001$], ali njegovo pristajanje je i dalje granično. U ovome slučaju, Model C, koji ima isti broj stupnjeva kao i Model B, pokazuje slabije parametre slaganja od Modela B [$\Delta\chi^2 = 4,246$, $\Delta CFI = -,007$, $\Delta TLI = ,012$, $\Delta RMSEA = ,007$, $\Delta SMRM = ,014$, $\Delta AIC = 9,6$, $\Delta BIC=9,6$], što može implicirati da ovakav način kontrole utjecaja formulacije iskaza ne mora biti univerzalan, već da ovisi od skale do skale i od uzorka do uzorka. Modeli D i E, makar pokazuju adekvatnije parametre slaganja, javljaju negativne varijance u jednom od latentnih dimenzija, čime ih moramo odbaciti, a u Modelu F matrica varijanci/kovarijanci latentnih varijabli nije pozitivno definitna, čime ona ne može biti adekvatno rješenje.

Tablica 4. Pristajanje s i bez kontroliranja utjecaja formulacije iskaza

UZORAK PARNIH MJESECI (N = 400)								
	χ^2	df	p	CFI	TLI	RMSEA	90% CI	SRMR
<i>Model original</i>	76,703	19	,000	,930	,897	,098	,075 - ,121	,059
<i>Model B</i>	69,367	18	,000	,940	,906	,093	,071 - ,117	,059
<i>Model C</i>	57,089	18	,000	,953	,927	,082	,059 - ,107	,054
<i>Model D</i>	49,648	17	,000	,962	,937	,077	,052 - ,102	,043
UZORAK NEPARNIH MJESECI (N = 391)								
	χ^2	df	p	CFI	TLI	RMSEA	90% CI	SRMR
<i>Model original</i>	88,606	19	,000	,930	,884	,100	,079 - ,121	,067
<i>Model B</i>	70,714	18	,000	,951	,924	,085	,065 - ,107	,051
<i>Model C</i>	74,960	18	,000	,944	,912	,092	,071 - ,114	,065
<i>Model D</i>	68,313	17	,000	,950	,917	,089	,068 - ,112	,055
<i>Model E</i>	44,016	16	,000	,972	,951	,069	,045 - ,093	,047
<i>Model F</i>	30,731	12	,002	,982	,959	,063	,036 - ,091	,027

Napomena: Svi indikatori su robusne procjene izračunate koristeći MLR

7.4. Zaključak i prijedlozi

Rezultati ove studije bili su u skladu s očekivanim te u skladu s ciljevima. Utjecaji formulacije iskaza vidljivi su u statistički značajnoj razlici u aritmetičkim sredinama, ali i u razinama mjerne invarijantnosti koje skala doseže. Postoje razni načini i prijedlozi kako kontrolirati utjecaj formulacije iskaza, a većina njih je provjerena u ovoj studiji. Makar ovakav način kontrole statistički značajno popravi pristajanje podacima, on se primarno može, ali i treba shvatiti kao matematička manipulacija podacima i specifikaciji modela. Problem je u tome, što ne možemo sa sigurnošću znati što dodane latentne dimenzije predstavljaju i makar je argumentacija kako se radi o efektu metode koji je rezultat formulacije iskaza logičan i realan, on nije definitivan. I zaista, u poglavlju gdje smo objašnjavali pristranost slaganja, vidljivo je da se isti/slični načini koji se koriste prilikom kontrole utjecaja formulacije iskaza koriste i prilikom kontrole pristranosti slaganja, samo što je dodana latentna dimenzija objašnjena putem drugačijeg teorijskog polazišta.

Potrebno je zaključiti da ovi prijedlozi dokazuju da se dodavanjem latentne dimenzije matematičkom manipulacijom dobivaju statistički značajno bolja pristajanja podacima, ali da postoje brojne teorije koje bi novostvorenu latentnu dimenziju mogle legitimno opravdati.

Generalno, postoji dovoljan broj radova čiji rezultati dokazuju negativne posljedice implementiranja negativno formuliranih iskaza u skale (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018), ali određena struja istraživača ih i dalje preporuča (Weijters i Baumgartner, 2012). Budući da je korištenje negativnih iskaza u budućnosti neizbježno, u obliku koraka dajemo sljedeće preporuke prilikom konstrukcije skala koje u sebi sadrže negativne iskaze:

1. korak – ovaj korak je neizmijenjen u odnosu na trenutnu praksu. Sastoji se od generacije velikog broja čestica koje su vezane uz temu istraživanja te njihovo postupno čišćenje. Čišćenje je poželjno obaviti na nekom temelju rezultata na nekom prigodnom uzorku, primjerice studentskom.
2. korak – nakon što se dobije određeni broj čestica nakon prvog čišćenja s kojim se planira na teren, potrebo je učiniti sljedeće: ako se na teren inicijalno planiralo otići s 30 čestica koje su odabrane nakon prvog koraka, potrebno je svakoj čestici stvoriti njezinu polarnu suprotnost. U ovom slučaju, na teren se ide s 30 pozitivno formuliranih i 30 negativno formuliranih čestica. Čestice bi trebale biti pokazane na različitim ekranima/stranicama, kako bi se smanjili takozvani *spill-over* utjecaji jedne vrste iskaza na drugu. Važno je napomenuti da, ako se već koriste negativni iskazi, da se oni koriste u skladu s naputcima koje su dali Weijters i Baumgartner (2012), a to je da u negativno formuliranim iskazima nastoje izbjegnuti formulacije koje u sebi sadrže negacije.
3. korak – nakon što su se prikupili podaci u skladu s drugim korakom, potrebno je konstruirati skalu koja se sastoji od pozitivnih i od negativnih iskaza. Negativne iskaze, osim u skladu s njihovim metrijskim karakteristikama i/ili saturacijama na svojstvenom faktoru, treba odabrati i na dodatan način: odabiru se negativni iskazi koji pokazuju najveće koeficijente korelacije sa svojim pozitivnim parom. Na primjer, ako iskaz „Uopće nisam sretan/na“ pokazuje koeficijent korelacije $r = -,02$ s iskazom „U potpunosti sam sretan/na“, njega se ne bi trebalo koristiti u skali čak ni u slučaju njegove visoke saturacije na latentnoj dimenziji te je u tom slučaju uputno koristiti iskaz „U potpunosti sam sretan“. Ako se između istog iskaza i njegovog pozitivno formuliranog para javlja snažna korelacija, npr. $r = -,08$, onda bi utjecaji formulacije iskaza trebali biti manji jer je

jedan iskaz u većoj mjeri antonim drugome te se negativni iskaz može koristiti umjesto pozitivnog.

4. korak – predstavlja validaciju konstruirane skale na nekom uzorku. Naravno, poželjno je koristiti i nezavisne konstrukte koji bi argumentirali valjanost skale.

Potrebno je napomenuti da ovakav pristup, makar u sebi sadrži korak koji bi trebao smanjiti utjecaj formulacije iskaza i time osnažiti metrijske karakteristike skale, neće eliminirati utjecaje formulacije iskaza, pogotovo ako u uzorku postoji osjetni udio sudionika sa slabijim vokabularnim vještinama ili manjim razinama motivacije. Iz ovih razloga, preporučamo da se u skalama ne koriste negativno formulirani iskazi, već da se koriste isključivo u svrhu validacije pažljivog odgovaranja, što je postupak koji smo koristili i opisali u studiji 2. U sljedećoj studiji smo detaljnije objasnili načine kontrole ostalih utjecaja efekta metode i stila odgovaranja na rezultate anketnih istraživanja, ako je u pitanju skala koja u sebi ne sadrži negativno formulirane iskaze.

8. Studija 2

8.1. Metode

Prikupljanje podataka

U drugoj studiji, koristili su se podatci tvrtke Equestris d.o.o. koji su prikupljeni u sklopu projekta „Istraživanje navika vlasnika kućnih ljubimaca“, koje se provodilo od 5. do 8. veljače 2021. godine. Podatci su prikupljeni putem zakupljenog *online* panela, a autor je sudjelovao u ovome procesu. Prosječno vrijeme ispunjavanja upitnika iznosilo je 25 minuta. Voditelj navedenog projekta dao je dopuštenje za implemetaciju pitanja koja se korištena u analizi u ovom radu u anketnom upitniku koji je korišten u sklopu projekta.

Uzorak

Istraživanje je provedeno na kvotnom uzorku vlasnika kućnih ljubimaca u Hrvatskoj veličine N=483. Brojka finalnog uzorka sastoji se od 50% od ukupnog broja sudionika koji su pristupili upitniku. 32% sudionika je odustalo prilikom ispunjavanja upitnika, a 18% sudionika je diskvalificirano na prvom pitanju jer nisu vlasnici kućnih ljubimaca. Kvote su postavljene u skladu s profilom kupaca jedne specijalizirane trgovine za kućne ljubimce. 64,7% uzorka činile su pripadnice ženskog spola, a prosječna dob svih sudionika bila je 37,9 godina [Min = 18; Max = 69; SD = 11,5]. 60,3% sudionika ima nekakav oblik višeg formalnog obrazovanja (viša stručna sprema i više), a 66,1% ih je zaposleno. 62,4% sudionika je oženjeno/udano ili žive s partnerom/icom, a 68,8% ih živi u gradu. Detaljniji prikaz sociodemografskih karakteristika prikazan je u tablici 5.

Tablica 5. Pregled sociodemografskih karakteristika uzorka

Varijabla	Sudionici		Muškarci	Žene
	Broj	%	%	
Spol				
<i>Muški</i>	170	35,2	100	0
<i>Ženski</i>	313	64,8	0	100
Stupanj obrazovanja				
<i>Osnovna škola</i>	12	2,5	2,9	2,2
<i>Srednja škola</i>	180	37,3	38,2	36,7
<i>Visoka stručna sprema</i>	111	23	23,5	22,7
<i>Fakultet</i>	171	35,4	32,4	37,1
<i>Doktorat</i>	9	1,9	2,9	1,3
Radni status				
<i>Zaposlen/a</i>	320	66,3	65,9	66,5
<i>Nezaposlen/a</i>	62	12,8	8,8	15
<i>Student/ica/Učenik/ica</i>	36	7,5	7,6	7,3
<i>U mirovini</i>	34	7	12,9	3,8
<i>Briga o kućanstvu</i>	14	2,9	1,2	3,8
<i>Nešto drugo</i>	17	3,5	3,5	3,5
Bračni status				
<i>Oženjen/udana ili živim s partnerom/icom</i>	302	62,5	55,9	66,1
<i>U vezi, ne živim s partnerom/icom</i>	57	11,8	8,2	13,7
<i>Razveden/a, rastavljen/a</i>	30	6,2	4,7	7
<i>Udovac/udovica</i>	6	1,2	0,6	1,6
<i>Nisam u vezi</i>	77	15,9	28,8	8,9
<i>Nešto drugo</i>	5	1	0,6	1,3
<i>Ne želim odgovoriti</i>	6	1,2	1,2	1,3
Tip kućanstva				
<i>S partnerom/icom i punoljetnom djecom</i>	51	10,6	11,8	9,9
<i>S partnerom/icom i maloljetnom djecom</i>	150	31,1	26,5	33,5
<i>S partnerom/icom – bez djece</i>	100	20,7	17,6	22,4
<i>S vlastitim roditeljima ili roditeljima partnera/ice</i>	98	20,3	25,3	17,6
<i>Sam/a – s djecom</i>	37	7,6	3,4	9,9
<i>Sam/a</i>	35	7,2	11,8	4,8
<i>S nekim drugim</i>	12	2,5	3,5	1,9
Tip naselja				
<i>Grad</i>	333	68,9	70	68,4
<i>Predgrađe</i>	50	10,4	10,6	10,2
<i>Manje mjesto</i>	48	9,9	9,4	10,2
<i>Selo</i>	52	10,8	10	11,2
Mjesečni prihodi kućanstva				
<i>2001 - 4000 kn</i>	34	7	6,5	7,3
<i>4001 – 6000 kn</i>	50	10,4	7,6	11,8
<i>6001 – 8000 kn</i>	78	16,1	16,5	16
<i>8001 – 10000 kn</i>	72	14,9	12,9	16
<i>10001 – 12000 kn</i>	65	13,5	14,7	12,8
<i>12001 – 14000 kn</i>	48	9,9	11,8	8,9
<i>14001 – 16000 kn</i>	36	7,5	10	6,1
<i>16001 – 18000 kn</i>	26	5,4	3,5	6,4
<i>Više od 18000 kn</i>	31	6,4	8,8	5,1
<i>Ne znam/Ne želim odgovoriti</i>	43	8,9	7,6	9,6

Analitički pristup

Odlučili smo testirati utjecaj svih efekata metode uključenih u anketni upitnik koristeći skalu *suvremenog seksizma* (Šimac i Klasnić, 2021). Ova studija je ujedno služila kao dodatna provjera dimenzionalnosti i valjanosti skale. Skala se sastojala od 12 čestica, a sve čestice su se bodovale na skali procjene slaganja od 1 do 5. Budući da smo jedan od iskaza iz ove skale duplicirali te ga preformulirali u njegovu polarnu suprotnost, iskazi skale su bili podijeljeni na dva ekrana. Više o dupliciranju i preformulaciji odabranog iskaza u sekciji o nemarnom odgovaranju. ne navodi ako nije značajno.

Nemarno odgovaranje

Kako bismo mogli mjeriti utjecaj nemarnog odgovaranja, uključili smo dva posebna iskaza (IMC), od kojih je jedan postavljen prije svih iskaza na prvom ekranu skale *suvremenog seksizma*, a drugi prije svih iskaza na drugom ekranu skale *suvremenog seksizma*. Odlučili smo koristiti dva odvojena pristupa prilikom formulacije ovih iskaza. Prvi iskaz formuliran je u obliku direktne naredne koju su sudionici trebali pratiti kako ih ne bismo klasificirali kao nemarne: „*Molimo Vas da odaberete opciju 1 za ovo pitanje*“. Iskazi su se bodovali na skali 1-5, gdje je iznad svakog broja bila navedena labela što on predstavlja. Drugi iskaz formuliran je tako da je iz logičke perspektive na njega moguć samo jedan ispravan odgovor: „*Trenutno se koristim elektroničkim uređajem*“. Klasifikacija sudionika kao pažljivima se u ovome slučaju može napraviti na dva načina, gdje se u prvom slučaju kao pažljive klasificira samo one koji su odabrali maksimalno slaganje s ovim iskazom, a u drugom se ovaj način i uspoređi sa sudionicima koji su odabrali bilo koju opciju slaganja (u peterotomnoj bipolarnoj skali bile to bi bile gornje dvije opcije) (npr. Kim, McCabe, Yamasaki, Louie i King, 2017). Budući da smo skalu procjene posebno prilagodili ovome problemu, sudionici koji su odabrali opciju 4 - „*Donekle se slažem*“, bit će klasificirani kao nemarni. U ovome kontekstu, smatramo da je korištenje elektroničkim uređajem proces koji u sebi ne sadrži stupnjevanje: ili se elektroničkim uređajem koristi prilikom

ispunjavanja anketnog upitnika ili ne, bez obzira što je u ovome slučaju peterotomna skala umjetno proširivanje dihotomne da/ne skale u vidu korištenja elektroničkog uređaja.

Osim što smo provjerili utjecaj nemarnog odgovaranja na pristajanje mjernih i strukturalnih modela, provjerili smo postoji li razlika u korelaciji između dva međusobno polarno obrnuta iskaza s obzirom na klasifikaciji sudionika kao nemarnih ili pažljivih. Prvi iskaz je glasio: „*U posljednje vrijeme mediji posvećuju premalo prostora nasilju nad ženama*“, a drugi „*U posljednje vrijeme mediji posvećuju previše prostora nasilju nad ženama*“. Odabrali smo upravo ovaj iskaz jer je njegovo obrtanje rezultiralo u nama najjasnije vidljivoj polarnoj suprotnosti dvaju iskaza. Ako su ova dva iskaza zaista polarne suprotnosti i ako je pristup klasifikaciji sudionika na pažljive i nemarne ispravan, korelacije između ova dva iskaza za pažljive sudionike trebao bi se približiti iznosu -1 ili bi se barem trebao statistički značajno razlikovati s obzirom na grupu u koju su sudionici klasificirani. Upravo se na temelju statistički (ne)značajne razlike u korelacijama između grupa konstruirala završna mjera koja se koristila u analizi. Prvo se za svaki od dva posebna iskaza zasebno provjerilo uspješnost razlikovanja i klasifikacije sudionika, tako da su se usporedile njihove korelacije u odnosu na par polarno suprotnih iskaza. Valjani IMC iskaz klasificira sudionike tako da se među dvjema grupama pronalaze statistički značajne razlike u korelacijama. Ako su oba iskaza valjana, sljedeći korak je upotreba oba iskaza u finalnoj klasifikaciji, gdje će sudionici biti pažljivi samo ako su slijedili obje upute.

Naravno, ovakav pristup ima određena ograničenja. Prvo ograničenje vezano je uz nacrt istraživanja. Naime, ne možemo sa sigurnošću *a priori* znati da su odabrana dva međusobno obrnuta iskaza stvarno polarno suprotni jedan drugome. Idealno, za sve iskaze korištene u originalnoj skali *suvremenog seksizma* bi se konstruirali njihove polarne suprotnosti te bi se za validaciju IMC iskaza koji se koristi za potrebe mjerenja nemarnog odgovaranja odabrala kombinacija iskaza čija je korelacija najbliža -1. Preporučamo obrtanje jednog od iskaza koji se već koristi u skali, jer bi implementacija iskaza sadržajno nevezanih za temu mogla utjecati na ishod klasifikacije. Na primjer, ako sudionik daje odgovore na 15 čestica koje su sadržajno vezane uz *suvremeni seksizam*, veća je šansa da će primijetiti dva polarno suprotna iskaza koji se sadržajno ističu u odnosu na druge korištene iskaze te na primjer, mjere *političku orijentaciju*.

Drugo, čak i korištenjem dva posebna iskaza, vjerojatnost da sudionik slučajno ispravno odgovori na oba je 4%, ako se radi o peterotomnoj skali.

Pristranost slaganja s iskazima

Pristranost slaganja s iskazima odlučili smo mjeriti takozvanim indirektnim pristupom, koji nalaže da se treba specificirati posebna skala namijenjena isključivo mjerenu pristranosti slaganja. Kako bismo efikasno uključili pristranost slaganja s iskazima u model, specificiran je indeks pristranosti slaganja s iskazima bez obzira na njihov sadržaj odnosno NARS (engl. *net acquiescence response style*). Rezultat pojedinog sudionika na tom indeksu je aritmetička sredina svih odgovora na skali od pet heterogenih iskaza korištena u anketnom upitniku, za koje smo pretpostavili kako ne dijele zajednički sadržaj. Skala je blago modificirana verzija one koju je Greenleaf (1992) koristio kako bi mjerio ekstremni stil odgovaranja, te ju je sastavio s ciljem da iskazi budu heterogeni i da ne dijele zajednički sadržaj. Vrijednost Cronbachovog koeficijenta je iznosila 0,445, što je bio prvi korak u potvrđivanju pouzdanosti korištene skale. Iako je među nekim varijablama pronađena statistički značajna korelacija, vrijednost koeficijenta nikada ne prelazi $r = 0,219$.

Osim ovog pristupa, koji je primarno usmjeren na mjerne i strukturne modele, odlučili smo koristiti i ipsatizaciju rezultata. Ipsatizacija se vršila tako da se uzela aritmetička sredina svih odgovora na skali *svremenog seksizma*. Nakon toga, od svakog pojedinog rezultata na skali *svremenog seksizma*, oduzet je iznos koji je ipsatizacijom dobiven za svakog sudionika. Nakon ovoga, također je konstruiran mjerni model skale *svremenog seksizma* te su se usporedili podatci modela.

Indeks pristranosti slaganja i nemarnog odgovaranja

Budući da određeni broj istraživača i dalje nije upoznat sa strukturalnim modeliranjem i/ili ne koriste ovu metodologiju, odlučili smo pokušati konstruirati indeks pomoću kojeg će biti moguće otežati varijable skale *svremenog seksizma*, čime se pristranost slaganja i nemarno odgovaranje pokušalo kontrolirati. Ovakav pristup mogao bi biti pristupačniji jer ne zahtjeva znanje strukturalnog modeliranja, ali na temelju rezultata samo jedne studije ne možemo sa sigurnošću vjerovati njegovu ispravnost. Budući da ne smatramo da sve pristranosti i efekti metode u istoj

mjeri utječu na rezultate skale pa ni da su ti utjecaji jednaki za sve varijable neke skale, odlučili smo koristiti pristup koji je u obzir uzeo ovu pretpostavku. Konstrukciji indeksa pristupili smo zasebno za svaku varijablu, jer određene pristranosti i efekti metode mogu jače ili slabije djelovati ovisno o pojedinoj varijabli. U obzir smo uzimali korelacije koju varijable skale *suvremenog seksizma* imaju s određenim indikatorima pristranosti slaganja i nemarnog odgovaranja te u skladu s time otežali rezultate. Odabrali smo koeficijente korelacija kao element otežavanja rezultata jer smatramo da je on najbolji pokazatelj bivarijatnog suodnosa pojedine varijable i indikatora određene pristranosti.

8.2.Rezultati

Prvu instruktivnu manipulaciju, „*Molimo Vas odaberite opciju 1 za ovo pitanje*“, slijedilo je 83,4% sudionika. Korelacije između dva suprotno obrnuta iskaza, „*U posljednje vrijeme mediji posvećuju previše prostora nasilju nad ženama*“ i „*U posljednje vrijeme mediji posvećuju premalo prostora nasilju nad ženama*“, bili su statističke značajne za grupu pažljivih sudionika [$r = -,334$; $p < ,01$], ali i za grupu nemarnih sudionika [$r = -,256$; $p < ,05$]. Test nije rezultirao statistički značajnom razlikom među korelacijama dvaju grupa [$p = ,49$], a i njihovi koeficijenti korelacije daleko su od iznosa -1, čime moramo zaključiti kako sudionici iskaze nisu percipirali kao međusobne antonime/suprotnosti. Upravo ovdje leži prednost implementacije višestrukog broja antonima, jer se *a priori* ne može sa sigurnošću znati njihova korelacija. Zbog ograničenog broja varijabli koje su se mogle implementirati u upitnik, bila nam je omogućena implementacija samo jednog ovakvog para obrnutih iskaza. Među dvjema grupama nije pronađena statistički značajna razlika između njihovog prosjeka na NARS skalama [$p = ,148$]. Nalaz koji ide u prilog validacije ove klasifikacije jest statistički značajna razlika unutar kojeg su dvije grupe sudionika ispunile upitnik [$t_{pažljivi} = 26,20$ minuta, $t_{nemarni} = 18,28$ minuta; $F = 1,151$; $p = ,284$; $t = 2,170$; $p = ,031$]. Također, pronađena je statistički značajna razlika među grupama između njihovog prosjeka na svim iskazima skale pristranosti slaganja i *suvremenog seksizma* [$p = ,003$]. Navedeni prosjek je zapravo uprosječni rezultat grupe na svim relevantnim iskazima, a pokazalo se da grupa nemarnih sudionika u prosjeku daje više rezultate na iskazima od grupe pažljivih sudionika. Sukladno ovome, pronađena je očekivana statistički značajna razlika između uprosječenih rezultata dvaju grupa nakon ipsatizacije [$p = ,018$]. Ovaj podatak govori da su nemarni sudionici, čak i nakon ipsatizacije, davali statistički značajno različite odgovore na skali

suvremenog seksizma u odnosu na grupu pažljivih sudionika. Na varijablama skale *suvremenog seksizma*, opciju „*U potpunosti se slažem*“ odabralo je 6,2% pažljivih sudionika, dok je istu opciju odabralo 10% nemarnih sudionika, što znači da je grupa nemarnih sudionika nešto češće odabrala ekstremni odgovor, makar razlika nije statistički značajna [$p = ,213$].

Drugi iskaz namijenjen praćenju nemarnog odgovaranja, „*Trenutno se koristim elektroničkim uređajem*“, uspješno je slijedilo 83,2% sudionika. 91,5% sudionika odabralo je opcije „*Donekle se slažem*“ ili „*U potpunosti se slažem*“, ali smo iz ranije navedenih razloga odlučili priznati samo opciju „*U potpunosti se slažem*“, što je slijedilo 83,2% sudionika. Korelacija među polarno obrnutim iskazima je statistički značajna za skupinu pažljivih sudionika [$r = -,323$; $p < ,01$], ali nije značajna za skupinu nemarnih sudionika [$r = -,211$; $p = ,059$]. Razlika među koeficijentima korelacija nije se pokazala statistički značajnom [$p = ,033$], makar se to može pripisati razlici u veličini grupa, jer je razlika u koeficijentima vidljiva. Kao i kod prvog IMC iskaza, koeficijenti korelacije dvaju polarno obrnuta iskaza nisu blizu iznosa -1, čime zaključujemo da ih sudionici nisu percipirali kao međusobne antonime, makar je ovdje razlika među grupama veća. Ovaj pronalazak definitivno utječe na ispravnost klasifikacije, jer nam je nezavisni kriterij validacije kontaminiran. Nadalje, među grupama koje su dobivene na temelju klasifikacije drugog IMC iskaza nije pronađena statistički značajna razlika s obzirom na vrijeme ispunjavanja upitnika [$t_{\text{pažljivi}} = 26,02$ minuta, $t_{\text{nemarni}} = 20,12$ minuta; $F = 0,33$; $p = ,857$; $t = 2,170$; $p = ,106$], makar na prvi pogled razlika postoji. Koeficijent korelaciju između dva korištena IMC iskaza [$r = ,307$; $p < ,01$] govori u prilog tome da postoji razlika s obzirom na to koji sudionici su klasificirani kao nemarnima u dva slučaja. Među dvjema grupama su pronađene statistički značajne razlike s obzirom na njihov rezultat na NARS indeksu [$p < ,001$], ali smjer razlike nije u skladu s očekivanjima, jer grupa pažljivih sudionika ima viši prosjek. No, grupa nemarnih sudionika ima statistički značajno viši rezultat na uprosječnoj vrijednosti sastavljenoj od svih relevantnih iskaza [$p < ,001$], što je u skladu s očekivanjima. Sukladno tome, pronađena je očekivana statistički značajna razlika između uprosječenih rezultata dvaju grupe nakon ipsatizacije [$p < ,001$], gdje grupa nemarnih sudionika ima viši prosjek. Na varijablama skale *suvremenog seksizma*, opciju „*U potpunosti se slažem*“ odabralo je 6,8% pažljivih sudionika, dok je istu opciju odabralo 7,1% nemarnih sudionika, što znači da je grupa nemarnih sudionika u gotovo istoj mjeri odabrala ekstremni odgovor.

Ovaj podatak, zajedno s time da nisu pronađene statistički značajne razlike u koeficijentima korelacija s obzirom na grupe u oba slučaja, nas je naveo da provjerimo klasifikaciju za oba iskaza istovremeno: samo oni koji su slijedili obje upute bit će klasificirani kao pažljivi. 73,7% sudionika slijedilo je obje upute, a ovaj podatak sugerira da, ako se radi o ispravnoj klasifikaciji, četvrtinu ukupnog uzorka čine nemarni sudionici. Koeficijenti korelacije između dva polarno suprotna iskaza su statistički značajna za skupinu pažljivih sudionika [$r = -,329$; $p < ,01$], ali i za skupinu nemarnih sudionika [$r = -,278$; $p < ,01$]. Kao ni u prijašnjim slučajevima, razlika u njihovim koeficijentima korelacije nije bila statistički značajna [$p = ,59$], s time da je razlika u apsolutnom smislu i najmanja. Grupa nemarnih sudionika je statistički značajno kraće ispunjavala upitnik [$p = ,028$]. Među dvjema grupama pronađena je statistički značajna razlika u njihovom rezultatu na NARS indeksu [$p = ,005$], s time da grupa pažljivih sudionika ima viši rezultat, što nije u skladu s očekivanjima. Pronađene su i razlike na uprosječnoj vrijednosti svih relevantnih iskaza [$p < ,001$] te na istoj vrijednosti nakon ipsatizacije [$p < ,001$], a smjer razlike je u skladu s očekivanjima. Svi podatci su prikazani u tablici 6.

Generalno, niti jedan od IMC iskaza nije zadovoljio sve preduvjete, ali je među grupama određenim putem drugog IMC iskaza pronađena najveća razlika u koeficijentima korelacija između antonima, vidljiva, iako statistički neznčajna razlika u trajanju ispunjavanja upitnika te statistički značajna razlika u uprosječnoj vrijednosti skale *suвременog seksizma*. Kako bismo dodatno provjerili koeficijente korelacija, izračunali smo 95% intervale pouzdanosti koristeći *bootstrap* metodu na 2000 poduzoraka te pronašli da je standardna pogreška procjene koeficijenta korelacije nemarnih sudionika ($e = ,121$) više nego duplo veća od standardne pogreške procjene koeficijenta korelacije pažljivih sudionika ($e = ,051$), a i grupa nemarnih sudionika ima puno veći raspon 95% intervala pouzdanosti [$,037$ do $-,431$] u odnosu na grupu pažljivih sudionika [$-,221$ do $-,422$]. Navedeno možemo pripisati razlikama u veličinama grupa pažljivih sudionika [$N = 402$] i nemarnih sudionika [$N = 81$]. Kada smo specificirali da su obje grupe iste veličine [$N = 402$], razlika među koeficijentima korelacija postala je statistički značajna [$p = ,009$], a isto se dogodi ako su obje grupe veličine grupe nemarnih sudionika [$N = 81$; $p = ,045$]. Iz ovih razloga, klasifikaciju sudionika na temelju ovog IMC iskaza smatramo najvaljanijom, makar ne možemo tvrditi da je u potpunosti ispravna.

Tablica 6. Usporedba rezultata klasifikacije IMC iskaza

	<i>Pažljivi</i>	<i>Nemarni</i>	<i>p</i>	<i>U skladu s očekivanim</i>
IMC iskaz 1	N = 403	N = 80		
<i>Klasifikacija</i>	83,4%	16,6%		
<i>Korelacija antonima</i>	-,334*	-,256*	,492	Ne
<i>Vrijeme ispunjavanja</i>	26:20	18:28	,031	Da
<i>NARS</i>	3,58	3,46	,091	Ne
<i>Uprosječna vrijednost</i>	2,72	2,96	,003	Da
<i>Uprosječna vrijednost nakon ipsatizacije</i>	-,09	-,06	,018	Ne
IMC iskaz 2	N = 402	N = 81		
<i>Klasifikacija</i>	83,2%	16,8%		
<i>Korelacija antonima</i>	-,323*	-,211	,329	Ne
<i>Vrijeme ispunjavanja</i>	26:02	20:12	,106	Ne
<i>NARS</i>	3,61	3,33	<,001	Da
<i>Uprosječna vrijednost</i>	2,71	3,33	<,001	Da
<i>Uprosječna vrijednost nakon ipsatizacije</i>	-,09	-,03	<,001	Ne
Spojeni IMC iskazi	N = 356	N = 127		
<i>Klasifikacija</i>	73,7%	22,3%		
<i>Korelacija antonima</i>	-,329*	-,278*	,590	Ne
<i>Vrijeme ispunjavanja</i>	26:48	20:06	,028	Da
<i>NARS</i>	3,61	3,42	,005	Da
<i>Uprosječna vrijednost</i>	2,69	2,94	<,001	Ne
<i>Uprosječna vrijednost nakon ipsatizacije</i>	-,09	-,05	<,001	Ne

8.3. Razlike u nemarnom odgovaranju s obzirom na sociodemografske karakteristike

Provjerili smo rezultate klasifikacije nemarnih i pažljivih sudionika s obzirom na njihove sociodemografske karakteristike. S obzirom na sve tri klasifikacije, veći udio pripadnica ženskog spola klasificirano je kao pažljivima u odnosu na pripadnike muškog spola, a ta razlika je statistički značajna s obzirom na drugi IMC iskaz. Nadalje, sudionici koji imaju završenu višu stručnu spremu i više, u većoj mjeri su pažljivi od sudionika koji imaju završenu srednju školu i niže, a ta razlika nije statistički značajno jedino za drugi IMC iskaz. Rezultati su prikazani u tablici 7. S obzirom na stupanj obrazovanja, ali i dob, nisu pronađene statistički značajne razlike među grupama u njihovim koeficijentima korelacija među dva polarno suprotna iskaza.

Tablica 7. Usporedba sociodemografskih karakteristika i rezultata klasifikacije

	Nemarni	Pažljivi	p
SPOL			
IMC iskaz 1	N = 403	N = 80	
<i>Muško</i>	18,8%	81,2%	,370
<i>Žensko</i>	15,3%	84,7%	
IMC iskaz 2	N = 402	N = 81	
<i>Muško</i>	25,3%	74,7%	< ,001
<i>Žensko</i>	12,1%	87,9%	
Spojeni IMC iskazi	N = 356	N = 127	
<i>Muško</i>	30,6%	69,4%	,071
<i>Žensko</i>	24%	76%	
OBRAZOVANJE			
IMC iskaz 1	N = 403	N = 80	
<i>Završena srednja škola i niže</i>	20,8%	79,2%	,028
<i>Završena Viša stručna sprema i više</i>	13,7%	86,3%	
IMC iskaz 2	N = 402	N = 81	
<i>Završena srednja škola i niže</i>	18,2%	81,8%	,282
<i>Završena Viša stručna sprema i više</i>	15,8%	84,2%	
Spojeni IMC iskazi	N = 356	N = 127	
<i>Završena srednja škola i niže</i>	32,3%	67,7%	,010
<i>Završena Viša stručna sprema i više</i>	22,3%	77,7%	
DOB			
IMC iskaz 1	N = 403	N = 80	
<i>Do 35 godina</i>	16,5%	83,5%	,735
<i>36-50 godina</i>	17,7%	82,3%	
<i>Više od 50 godina</i>	13,7%	86,3%	
IMC iskaz 2	N = 402	N = 81	
<i>Do 35 godina</i>	18,8%	81,2%	,299
<i>36-50 godina</i>	16,7%	83,3%	
<i>Više od 50 godina</i>	11%	89%	
Spojeni IMC iskazi	N = 356	N = 127	
<i>Do 35 godina</i>	27,5%	72,5%	,112
<i>36-50 godina</i>	28,6%	71,4%	
<i>Više od 50 godina</i>	16,4%	83,6%	

8.4. Dimenzionalnost skale suvremenog seksizma

Prije nego što smo implementirali kontrolu utjecaja nemarnog odgovaranja i pristranosti slaganja, odlučili smo provjeriti dimenzionalnost skale *suvremenog seksizma*. Na temelju prethodnih istraživanja (Šimac i Klasnić, 2021) specificirali smo jednofaktorski, ali i dvofaktorski model skale *suvremenog seksizma*. Zbog asimetrične prirode uključenih varijabli, koristili smo procjenu maksimalne vjerojatnosti (engl. *maximum likelihood estimator*) s robusnim standardnim pogreškama (MLR; White, 1980) i skaliranim statistikom koji je asimptotičko jednak Yuan-Bentlerovom (2007) testovnom statistiku. Pristajanje jednofaktorskog modela nije bilo zadovoljavajuće [$\chi^2 = 276,06$; $df = 54$; $p < ,001$; CFI = ,886; TLI = ,860; RMSEA = ,105; 90% CI = ,093-,118; SRMR = ,059], pogotovo jer bi vrijednosti parametara slaganja za jednostavne modele generalno trebale biti znatno više. Pristajanje dvofaktorskog modela bilo je adekvatno [$\chi^2 = 146,38$; $df = 53$; $p < ,001$; CFI = ,952; TLI = ,940; RMSEA = ,06; 90% CI = ,056-,082; SRMR = ,041] te je dvofaktorska solucija statistički značajno bolje pristajala podacima [$p < ,001$]. Svi iskazi visoko saturiraju na prvom faktoru [,616-,795] i na drugom faktoru [,703-,756], a latentne dimenzije su snažno povezane [,796]. Iz ovih razloga, kontrolu utjecaja nemarnog odgovaranja i pristranosti slaganja analizirali smo u usporedbi s dvofaktorskom solucijom skale *suvremenog seksizma*.

8.5. Kontroliranje utjecaja nemarnog odgovaranja i pristranosti slaganja

U nastojanju da uspješno kontroliramo utjecaje nemarnog odgovaranja i pristranosti slaganja, specificirali smo osam modela, od kojih su njih šest međusobno alternativni. Dva modela koja nisu direktno usporediva s ostalima su oni koji kao manifestne varijable koriste indikatore na kojima je izvršena ipsatizacija, što znači da ta dva modela ne sadrže iste manifestne varijable kao prvih šest modela. Svi modeli se sastoje od dvofaktorske strukture *suvremenog seksizma*, te ovisno o inačici, dodatno sadrže jednu ili dvije latentne dimenzije koje reprezentiraju pristranosti/efekte metode koji su međusobno (ne)korelirani, ali i (ne)korelirani sa sadržajnim dimenzijama *suvremenog seksizma*. Svi modeli su predstavljeni u prikazu 4.

Model A se sastoji od dvije dimenzije *suvremenog seksizma*, kao i od jedne latentne dimenzije koja u sebi sadrži indikatore nemarnog odgovaranja i pristranosti slaganja. Sadržajne

dimenzije su u nultoj korelaciji s latentnom dimenzijom efekta metode, ali su u svim modelima dopuštene njihove međusobne korelacije. Nemarno odgovaranje predstavljeno je s dvije varijable pomoću koje smo vršili klasifikaciju sudionika. Budući da niti jedan od IMC iskaza nisu rezultirali klasifikacijom u koju možemo biti u potpunosti sigurni te da se isto dogodilo s njihovim kompozitom, odlučili smo koristiti oba pojedinačna IMC iskaza kao indikatore nemarnog odgovaranja. Svjesni smo ograničenja i potencijalnog problema ovakvog pristupa, ali smatramo da je implementacija dvije pojedinačne varijable bolja od odabira jedne od njih ili odabira njihovog kompozita. Kompozit nije uključen jer je on linearna kombinacija dva posebna IMC iskaza, što bi stvaralo probleme unutar modela. Dodatno, kao indikatore pristranosti slaganja specificirali smo pet varijabli koje su korištene za mjerenje pristranosti slaganja. Model B je identičan kao Model A, s time da su sadržajnim dimenzijama omogućene korelacije s latentnom dimenzijom efekta metode.

Model C odvaja latentne dimenzije efekta metode na latentnu dimenziju pristranosti slaganja i latentnu dimenziju nemarnog odgovaranja. Sadržajne latentne dimenzije su u nultoj korelaciji s dimenzijama efekta metode, s time da su dimenzije efekta metode međusobno također u nultoj korelaciji. Svi indikatori su isti kao i u prethodnim modelima. Model D se od modela C razlikuje po tome da se dopustila korelacije između dvije sadržajne dimenzije i latentnih dimenzija efekata metode, a unutar modela E dopušta korelacije među svim latentnim dimenzijama. Model F dopušta korelacije među latentnim dimenzijama efekta metode, ali ne i njihovu korelaciju sa sadržajnim dimenzijama. Unutar Modela G koriste se indikatori *suvremenog seksizma* nakon ipsatizacije, čime više nije potrebno koristiti varijable koje su korištene unutar skale pristranosti slaganja. Samim time, specificirala se samo jedna latentna dimenzija efekta metode koja se odnosila na nemarno odgovaranje. Unutar Modela G latentna dimenzija efekta metode je u nultoj korelaciji sa sadržajnim dimenzijama, a u Modelu H ta je korelacija dopuštena.

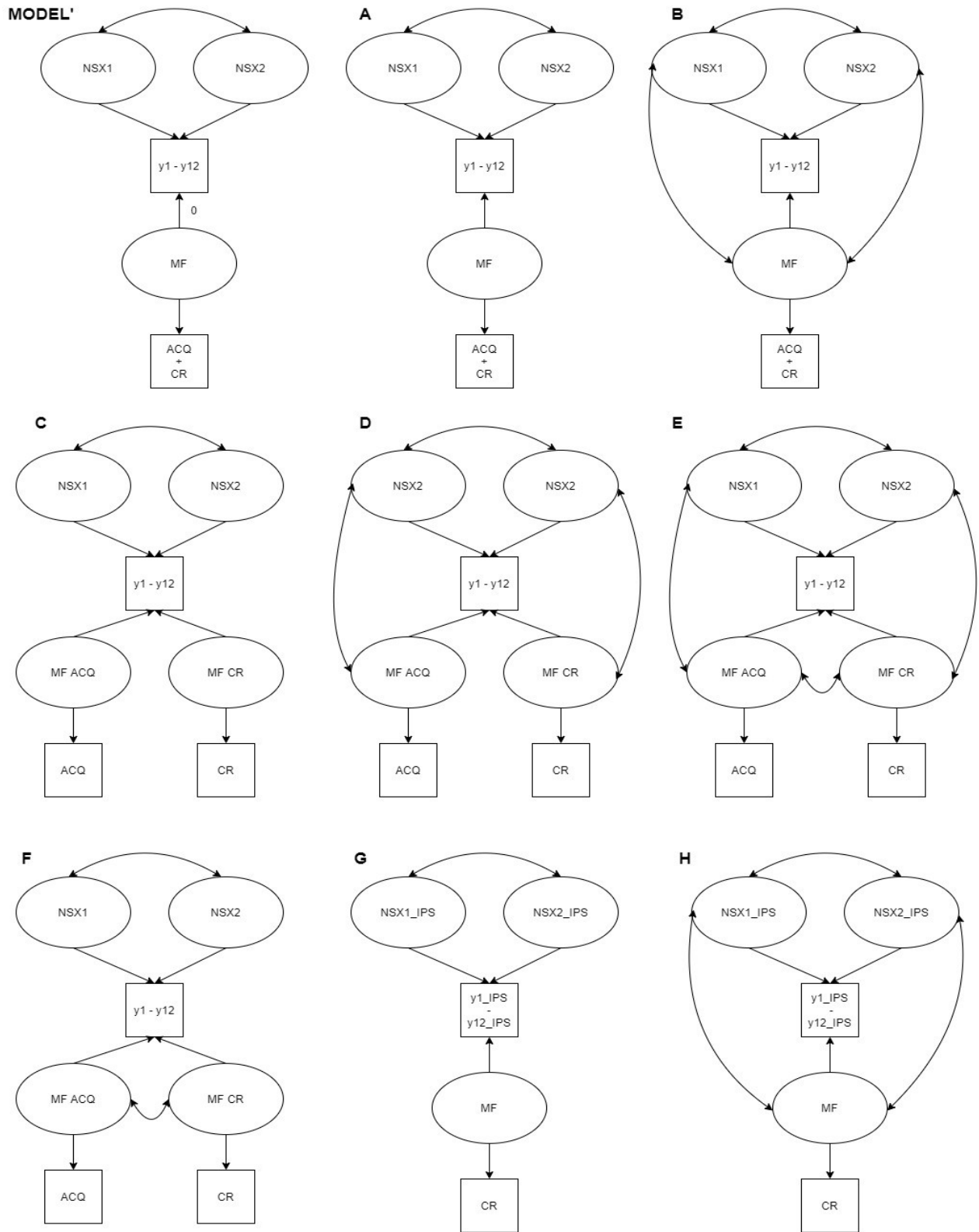
Za Modele G i H, na varijablama koje su mjerile *suvremeni seksizam* smo napravili ipsatizaciju. Kreirali smo varijablu koja je predstavljala prosječnu vrijednost na svih 12 čestica koja je predstavljala uprosječenu vrijednost pristranosti slaganja na skali *suvremenog seksizma* za svakog ispitanika. Potom smo od svake od 12 varijabli oduzeli tu vrijednost čime je kreiran novi sklop 12 varijabli. Koristili smo uprosječenu vrijednost na 12 varijabli skale *suvremenog*

seksizma, a ne na svim varijablama, zbog pretpostavke da se razne pristranosti i utjecaji efekata metode mogu razlikovati od skale do skale.

$$Ips = \frac{V1 + V2 + V3 \dots + Vn}{n}$$

$$V_1_Ips = V_1 - Ips$$

U tablici 8 nalaze se rezultati pristajanja svih objašnjenih modela, uključujući i model koji se sastoji samo od dvije korelirane dimenzije *suвременog seksizma* i njihovih indikatora. Modeli A-H nisu direktno usporedivi s originalnim modelom, jer sadrže dodatne ili drugačije manifestne varijable. Upravo iz tog razloga, specificirali smo model koji je duplikacija Modela A (najjednostavnijeg modela kontrole), ali gdje su korelacije između latentne dimenzije kontrole i manifestnih varijabli *suвременog seksizma* specificirane na 0, čime se s obzirom na okolnosti maksimalno sprječava njihov utjecaj i rezultati parametara slaganja su najsličniji dvofaktorskom modelu nekontroliranja bilo kakvih utjecaja iz poglavlja o dimenzionalnosti skale *suвременog seksizma*. Ovaj model smo nazvali Model'. Iako ovakva specifikacija nije idealna i rezultira nešto drugačijim parametrima slaganja, omogućuje usporedbu sa svim alternativnim modelima koji uključuju kontrolu pristranosti slaganja i nemarnog odgovaranja.



Prikaz 4: Pojednostavljeni modeli kontroliranja pristranosti slaganja i nemarnog odgovaranja; NSX1-NSX2 = latentne dimenzije neoseksizma; MF = latentna dimenzija pristranosti ili efekta metode; ACQ = pristranost slaganja; CR = nemarno odgovaranje; IPS = varijable nakon ipsatizacije

Između Modela' i Modela A pronađena je statistički značajna razlika u slaganju [$p < ,001$], što znači da Model A bolje odgovara podacima. Ovaj nalaz ide u prilog kontroliranju utjecaja raznih pristranosti i efekata metode. Model B po parametrima slaganja ne daje bolje rezultate od Modela A, tako da ne čudi što je skalarni faktor negativan, što znači da Model B nije unaprjeđenje u pristajanju podacima u odnosu na Model A. Model C je pokazao statistički značajne bolje pristajanje podacima u odnosu na Model A [$p = ,011$], a između Modela D i Modela C se dogodila ista situacija kao između Modela A i Modela B, čime se trenutno Model C pokazao kao specifikacijom koja pokazuje najbolje pristajanje. Između Modela C i Modela E te između Modela C i Modela F se također javljaju negativni skalarni faktori, čime na temelju rezultata zaključujemo da od svih međusobno alternativnih modela, Model C najbolje pristaje podacima te da je njegovo pristajanje statistički značajno bolje od „originalnog“ modela odnosno Modela'. Negativni skalarni faktor označava da kompleksniji model po svojim opaženim parametrima slaganja pokazuje slabije rezultate u odnosu na jednostavniji model, čime se usporedba modela ne može provesti, jer je pretpostavka da kompleksniji model pokazuje bolje parametre pristajanja podacima. Iako ovakva usporedba nije idealna, rezultati pokazuju kako se kontrolom i dalje može dobiti statistički značajno pristajanje podacima. Ovaj nalaz je u skladu s rezultatima nekih istraživanja koja su pokazala da upravo ovakvi modeli najbolje pristaju podacima (Reise et al., 2007).

Tablica 8. Pristajanje modela nakon kontrole utjecaja pristranosti slaganja i nemarnog odgovaranja
UZORAK (N = 483)

	χ^2	df	p	CFI	TLI	RMSEA	90% CI	SRMR
<i>Model original</i>	146,380	53	,000	,952	,940	,069	,056 - ,082	,041
<i>Model original'</i>	283,395	103	,000	,921	,908	,065	,056 - ,074	,077
<i>Model A</i>	210,079	91	,000	,951	,935	,055	,045 - ,064	,052
<i>Model B</i>	205,462	89	,000	,951	,934	,055	,045 - ,065	,052
<i>Model C</i>	185,938	79	,000	,961	,941	,053	,043 - ,063	,039
<i>Model D</i>	197,752	75	,000	,957	,932	,056	,047 - ,066	,039
<i>Model E</i>	145,598	74	,000	,970	,951	,047	,036 - ,059	,030
<i>Model F</i>	153,469	78	,000	,970	,954	,046	,035 - ,057	,030
<i>Model G</i>	154,949	64	,000	,858	,798	,058	,046 - ,070	,057
<i>Model H</i>	150,095	62	,000	,858	,791	,059	,047 - ,071	,057

Napomena: Svi indikatori su robusne procjene izračunate koristeći MLR

Modeli G i H jasno pokazuju da ipsatizacija rezultata odnosno nastojanje uklanjanja utjecaja pristranosti slaganja na ovaj način rezultira modelima koji pokazuju slabije parametre slaganja u odnosu na originalni model. Iako na temelju korištenog nacrt istraživanja ne možemo

sa sigurnošću odrediti je li ipsatizacija (ne)ispravan način kontrole pristranosti slaganja, rezultati njezine provedbe nisu u skladu s metodom implementacije latentnih dimenzija u mjerne modele, koji su se pokazali korisnima. Odlučili smo provjeriti utjecaj ipsatizacije na rezultate skala *savjesnosti* i *ugodnosti* iz prve studije, kako bismo vidjeli utječe li ipsatizacija negativno na pristajanje modela. Nakon što je ipsatizacija na varijablama provedena, niti jedan od četiri specificirana jednodimenzionalna modela nije konvergirao, što znači da im strukture značajno odstupaju od jednodimenzionalnih. No, budući da su na skalama rađene adaptacije za potrebe mjerenja utjecaja formulacije iskaza, ovo se djelomično može pripisati i upravo tome. Iz tih razloga, odlučili smo provesti ipsatizaciju na skali *intelekta*, koja je pokazala maksimalni stupanj mjerne invarijantnosti s obzirom na poduzorke. Nakon ipsatizacije, model nije konvergirao niti u jednom od oba poduzorka. Ovo je definitivni pokazatelj, na više skala iz međusobno različitih dimenzija, da je utjecaj ipsatizacije negativan na pokazatelja pristajanja modela. Iz ovih razloga, predlažemo kontrolu putem strukturalnog modeliranja.

Kako bismo dodatno argumentirali strukturalni pristup validaciji podataka, proveli smo multigrupne analize gdje smo dodatno provjerili utjecaje efekata metode, čime smo pokušali opravdati potrebu za njihovom kontrolom. Kada se specificirao jednostavni dvofaktorski multigrupni model, koji se sastojao samo od dvije međusobno povezane korelirane dimenzije *suvremenog seksizma*, a gdje su grupe bile nemarni i pažljivi sudionici, konfiguralni model nije konvergirao, što znači da se dimenzionalnost ovog konstrukta razlikuje s obzirom na grupe. Ovakav ishod dogodio se i za prvi IMC iskaz, i za drugi IMC iskaz, i za njihov kompozit. Potrebno je napomenuti da ovaj rezultat jednim dijelom može biti i posljedica velike razlike u veličinama grupa, makar veličina manje grupe u niti jednom slučaju nije toliko mala da se multigrupna analiza ne bi mogla provesti.

Kako bismo multigrupnu analizu proveli i za sudionike koji su pokazali više ili manje mjere pristranosti slaganja, odredili smo granice po kojima su se sudionici podijelili u jednu od dvije grupe. Budući da se od opaženih rezultata na 12 varijabli skale *suvremenog seksizma* oduzeo uprosječen rezultat na tih 12 varijabli, rezultati na novodobivenih varijablama su pozitivni ili negativni u svom predznaku, ovisno o razini pristranosti slaganja. Isto tako, novoizračunati uprosječen rezultat na toj skali bio je pozitivan ili negativan u svome predznaku te je predstavljao više ili manje razine pristranosti slaganja, ovisno o sudioniku. Novoizračunati

uprosječeni rezultat bio je iznosa -0,087. Ako se slijedi premisa da je ovakav način mjerenja pristranosti slaganja validan, zaključujemo da je pristranost slaganja bitno utjecala na rezultate. Kao potvrdu ovog argumenta, proveli smo t-test, gdje smo inicijalni prosjek na 12 varijabli skale *suvremenog seksizma* usporedili s istim rezultatom nakon ipsatizacije [$X_1 = 2,509$; $X_2 = 2,422$; $p = ,016$]. Rezultat $X_2 = 2,422$ jednostavno predstavlja rezultat oduzimanja novoizračunatog uprosječenog rezultata iznosa 0,087 od vrijednosti $X_1 = 2,509$. Nakon ovoga, pronašli smo granicu razlike X_1 i X_2 koja bi i dalje rezultirala statistički značajnom razlikom u prosjecima na ovom uzorku te dobili veličinu te razlike koja je iznosila ,071.

$X_1 = 2,509$	$X_2 = 2,438$	$p = ,049$
$X_1 = 2,509$	$X_2 = 2,439$	$p = ,052$
$X_1 = 2,509$	$X_2 = 2,580$	$p = ,050$
$X_1 = 2,509$	$X_2 = 2,579$	$p = ,053$

Ovaj smo korak proveli jer smo htjeli odvojiti sudionike kojima se nakon uračunavanja njihove pristranosti slaganja na skali *suvremenog seksizma* novoizračunati rezultati statistički značajno razlikuju od opaženih rezultata. Budući da dobivena granična vrijednost iznosi ,071, sve sudionike čiji se rezultat razlikovao iznad ili jednako $\pm,071$ svrstali smo u grupu gdje je izražena pristranost slaganja, a sve čiji se rezultati razlikuju manje od tog iznosa kao grupu gdje je manje izražena pristranost slaganja. Izabrali smo $\pm,071$ varijantu, jer kao što je moguće slaganje s iskazima bez obzira na njihov sadržaj, moguće i takozvano neslaganje s rezultatima bez obzira na njihov sadržaj, a smatramo da je izvorište utjecaja isto za obje inačice. 10,1% sudionika klasificirano je u grupu jakog utjecaja pristranosti slaganja, a 89,9% u grupu slabog utjecaja pristranosti slaganja. Konfiguralni multigrupni model pokazuje granično pristajanje podacima, pogotovo ako se u obzir uzme jednostavna dvofaktorska struktura specificiranog modela skale *suvremenog seksizma* [CFI = ,924; RMSEA = ,082], što je i osjetni pad s obzirom na pristajanje originalnog modela [Δ CFI = -,38; Δ CFI = +,13], čime se vidi suptilni utjecaj u odnosu na nemarno slaganje, ali i dalje prisutan utjecaj pristranosti slaganja na rezultate.

8.6. Indeks pristranosti slaganja i nemarnog odgovaranja

Konstrukciji indeksa smo pristupili individualno za svaku varijablu skale *suvremenog seksizma*. Kako bismo osigurali otežavanje s obzirom na obje pojave, koristili smo dizajn koji se može opisati kao svojevrsna križna validacija. Prvo smo grupama nemarnih/pažljivih sudionika

kontrolirali pristranost slaganja, a zatim smo grupama visoke/niske prisutnosti pristranosti slaganja kontrolirali nemarno odgovaranje.

Prvo smo uzorak podijelili na nemarne i pažljive sudionike. Budući da niti jedna od dvaju IMC iskaza nisu u potpunosti zadovoljili preduvjete, koristili smo onaj pomoću kojih je klasifikacija bila bolja s obzirom na relevantne pokazatelje: IMC iskaz 2. Potom smo korelirali varijable skale *suvremenog seksizma* s indeksom NARS, koji je služio kao mjera pristranosti slaganja, a te smo korelacije podijelili s obzirom na nemarnu/pažljivu skupinu. Potom smo opaženi rezultat sudionika na pojedinoj varijabli pomnožili s korelacijom koju ta varijabla ima s indeksom NARS, ovisno o grupi sudionika kojoj pripadaju. Ovime smo za obje skupine sudionika otežali rezultate s obzirom na korelaciju varijable s NARS indikatorom, čime su rezultati otežani s obzirom na pristranost slaganja. Ovaj postupak se ponovio za svih 12 varijabli skale *suvremenog seksizma*. Nakon ovog koraka, podijelili smo sudionike s obzirom na to jesu li bili svrstani u grupi u kojoj je visoko ili nisko izražena pristranost slaganja. Rezultate i logiku ove klasifikaciji opisali smo prethodnom poglavlju gdje smo opisivali rezultate multigrupnih analiza. U skladu s logikom prijašnjeg koraka, izračunali smo korelaciju svake varijable s indikatorom nemarnog odgovaranja (IMC iskaz 2), s obzirom na grupu visoko ili nisko izražene pristranosti slaganja te smo varijable otežali na isti način kao u prvom koraku.

U skladu s prethodna dva koraka, dobivene su dvije vrijednosti, koje predstavljaju otežanja s obzirom na pristranost slaganja i nemarno odgovaranje. Koristili smo sljedeći način implementacije ovih vrijednosti u kontrolirane rezultate: zbrajanje opaženog rezultata s dvije vrijednosti otežanja. Naime, navedene dvije vrijednosti imale su pozitivan ili negativan predznak, ovisno o korelaciji s indikatorima pristranosti slaganja i nemarnog odgovaranja. Ovo znači da su na određene varijable/grupe ove pristranosti djelovale na drugačiji način. Množenjem dvaju negativnih vrijednosti negativni utjecaji bi se međusobno poništili, čime finalni kontrolirani rezultat ne bi na valjani način predstavljali kontrolu oba fenomena zbog promjene finalnog predznaka. Nadalje, ovakav način ne mijenja drastično varijancu finalne vrijednosti u usporedbi s opaženom vrijednosti, kao što bi to napravilo množenje. Upravo iz ovih razloga, koristili smo sukcesivno zbrajanje, jer je takav pristup u obzir uzimao predznake svih vrijednosti.

$$V1_{ACQ} = V1 * cor(V1, NARS)$$

$$V1_{CR} = V1 * cor(V1, IMC_2)$$

$$V1_{KONTROLIRANA} = V1 + V1_{ACQ} + V1_{CR}$$

Tablica 9. Koeficijenti korelacija na temelju kojih su otežani rezultati varijabli

	Koeficijent korelacije s NARS indikatorom		Koeficijent korelacije s IMC iskazom 2	
	Pažljivi	Nemarni	Niska razina pristranosti slaganja	Visoka razina pristranosti slaganja
ssx1	,112	,177	,145	-,105
ssx2	,114	,087	,178	,071
ssx3	,115	-,046	,216	,124
ssx4	,101	,226	,247	-,184
ssx5	,106	,179	,130	-,201
ssx6	,113	,044	,216	-,247
ssx7	,112	,062	,244	,009
ssx8	,127	,240	,107	-,325
ssx9	,076	,083	,215	-,141
ssx10	,055	,253	,170	-,034
ssx11	,025	,173	,158	-,241
ssx12	,083	,141	,271	,005

Napomena: ssx1 – ssx12 su varijable skale *suвременog seksizma*

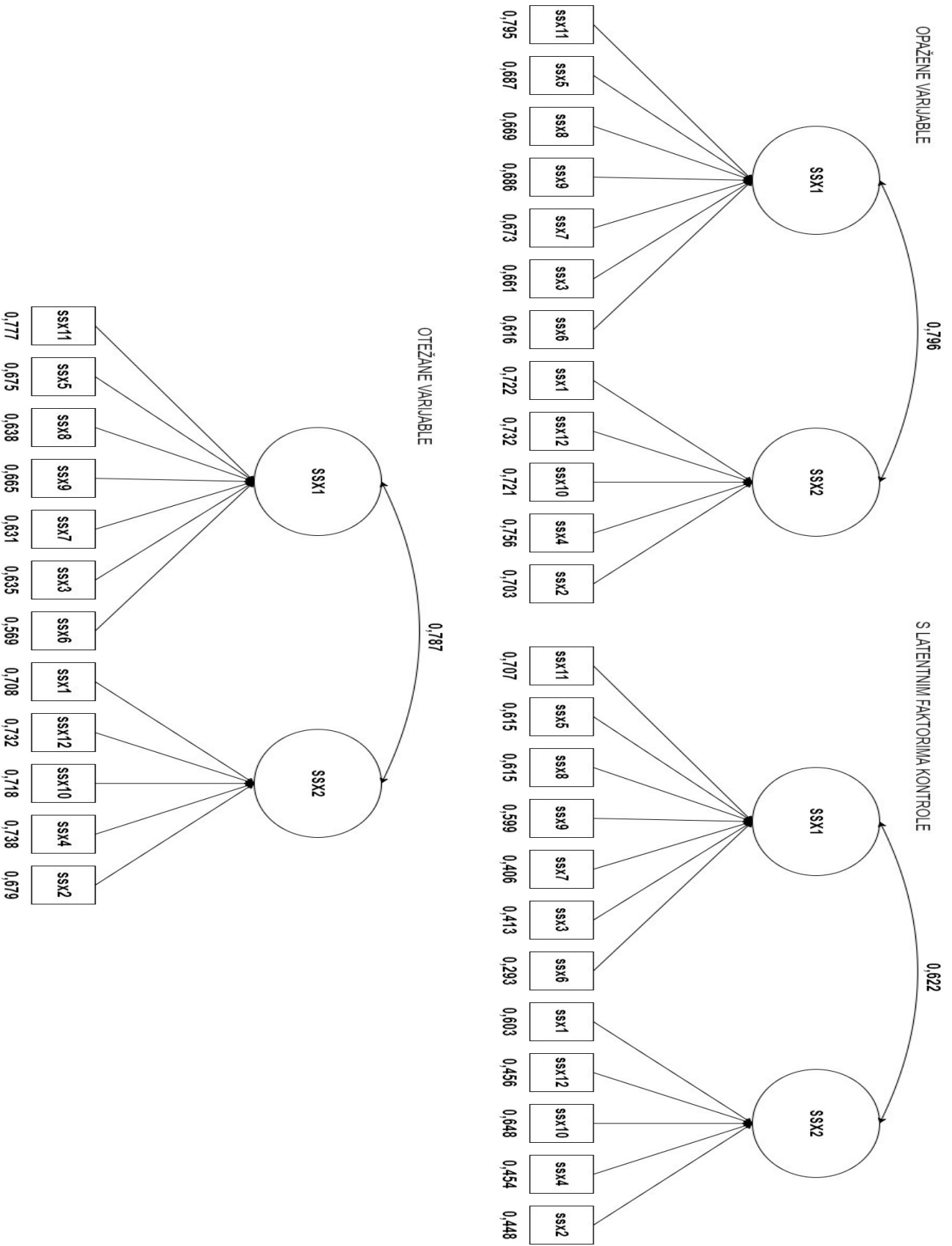
Nakon što smo otežali rezultate, proveli smo konfirmatornu faktorsku analizu na dvofaktorskom modelu s povezanim latentnim dimenzijama *suвременog seksizma*. Parametri pristajanja modela bili su nešto niži od poželjnih za ovako jednostavan model [$\chi^2 = 156,88$; $df = 53$; $p < ,001$; CFI = ,942; TLI = ,928; RMSEA = ,072; 90% CI = ,059-,085; SRMR = ,043]. Varijabla koja ima najnižu saturaciju na sebi nadređenu dimenziju, *ssx6* [,569], je ujedno jedina varijabla čiji koeficijent korelacije s IMC iskazom 2 prelazi $\pm,200$ za obje grupe, što sugerira da su odgovori na taj iskaz pod velikim utjecajem nemarnog odgovaranja i da su ti utjecaji suprotnog predznaka s obzirom na grupu, što je vidljivo u tablici 9. Ako bismo izbacili ovu varijablu iz inicijalnog modela, pod argumentom da se njezina suboptimalna saturacija jednim dijelom može pripisati nemarnom odgovaranju, dobili bismo dvije stvari: a) dodatan uvid u rezultate koji ne bismo dobili inicijalnim pristajanjem modela, koje je adekvatno i b) podatak da pristajanje modela i dalje nije u potpunosti zadovoljavajuće s obzirom na jednostavnost modela [$\chi^2 = 129,14$; $df = 43$; $p < ,001$; CFI = ,947; TLI = ,932; RMSEA = ,073; 90% CI = ,059-,088;

SRMR = ,042], što bi mogao biti argument za dodatnu provjeru valjanosti korištene skale *suvremenog seksizma* i njenu potencijalnu adaptaciju. Ovakav pristup služi dodatnom razumijevanju korištenih varijabli i skala, pogotovo u vidu njihove korelacije s indikatorima određenih pristranosti i/ili efekata metode, što može biti posebice korisno pri konstrukciji novih skala. No, potrebno je napomenuti očitu validaciju ovakvog pristupa, a to je da ne možemo u potpunosti biti sigurni u njegovu ispravnost i pouzdanost. Kako bi se ovakav pristup uopće validirao, poželjno bi bilo provjeriti ga na većem broju nezavisnih uzoraka, na skalama koje sadržajno nisu slične. Usporedili smo i aritmetičke sredine otežanih i opaženih varijabli. Budući da je otežavanje rezultiralo proširenim rasponom vrijednosti u odnosu na opažene varijable, proveli smo standardizaciju kako bi i otežane varijable imale isti raspon kao i opažene. Novoizračunati otežani rezultati varijabli *suvremenog seksizma* se u velikoj mjeri statistički značajno razlikuju od opaženih rezultata na istoj skali. Aritmetička sredina samo jedne varijable niža je u odnosu na opaženi rezultat, što definitivno sugerira potrebu za doradom korištenog pristupa. Bez obzira što smo u samoj formuli koristili zbrajanje, dvije otežane vrijednosti mogle su biti negativne u svom predznaku te nismo očekivali da će gotovo sve vrijednosti biti promijenjene u istom smjeru. Ovakvi nalazi sugeriraju da su aritmetičke sredine svih rezultata zapravo više, ali su supresirane utjecajem nemarnog odgovaranja i pristranosti slaganja, što nije u skladu s literaturom. Pristranost slaganja trebala bi rezultirati višim aritmetičkim sredinama, a za nemarno odgovaranje nema razloga da mijenja aritmetičke sredine isključivo u jednom smjeru. Rezultati su prikazani u tablici 10.

Tablica 10. Usporedba opaženih i otežanih rezultata

Varijabla	Opazena aritmetička sredina	Otežana aritmetička sredina	Opazena standardna devijacija	Otežana standardna devijacija
ssx1	2,53*	2,84*	1,36	1,52
ssx2	2,57*	3,09*	1,34	1,59
ssx3	1,93*	2,26*	1,16	1,33
ssx4	2,33	2,37	1,28	1,31
ssx5	2,88*	3,19*	1,28	1,44
ssx6	2,53*	2,71*	1,28	1,35
ssx7	2,13*	2,44*	1,19	1,31
ssx8	2,87	2,96	1,12	1,27
ssx9	2,41*	2,69*	1,12	1,3
ssx10	2,27	2,29	1,26	1,28
ssx11	2,55	2,50	1,35	1,30
ssx12	2,06*	2,29*	1,19	1,32

Napomena: parovi koji su podebljani i označeni s * označava statistički značajnost



9. Sveukupni zaključak i prijedlozi

U prvoj studiji smo pokazali moguće utjecaje koju formulacija iskaza može imati na rezultate anketnih istraživanja. Generalno, postoji dovoljan broj radova čiji rezultati dokazuju negativne posljedice implementiranja negativno formuliranih iskaza u skale (Dunbar, Ford, Hunt, i Der, 2000; Horan, DiStefano, i Motl, 2003; Marsh, 1986; Suárez-Alvarez et al., 2018). Bez obzira na navedeno, smatramo da će se negativni iskazi i dalje koristiti u skalama, tako da smo u skladu s time dali prijedlog konstrukcije skala koje će ih u sebi sadržavati. Potrebno je napomenuti da spomenuti pristup, makar u sebi sadrži korak koji bi trebao smanjiti utjecaj formulacije iskaza i time osnažiti metrijske karakteristike skale, neće eliminirati utjecaje formulacije iskaza, pogotovo ako u uzorku postoji osjetni udio sudionika sa slabijim vokabularnim vještinama ili manjim razinama motivacije. Iz ovih razloga, preporučamo da se u skalama ne koriste negativno formulirani iskazi, već da se koriste isključivo u svrhu validacije pažljivog odgovaranja.

Ciljevi druge studije nisu u potpunosti ostvareni. Makar su utjecaji nemarnog odgovaranja i pristranosti slaganja vidljivi na rezultatima, primarno u rezultatima multigrupnih analiza, načini njihove kontrole pa i načini njihovog mjerenja, nisu pokazali zadovoljavajuće razine valjanosti. Nisu pronađene očekivane statistički značajne razlike među grupama nemarnih i pažljivih sudionika, a i način na koji se nemarno odgovaranje mjerilo ne mora biti ispravan. Na primjer, sudionici su mogli iskaze interpretirati kao da se mjeri njihova razina pokornosti ili konformizma te u skladu s time odgovoriti. Nadalje, prisutna je mogućnost „slučajnih pogodaka“, koja je za klasifikaciju na temelju jednog iskaza iznosila 20%. Isto tako, nacrt istraživanja nije uključio nikakve izvanjske skale pomoću kojih bismo mogli opravdati valjanost napravljenih klasifikacija. IMC iskazi korišteni u drugoj studiji nisu pokazali statistički značajnu međusobnu razliku s obzirom na koeficijent korelacije dvaju polarno suprotna iskaza, ali smatramo da je to jednim dijelom rezultat toga što sudionici nisu dvije formulacije percipirali kao polarne suprotnosti, ali i rezultat veličine uzorka/grupa. Utjecaji nemarnog odgovaranja su vidljivi po tome što multigrupnom analizom, bez obzira na korišteni IMC iskaz, nije došlo do konvergiranja konfiguralnog modela *suvremenog seksizma*, koji je pokazao zadovoljavajuće pristajanje na razini cijelog uzorka. Iskazi koji se inače koriste za mjerenje nemarnog odgovaranja slični su onima korištenim u ovoj studiji, a njihova dodatna validnost leži u korelaciji dvaju polarno suprotna iskaza. Korištenje dvaju suprotna polarna iskaza koja su sadržajno nevezana za temu

može biti opcija, ali postoji vjerojatnost da će i nemarni sudionici primijetiti iskaze koji se značajno sadržajno razlikuju u odnosu na ostale koji ih okružuju. Budući da se ne može sa sigurnošću tvrditi da je nemarni sudionik nemaran ili pažljiv kroz cijeli upitnik, IMC iskaze je potrebno implementirati prije skala od posebnog interesa te imati u vidu da ih je potrebno minimalno dva u anketnom upitniku. Generalno, poželjno ih je imati što više, s time da ih ne bi trebalo biti previše kako ne bi postali previše uočljivi

Također, utjecaj pristranosti slaganja je statistički značajan na razini uprosječenog rezultata 12 varijabli skale *svremenog seksizma*, ali rezultati multigrupne analize su pokazali da njegovi utjecaji nisu snažni kao oni nemarnog odgovaranja. Ipsatizacija se nije pokazala korisnom metodom kontroliranja pristranosti slaganja, jer rezultira modelima koji značajno slabije pristaju podacima u odnosu na originalne modele te njeni rezultati nisu u skladu s rezultatima kontrole pomoću strukturalnih modela, stoga ne preporučamo korištenje ove metode pri kontroli pristranosti slaganja. Na temelju naše analize, model koji odvaja latentne dimenzije nemarnog odgovaranja i pristranosti slaganja, koje su u nultoj korelaciji međusobno i sa sadržajnim dimenzijama, pokazuje najbolje slaganje, što je u skladu s nekim drugim istraživanjima (Reise et al., 2007). Ovaj nalaz ima sljedeće implikacije: utjecaji raznih efekata metode i/ili pristranosti mogu imat međusobno različit izvor, što znači da su bilo kakvi RIFA modeli inherentno krivi, čime dodatno zagovaramo izbjegavanje njihovog korištenja. Nadalje, s obzirom na međusobno različite nadređene uzroke, za njihovo mjerenje se trebaju koristiti različiti instrumenti/pristupi.

Postoje brojna ograničenja korištenog istraživačkog nacrt. Prvo ograničenje nalazi se upravo u uzorcima, koji nisu reprezentativni za generalnu populaciju, već su prigodni/kvotni i potencijalno nedovoljne veličine za provedene analize. Unutar obje studije, ispitanici javljaju nešto više razine obrazovanja u odnosu na prosjek Hrvatske, što može značiti da se utjecaji istraživanih efekata metode u većoj ili manjoj mjeri manifestiraju u populaciju, pogotovo jer određeni radovi vežu utjecaje efekata metode uz stupanj obrazovanja (npr. Suárez-Álvarez et al., 2018). Na temelju ove studije ne možemo pretpostaviti koliki bi udio nemarnih sudionika bio u reprezentativnom uzorku, makar nas s obzirom na razine obrazovanja korištenih uzoraka ne bi čudilo da su one u populaciji još više. Drugo ograničenje vezano je uz valjanost IMC iskaza, gdje postoje legitimni argumenti protiv nje (npr. odsutnost statistički značajne razlike u koeficijentima korelacija antonima s obzirom na grupe), ali i argumenti protiv ovakvog načina mjerenja

nemarnog odgovaranja koje smo spomenuli u prethodnom ulomku. Treće ograničenje je što ne možemo tvrditi da će najbolje pristajući model po rezultatima druge studije biti najbolje pristajući model u nekom budućem istraživanju, jer postoje dokazi da utjecaji efekata metode i raznih pristranosti ovise upravo o korištenoj skali (Kam, 2016). Četvrto ograničenje je da u društvenim znanostima i dalje jako veliki broj istraživača nije upoznato sa strukturalnim modeliranjem, čime je ovakav način kontrole realno dostupan određenom postotku onih koji su upoznati s ovom metodologijom. Iako se utjecaji raznih efekata metode i/ili pristranosti mogu jednostavno provjeriti na univarijantnoj razini, provjera u multidimenzionalnom prostoru često znači korištenje upravo ovakve metodologije. Peto ograničenje je da ne možemo biti sigurni u ispravnost korištenog indeksa i otežanih rezultata te je potrebno više razvoja u tom području, a ova studija se može smatrati pilot studijom u stvaranju indeksa kontrole pristranosti. Iako su po formuli rezultati otežanih varijabli mogli biti i veći i manji u odnosu na opažene, oni su gotovo isključivo bili veći, što je jasan znak da je adaptacija ovog pristupa potrebna.

Generalno, smatramo da proučavanje teorije mjerenja, posebice utjecaja efekata metode i raznih pristranosti predstavlja područje puno potencijala koje se i dalje nastavlja razvijati te koje ima potencijal davanja validnijih rezultata istraživanja, što bi moglo pomoći trenutnom stanju replikacijske krize unutar društvenih znanosti. Smatramo da bi daljnjem razvoju ove grane pridonijela longitudinalna istraživanja, koja bi dodatno rasvijetlila ovise li njihove jačine/prisutnosti o skalama, osobnim karakteristikama, raspoloženju u trenutku ispunjavanja upitnika ili nečemu drugome.

10.Literatura

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality*, 53, 1–4. doi: <https://doi.org/10.1016/j.jrp.2014.07.001>

Allen, M. i Yen, W. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Pub. Co.

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A. i Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*. 52(6), 2489-2505. doi:10.3758/s13428-020-01401-8

- Baumgartner, H. i Steenkamp, J.-B. E. M. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 143-156. doi:10.1509/jmkr.38.2.143.18840
- Beck, M. F., Albano, A. D. i Smith, W. M. (2018). Person-Fit as an Index of Inattentive Responding: A Comparison of Methods Using Polytomous Survey Data. *Applied Psychological Measurement*, 43(5), 374-387. doi:10.1177/0146621618798666
- Bentler, P. M., Jackson, D. N. i Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76(3), 186-204. doi: <https://doi.org/10.1037/h0031474>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. doi: 10.1037/0033-2909.107.2.238
- Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, 36(4), 542–562. doi: <https://doi.org/10.1177/0049124107313901>
- Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M. i Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, 111(2), 218–229. doi:10.1037/pspp0000085
- Brown, T. A. (2015). *Methodology in the social sciences. Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. doi: <https://doi.org/10.1037/h0046016>
- Cappelleri, J. C., Jason Lundy, J. i Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clinical Therapeutics*, 36(5), 648–662. doi:10.1016/j.clinthera.2014.04.006
- Chang, E. C., Maydeu-Olivares, A. i D'Zurilla, T. J. (1997). Optimism and pessimism as partially independent constructs: Relationship to positive and negative affectivity and psychological well-being. *Personality and Individual Differences*, 23(3), 433–440. doi: [https://doi.org/10.1016/S0191-8869\(97\)00040-8](https://doi.org/10.1016/S0191-8869(97)00040-8)
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. doi:10.1080/10705510701301834

Cheung, G. W. i Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. doi: https://doi.org/10.1207/S15328007SEM0902_5

Costa, P. T., Jr., Terracciano, A. i McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322–331. doi: <https://doi.org/10.1037/0022-3514.81.2.322>

Costello, A. B. and Osborne, J. W. (2005) Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1-9.

Credé, M., Chernyshenko, O. S., Bagraim, J. i Sully, M. (2009). Contextual Performance and the Job Satisfaction–Dissatisfaction Distinction: Examining Artifacts and Utility. *Human Performance*, 22(3), 246–272. doi:10.1080/08959280902970427

Credé, M. (2010). Random Responding as a Threat to the Validity of Effect Size Estimates in Correlational Research. *Educational and Psychological Measurement*, 70(4), 596–612. doi:10.1177/0013164410366686

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19. doi:10.1016/j.jesp.2015.07.006

DeSimone, J. A., DeSimone, A. J., Harms, P. D. i Wood, D. (2017). The Differential Impacts of Two Forms of Insufficient Effort Responding. *Applied Psychology*, 67(2), 309–338. doi:10.1111/apps.12117

DeSimone, J. A. i Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, 33(5), 559–577. <https://doi.org/10.1007/s10869-017-9514-9>

de la Fuente, J. i Abad, F.J. (2020). Comparing Methods for Modeling Acquiescence in Multidimensional Partially Balanced Scales. *Psicothema*, 32(4), 590-597. doi: 10.7334/psicothema2020.96

DiStefano, C. i Motl, R. W. (2006). Further Investigating Method Effects Associated With Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3), 440–464. doi:10.1207/s15328007sem1303_6

DiStefano, C. i 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(3), 309–313. doi: <https://doi.org/10.1016/j.paid.2008.10.020>

Donnellan, M. B., Oswald, F. L., Baird, B. M., i Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192-203. doi: 10.1037/1040-3590.18.2.192

Dunbar, M., Ford, G., Hunt, K. i Der, G. (2000). Question Wording Effects in the Assessment of Global Self-Esteem. *European Journal of Psychological Assessment*, 16(1), 13–19. doi: <https://doi.org/10.1027//1015-5759.16.1.13>

Eden, D. i Leviatan, U. (1975). Implicit Leadership Theory as a Determinant of the Factor Structure Underlying Supervisory Behavior Scales. *Journal of Applied Psychology*, 60, 736-741. doi: <http://dx.doi.org/10.1037/0021-9010.60.6.736>

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. i Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. doi:10.1037/1082-989x.4.3.272

Ferrando, P. J. i Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, 63(2), 427–448. doi:10.1348/000711009x470740

Fischer, R. (2004). Standardization to Account for Cross-Cultural Response Bias. *Journal of Cross-Cultural Psychology*, 35(3), 263–282. doi:10.1177/0022022104264122

Greenleaf, E. A., Measuring Extreme Response Style. *Public Opinion Quarterly*, 56(3), 328–351. doi: <https://doi.org/10.1086/269326>

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19, 149–161. doi: <https://doi.org/10.1007/BF02289162>

Harris, E. F. i Smith, R. N. (2009). Accounting for measurement error: A critical but often overlooked process. *Archives of Oral Biology*, 54, 107–117. doi:10.1016/j.archoralbio.2008.04.010

Hidalgo-Rasmussen, C.A. i González-Betanzos, F. (2019). The Treatment of Acquiescence and the Factorial Structure of the Brief Resilience Scale (BRS) in Mexican Chilean University Students. *Anales de Psicología*, 35(1), 26-32. doi: <https://dx.doi.org/10.6018/analesps.35.1.297781>.

Horan, P. M., DiStefano, C. i Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, 10(3), 435–455. doi: https://doi.org/10.1207/S15328007SEM1003_6

Hu, L., i Benter, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. doi: <https://doi.org/10.1080/10705519909540118>

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M. i DeShon, R. P. (2011). Detecting and Detering Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, 27(1), 99–114. doi:10.1007/s10869-011-9231-8

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M. i DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114. doi: <https://doi.org/10.1007/s10869-011-9231-8>

Huang, J. L., Liu, M. i Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828–845. doi:10.1037/a0038510

Husremović, Dž. (2016). *Osnove psihometrije*. Sarajevo: Filozofski fakultet Univerziteta u Sarajevu.

Kam, C., Schermer, J. A., Harris, J. i Vernon, P. A. (2013). Heritability of acquiescence bias and item keying response style associated with the HEXACO Personality Scale. *Twin Research and Human Genetics*, 16(4), 790–798. doi: <https://doi.org/10.1017/thg.2013.38>

Kam, C. C. S., i Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: the case of job satisfaction. *Organizational Research Methods*, 18(3), 512–541. doi: 10.1177/1094428115571894

Kam, C. C. S. (2016). Why Do We Still Have an Impoverished Understanding of the Item Wording Effect? An Empirical Examination. *Sociological Methods & Research*, 47(3), 574–597. doi:10.1177/0049124115626177

Kaiser, H. F. *The varimax method of factor analysis*. neobjavljena doktorska disertacija, University of California, 1956.

Kunda, Z., Fong, G. T., Santioso, R., i Reber, E. (1993). Directional questions direct self-conceptions. *Journal of Experimental Social Psychology*, 29(1), 63–86. doi: 10.1006/jesp.1993.1004

Lord, F.M., Novick, M.R. i Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Mandić, L. (2020). Testiranje utjecaja efekata metode na rezultate anketnih istraživanja. Rad nagrađen rektorovom nagradom. Zagreb: Sveučilište u Zagrebu.

Mandić, L. i Klasnić, K. (2021). Is it biased? Empirical analysis of various phenomena that effect survey results. U postupku recenzije u časopisu „Revija za sociologiju“.

- Maniaci, M. R. i Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83. doi: <https://doi.org/10.1016/j.jrp.2013.09.008>
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22, 37–49. doi:10.1037/0012-1649.22.1.37
- Marsh, H., Hau, K., i Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal* 11(3), 320-341. doi: 10.1207/s15328007sem1103_2
- Maul, A. (2013). Method Effects and the Meaning of Measurement. *Frontiers in Psychology*, 4, Article 169. doi: <https://doi.org/10.3389/fpsyg.2013.00169>
- Maydeu-Olivares, A., i Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362. doi: <https://doi.org/10.1037/1082-989X.11.4.344>
- McClendon, M. J. (1991). Acquiescence: Tests of the cognitive limitations and question ambiguity hypotheses. *Journal of Official Statistics*, 7, 153–166.
- McCrae, R. R., Herbst, J. H. i Costa, P. T., Jr. (2001). Effects of acquiescence on personality factor structures. U R. Riemann, F. Ostendorf, i F. Spinath (Ur.), *Personality and temperament: Genetics, evolution, and structure* (str. 217–231). Berlin, Germany: Pabst Science.
- Meade, A. W. i Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. doi: <https://doi.org/10.1037/a0028085>
- Mestre, J. P. (1988). The role of language comprehension in mathematics and problem solving. U R. R. Cocking i J. P. Mestre (Ur.), *Linguistic and cultural influences on learning mathematics* (str. 201–220). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Neuringer, A. (1986). Can people behave "randomly?": The role of feedback. *Journal of Experimental Psychology: General*, 115(1), 62–75. doi: <https://doi.org/10.1037/0096-3445.115.1.62>
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. doi: [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Paulhus, D. L. (1991). Measurement and control of response bias. U J. P. Robinson, P. R. Shaver, i L. S. Wrightsman (Ur.), *Measures of personality and social psychological attitudes* (17–59). San Diego, CA: Academic Press. doi: <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>

Podsakoff, P. M., MacKenzie, S. B. i Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, 63(1), 539–569. doi:10.1146/annurev-psych-120710-100452

Primi, R., Santos, D., De Fruyt, F. i John, O. P. (2019). Comparison of classical and modern methods for measuring and correcting for acquiescence. *British Journal of Mathematical and Statistical Psychology*. 72(3), 447-465. doi:10.1111/bmsp.12168

Putnick, D. L. i Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. doi:10.1016/j.dr.2016.06.004

R Development Core Team. (2020). R: *A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: Author.

Reise, S. P., Morizot, J. i Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19–31. doi:10.1007/s11136-007-9183-7

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.9.12, <https://CRAN.R-project.org/package=psych>.

Rosseel, Y. (20129). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36. doi: 10.18637/jss.v048.i02

Savalei, V. i Falk, C. F. (2014). Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches. *Multivariate Behavioral Research*, 49(5), 407–424. doi:10.1080/00273171.2014.931800

Schmitt, N., i Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367–373. doi: <https://doi.org/10.1177/014662168500900405>

Steiger, J.H. i Lind, J.C. (1980, May 30, 1980). *Statistically-based tests for the number of common factors*. Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City.

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., i Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema* 30(2), 149-158. doi: 10.7334/psicothema2018.33

Šimac, B. i Klasnić, K. (2021). Što još one žele? Razvoj nove skale za mjerenje suvremenih oblika seksizma. *Socijalna ekologija*, prihvaćeno za objavljivanje u 2021. godini.

- Tomas, J. M. i Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 84–98. doi:10.1080/10705519909540120
- Tucker, L.R., i Lewis, C. A. (1973). reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. doi: <https://doi.org/10.1007/BF02291170>
- Watson, D. (1992). Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale. *Sociological Methods & Research*, 21(1), 52–88. doi:10.1177/0049124192021001003
- Weijters, B., Geuens, M., i Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods*, 15(1), 96-110. doi: 10.1037/a0018721.
- Weijters, B., i Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. *Journal of Marketing Research*, 49(5), 737-747. doi: <https://doi.org/10.1509/jmr.11.0368>
- Weijters, B., Baumgartner, H., i Schillewaert, N. (2013). Reverse item bias: An integrative model. *Psychological Methods*, 18(3), 320-334. doi: 10.1037/a0032121
- Wise, S. L. i Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18(2), 163–183. doi:10.1207/s15324818ame1802_2
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817–38.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186-191.
- Yuan, K., i Bentler, P. M. (2007). Multilevel Covariance Structure Analysis By Fitting Multiple Single-Level Models. *Sociological Methodology*, 37(1), 53-82. doi: <https://doi.org/10.1111/j.1467-9531.2007.00182.x>
- Zeng, B., Wen, H. i Zhang, J. (2020). How Does the Valence of Wording Affect Features of a Scale? The Method Effects in the Undergraduate Learning Burnout Scale. *Frontiers in Psychology*, 11. doi:10.3389/fpsyg.2020.585179

Efekti metode - analiza odabranih karakteristika grupe fenomena koji utječu na rezultate anketnih istraživanja

SAŽETAK

Prema klasičnoj teoriji testova, bruto rezultat se sastoji od dva djela: pravog rezultata i komponente pogreške, koja se dijeli na dva tipa: sustavnu i slučajnu pogrešku. U anketnim istraživanjima postoji dodatna grupa efekata koji utječu na rezultate, koji po svojoj definiciji ne spadaju u sustavne pogreške, ali im po svojim utjecajima slične, a to su efekti metode. U studiji 1, fokusirali smo se isključivo na utjecaje jednog efekta metode, formulacije iskaza te smo koristili dvije inačice anketnog upitnika. Uzorak se sastojao od $N = 791$ sudionika. Utjecaji formulacije iskaza vidljivi su u statistički značajnoj razlici u aritmetičkim sredinama, ali i u razinama mjerne invarijantnosti koje skale dosežu. Na kraju smo dali prijedlog konstruiranja skale koje u sebi sadrže i pozitivne i negativne iskaze. U studiji 2, fokusirali smo se na utjecaje nemarnog odgovaranja i pristranosti slaganja. Istraživanje je provedeno na kvotnom uzorku veličine $N=483$. Kako bismo mogli mjeriti utjecaj nemarnog odgovaranja, uključili smo dva posebna iskaza (IMC). Pristranost slaganja s iskazima mjerili smo tako što smo specificirali indeks pristranosti slaganja s iskazima bez obzira na njihov sadržaj odnosno NARS. Osim ovog pristupa, koji je primarno usmjeren na mjerne i strukturne modele, odlučili smo koristiti i ipsatizaciju rezultata. Konstruirali smo i indeks pomoću kojeg smo otežali varijable s obzirom na efekte metode. IMC iskazi nisu pokazali zadovoljavajuće razine valjanosti, ali smatramo da je to jednim dijelom rezultat veličine uzorka/grupa. Utjecaji nemarnog odgovaranja su vidljivi po tome što multigrupnom analizom nije došlo do konvergiranja konfiguralnog modela *suвременog seksizma*, koji je pokazao zadovoljavajuće pristajanje na razini cijelog uzorka. Ipsatizacija se nije pokazala korisnom metodom kontroliranja pristranosti slaganja, dok se SEM pristup pokazao korisnim. Na temelju naše analize, model koji odvaja latentne dimenzije nemarnog odgovaranja i pristranosti slaganja, koje su u nultoj korelaciji međusobno i sa sadržajnim dimenzijama, pokazuje najbolje slaganje.

Ključne riječi: *efekti metode, formulacija iskaza, nemarno odgovaranje, pristranost slaganja, indeks pristranosti*

Method Effects: Analysis of Selected Characteristics of a Group of Phenomena That Affect Survey Results

ABSTRACT

The classical test theory defines that the raw result is made up of two parts: true score and its error, which can be random or systematic. There is a group of phenomena in survey research, which by definition aren't a part of systematic errors, but are similar to them in regard to their effects: method effects. In the first study, we tested the effects of one method effect: item wording, using a split-ballot design. The sample size was $N = 791$ participants. The effects of item wording were noticeable in item means, but also in levels of measurement invariance. We presented a way of handling item wording effects when constructing balanced scales. In the second study, we tested the effects of careless responding and acquiescence bias. We used a quota sample [$N = 485$]. To classify respondents as either careful or careless, we used IMC items. To test levels of acquiescence, we specified a NARS index. These ways were primarily SEM oriented, so we decided to ipsatize the results as well. We also constructed a bias index which should remove the biases from the raw data. The IMC items didn't show adequate levels of validity, but we think that it is in part due to the sample size. Careless responding effects were noticeable in the multigroup model not converging, whereas the whole-sample model showed adequate fit parameters. Ipsatization did not prove to be a useful method for controlling acquiescence, whereas the SEM approach was. The model which separated method effects and traits showed best fit parameters.

Key words: method effects, item wording, careless responding, acquiescence bias, bias index