

Usporedba metoda za otkrivanje različita funkcioniranja čestica u multidimenzionalnim instrumentima s politomnim česticama

Ćosić Pilepić, Ana

Doctoral thesis / Doktorski rad

2025

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

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:186:708893>

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

Download date / Datum preuzimanja: **2025-03-31**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET

Ana Čosić Pilepić

**USPOREDBA METODA ZA
OTKRIVANJE RAZLIČITA
FUNKCIONIRANJA ČESTICA U
MULTIDIMENZIONALNIM
INSTRUMENTIMA S POLITOMNIM
ČESTICAMA**

DOKTORSKI RAD

Mentor: prof. emer. Vladimir Takšić

Komentor: izv. prof. dr. sc. Tamara Mohorić

Rijeka, 2025

UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Ana Ćosić Pilepić

**A COMPARISON OF DETECTION
METHODS FOR MULTIDIMENSIONAL
POLYTOMOUS DIFFERENTIAL ITEM
FUNCTIONING**

DOCTORAL THESIS

Rijeka, 2025

Mentor doktorskog rada: prof. emer. Vladimir Takšić

Komentor doktorskog rada: izv. prof. dr. sc. Tamara Mohorić

Doktorski rad obranjen je dana _____ u/na
_____, pred povjerenstvom u sastavu:

1. _____
2. _____
3. _____
4. _____
5. _____

SAŽETAK

Metode za otkrivanje DIF-a tradicionalno su se koristile za jednodimenzionalne mjerne instrumente, pretpostavljajući da sama prisutnost višedimenzionalnosti u podacima upućuje na prisustvo DIF-a. Međutim, brojni psihološki instrumenti su inherentno višedimenzionalni jer mjere složene konstrukte kroz različite skupine čestica, što može dovesti do netočnih zaključaka o prisutnosti DIF-a kada se koriste jednodimenzionalni modeli. Dosad je nekoliko metoda za detekciju DIF-a adaptirano za upotrebu na višedimenzionalnim podacima, no njihova učinkovitost provjeravana je samo na višedimenzionalnim dihotomnim podacima.

Stoga se u ovom istraživanju učinkovitost željela provjeriti na višedimenzionalnim politomnim podacima. Pritom su uspoređivane 4 metode: omjer izglednosti iz okvira teorije odgovora na zadatak (IRT-LR), MIMIC model s interakcijom, logistička regresija s latentnom procjenom i logistička regresija s bruto rezultatima.

Učinkovitost ovih metoda provjeravana je na računalno simuliranim, ali i na realnim podacima koji su dobiveni pomoću kraće verzije Upitnika emocionalne kompetentnost UEK-27. Evaluacija metoda obuhvatila je analizu stope pogrešaka tipa I i statističke snage (osjetljivosti) svake metode.

Rezultati istraživanja pokazuju da je logistička regresija s procjenom latentne varijable bila najučinkovitija u otkrivanju DIF-a na politomnim višedimenzionalnim podacima, posebno efikasna u otkrivanju niskog DIF-a te otporna na stvarne razlike u grupnim sposobnostima i razlike u veličini grupa. S druge strane, MIMIC metoda s interakcijom pokazala je tendenciju povećanja stope pogrešaka tipa I.

Empirijski nalazi djelomično su potvrdili rezultate simulacije, no ograničenje na jedan uzorak u empirijskom dijelu upozorava na potrebu opreznog tumačenja i generalizacije rezultata. Buduća istraživanja trebala bi nastaviti razvijati i testirati metodologije za detekciju DIF-a, proširujući njihovu primjenjivost i adresirajući postojeća ograničenja u različitim testnim situacijama i populacijama.

SUMMARY

This dissertation evaluates the effectiveness of various Differential Item Functioning (DIF) detection methods in multidimensional testing environments, particularly focusing on their application to polytomous data settings. The study aims to address the complexities that arise when traditional unidimensional approaches are used for inherently multidimensional constructs. Data for the study were gathered through both computer-simulated scenarios and real data collected via the shortened form of Emotional Skills and Competence Questionnaire (ESCQ-27).

Four DIF detection methods were compared: Item Response Theory-Likelihood Ratio (IRT-LR), MIMIC with interaction, logistic regression with latent trait estimation, and logistic regression with observed total scores. The efficacy of these methods was assessed through their Type I error rates and statistical power (sensitivity).

The findings revealed that logistic regression with latent trait estimation was most effective in identifying DIF in multidimensional polytomous data, efficiently detecting low DIF and demonstrating robustness against actual differences in group abilities and group sizes. Conversely, the MIMIC method with interaction tended to increase Type I error rates, indicating potential issues with false positives. Although simulated data supported the effectiveness of the logistic regression methods, the empirical results from the ESCQ-27 questionnaire emphasized the need for cautious interpretation and generalization due to the sample limitation in the empirical component of the study.

The research highlights the necessity of using appropriate multidimensional DIF detection methods for complex constructs measured by psychometric instruments and recommends logistic regression with latent trait estimation for its superior performance in diverse testing scenarios. Future studies should focus on further developing and testing DIF detection methodologies to enhance their applicability and address existing limitations in different test situations and populations. This aim would improve the accuracy and reliability of psychometric evaluations across varied demographic groups.

SADRŽAJ

1. UVOD	1
1.1. Teorija odgovora na zadatak (IRT)	1
1.1.1. Pretpostavke teorije odgovora na zadatak	12
1.1.2. Modeli za dihotomne čestice	14
1.1.3. Modeli za politomne čestice	20
1.1.4. Višedimenzionalna teorija odgovora na zadatak (MIRT)	26
1.2. Diferencijalno funkcioniranje čestica (DIF)	32
1.2.1. Diferencijalno funkcioniranje čestica, valjanost i pravednost	32
1.2.2. Mjerna invarijantnost i pristranost	33
1.2.3. Definicija DIF-a	36
1.2.4. Jednoliki i nejednoliki DIF	37
1.3. Metode detekcije DIF-a	39
1.3.1. Zajednička obilježja metoda detekcije DIF-a	39
1.3.2. Različite klasifikacije metoda detekcije DIF-a	40
1.3.3. Metode za višedimenzionalne instrumente s politomnim česticama	43
2. CILJ ISTRAŽIVANJA	53
2.1. Problemi i hipoteze istraživanja	53
3. METODA	55
3.1. Simulirani podaci	55
3.1.1. Monte Carlo simulacije	55
3.1.2. Generiranje podataka	57
3.1.3. Simulacijski faktori	57
3.1.4. Evaluacijski kriteriji	60
3.2. Empirijski podaci	63
3.2.1. Izvor podataka	63
3.2.2. Instrument	64
4. REZULTATI	65
4.1. Rezultati simulacijskog istraživanja	65
4.1.1. Rezultati pogreške tipa I	65

4.1.2. Rezultati statističke snage	67
4.2. Rezultati DIF-analiza na realnim podacima	73
4.2.1. Mjerna invarijantnost Upitnika emocionalne kompetentnosti UEK-27	73
4.2.2. Dimenzionalnost Upitnika emocionalne kompetentnosti UEK-27	78
4.2.3. Diferencijalno funkcioniranje čestica	88
4.2.4. Usporedba rezultata simulacijskog i empirijskog istraživanja	93
5. RASPRAVA	95
6. DOPRINOSI I OGRANIČENJA ISTRAŽIVANJA	104
7. ZAKLJUČAK	107
8. LITERATURA	110
9. POPIS SLIKA	128
10. POPIS TABLICA	129
11. POPIS JEDNADŽBI	130
12. PRIVITCI	131
Privitak 1: Kodovi za simulaciju podataka i provedene analize	132
Privitak 2: Upitnik emocionalne kompetentnosti UEK-45	306
Privitak 3: Grafički prikazi čestica UEK-27	308
Privitak 4: Usporedni prikaz EGA rješenja svih 13 zemalja	314

1. UVOD

1.1. Teorija odgovora na zadatak (IRT)

„Moramo opisati čestice koristeći parametre čestica a ispitanike koristeći parametre ispitanika na način koji nam omogućuje probabilističko predviđanje odgovora bilo kojeg ispitanika na bilo koju česticu, čak i kada slični ispitanici nikada nisu odgovarali na slične čestice.“ (Lord, 1980).

Teorija odgovora na zadatak (engl. *item response theory*, *IRT*) je primijenjena statistička i psihometrijska disciplina čiji je glavni cilj objasniti:

- 1) interakciju između ispitanika i elemenata korištenog mjernog instrumenta, te
- 2) količinu informacija koju sadržavaju podaci dobiveni tim instrumentom.

Kako bi postigla svoj cilj, teorija odgovora na zadatak se koristi probabilističkim funkcijama (Reckase, 2009).

Općenito govoreći, IRT se temelji na dva postulata. Prvi od njih smatra da se odgovori ispitanika na zadanom skupu čestica ili na cjelokupnom mjernom instrumentu mogu objasniti korištenjem jedne ili više latentnih osobina (konstrukata, dimenzija, kontinuuma ili faktora). Prema drugom postulatu, odnos između odgovora na čestici i tih latentnih osobina može se opisati upotrebom odgovarajuće probabilističke funkcije, koju u jednodimenzionalnoj teoriji odgovora na zadatak predstavlja karakteristična krivulja čestice. Različiti IRT modeli različito prikazuju ovaj odnos, no svima im je zajedničko da se pritom zadržavaju na razini pojedinačne čestice, a ne na razini skale ili instrumenta kako to čini klasična teorija testova.

U IRT-u se odgovori na česticu smatraju zavisnom varijablom, dok su sposobnost ispitanika i karakteristike čestica latentne nezavisne, odnosno prediktorske varijable.

Povijesni razvoj teorije odgovora na zadatak

Klasična teorija testova (engl. *classical test theory*, *CTT*) bila je dominantan psihometrijski pristup sve do 1953. godine kada je Frederic Lord objavio svoju doktorsku disertaciju o latentnoj teoriji osobina (engl. *Latent Trait Theory*, *LTT*) (Lord, 1953). Njegov rad izazvao je veliko zanimanje, pa su stoga pedesete i šezdesete godine prošlog stoljeća bile plodonosne za teorijski razvoj IRT-a. Allen Birnbaum (1957, 1958a, 1958b) je napisao niz tehničkih izvještaja o logističkim modelima testova i procjeni njihovih parametara. George Rasch (1960/1980) je objavio knjigu s nekoliko modela odgovora na čestice, dok je Baker (1961) empirijski usporedio logističku i normalnu funkciju. Lord i Novick (1968) te Wright (1968) su predstavili modele za dihotomne čestice, dok je Samejima (1969) prvi istraživao politomne čestice i opisao model stupnjevitih odgovora (engl. *graded response model*, *GRM*).

Do sedamdesetih godina, zbog matematičke zahtjevnosti, IRT je bila uglavnom konceptualno područje, no tada se, uz pomoć uznapredovale računalne tehnologije koja je omogućila prve ozbiljnije analize podataka, počinje i praktično primjenjivati. Osamdesetih pak doživljava procvat, te se razvijaju prvi računalni programi za analizu podataka s odgovorima na čestice testova i upitnika (npr. BILOG, MULTILOG, WINSTEPS, IRTPRO, MPLUS, HLM), a krajem 20. stoljeća IRT procedure postaju dominantne u obrazovnim testiranjima širokog obuhvata. Daljnjim tehnološkim napretkom u 21. stoljeću pojavljuje se i niz paketa za statistički program otvorenog koda R, među kojima su u širokoj upotrebi *ltm* za jednodimenzionalnu IRT (Rizopoulos, 2006), *eRm* za ekstenzije Raschevih modela (engl. *extended Rasch models*) (Mair i Hatzinger, 2007), *mlirt* za višerazinske i bayesijanske procjene nekih IRT modela (engl. *multilevel and Bayesian IRT*) (Fox, 2007), *gpcm* (Johnson, 2007) za bayesijansku procjenu generaliziranog modela djelomičnog bodovanja (engl. *generalized partial credit model, GPCM*), *MCMCpack* za bayesijansku IRT (Martin i sur., 2011) i *mirt* za višedimenzionalnu IRT (Chalmers, 2012).

Dvije su temeljne grane teorije odgovora na zadatak: jednodimenzionalna i višedimenzionalna teorija odgovora na zadatak (IRT). Jednodimenzionalna IRT pretpostavlja da postoji jedna latentna osobina koja leži u osnovi svih čestica u instrumentu. S druge strane, višedimenzionalna IRT (MIRT) smatra da su čestice grupirane u klastere koji pokrivaju različite domene, pri čemu instrumenti mjere te višestruke domene istog konstrukta. U MIRT-u, ne samo da su čestice međusobno povezane unutar domena, već i ispitanici mogu biti međusobno povezani jer se nalaze u sličnim okruženjima kao što su razredi ili škole. Ova ideja o hijerarhijskoj strukturi ispitanika proširuje MIRT u višerazinsku višedimenzionalnu teoriju odgovora na zadatak (engl. *multilevel MIRT*), kao najnovije područje teorije odgovora na zadatak.

Teorija odgovora na zadatak je tako evoluirala od svojih početaka kada je korištena isključivo za jednodimenzionalno modeliranje i mjerenje latentnih osobina pojedinaca do višedimenzionalne analize koja se trenutno smatra podvrstom generaliziranog linearnog mješovitog modela (engl. *generalized linear mixed model*) (de Boeck i Wilson, 2004). IRT je tako postao sveobuhvatniji i prilagodljiviji za različite složene situacije u kojima se mjere latentne osobine i njihove međuovisnosti.

Danas su IRT modeli i metode u svakodnevnoj primjeni u području procjene znanja, a pogotovo prilikom testiranja širokog obuhvata (engl. *large-scale testing*), no u područjima koja koriste testove tipičnog učinka tek se počinju koristiti (Reise i Revicki, 2014).

Usporedba klasične teorije testova i teorije odgovora na zadatak

Uobičajeni način konstrukcije, evaluacije i upotrebe testova u području obrazovanja i psihologije temeljen na klasičnoj teoriji testova iskazuje određene nedostatke, a teorija odgovora na zadatak pojavila se kao pokušaj da se ti nedostaci otklone. Klasična teorija testova (engl. *classical test theory, CTT*) počiva na klasičnom modelu testa (engl. *classical test model*). Njegove su pretpostavke slabe, pa ih većina skupova podataka vrlo lako zadovoljava. Iz tog

razloga se klasični model testa vrlo lako primjenjuje, te se često koristi pri konstrukciji testova kao i analizi rezultata.

Međutim, glavni nedostatak klasične teorije testova je da njeni pokazatelji nisu neovisni o uzorku. Karakteristike ispitanika i karakteristike testa su međusobno povezane i njihova interpretacija je moguća samo u zajedničkom kontekstu. Vrijednosti parametara čestica se razlikuju ovisno o skupini ispitanika na kojima su dobiveni, a procjene sposobnosti ispitanika ovise o tome koje su čestice odabrane za sastavljanje mjernog instrumenta. To znači da kada bismo procjenjivali efikasnost skale za mjerenje mentalnog zdravlja u okviru CTT-a, dobili bismo različite parametre čestica i skale (npr. težinu čestice, diskriminativnost čestice, pouzdanost skale) na uzorcima zdravih osoba i osoba s depresijom.

Iako su u teoriji odgovora na zadatak pretpostavke o opaženim i pravim rezultatima testa iste kao i kod klasične teorije testova, pokazatelji koji se koriste u IRT analizama za ispitivanje funkcioniranja čestica i skala nisu temeljeni na prvom i drugom momentu (aritmetička sredina, varijanca, kovarijanca) koji su osjetljivi na karakteristike uzorka, već na momentima višeg reda, kao što su parametri praga i nagiba, koji su na karakteristike uzorka manje osjetljivi i stoga stabilniji (Baker, 2001). Iz tog razloga za psihometrijske karakteristike koje potječu iz teorije odgovora na zadatak kažemo da su invarijantne odnosno ne ovise o odabiru uzorka ispitanika (Presser i dr., 2004). Invarijantnost parametara je preduvjet za tzv. *specifičnu objektivnost*, svojstvo IRT modela koje omogućava mjerenje latentnih osobina ispitanika neovisno o specifičnim zadacima koji se koriste u testu, te procjenu karakteristika čestica neovisno o sposobnostima ispitanika u uzorku.

Kod parametara klasične teorije testova situacija je drugačija. Naime, ukoliko jedan mjerni instrument sadrži teže čestice čini se da je sposobnost ispitanika (pravi rezultat u terminima CTT-a) koji su ga rješavali niža u odnosu na one koji su rješavali test sa lakšim česticama. Također, ako je sposobnost ispitanika visoka, čestice će se činiti lakšima u usporedbi s mjernim instrumentom koji su rješavali ispitanici niže sposobnosti, iako su objektivno oba mjerna instrumenta jednako teška za ispitanike iste sposobnosti (Hambleton, 1989).

Osim težine čestica i diskriminativnosti čestica, i pouzdanost i valjanost rezultata testa dobiveni u okviru klasične teorije testova ovise o skupini ispitanika na kojoj su procijenjeni. Ovo svojstvo klasične teorije testova se naziva "ovisnost o uzorku" (engl. *sample-dependance* ili *group-dependance*).

Tako je, na primjer, diskriminativnost čestice viša ukoliko su ispitanici heterogeni po sposobnostima, što proizlazi iz dobro poznatog utjecaja heterogenosti skupine na veličinu koeficijenta korelacije (Lord i Novick, 1968). Diskriminativnost čestice ovisi i o težini čestice, na način da će čestice imati niže point-biserijalne korelacije u populacijama u kojima je težina čestice ekstremno visoka ili ekstremno niska. To praktično znači da će diskriminativnost čestice biti najveća za ispitanike prosječne sposobnosti.

Pouzdanost (a onda i valjanost) mjernog instrumenta je također ovisna o varijabilnosti rezultata ispitanika. Ako rezultati ispitanika pokazuju malu varijabilnost, to može rezultirati

nižom pouzdanosti i valjanošću testa. Ova varijabilnost rezultata ispitanika može biti posljedica različitih faktora, kao što su pogreške u mjerenju ili stvarne razlike u sposobnostima među ispitanicima.

Iz tih je razloga je vrlo teško uspoređivati ispitanike koji su rješavali različite mjerne instrumente, kao i čestice čiji su parametri θ dobiveni na različitim skupinama ispitanika, a kako bi se rezultati ipak mogli generalizirati pretpostavlja se da su dobiveni na slučajnom uzorku (Crocker i Algina, 2008).

Procjene sposobnosti dobivene klasičnim psihometrijskim načinom su ovisne i o mjernom instrumentu (engl. *test-dependent*), što znači da rezultati ispitanika dobiveni upotrebom različitih mjernih instrumenata nisu usporedivi (Hambleton i Jones, 1993). Tako dobiveni rezultati ispitanika mogu se nalaziti na različitim skalama, što znači da ne postoji funkcionalna povezanost između rezultata na različitim testovima.

Čak i kada su dvije skupine ispitanika testirane istim ili paralelnim testovima, ovo ograničenje je i dalje prisutno, s obzirom da se kod ispitanika različitih sposobnosti mogu javiti različite razine pogreške mjerenja, što dodatno otežava usporedbu rezultata između skupina.

No, ovu razliku između IRT-a i CTT-a po pitanju invarijantnosti parametara ne treba precijeniti. Ako čestice nisu ekstremne i populacije nisu previše različite, klasični parametri čestica pokazuju uglavnom linearni odnos među populacijama. Fan je 1998. otkrio da su rezultati dobiveni upotrebom IRT-a i CTT-a međusobno usporedivi i da se ne može govoriti o superiornosti IRT-a nad CTT-om (Fan, 1998).

Osim toga, kad se u osnovi opaženih rezultata pretpostavlja normalno distribuirana osobina i kad je mogućnost pogađanja minimalna, nelinearnom je transformacijom moguće skalirati indekse težine dobivene u dvije različite populacije na istu ljestvicu. Primjerice, Educational Testing Service (ETS), najveća svjetska privatna organizacija za testiranje i ocjenjivanje u obrazovanju, je prije nego što je IRT postala uobičajena za analizu čestica u tu svrhu koristio delta skalu (Holland i Thayer, 1985).

Uz iste pretpostavke normalne distribucije i minimalnog pogađanja, point-biserijalne korelacije mogu se transformirati u biserijalne korelacije, koje teoretski nisu osjetljive na prosječnu sposobnost u populaciji, iako, kao i svi koeficijenti korelacije, ovise o varijanci sposobnosti. Biserijalna korelacija je procjena kolika bi bila korelacija čestice i ukupnog rezultata (engl. *item-total*) da je čestica mjerena na kontinuiranoj skali, a ne na dihotomnoj skali. Ovakve nelinearne transformacije zahtijevaju vrlo velik uzorak, što nije relativni nedostatak jer i IRT modeli zahtijevaju velike uzorke, ali IRT modeli pritom ne zahtijevaju ispunjavanje pretpostavki o normalnosti distribucije i minimalnoj mogućnosti pogađanja.

Važno je naglasiti da iako je u nekim situacijama moguće postići usporedivost i skaliranje između IRT-a i CTT-a, IRT pruža i neke dodatne prednosti, poput mogućnosti modeliranja latentnih osobina, preciznijeg mjerenja, razumijevanja strukture testa i procjene parametara čestica na temelju statističkih modela. Stoga, iako se mogu postići određene razine usporedivosti, IRT pruža naprednije i sofisticiranije metode za analizu podataka u psihometriji

(Embretson i Reise, 2000). Ove prednosti IRT-a dovode do mogućnosti formiranja standardiziranih mjernih instrumenata i računalno adaptivnog testiranja (engl. *computer adaptive testing, CAT*).

Što se invarijantnosti parametara tiče, bitno je naglasiti da ta osobina vrijedi isključivo za parametre populacije, a ne i za procjene parametara u uzorku. Zbog pogreške mjerenja, čak i uz najbolje transformacije, dva skupa parametara dobivena na dva različita uzorka ispitanika, bit će u konačnici samo slična, nikada identična (deMars, 2010).

Osnovni pojmovi u teoriji odgovora na zadatak

Teorija odgovora na zadatak ima za cilj modelirati povezanost između odgovora ispitanika na pojedinoj čestici/zadatku i latentne varijable koja se nalazi u pozadini, a koju instrument nastoji izmjeriti. Ova latentna varijabla se često naziva konstrukt, osobina, domena ili sposobnost i simbolički je označena grčkim slovom theta, θ (Crocker i Algina, 2008).

IRT pretpostavlja da je količina latentne varijable koju pojedinac posjeduje ili iskazuje jedini faktor koji utječe na ispitanikov odgovor na pojedinoj čestici skale. Na primjer, osoba s visokom razinom emocionalne inteligencije će vjerojatno odgovoriti da "dobro raspoloženje može zadržati čak i kad mu/joj se dogodi nešto loše" i označiti odgovor "5 - u potpunosti DA" na toj čestici Upitnika emocionalne kompetencije. S druge strane, osoba s niskom emocionalnom inteligencijom će svojim odgovorom na tu česticu pokazati da nema ili samo djelomično ima sposobnost za takvo ponašanje.

Karakteristična krivulja čestice

Za modeliranje odgovora na čestici u teoriji odgovora na zadatak koristi se nekoliko parametara koji opisuju funkcioniranje te čestice. Odnos između odgovora ispitanika na čestici i latentne varijable modelira se putem karakteristične krivulje čestice (engl. *item characteristic curve, ICC*). Ova krivulja poznata je još i kao pravac traga čestice (engl. *item trace line*), krivulja funkcije odgovora na česticu (engl. *item response function, IRF*) ili krivulja kategorije odgovora (engl. *category response curve, CRC*) u slučaju politomnih čestica (deMars, 2010).

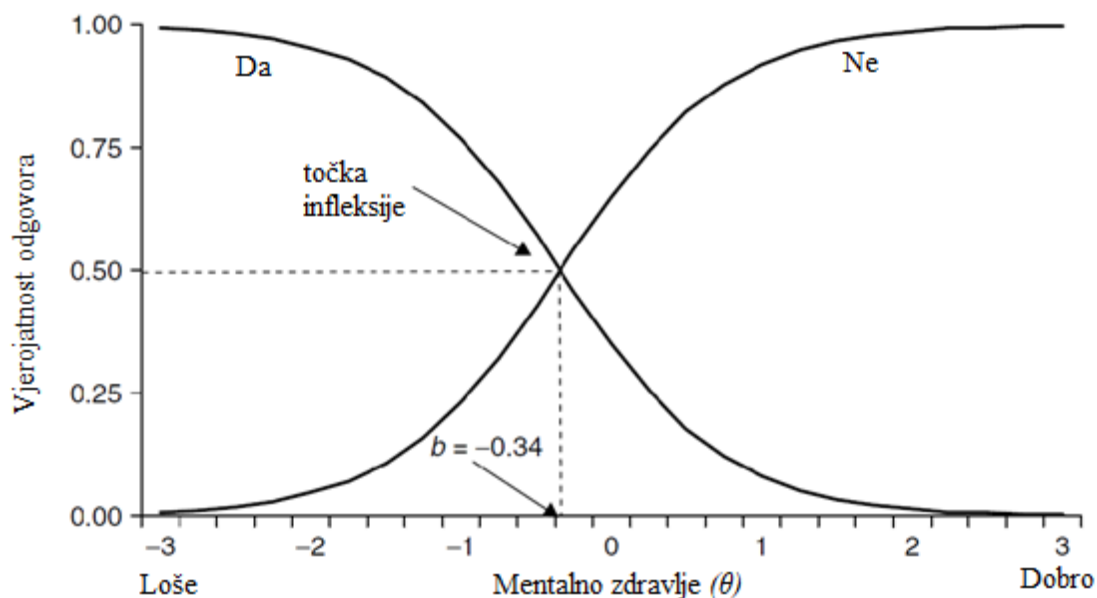
Iako se u hrvatskom jeziku često koristi izraz „karakteristična krivulja čestice“, smatram da bi prijevod „krivulja parametara čestice“ tj. „krivulja karakteristika (svojstava) čestice“ bio precizniji. Naime, svrha te krivulje je prikazati parametre čestice i njihov odnos s latentnom varijablom. No, s obzirom na uvriježenost pojma „karakteristična krivulja čestice“, on će se koristiti u ostatku rada.

Na Slici 1. prikazan je jedan primjer karakteristične krivulje čestice. Riječ je o čestici iz subskale psihičkog zdravlja Upitnika zdravstvenog statusa SF-36 koja glasi „Jeste li u protekla 4 tjedna zbog utjecaja bilo kakvih emocionalnih problema (npr. osjećaj depresije ili tjeskobe) na poslu ili pri obavljanju nekih drugih svakodnevnih aktivnosti obavili manje nego

ste željeli?“. Ovom se subskalom koja sadrži 14 čestica mjeri latentna varijabla psihičko zdravlje (θ), koja je prikazana na horizontalnoj osi grafa. Na lijevoj strani osi kontinuuma nalaze se osobe vrlo lošeg samoprocijenjenog psihičkog zdravlja, a na desnoj se strani nalaze oni pojedinci koji su svoje psihičko zdravlje procijenili jako dobrim.

Brojke na apscisi su standardizirane jedinice, a za potrebu ilustracije, prosječno je psihičko zdravlje istraživane populacije centrirano na vrijednost 0 uz standardnu devijaciju od 1. Ordinata na grafu prikazuje vjerojatnost (u rasponu od 0 do 1) da će osoba odabrati jednu od kategorija odgovora na čestici, a s dvjema karakterističnim krivuljama čestice prikazana je vjerojatnost da ispitanik na česticu odgovori s „da“ ili „ne“ ovisna o razini njegova samoprocijenjenog psihičkog zdravlja. Psihički zdraviji pojedinci imaju veću vjerojatnost da će na ovoj čestici odgovoriti „ne“, dok osobe slabijeg psihičkog zdravlja imaju veću vjerojatnost dati odgovor „da“.

Slika 1. Karakteristične krivulje odgovora na česticu „Jeste li u protekla 4 tjedna zbog utjecaja bilo kakvih emocionalnih problema (npr. osjećaj depresije ili tjeskobe) na poslu ili pri obavljanju nekih drugih svakodnevnih aktivnosti obavili manje nego ste željeli?“



Napomena: Težina čestice od -0.34 predstavlja razinu mentalnog zdravlja pri kojoj je vjerojatnost odgovora „da“ i „ne“ jednaka (tj. 50%). Prilagođeno prema Changu i Reeves, 2005.

Krivulje parametara čestice na Slici 1. su logističkog oblika i njima se modelira vjerojatnost (P) da će osoba odgovoriti „ne“ pomoću sljedeće matematičke funkcije, koja se monotono povećava:

$$P(\theta, a_i, b_i) = \frac{1}{1 + e^{a_i(\theta - b_i)}} \quad (1)$$

Kako je vidljivo iz gornje formule (koja predstavlja u ovom slučaju korišten dvoparametrijski model), vjerojatnost da će osoba odgovoriti „ne“ na ovu česticu je funkcija razine njezinog samoprocijenjenog psihičkog zdravlja (θ), povezanosti čestice s mjerenim konstruktom (a) i praga tj. limena čestice na skali (b).

U teoriji odgovora na zadatak, a i b se nazivaju parametrima čestice. U situaciji kakva je ova, kada čestica ima samo dvije kategorije odgovora, formula za monotono opadajuću funkciju (kojom je prikazan odgovor „da“) je jednostavna linearna transformacija prethodne formule:

$$1 - \frac{1}{1 + e^{a_i(\theta - b_i)}} \quad (2)$$

S obzirom na komplementarnost funkcija kategorija kod dihotomnih čestica, u praksi se funkcija odgovora obično eksplicitno modelira i grafički prikazuje samo za jednu kategoriju odgovora (tipično je to monotono rastuća funkcija kojom se modelira pozitivno povezan odgovor).

Težina čestice

Težina čestice i u okviru klasične i u okviru moderne teorije testova predstavlja vjerojatnost točnog odgovora, a ne subjektivni doživljaj težine ili uložene količine truda. U CTT je indeks težine, p , jednak proporciji ispitanika koji su točno odgovorili na zadatak. Za politomne (ili polihotomne) čestice, težina se računa kao prosječna vrijednost kategorije odgovora. Recimo, ukoliko imamo česticu s Likertovom skalom od 5 stupnjeva na kojoj su dva ispitanika zaokružila 5. stupanj, a dva 4. stupanj, prosječna vrijednost iznosi 4.5 i to je ujedno težina čestice.

U okviru IRT-a, indeks težine, b , označava količinu mjerene osobine pri kojoj se očekuje da će otprilike 50% ispitanika (ili nešto više, ukoliko je u model uključen parametar pogađanja c) točno odgovoriti na česticu. Indeks težine b mjeri se istom skalom kao i latentna osobina. Ova je skala arbitrarna, ali vrlo često se centrirana na način da distribucija latentne osobine ima aritmetičku sredinu 0 i standardnu devijaciju 1. Za razliku od CTT-a, teže čestice u IRT-u imaju viši indeks težine (u CTT-u je indeks težine zapravo niži što je čestica teža, pa ga neki nazivaju i indeksom lakoće (engl. *facility index*)).

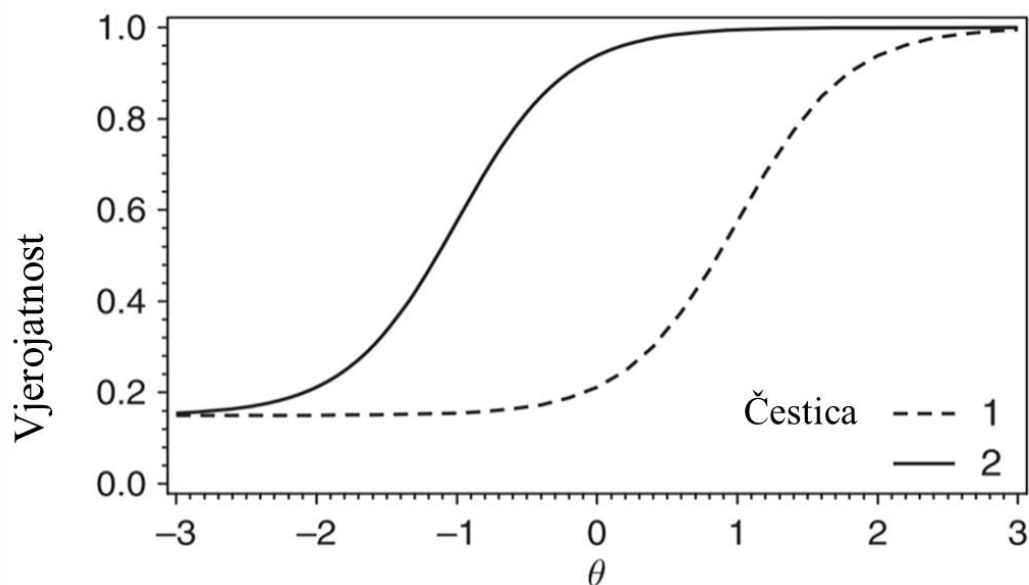
Parametar čestice b naziva se parametrom težine, lokacije ili limena i kod dihotomnih čestica je analogan klasičnom pokazatelju težine čestice. To je točka na latentnoj skali θ u kojoj osoba ima 50% vjerojatnosti da odgovori „ne“ (ili „da“) na danu česticu.

Na Slici 1. vrijednost limena čestice je $b = -0.34$, a to predstavlja razinu samoprocijenjenog psihičkog zdravlja u standardiziranim jedinicama pri kojoj je vjerojatnost davanja odgovora „da“ i „ne“ jednaka. Odnosno, za osobu čija je samoprocijenjena z -vrijednost psihičkog zdravlja jednaka -0.34 ne možemo biti sigurni kako će odgovoriti na ovu česticu, jer je vjerojatnost i jednog i drugog odgovora jednaka, ali zato znamo da je za osobe slabijeg psihičkog zdravlja vjerojatnost potvrdnog odgovaranja veća, dok je za osobe boljeg psihičkog zdravlja veća vjerojatnost niječnog odgovora na ovu česticu.

Vrijednosti parametara limena razlikuju se za svaku česticu subskale, te je usporedbom ovih vrijednosti moguće odrediti one čestice s kojima će se vjerojatnije složiti osobe slabijeg odnosno one s kojima će se složiti osobe boljeg psihičkog zdravlja.

Na Slici 2. prikazane su dvije čestice različite težine. Povlačenjem imaginarnog pravca s mjesta na ordinati na kojem je vjerojatnost točnog odgovaranja na česticu 0.5 odnosno 50% do sjecišta s krivuljom čestice, te zatim spuštanjem drugog pravca do apscise, možemo dobiti vrijednost b parametra za obje čestice, te zaključiti da je čestica 1 teža od čestice 2. Osim toga, za bilo koju vrijednost latentne osobine vjerojatnost točnog odgovaranja na česticu 1 je niža od vjerojatnosti točnog odgovaranja na česticu 2.

Slika 2. Prikaz čestica različite težine



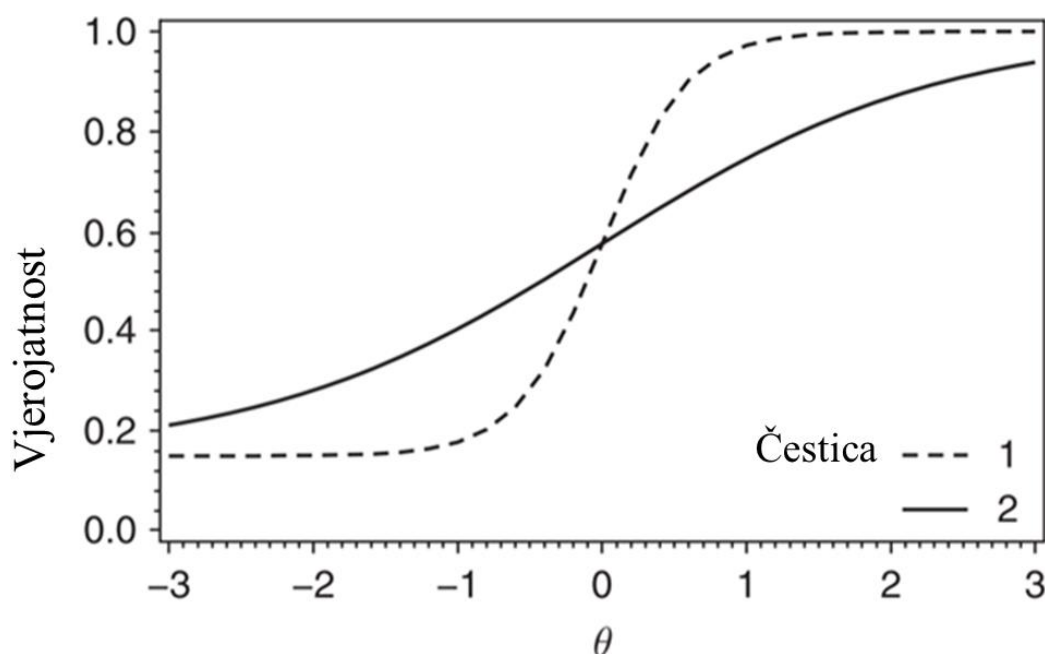
Napomena: Prilagođeno prema deMars, 2010.

Diskriminativnost čestice

Parametar a je parametar diskriminativnosti ili nagiba krivulje i njime se definira koliko je neka čestica sposobna diferencirati pojedince koji posjeduju različite količine latentne varijable. Čestica optimalno diskriminira pojedince koji posjeduju količinu latentne varijable na razini i oko limena čestice b . Parametar diskriminativnosti se obično kreće u rasponu vrijednosti od 0.5 do 2.5.

Na slici 1. nagib krivulje u točki infleksije odnosno preokreta (tj. točki u kojoj se nagib mijenja iz kontinuiranog povećanja u kontinuirano opadanje) iznosi $a=1.82$. Što je ova vrijednost veća, krivulja parametara čestice je strmija u točki infleksije, a strmiji nagibi ukazuju da se krivulja povećava relativno brzo, to jest da malene promjene u latentnoj varijabli (npr. male promjene statusa psihičkog zdravlja) vode do velikih promjena u vjerojatnosti slaganja s česticom. Parametar a se još interpretira i kao povezanost između čestice i osobine koju skala mjeri, te je tako direktno povezan, uz pretpostavku normalne distribucije latentne osobine θ , s biserijalnom¹ *item-total* korelacijom kao pokazateljem iz klasične teorije testova (van der Linden i Hambleton, 1997). Čestice koje imaju veći parametar nagiba snažnije su povezane s latentnim konstruktom i više doprinose ukupnom rezultatu pojedinca ($\hat{\theta}$). Na slici 3. prikazane su dvije čestice s različitim parametrom diskriminativnosti, među kojima čestica 1 ima veći nagib, te je istovremeno i diskriminativnija.

Slika 3. Prikaz čestica različite diskriminativnosti



Napomena: Prilagođeno prema deMars, 2010.

¹ Biserijalna korelacija je korelacija između dihotomne varijable (rezultat na čestici) i kontinuirane varijable (rezultat na subskali).

U okviru CTT-a se kao indeks diskriminativnosti čestice koristi korigirani point-biserijalni koeficijent korelacije zadatka s ukupnim rezultatom na testu (tzv. *item-total* korelacija). Kada je ovaj koeficijent pozitivan, ispitanici koji su točno riješili zadatak postići će veći broj bodova i na sumi svih ostalih zadataka u testu, nego oni ispitanici koji su netočno odgovorili na dotični zadatak.

Pouzdanost i standardna pogreška mjerenja

Klasična teorija testova definira pouzdanost kao omjer varijance pravih rezultata i varijance opaženih rezultata $\left(\frac{\sigma_T^2}{\sigma_X^2}\right)$ ili kao kvadriranu korelaciju između pravih i opaženih rezultata (ρ_{XT}^2) , a pritom pretpostavlja da je pravi rezultat hipotetska prosječna vrijednost opaženih rezultata koja bi bila dobivena da se mjerenje ponovi neograničeni broj puta i to u sličnim uvjetima.

Standardna pogreška mjerenja ili standardna pogreška procijenjenog rezultata, σ_e , temelji se na pretpostavci da je varijanca opaženih rezultata jednaka zbroju varijance pravih rezultata i varijanci pogrešaka ($\sigma_X^2 = \sigma_T^2 + \sigma_e^2$, to jest $\sigma_e = \sqrt{(1 - \rho_{XT}^2)}$). Kad se koristi ova definicija standardne pogreške, dobiva se jedinstvena procjena standardne pogreške mjerenja neovisno o vrijednosti opaženog rezultata. Pritom nam je za izračunavanje standardne pogreške potrebna standardna devijacija rezultata u uzorku, σ_X , i neka procjena pouzdanosti (koeficijent korelacije paralelnih testova, test-retest koeficijent korelacije ili Cronbach α koeficijent (KR-20 kod dihotomnih čestica)).

Drugi način računanja standardne pogreške mjerenja je putem formule

$\sigma_e = \sqrt{\frac{X(n-X)}{n-1}}$, u kojoj je X ukupni rezultat izračunat kao suma pozitivnih odgovora (engl. *number-correct score*) a n je broj čestica. Temeljem ove formule vidimo da ekstremni rezultati (oni bliži 0 ili maksimalnoj vrijednosti) imaju niže standardne pogreške od rezultata koji su bliži prosjeku. Kad se koristi ova formula nije potrebno imati prethodnu procjenu pouzdanosti, nego se prvo izračunava standardna pogreška, a zatim, kada je ona poznata, računa se koeficijent pouzdanosti.

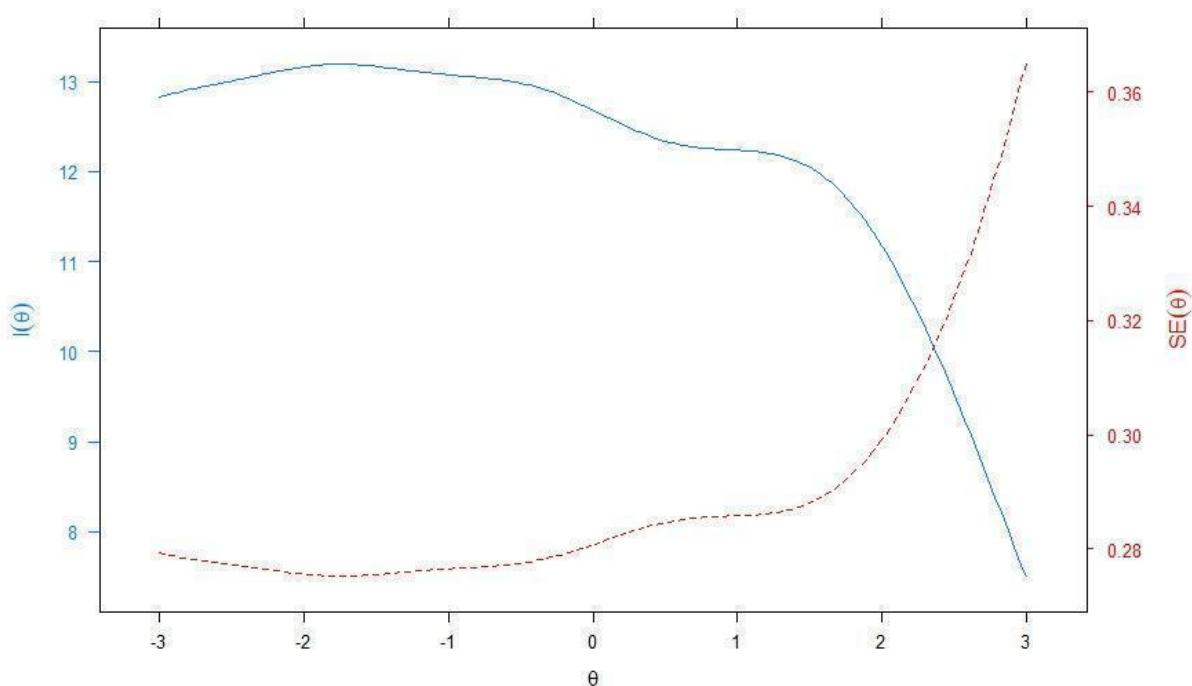
Klasična teorija testova izračunava samo jedan koeficijent pouzdanosti i jednu pogrešku mjerenja, neovisno o vrijednosti opaženog rezultata, X , tj. neovisno o razini mjerene osobine ispitanika.

Teorija odgovora na zadatak koristi informacijsku funkciju (engl. *information function*) za izračun pouzdanosti i standardne pogreške mjerenja. Informativnost testa (engl. *test information*) je funkcija kompetentnosti (engl. *proficiency*) (odnosno osobine, sposobnosti, vještine ili stava koji test mjeri) i čestica na testu.

Informativnost testa ($I(\theta)$) je različita na različitim razinama latentne osobine (Slika 4.). Standardna pogreška mjerenja je inverzna vrijednost drugog korijena iz informativnosti:

$$SE(\theta) = 1/\sqrt{I(\theta)} \quad (3)$$

Slika 4. Informativna funkcija



Napomena: Informativnost testa ($I(\theta)$) prikazana je plavom linijom, a standardna pogreška mjerenja ($SE(\theta)$) crvenom isprekidanom linijom.

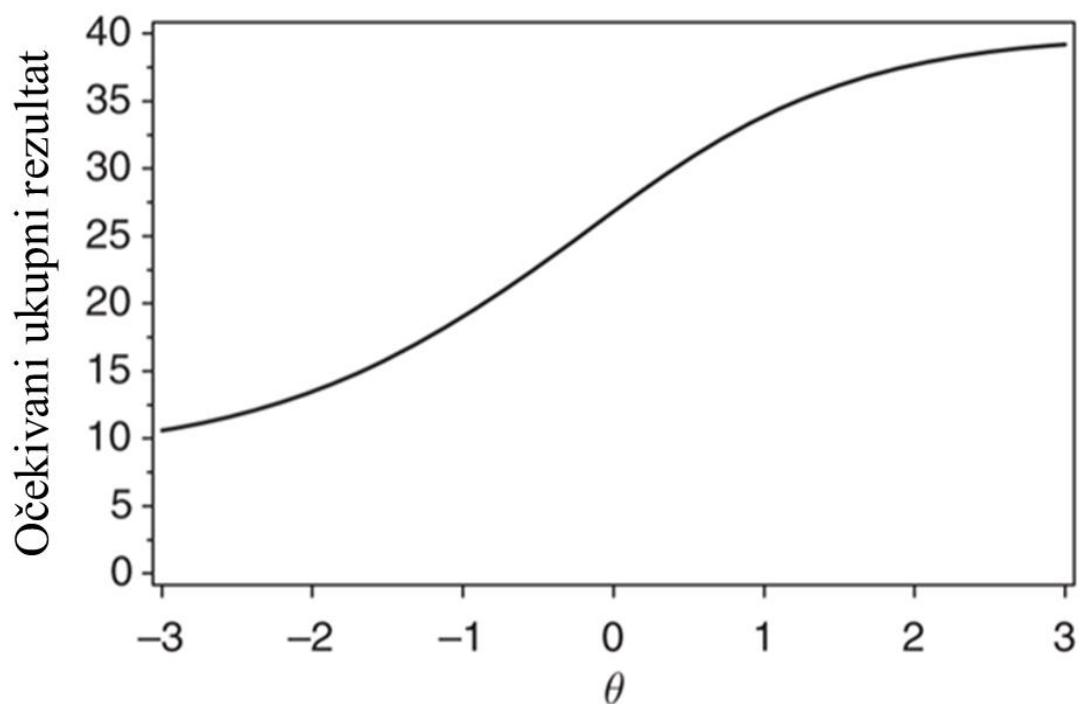
Prednost IRT-a nad CTT-om je u činjenici da se informativna funkcija može definirati i na razini pojedine čestice, a pojedinačne informativne funkcije čestica mogu se sumirati u informativnu funkciju testa. Ta osobina je vrlo korisna jer se na taj način čestice iz različitih verzija testa mogu koristiti za formiranje različitih konfiguracija testa, a zatim se informativnost testa može izračunati za svaku novu verziju testa i prije njegove primjene. Kod klasične teorije testova, nove bi se verzije testa morale primijeniti na novim uzorcima kako bi se izračunao novi koeficijent pouzdanosti.

Također, kod IRT-a je moguće isključiti dio čestica i vrlo brzo izračunati novu informativnost testa jednostavnim oduzimanjem informativnih funkcija isključenih čestica. Kod CTT-a je ponovna procjena pouzdanosti u tom slučaju matematički dosta kompleksnija. Osim toga, procjene pouzdanosti IRT rezultata za određenu skupinu ispitanika moguće je dobiti i unaprijed (tj. prije samog testiranja) samo temeljem poznatih parametara čestica i distribucije latentne osobine u toj populaciji.

Karakteristična krivulja testa

Karakteristična krivulja testa predstavlja zbroj karakterističnih krivulja pojedinih čestica i prikazuje očekivanu vrijednost ukupnog rezultata ispitanika u ovisnosti o razini mjerene osobine θ . Ova krivulja također ima oblik slova "S" (tzv. sigmoidna krivulja) te služi za prikazivanje trenda kojim se očekivani rezultat mijenja s promjenom razine latentne osobine. Karakteristična krivulja testa može se formirati i kod dihotomnih i kod politomnih čestica (tada se sumiraju karakteristične krivulje kategorija).

Slika 5. Karakteristična krivulja testa



Napomena: Prilagođeno prema deMars, 2010.

1.1.1. Pretpostavke teorije odgovora na zadatak

Glavna pretpostavka IRT-a je da instrument mjeri samo jednu latentnu osobinu, odnosno da je jednodimenzionalan. Kadgod mjerni instrument daje samo jedan ukupni rezultat pretpostavlja se da sve čestice dijele zajednički primarni konstrukt. **Jednodimenzionalnost** znači da model ima samo jednu *thetu* za pojedinog ispitanika, a svi drugi faktori koji eventualno utječu na odgovaranje na česticu smatraju se ili slučajnom pogreškom ili ometajućom dimenzijom jedinstvenom samo za tu česticu. Kršenje ove temeljne pretpostavke može dovesti do pogrešnih procjena parametara ili standardnih pogrešaka.

Iako neki testovi evidentno mjere više od jednog konstrukta, oni i dalje mogu tehnički biti jednodimenzionalni. Tako, primjerice, u matematičkim problemskim zadacima koji su predstavljeni tekstualno, očito je da se za odgovor na takve zadatke od ispitanika očekuje uz znanje matematike određena razina čitalačke sposobnosti. No, ukoliko sve čestice testa mjere oba ova konstrukta u jednakom relativnom omjeru, onda će oni matematički i dalje mjeriti jedinstvenu latentnu osobinu koja je hibrid ova dva konstrukta. Također, ukoliko se ispitanici ne razlikuju po jednom od ova dva konstrukta (recimo, ista je razina čitalačke sposobnost), sve individualne razlike mogu se pripisati ovom drugom konstrukt, a odgovori su u tom slučaju matematički jednodimenzionalni.

Slično je to i kod zadataka koji imaju ograničeno vrijeme za odgovor, jer se u tom slučaju mjeri i znanje i brzina odgovaranja. Ili primjerice kod selekcijskih ispita kod kojih se osim sposobnosti mjeri i motivacija za rješavanjem testa.

Da bi mjerni instrumenti bili višedimenzionalni, različite čestice trebaju pokrivati različite kombinacije konstrukta i ispitanici trebaju varirati na oba (ili više) konstrukta koje test mjeri. Primjerice, kod već navedenih testova s ograničenim vremenom za odgovaranje, čestice pri kraju testa mjerit će više drugi konstrukt, brzinu odgovaranja, i zbog različitog omjera ova dva konstrukta u različitim česticama, ovaj je test višedimenzionalan. Postoji mnogo metoda za provjeru jednodimenzionalnosti mjernog instrumenta, no najčešće se koriste svojstvene vrijednosti korelacijske matrice čestica (npr. *scree* test ili paralelna analiza), Stoutov test esencijalne jednodimenzionalnosti (tzv. DIMTEST procedura) i analize temeljene na rezidualima.

Sljedeća pretpostavka je usko povezana s prethodnom. Ukoliko je test zaista jednodimenzionalan, odgovor na jednu česticu bit će neovisan o odgovoru na bilo koju drugu česticu testa, uz kontrolu razine latentne varijable. Ova pretpostavka naziva se **lokalna nezavisnost** (engl. *local independence*).

Lokalno nezavisne čestice trebaju biti nekorelirane nakon što se ekstrahira utjecaj *thete*. No, ukoliko čestice testa za kojeg se smatra da je jednodimenzionalan nisu lokalno nezavisne, postoji mogućnost da njihovu povezanost izaziva dodatna dimenzija. Fokus je kod ove pretpostavke na provjeri ovisnosti (engl. *dependencies*) između parova čestica. Ako one ne utječu na velik broj čestica, te pojedinačne međuovisnosti ne moraju rezultirati odvojenim dimenzijama, pa ih iz tog razloga provjere jednodimenzionalnosti mogu previdjeti.

Upravo zbog toga su razvijeni dodatni postupci za provjeru lokalne nezavisnosti. Jedan od češće korištenih je Yenov Q_3 test (Yen, 1984) koji se temelji na korelaciji reziduala parova čestica. Reziduali koji se uspoređuju računaju se kao odstupanja između odgovora predviđenog odabranim IRT modelom i opaženog odgovora. Sve korelacije reziduala parova čestica veće od .20 potencijalno narušavaju pretpostavku lokalne nezavisnosti.

Zadnja je pretpostavka da je odabrani model ispravno specificiran. **Slaganje između podataka i modela** provjerava se na više načina. Uobičajene provjere uključuju slaganje čestica (engl. *item fit*), generalno slaganje ili slaganje modela (engl. *overall fit*; *model fit*) te, pretežno kod Raschevih sljedbenika, slaganje osoba (engl. *person fit*).

Dvije su kategorije indeksa slaganja: Pearsonov χ^2 ili χ^2 log-vjerojatnosti, odnosno G^2 , a postoje i modificirane verzije oba, $S - \chi^2$ i $S - G^2$. Osim toga, i dalje se preporuča vizualna inspekcija grafičkog prikaza reziduala, koja iako ne daje statistički test slaganja, još uvijek može biti sasvim adekvatna za donošenje zaključka.

Za procjenu slaganja osoba, Raschevi sljedbenici koriste drugačije indekse, koji se nazivaju Outfit MS (engl. *mean square*) i Infit MS (engl. *mean square*), no s obzirom da oni nisu rutinski dio IRT analiza, o njima ovdje više neće biti govora.

1.1.2. Modeli za dihotomne čestice

Modelima za dihotomne čestice se modelira se odnos između vjerojatnosti točnog odgovaranja na čestice koje imaju samo dvije kategorije (točno-netočno) i latentne osobine koja se mjeri tim instrumentom.

Modeli za dihotomne čestice naziv su dobivali prema broju parametara koje uključuju, te prema vrsti funkcije kojom se modelira odnos sa latentnom varijablom. Tako imamo: jednoparametarski (engl. *one-parameter logistic model*, *IPL model* ili *IPLM*), dvoparametarski, troparametarski, četveroparametarski logistički model, ali isto tako i jednoparametarski (engl. *one-parameter normal ogive model*, *IPNO model* ili *IPNOM*), dvoparametarski, troparametarski, četveroparametarski model normalne ogive².

Jedan od nedostataka modela normalne ogive u usporedbi s logističkim modelima je potreba za korištenjem integrala za izračunavanje vjerojatnosti odgovora. Budući da su integrali računski zahtjevniji, logistički modeli su se proširili kao jednostavnija alternativa, pogotovo u vrijeme dok su računalne tehnologije bile manje razvijene. Međutim, s razvojem računalnih tehnologija, interes za modele temeljene na normalnoj ogivi ponovno se javlja. Važno je napomenuti da je prvi model koji je Lord koristio još davne 1953. godine bio dvoparametarski model normalne ogive (2PNOM) (Lord, 1953). Danas, uz parametrijske modele, postoje i rjeđe korišteni neparametrijski modeli. Međutim, detaljnije rasprave o neparametrijskim modelima neće biti obuhvaćene u ovom kontekstu.

Jednoparametarski model za predviđanje vjerojatnosti odgovora koristi isključivo b -parametar odnosno težinu čestice, dvoparametarski u modeliranje uključuje i a -parametar odnosno diskriminativnost čestice, a troparametarski dodaje i c -parametar odnosno parametar donje asimptote ili pseudo-pogađanja. Relativno rjeđe u upotrebi je i četveroparametarski model, koji u modeliranje uključuje i parametar gornje asimptote, d -parametar.

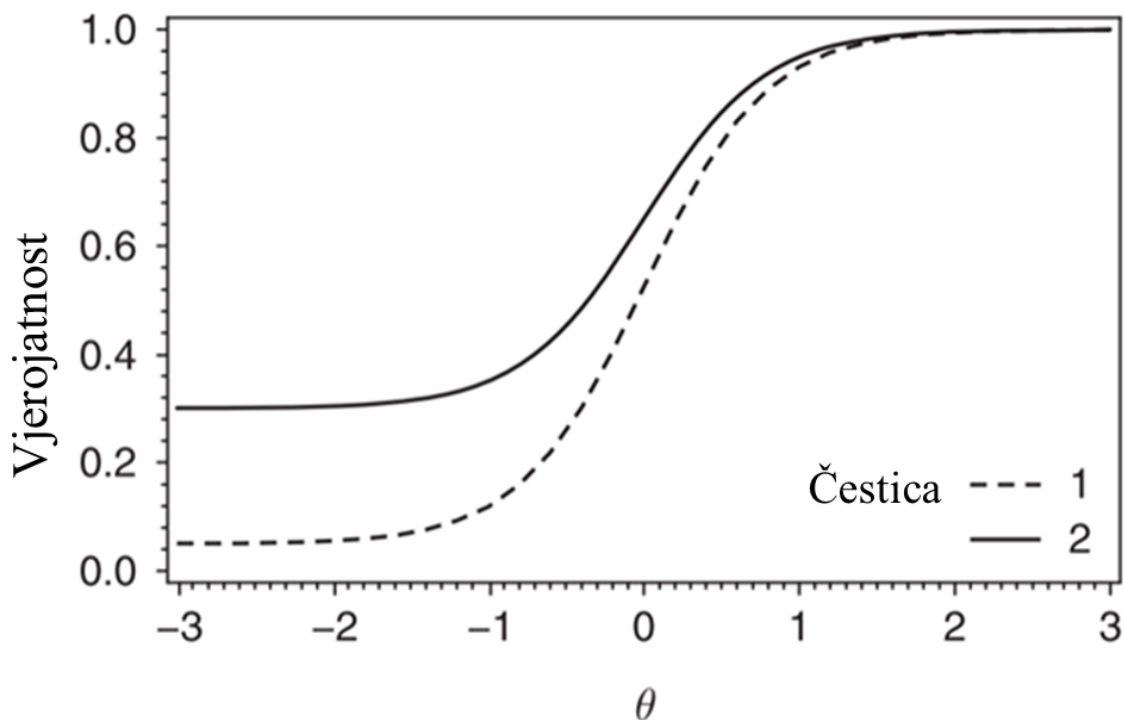
Parametar pseudo-pogađanja se koristi kako bi se objasnila činjenica da na testovima s višestrukim izborom rijetko dobivamo nulte rezultate, čak i kada ispitanici imaju nisku razinu sposobnosti, jer samim pogađanjem mogu postići određeni broj bodova. Slično tome,

² Ogiva je drugi naziv za funkciju kumulativne raspodjele (eng. *cumulative distribution function*, *CDF*) standardne normalne distribucije.

parametar gornje asimptote se koristi kako bi se objasnila rijetkost potpuno točno riješenih testova, uz pretpostavku da čak i ispitanici visoke sposobnosti mogu pogriješiti iz različitih razloga kao što su primjerice nemar, nedostatak motivacije ili umor.

Za ispravnu procjenu sposobnosti ispitanika lošeg postignuća 3PLM je korisniji od modela s manje parametara. S druge strane, on ne uzima u obzir da i vrlo uspješni ispitanici mogu također na lagani zadatak odgovoriti netočno, bilo iz nemara, nedostatka motivacije ili radi stresa. Iz tog razloga su Barton i Lord 1981. predložili model s gornjom asimptomom koja nije jednaka 1 kao što je to slučaj u jednostavnijim modelima (Barton i Lord, 1981).

Slika 6. Prikaz čestica s različitim parametrom pseudo-pogađanja



Napomena: Prilagođeno prema deMars, 2010.

Donja asimptota je dio krivulje koji sadrži vrijednosti koje funkcija vjerojatnosti odgovora postiže kako se θ približava negativnoj beskonačnosti. Gornja asimptota je dio krivulje koji sadrži vrijednosti koje funkcija vjerojatnosti odgovora postiže kako se θ približava pozitivnoj beskonačnosti. Parametar donje asimptote nam govori koja je vjerojatnost da ispitanik s vrlo niskom razinom latentne osobine točno odgovori na tu česticu.

Na Slici 6. prikazane su dvije čestice s različitim parametrom donje asimptote, odnosno parametrom pseudo-pogađanja. Donja asimptota niža je kod čestice 1 nego kod čestice 2, što znači da je parametar pseudo-pogađanja niži kod čestice 1. To znači da ispitanici s niskom razinom sposobnosti imaju manju vjerojatnost točnog odgovora na česticu 1 nego na česticu 2, odnosno da je odgovor na čestici 2 lakše pogoditi.

Iako su modeli s manje parametara ranije razvijeni, danas se na njih gleda kao na posebne slučajeve odnosno ograničene modele 4PL modela. Ograničavanje modela znači da se jedan ili više parametara fiksira na određenu vrijednost (uobičajeno 0 ili 1).

Matematički izraz za 4PL model je sljedeći:

$$P(\theta) = c_i + (d_i - c_i) \frac{e^{1.7a_i(\theta-b_i)}}{1 + e^{1.7a_i(\theta-b_i)}} \quad (4)$$

gdje $P(\theta)$ predstavlja vjerojatnost točnog odgovora uz zadanu razinu latentne osobine θ i parametre čestice (potpuniji oblik za $P(\theta)$ je $P(x=I | \theta, a, b, c, d)$). Indeks i ukazuje na određenu česticu, a ponekad se uz latentnu osobinu θ dodaje indeks j , kako bi se označio određeni ispitanik, θ_j . No, češće se oba ova indeksa samo podrazumijevaju kako bi se olakšalo pisanje jednadžbe.

Slovo e u jednadžbi označava iracionalni broj e ili matematičku konstantu koja se još naziva i Eulerov broj ili Napierova konstanta i kojim se definira eksponencijalna funkcija. To je baza prirodnog logaritma čija vrijednost iznosi 2.718, a logaritam od e je jednak 1. U gornju jednadžbu je uključena i konstanta 1.7 koja se naziva parametrom skaliranja, D , a njegova je upotreba proizvoljna, ali njegovim se isključivanjem mijenja mjerna ljestvica parametra a . Parametar skaliranja omogućava direktne usporedbe između logističkog modela i modela normalne ogive. Bez parametra skaliranja a parametar postaje veći, a ova dva modela više nisu usporediva.

3PL model se od prethodnog razlikuje samo po tome što se d parametar fiksira na vrijednost 1, pa je matematički izraz sljedeći:

$$P(\theta) = c_i + (1 - c_i) \frac{e^{1.7a_i(\theta-b_i)}}{1 + e^{1.7a_i(\theta-b_i)}} \quad (5)$$

Kod 2PL modela se, osim parametra gornje asimptote, fiksira i parametar donje asimptote na vrijednost 0, pa sad ova jednadžba izgleda ovako:

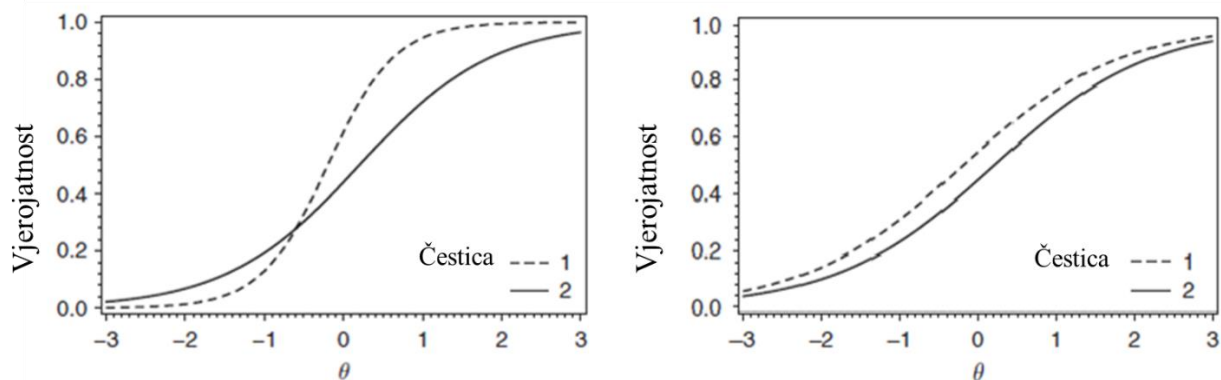
$$P(\theta) = \frac{e^{1.7a_i(\theta-b_i)}}{1 + e^{1.7a_i(\theta-b_i)}} \quad (6)$$

1PL model je još jednostavniji, te osim što su parametri asimptote fiksirani, i variranje parametra diskriminativnosti se ograničava na način da je dopušteno samo da a ima istu vrijednost za sve čestice. Matematički se to može prikazati na sljedeći način:

$$P(\theta) = \frac{e^{1.7a(\theta-b_i)}}{1 + e^{1.7a(\theta-b_i)}} \quad (7)$$

U usporedbi s prethodnom formulom, u gornjoj više nema indeksa i , nego se samo jedna vrijednost a koristi za sve čestice.

Slika 7. Usporedba dvoparametarskog (lijevo) i jednoparametarskog modela (desno)



Napomena: Prilagođeno prema deMars, 2010.

Često se 1PL model izjednačava s Raschevim modelom (1960, 1980). Iako su oni matematički ekvivalentni, Georg Rasch je svoj model razvio odvojeno, a ne kao poseban slučaj Lordovog 2PL modela (tj. 2PNOM). Iako matematički jednak 1PL modelu, on se konceptualno razlikuje, a temeljem njega razvilo se cijelo jedno područje psihometrije zvano Raschova analiza (engl. *Rasch measurement* ili *Rasch analysis, RA*), koja je, za razliku od IRT-a, konfirmatorna.

Naime, sljedbenici Rascha smatraju da bi sve čestice trebale slijediti 1PL model (tj. Raschov model), odnosno mjerenje latentne osobine bi trebalo biti ovisno samo o težini čestice. Oni se posvećuju konstrukciji upravo takvih čestica, te sve čestice koje ne slijede ovaj model eliminiraju iz mjernog instrumenta. Zagovornici teorije odgovora na zadatak objašnjavaju da bi „čestice trebale mjeriti onako kako čestice mjere, a ne onako kako bi to trebale činiti“ (Thissen i Orlando, 2001), te da je uloga psihometričara eksplorativna, odnosno da je potrebno pronaći model koji odgovara česticama, a ne obrnuto.

Raschov model je prvotno bio specificiran u terminima izgleda (engl. *odds*) ili logaritama izgleda (engl. *log-odds*) tj. logita, no danas se, kao i IRT, češće specificira u terminima vjerojatnosti. Kada se model specificira u terminima vjerojatnosti, matematički je identičan prethodnoj jednadžbi, s izuzetkom da se obično ne zadržava parametar skaliranja, a parametar a se ispušta iz modela ili fiksira na vrijednost 1. Posljedica ovih promjena je da se mijenja veličina mjernih jedinica.

Također, u Raschovom modelu razlikuje se i sustav notacija: umjesto b koristi se δ , a umjesto θ uobičajeni je simbol β . S obzirom na to, matematički se Raschov model može prikazati na sljedeći način:

$$P(\beta) = \frac{e^{(\beta-\delta_i)}}{1+e^{(\beta-\delta_i)}} , \quad (8)$$

odnosno

$$\ln\left(\frac{P(\beta)}{1-P(\beta)}\right) = \beta - \delta_i . \quad (9)$$

Vjerojatnost točnog odgovora ispitanika je na taj način uvjetovana samo jednim faktorom, a to je udaljenost između sposobnosti ispitanika i težine čestica na Raschovom kontinuumu.

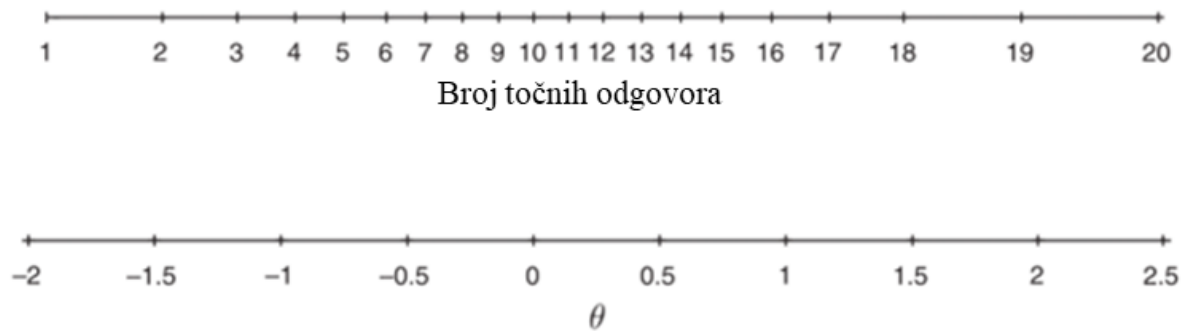
Raschov model i 1PL model imaju određena poželjna obilježja koja drugi modeli s više parametara nemaju. Za određivanje θ dovoljno je poznavati broj točnih odgovora, odnosno svi ispitanici s istim brojem točnih odgovora imaju istu procjenu θ .

Kod 2PL i 3PL modela tome nije tako. Ispitanici mogu imati isti broj točnih odgovora i posljedično isti ukupni rezultat, ali na procjenu θ će utjecati i obrazac danih odgovora.

Prema ovom obilježju, Raschov i 1PL model su ekvivalentni klasičnoj teoriji testova, kod koje je proporcija točnih odgovora dovoljan pokazatelj za određivanje razine latentne osobine. Ova karakteristika „dostatnosti“ (engl. *sufficiency*) ipak ne znači da postoji linearan odnos između proporcije točnih odgovora i latentne osobine (intervali između ukupno postignutih bodova nisu jednaki na svim razinama *theta* – u sredini su manji razmaci, a veći na krajevima, što znači da ispitanici prosječnih sposobnosti ostvaruju veći raspon različitih ukupnih rezultata nego ispitanici s relativno niskom i relativno visokom osobinom).

Drugim riječima broj bodova se mjeri na ordinalnoj skali, a *theta* na intervalnoj (Slika 8.).

Slika 8. Grafički prikaz odnosa broja točnih odgovora (CTT) i latentne osobine



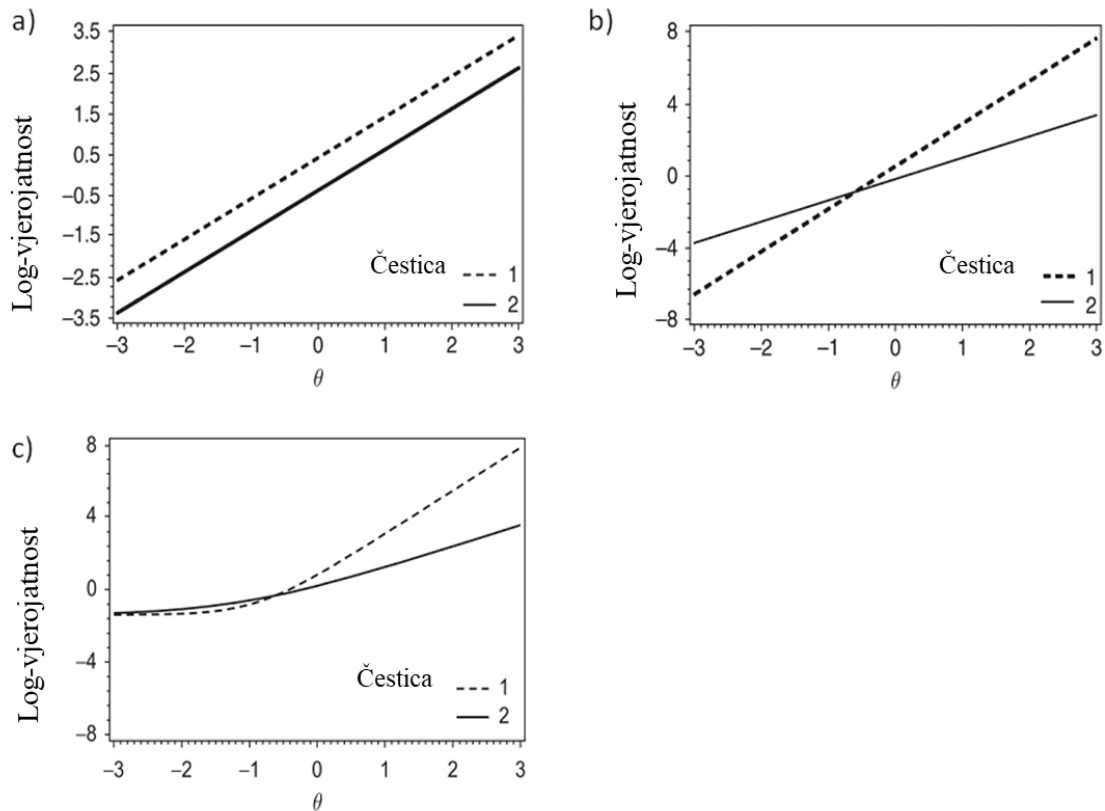
Napomena: U odnosu na skalu latentne osobine intervali između ukupnog broja točnih odgovora nisu jednaki. Prilagođeno prema deMars, 2010.

Prednost Raschova i 1PL modela nad CTT-om je u upotrebi logita, koji se nalaze na intervalnoj skali, pa je stoga odnos između vjerojatnosti točnog odgovora izražene u logitima i latentne osobine linearan. To praktično znači da ukoliko se dva ispitanika po latentnoj osobini razlikuju za 1 logit, razlika između logaritama njihovih izgleda za davanje točnog odgovora je također 1, neovisno o težini čestice i relativnom položaju ispitanika na kontinuumu *thete*.

No, prema Slici 8, kada je riječ o broju točnih odgovora (ukupnom rezultatu) u okviru CTT-a razlika vjerojatnosti točnog odgovora u logitima između ispitanika koji je postigao 10 i ispitanika koji je postigao 11 nije jednaka razlici u vjerojatnosti davanja točnog odgovora u logitima između ispitanika koji je postigao 18 i onog koji je postigao 19 bodova.

Još je jedno matematičko svojstvo prednost Raschovog i 1PL modela nad ostalim modelima za dihotomne čestice. Karakteristične krivulje dviju čestica se u ovim modelima nikad ne križaju, za razliku od kompleksnijih modela u kojima se to može dogoditi. U ovim modelima, krivulje čestica su paralelne jer samo težina čestice utječe na oblik krivulje, dok je nagib (diskriminativnost) jednak za sve čestice. Nasuprot tome, u modelima koji dopuštaju različitu diskriminativnost čestica, krivulje se mogu međusobno križati. Na primjer, jedna čestica može biti teža za ispitanike niže sposobnosti, ali lakša za ispitanike više sposobnosti, dok je druga čestica teža za ispitanike više sposobnosti, ali lakša za ispitanike niže sposobnosti. U tom se slučaju postavlja pitanje: Koja je čestica zapravo teža? Ovo svojstvo prikazano je na Slici 9.

Slika 9. Usporedba 1PL/Raschovog, 2PL i 3PL modela kada se vjerojatnost točnog odgovora prikazuje u logitima



Napomena: Prilagođeno prema deMars, 2010.

Kod 3PL modela to postaje još kompleksnije jer se gubi linearan odnos između logita i *thete* (Slika 9.) što Raschovi sljedbenici smatraju velikim nedostatkom teorije odgovora na zadatak.

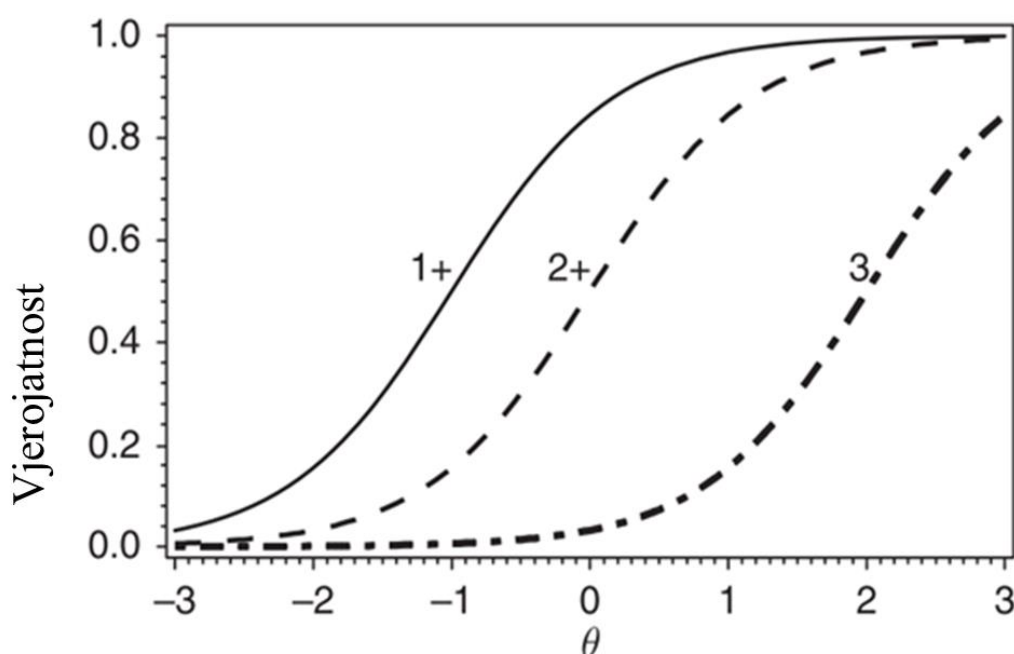
1.1.3. Modeli za politomne čestice

Politomne (odnosno polihotomne) čestice su čestice s više od dvije moguće kategorije odgovora. Među politomnim česticama razlikujemo nominalne i ordinalne čestice. Modeli za politomne čestice, za razliku od dihotomnih modela, obično nisu nazvani prema broju parametara i vrsti funkcije koju koriste za modeliranje, već se često nazivaju prema vrsti čestica za koje su najprikladniji. Ovi modeli se prilagođavaju specifičnostima politomnih čestica i pružaju metodologiju za analizu njihovih karakteristika.

Glavne vrste politomnih modela su: model nominalnih odgovora (engl. *nominal response model, NRM*), model stupnjevitih odgovora (engl. *graded response model, GRM*), (generalizirani) model djelomičnog bodovanja (engl. *(generalized) partial credit model, (G)PCM*), te model skala ocijenjivanja (engl. *rating scale model, RSM*).

Pomoću ovih modela se također može grafički prikazati odnos između pojedinih kategorija odgovora i latentne varijable koji je analogan karakterističnoj krivulji čestice. Nažalost još uvijek se ne koriste univerzalni pojmovi za ovakve krivulje, što može dovesti do nesporazuma, te je nužno pažljivo proučiti kontekst u kojem se pojedini termin javlja.

Slika 10. Prikaz politomne čestice modelirane pomoću modela stupnjevitih odgovora



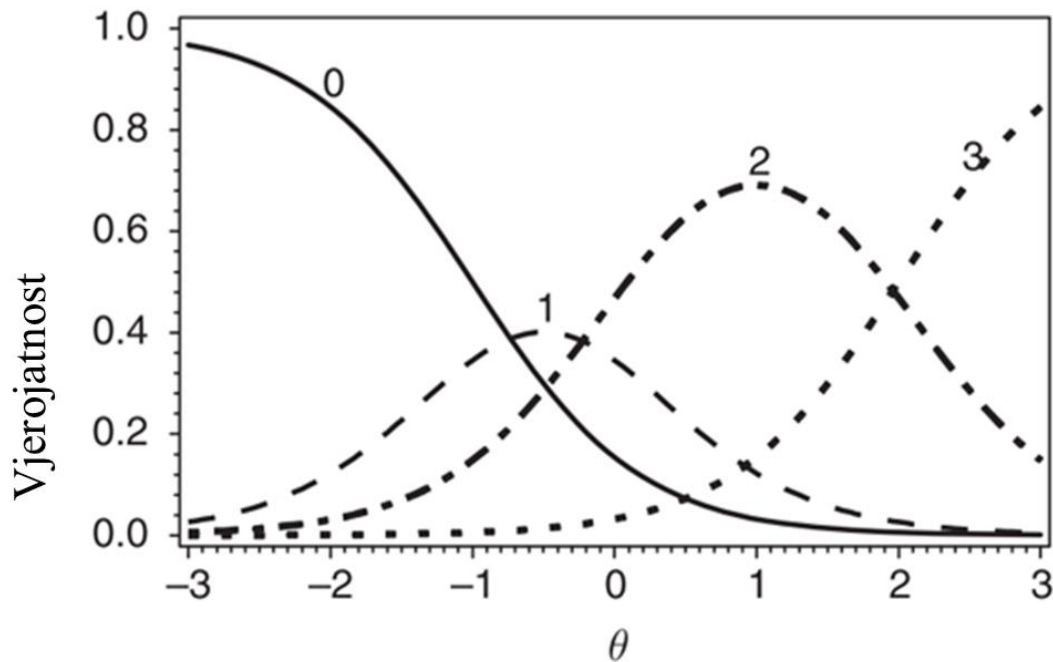
Napomena: Prilagođeno prema deMars, 2010.

Kod **modela stupnjevitih odgovora** (Samejima, 1969), koji se naziva još i modelom kumulativnih logita (engl. *cumulative logit model*), modelira se vjerojatnost da će ispitanik odabrati **određenu kategoriju ili kategorije iznad nje**.

Primjer modela za česticu s 4 kategorije dan je na Slici 10. Kategorije su označene s 0-3 (mogle su jednako tako biti označene i s 1-4). Vjerojatnost da će osoba odabrati odgovor 0 ili viši je naravno 1, tako da su grafički prikazane samo kategorije od 1 do 3. Njihove su funkcije paralelne, u suprotnom bi vjerojatnost davanja odgovora 1 ili višeg na nekim razinama latentne osobine bila niža nego vjerojatnost davanja odgovora 2 ili višeg, što se logički suprotstavlja cjelokupnoj poanti bodovanja pomoću stupnjevitih čestica.

Iz ovih funkcija koje prikazuju vjerojatnost odabira određene kategorije ili viših, može se izračunati vjerojatnost odabira konkretne kategorije, x . Ove se krivulje nalaze na Slici 11.

Slika 11. Prikaz vjerojatnosti odabira određene kategorije politomne čestice prema modelu stupnjevitih odgovora



Napomena: Prilagođeno prema deMars, 2010.

Samejima ih je zvao operativnim karakteristikama (engl. *operating characteristics*), a de Ayala (1993) operativnim karakterističnim krivuljama (dok je krivulje s prethodne slike nazvao karakterističnim krivuljama kategorija (engl. *category characteristic curves, CCC*). Upravo suprotno, Embretson i Reise (2000), krivulje sa Slike 10. nazivaju operativnim karakterističnim krivuljama (OCC), a one sa Slike 11. krivuljama kategorija odgovora (engl. *category response curves, CRC*). Sijtsma i Meijer (2007) upotrijebili su naziv funkcija kategorije odgovora (engl. *category response function, CRF*) za vjerojatnost odgovora x , a funkcija koraka u odgovoru na česticu (engl. *item step response function, ISRF*) za vjerojatnost odabira kategorije x ili više. Wilson (2005) je, kao i Sijtsma i Meijer, također koristio pojam funkcija kategorije odgovora, ali je za vjerojatnost odabira kategorije odgovora x ili više upotrijebio naziv kumulativna funkcija kategorija odgovora (engl. *cumulative category response function, CCRF*). Osim toga postoje još i pojmovi funkcija kategorije odgovora na česticu (engl. *item category response function, ICRF*) (Muraki, 1992) i funkcija opcije odgovora (engl. *option response function, ORF*) (de Ayala i Sava-Bolesta, 1999) za vjerojatnost odabira kategorije x , te funkcija granice odgovora (engl. *boundary response function, BRF*) (Oshima i Morris, 2008) za vjerojatnost odabira kategorije x ili više (deMars, 2010).

U osnovi je model stupnjevitih odgovora ekstenzija 2PL modela, stoga i njegova matematička funkcija izgleda slično. Jedina je razlika u tome da postoje višestruki b -parametri, po jedan za svaku kategoriju osim za prvu.

$$P_{ik}^*(\theta) = \frac{e^{1.7a_i(\theta-b_{ik})}}{1 + e^{1.7a_i(\theta-b_{ik})}}$$

(10)

U gornjoj formuli $P_{ik}^*(\theta)$ predstavlja vjerojatnost da će odgovor biti u kategoriji k čestice i (uz određenu razinu θ i zadane parametre čestice), a_i je kao i dosad nagib čestice, a b_{ik} je granica odnosno limen kategorije k (engl. *category boundary or threshold*) čestice i .

Primjerice na Slici 10. $b_{i1} = -1.0$ znači da 50% ispitanika s razinom latentne osobine $\theta = -1.0$ na čestici i postiže 1 ili više bodova (zaokružuje kategoriju 1 ili više).

Kod $P_{ik}^*(\theta)$ treba primijetiti da je dodan eksponent * (ponekad i $+$) kako bi se razjasnilo da je riječ o vjerojatnosti odabira kategorije k ili više, a ne kategorije k .

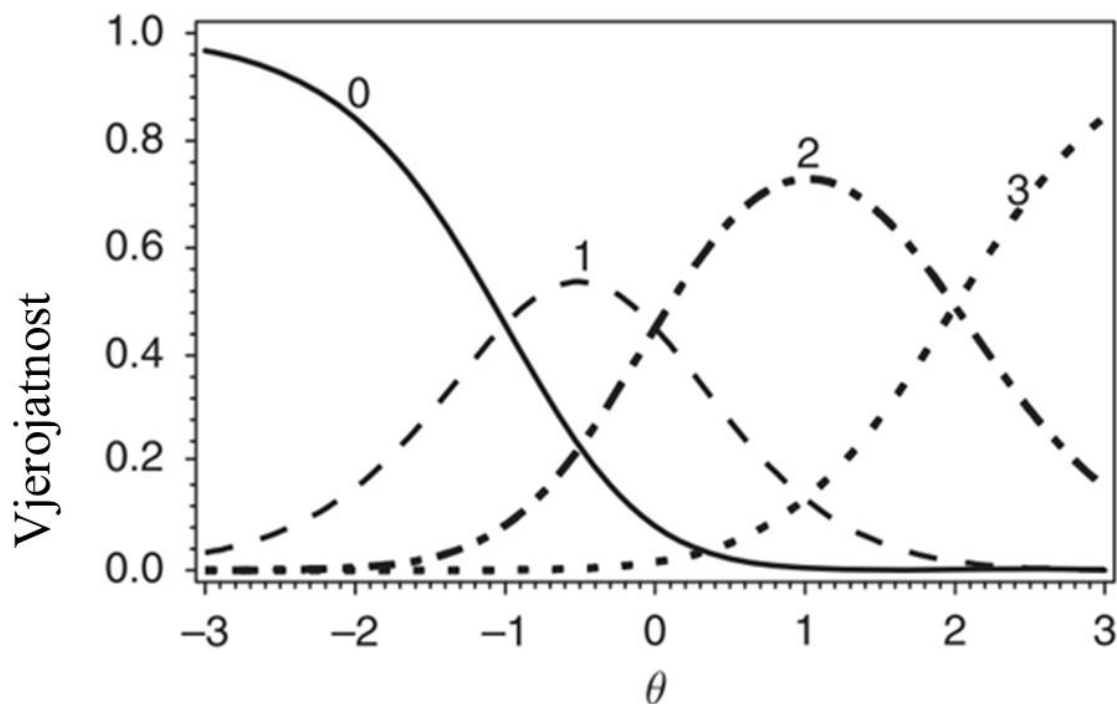
U modelu stupnjevitih odgovora b -parametri se obično nazivaju granicama ili limenima jer oni upućuju na graničnu vrijednost pri kojoj ispitanici imaju jednaku vjerojatnost izabrati kategoriju nižu od k ili kategoriju k i više. Broj graničnih vrijednosti je za 1 manji od broja opcija/kategorija čestice. Uz to, granične vrijednosti nisu fiksirane da budu međusobno jednako udaljene, tj. da postoje jednaki intervali među graničnim vrijednostima unutar jedne čestice. Na ovaj način model stupnjevitih odgovora uvažava činjenicu da opaženi odgovori nisu nužno na intervalnoj razini.

U posebnoj verziji modela stupnjevitih odgovora koji se naziva **modificiranim modelom stupnjevitih odgovora** (engl. *modified GR model*) (Embretson i Reise, 2000) ili **modelom skala ocjenjivanja** (Muraki, 1990) razmaci između kategorija se fiksiraju da budu jednaki između čestica (sve čestice imaju jednake razmake između kategorija, ali oni ne moraju biti jednaki unutar čestice, tj. za pojedinu česticu).

Središnja vrijednost svih graničnih vrijednosti čestice određuje težinu čestice odnosno parametar lokacije (b_i), a razmaci između graničnih vrijednosti kategorija su ekvidistantni, te postoji $k-1$ c -parametara (u ovom slučaju su to parametri kategorije, a ne pseudo-pogađanja). Za razliku od modela stupnjevitih odgovora koji ima parametar težine za svaku kategoriju svake čestice, u modificiranom modelu stupnjevitih odgovora ili modelu skala ocjenjivanja postoji samo po jedan parametar težine za svaku česticu. Iz tog razloga modificirani model stupnjevitih odgovora može biti prikladan za upotrebu kod mjernih instrumenata koji koriste čestice istog formata.

Generalizirani model djelomičnih odgovora (Muraki, 1992) osmišljen je za zadatke u kojima ispitanik može dobiti dio bodova ako djelomično točno odgovori na zadatak, no, kao i ostali modeli, može biti prikladan i za neke druge ordinalne čestice. U generaliziranom modelu djelomičnih odgovora vjerojatnost dobivanja određenog broja bodova ili odabira određene kategorije modelira se direktno, a ne putem operativnih krivulja.

Slika 12. Prikaz politomne čestice modelirane pomoću generaliziranog modela djelomičnih odgovora



Napomena: Prilagođeno prema deMars, 2010.

Krivulje prikazane na Slici 12. izgledaju vrlo slično onima na Slici 11. a nazivaju se operativnim karakterističnim krivuljama ili krivuljama kategorije odgovora, ovisno o autoru.

Matematička funkcija generaliziranog modela djelomičnih odgovora je nešto kompliciranija nego kod modela stupnjevitih odgovora:

$$P_{ik}(\theta) = \frac{e^{\sum_{x=0}^k 1.7a_i(\theta-b_{ix})}}{\sum_{j=0}^{m_i} e^{\sum_{x=0}^j 1.7a_i(\theta-b_{ix})}} \quad (11)$$

Kao i kod modela stupnjevitih odgovora i u ovom modelu postoji b -parametar za svaku kategoriju osim prve, a broj kategorija za česticu i označava se s m_i . U brojniku se nalazi suma već poznatog termina za sve kategorije do kategorije k , a u nazivniku se isti termini sumiraju za sve kategorije.

Parametri kategorija b_{ix} se u ovom modelu često nazivaju sekvencijalnim težinama (engl. *step difficulties*), te određuju točku u kojoj se krivulje za dvije susjedne kategorije sijeku, odnosno razinu latentne varijable u kojoj je odabir tih dviju susjednih kategorija jednako vjerojatan. Za razliku od modela stupnjevitih odgovora kod kojeg limeni kategorije predstavljaju lokaciju na kojoj vjerojatnost odabira te ili više kategorije iznosi 50%, u generaliziranom modelu djelomičnih odgovora, sekvencijalna težina kategorije je lokacija na

kojoj su vjerojatnosti odabira dviju susjednih kategorija jednake, odnosno nakon koje vjerojatnost sljedeće kategorije postaje veća. Iz tog razloga je drugi učestali naziv za ovaj model, **model logita susjednih kategorija**.

Značenje a -parametra je isto u oba ova modela. Neki psihometričari izbjegavaju a -parametar kod politomnih čestica nazivati diskriminativnošću, nego ga zovu samo parametrom nagiba. To je iz razloga što diskriminativnost politomnih čestica ne ovisi samo o a -parametru nego i o relativnim lokacijama b -parametara.

U generaliziranom modelu djelomičnih odgovora, za razliku od modela stupnjevitih odgovora, parametri kategorija ne moraju nužno biti strogo ordinalni, što znači da niže kategorije ne moraju imati niže vrijednosti parametara od viših kategorija.

Važno je napomenuti i to da se parametri dobiveni temeljem različitih modela ne bi trebali međusobno uspoređivati.

Generalizirani model djelomičnih odgovora ima termin „generalizirani“ u imenu jer je u biti ekstenzija jednog drugog modela, **modela djelomičnih odgovora** (Masters, 1982). Model djelomičnih odgovora je pak ekstenzija već opisanog Raschova modela za upotrebu na politomnim česticama. Kod modela djelomičnih odgovora postoji samo jedan a -parametar za sve čestice mjernog instrumenta (kao kod 1-PL modela), a često je i on fiksiran na vrijednost 1, zbog čega efektivno ispada iz modela (kao kod Raschova modela). U generaliziranom pak modelu djelomičnih odgovora, dopušta se da čestice imaju različite a -parametre.

Andrichov **model skala ocjenjivanja** (engl. *rating scale model*) (Andrich, 1978) za razliku od Murakijeva modela skala ocjenjivanja (Muraki, 1990) s kojim dijeli ime, a koji je modifikacija modela stupnjevitih odgovora, pripada skupini Raschovih modela i modifikacija je modela djelomičnih odgovora.

On je restriktivna verzija modela djelomičnih odgovora u kojoj su udaljenosti između susjednih sekvencijalnih težina jednake kod svih čestica (kao i kod modificiranog modela stupnjevitih odgovora). I ovdje svaka čestica ima jedan parametar lokacije koji je određen kao središnja vrijednost svih sekvencijalnih težina, a razmaci između sekvencijalnih težina su ekvidistantni, čime se smanjuje ukupni broj parametara u modelu. Kao i kod modela djelomičnih odgovora i ovdje su a -parametri jednaki za sve čestice. Primjena ove modifikacije modela djelomičnih odgovora preporuča se kad sve čestice mjernog instrumenta imaju isti format odgovora.

Bockov **model nominalnih odgovora** (engl. *nominal response model, NRM*) (Bock, 1972) je model koji se može koristiti za opisivanje odgovora na čestice koje nemaju ordinalnu dimenziju. Osnovna motivacija za razvoj ovog modela bila je u želji da se grafički prikaže vjerojatnost odgovora na distraktore u testovima s višestrukim izborom.

Nominalni model se danas koristi u tri svrhe: (1) kao metoda analize i ocjenjivanja čestica koje izazivaju isključivo nominalne odgovore, (2) kako bi pružio empirijsku provjeru da čestice za koje se očekuje da će rezultirati ordinalnim odgovorima zaista to i čine (Samejima,

1988, 1996), i (3) kao model za odgovore na testletima, tj. skupovima čestica koji se zajednički boduju (Wainer i Kiely, 1987).

Gore navedeni modeli nisu jedini koji se upotrebljavaju za modeliranje politomnih čestica (van der Linden i Hambleton, 1997), ali su svakako najpoznatiji i najčešće korišteni. Od ostalih modela vrijedi spomenuti model višestrukog izbora (engl. *multiple choice model*) (Thissen i Steinberg, 1984), model sukcesivnih intervala (engl. *successive intervals model*) (Rost, 1988) – koji kombinira karakteristike PCM-a i RSM-a, te Wilsonov (Wilson, 1992) model ordinalnog segmentiranja (engl. *ordered partitioned model*) koji se poput NRM-a, može primijeniti na čestice kod kojih kategorije odgovora nisu potpuno ordinalne.

S obzirom na tako velik broj postojećih modela, nameće se pitanje kako se odlučiti koji model primijeniti. Naravno, glavni kriterij pri tome bit će vrsta podataka s kojom istraživač radi. Ukoliko kategorije odgovora nisu ordinalne, u tom slučaju nominalni model odgovora bio bi najispravniji izbor. Ako se pretpostavlja da se čestice međusobno razlikuju po diskriminativnosti, što se može ustanoviti preliminarnom analizom čestica ili faktorskom analizom, tada su GRM ili jedan od Murakijevih (1990, 1992) modela u svakom slučaju prikladniji od bilo kojeg Raschevog modela, poput RSM-a ili PCM-a. Osim toga, ako mjerni instrumenti ne koriste istu ljestvicu bodovanja za sve čestice, tada također RSM nije primjenjiv (osim ukoliko se blokovi čestica s istom skalom ne analiziraju pojedinačno).

Za usporedbu prikladnosti modela za određeni skup podataka mogu se koristiti i pokazatelji slaganja modela. No, u teoriji odgovora na zadatak, za razliku od strukturalnog modeliranja, to područje je još uvijek u razvoju. Sve dok se još ne razvije čarobni statistički štapić koji bi odredio koji je model najispravniji, tu odluku će voditi i praktična razmatranja. Prvo među njima je broj parametara koji je potrebno procijeniti, s obzirom da je za modele s manje parametara potreban manji uzorak ispitanika. Drugo, lakoća bodovanja može biti od značaja. U Raschovim modelima, poput PCM-a, bruto rezultat na skali je dovoljan za procjenu razine sposobnosti ispitanika, uz pretpostavku da su svi ispitanici odgovorili na iste čestice. Konačno, i filozofija istraživača može odrediti upotrebu modela. Neki istraživači smatraju da nije prikladno zahtijevati da čestice imaju jednaku diskriminativnost, pa će koristiti GRM umjesto jednog od Raschovih modela. Drugi istraživači mogu preferirati da sve čestice u skali imaju istu diskriminativnost, kao u modelu RSM-a. Također, i dostupnost računalnih programa može igrati ulogu (Baker, 2001).

1.1.4. Višedimenzionalna teorija odgovora na zadatak (MIRT)

Pretpostavka jednodimenzionalnosti, koja je temelj jednodimenzionalne teorije odgovora na zadatke, pretjerano pojednostavljuje složenost ljudskih osobina koje se mjere. Uz to, istraživanja su pokazala da ova pretpostavka u praksi često nije ispunjena (Ackerman, 1994; Reckase, 1985) te da se broj latentnih dimenzija u instrumentima obično podcjenjuje (Reckase i Hirsh, 1991).

Korištenje jednodimenzionalnih modela za modeliranje čestica koje nisu strogo jednodimenzionalne rezultira pristranim procjenama parametara i lažno niskim standardnim pogreškama procjena sposobnosti (Ansley i Forsyth, 1985; Chang, 1992; de la Tore i Patz, 2005; Drasgow i Parsons, 1983; Reckase, 1979; Way i sur., 1988). Kada je za točan odgovor na čestici potrebno posjedovanje više od jedne sposobnosti, korištenje pristupa bodovanja koji se temelji na pretpostavci o jednodimenzionalnosti povećava pogrešku mjerenja, kao i vjerojatnost donošenja netočnih zaključaka o sposobnosti ispitanika (Walker i Beretvas, 2003). Stoga se u tim situacijama preporuča korištenje višedimenzionalnih modela koji omogućavaju bolju reprezentaciju stvarnosti i preciznije procjene.

Jedan od začetnika višedimenzionalne teorije odgovora na zadatak, Reckase (1997) je rekao da se na nju može gledati ili kao na ekstenziju jednodimenzionalne teorije odgovora na zadatak za upotrebu s višedimenzionalnim podacima ili kao poseban slučaj konfirmatorne faktorske analize (engl. *confirmatory factor analysis*, CFA). Nadalje, prema Reckaseu (2009) je realističnije odnos između vjerojatnosti odgovora i razine latentne osobine koju pojedinac posjeduje modelirati vektorom u višedimenzionalnom prostoru. Lokacija ispitanika j na theta kontinuumu tako je predstavljena vektorom $\theta_j = [\theta_{j1}, \theta_{j2}, \dots, \theta_{jM}]'$, gdje M predstavlja broj takvih kontinuumata odnosno dimenzija latentne osobine. Elementi ovog vektora predstavljaju koordinate tog ispitanika u višedimenzionalnom prostoru. Ovakvo modeliranje odgovora ispitanika u višedimenzionalnom prostoru omogućava donošenje zasebnih zaključaka o svakoj od korištenih vještina odnosno osobina (Walker i Beretvas, 2000).

Višedimenzionalni modeli teorije odgovora na zadatak mogu se koristiti i u eksplorativne i u konfirmativne svrhe. Eksplorativni MIRT modeli mogu se koristiti za utvrđivanje koje su latentne dimenzije u podlozi mjernog instrumenta, te za provjeru pretpostavke jednodimenzionalnosti. Kada je broj dimenzija i čestice koje definiraju pojedinu dimenziju unaprijed poznat, konfirmativni MIRT modeli se mogu koristiti za procjenu parametara čestica i parametara osoba za pojedinačne dimenzije.

I višedimenzionalni IRT modeli imaju određene pretpostavke. To su **monotonost** i **lokalna nezavisnost**. Monotonost znači da se vjerojatnost točnog odgovora ispitanika povećava kako se povećava razina latentnih sposobnosti. Lokalna nezavisnost je već definirana ranije kada je bila riječ o pretpostavkama jednodimenzionalne teorije odgovora na zadatak. Za razliku od jednodimenzionalne teorije odgovora na zadatak kod višedimenzionalne teorije odgovora na zadatak vjerojatnost točnog odgovora grafički se prikazuje pomoću karakteristične (hiper)površine čestice (engl. *item characteristic (hyper) surface*, ICS; *item response (hyper) surface*, IRS).

U odnosu na jednodimenzionalnu IRT, novi koncepti koji se javljaju u MIRT-u su i višedimenzionalna težina čestice (engl. *multidimensional item difficulty*, MDIFF) i višedimenzionalna diskriminativnost čestice (engl. *multidimensional item discrimination*, MDISC).

MDIFF predstavlja mjeru težine čestice u svakoj od latentnih dimenzija. U jednodimenzionalnoj IRT, težina čestice označava koliko je ta čestica teška u mjerenoj osobini.

Međutim, u MIRT-u, svaka čestica može imati različite težine u svakoj latentnoj dimenziji. MDIFF se odnosi na udaljenost od ishodišta do točke na vektoru čestice gdje je vektor najstrmiji u određenoj dimenziji. Ova udaljenost može se interpretirati kao težina čestice u toj dimenziji. Na primjer, što je veća udaljenost, to je čestica teža u toj dimenziji.

MDISC se odnosi na mjeru diskriminativnosti čestice u svakoj dimenziji. Grafički se može predstaviti dužinom vektora čestice, pri čemu duži vektor označava veću diskriminativnost čestice. To znači da čestica koja ima duži vektor ima veću sposobnost razlikovanja između različitih razina latentne osobine u toj dimenziji. Drugim riječima, takva čestica može bolje razlikovati pojedince s različitim sposobnostima u toj dimenziji.

Kompenzacijski i nekompenzacijski MIRT modeli

Dvije su velike kategorije MIRT modela. Prva vrsta se naziva kompenzacijskim MIRT modelima i temelji se na ponderiranoj linearnoj kombinaciji sposobnosti, te pretpostavlja da se manja razina jedne osobine može kompenzirati većom količinom druge osobine. Vjerojatnost točnog odgovora na čestici je u tom slučaju ponderirana **suma** pojedinačnih vjerojatnosti točnog odgovora ovisnih samo o jednoj od potrebnih sposobnosti (Ackerman i sur., 2003). Na taj način osobe s različitom količinom potrebnih osobina mogu postići jednak rezultat na pojedinoj čestici.

U ovoj vrsti modela, svaka osoba ima vektor vrijednosti θ , a svaka čestica ima vektor vrijednosti parametara diskriminativnosti a , te samo jednu vrijednost d , koja zamjenjuje parametre težine. Smjer vektora a govori nam o količini pojedinačnih sposobnosti/dimenzija koje se najbolje mjere česticom, a njegova duljina ukazuje na veličinu višedimenzionalne diskriminativnosti čestice (Ackerman, 1994). Osim toga, a -parametri pokazuju orijentaciju kontura jednake vjerojatnosti (engl. *equiprobable contours*), kao i stopu kojom se vjerojatnost točnih odgovora mijenja od jedne do druge točke u θ -prostoru.

Parametar d_i u modelu je povezan s težinom čestice. Međutim, vrijednost tog parametra se ne može tumačiti na isti način kao b -parametar u modelima jednodimenzionalne teorije odgovora na zadatak gdje on daje jedinstvenu procjenu težine čestice. Naime, u kompenzacijskim modelima, težina čestice ne ovisi samo o d_i -parametru nego i o parametrima diskriminativnosti (a -vektor). Da bi se dobila relativna težina čestice za određenu dimenziju povezanu s odgovarajućim a -parametrom, d_i -parametar se dijeli s pojedinačnim elementima a -vektora. Drugim riječima, težina čestice u kompenzacijskim modelima ovisi o više faktora, uključujući d_i i a -parametre, te se izračunava kao omjer dijela i -parametra povezanog s određenom dimenzijom i pripadajućeg elementa a vektora. Stoga, vrijednost parametra i ne pruža izravnu jedinstvenu procjenu težine čestice kao što to čini b -parametar u jednodimenzionalnim modelima, već se koristi za izračunavanje relativne težine čestice u odnosu na svaku pojedinu dimenziju.

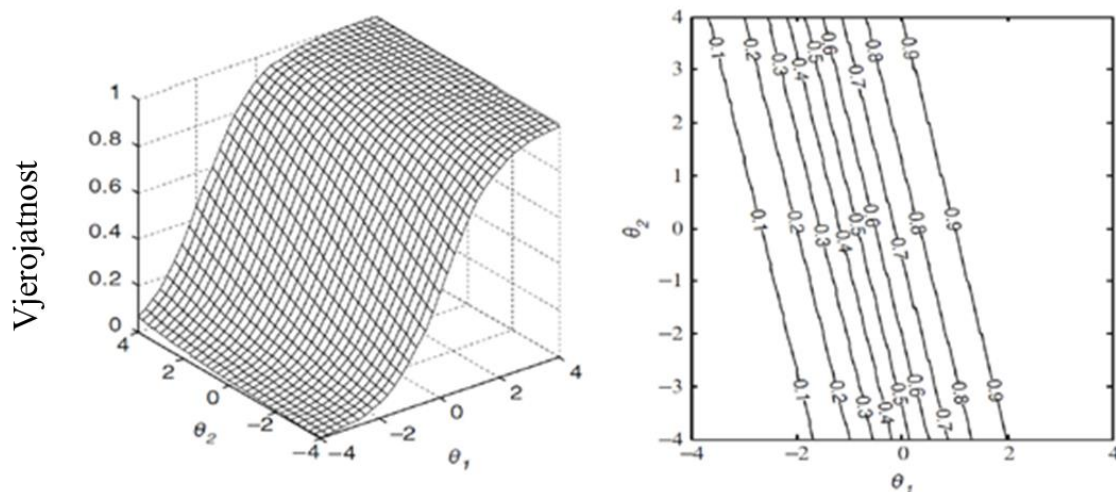
Slika 13. prikazuje kompenzacijski model za dvodimenzionalnu česticu na dva načina. Na lijevoj strani je prikazan plošni grafikon, koji je trodimenzionalan i prikazuje vjerojatnost

ispravnog odgovora na česticu kao funkciju karakteristične površine čestice. Na desnoj strani je prikazan konturni grafikon, koji prikazuje vjerojatnost ispravnog odgovora na česticu pomoću odabranih segmenata kontura karakteristične površine čestice. Pritom konture jednake vjerojatnosti tvore ravne linije.

Oba grafikona jasno ilustriraju da visoka razina θ na jednoj dimenziji može nadoknaditi nisku razinu θ na drugoj dimenziji. Na primjer, na Slici 13, osoba s $\theta = -2.0$ na drugoj dimenziji čestice može i dalje imati vjerojatnost točnog odgovora od 0.9 ako istovremeno ima vrijednost θ na prvoj dimenziji blizu 3.0.

Druga vrsta modela su nekompensacijski (neki ih nazivaju i djelomično kompenzacijskim modelima) MIRT modeli koji pretpostavljaju da mora postojati određena minimalna razina svih latentnih osobina potrebnih za rješavanje zadatka kako bi uopće postojala vjerojatnost za davanje točnog odgovora. U tim modelima vjerojatnost točnog odgovora je ponderirani **produkt** pojedinačnih vjerojatnosti točnog odgovora ovisnih samo o jednoj od potrebnih sposobnosti (Reckase i McKinley, 1982; Simpson, 1978). S obzirom da se vjerojatnost točnog odgovora temeljena na kompozitu sposobnosti formira kao umnožak, ona ne može biti viša od najviše vjerojatnosti koja je element tog umnoška.

Slika 13. Plošni grafikon (lijevo) i konturni grafikon (desno) za vjerojatnost točnog odgovora na dvodimenzionalnoj kompenzacijskoj čestici s parametrima $a_1=1.5$, $a_2=.5$, $d=.7$



Napomena: Prilagođeno prema Reckaseu, 2009.

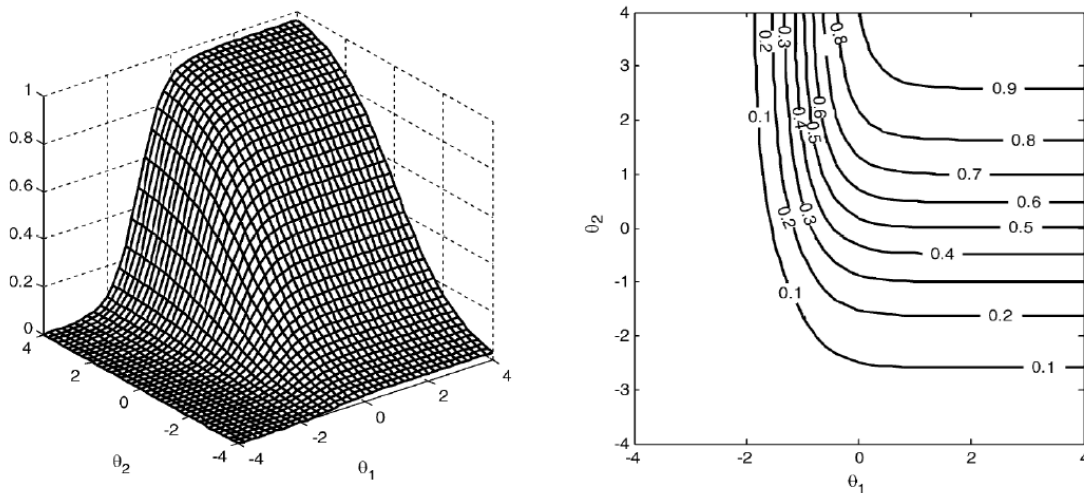
Za razliku od kompenzacijskih modela, u nekompensacijskim modelima svaka uključena sposobnost odnosno dimenzija latentne osobine ima osim vlastitog parametra težine i vlastiti parametar diskriminativnosti. Temeljem tih vrijednosti izračunavaju se pojedinačne vjerojatnosti za svaku dimenziju, a zatim se računa njihov produkt, koji predstavlja ukupnu vjerojatnost točnog odgovaranja na česticu (Reckase i McKinley, 1982).

Na Slici 14. su prikazani plošni (lijevo) i konturni (desno) grafički prikaz dvodimenzionalne nekompensacijske čestice. Osoba koja posjeduje nisku razinu jedne od sposobnosti ima nisku vjerojatnost točnog odgovora na česticu bez obzira koliko visoku razinu druge sposobnosti posjeduje. Vjerojatnost točnog odgovora u slučajevima u kojima je razina bilo koje sposobnosti niska je blizu 0. Samo kad su razine obje sposobnosti visoke, postoji visoka vjerojatnost točnog odgovora. Na primjer, osoba koja vrlo slabo razumije jezik na kojem je postavljen matematički zadatak ima nisku vjerojatnost točnog rješavanja, neovisno o tome što su njene matematičke sposobnosti visoke.

Kompensacijski modeli su prikladniji za čestice koje koriste tzv. disjunktivni komponentni proces (Maris, 1999). Recimo, zadaci koje se mogu riješiti na više načina vjerojatno sadrže u sebi kompensacijsku višedimenzionalnost.

S druge strane, nekompensacijski su modeli prikladniji za zadatke koji koriste tzv. konjuktivni komponentni proces. Primjerice, problemni zadaci na testu matematike vjerojatno zahtijevaju određenu razinu sposobnosti čitanja kao preduvjet za korištenje matematičke sposobnosti koja je potrebna za rješavanje zadatka. Kod takvog zadatka je malo vjerojatno da se nedostatak jedne od sposobnosti može nadomjestiti drugom (Bolt i Lall, 2003).

Slika 14. Plošni grafikon (lijevo) i konturni grafikon (desno) za vjerojatnost točnog odgovora na dvodimenzionalnoj nekompensacijskoj čestici s parametrima $a_1=1.5$, $a_2=.5$, $b_1=-1.0$, $b_2=0.0$



Napomena: Prilagođeno prema Reckaseu, 2009.

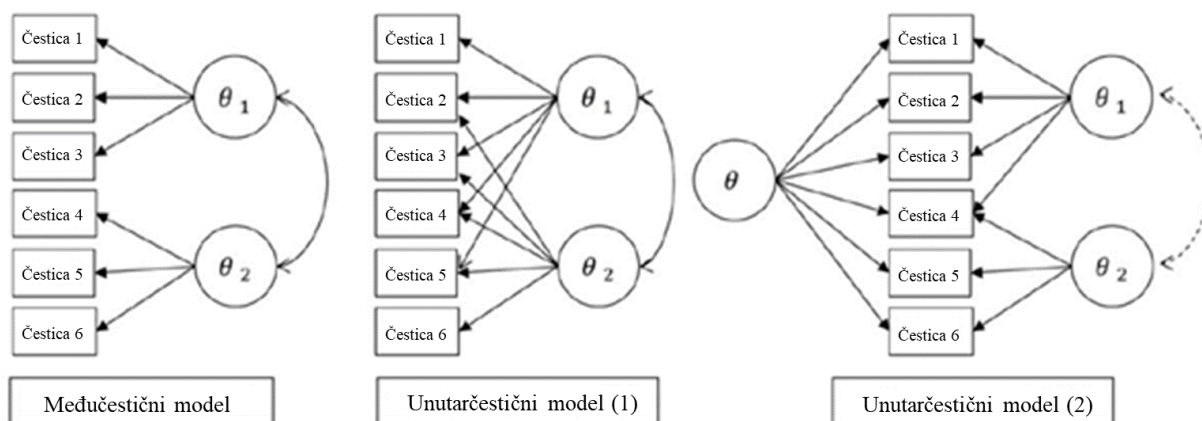
Dosad su se, zbog svoje jednostavnosti i dostupnih programa, u istraživanjima češće koristili kompensacijski modeli, s obzirom da multiplikativna priroda nekompensacijskih modela predstavlja ozbiljne izazove u izračunu potrebnih parametara.

Unutar ove dvije velike kategorije modela, postoji niz MIRT modela za dihotomne i politomne čestice (Bock i Aitkin, 1981; Bock i Liberman, 1970; McDonald, 1982; Sympson, 1978). Oni su obično ekstenzije jednodimenzionalnih modela koji su već prethodno opisani.

Unutarčestična i međučestična višedimenzionalnost

S obzirom na strukturu mjernog instrumenta koji modeliraju, MIRT modeli se mogu podijeliti u dvije skupine: na međučestične (eng. *between-item models*) i unutarčestične modele (engl. *within-item models*) (Adams i sur., 1997; Bulut, 2013; Wang i sur., 2004). Višedimenzionalni međučestični modeli namijenjeni su mjernim instrumentima u kojima postoje međusobno isključive subskale koje mjere različite latentne varijable. Ovaj tip strukture mjernih instrumenata poznat je i kao "jednostavna struktura" u kojoj je svaka čestica povezana samo s jednom latentnom dimenzijom. Nasuprot međučestičnim modelima, višedimenzionalni unutarčestični modeli prikladni su za mjerne instrumente u kojima čestice mogu biti indikatori više različitih latentnih dimenzija. Ovaj tip strukture testa poznat je kao "složena struktura" (engl. *nonsimple structure* ili *complex structure*). Primjer unutarčestičnih višedimenzionalnih MIRT modela su konfirmatorni modeli sa složenom strukturom ili bi-faktorski modeli. Slika 15. prikazuje međučestične i unutarčestične modele.

Slika 15. Primjer međučestičnih i unutarčestičnih modela



Napomena: θ_1 & θ_2 = pojedinačne sposobnosti/dimenzije, θ = ukupna/opća sposobnost. Prilagođeno prema Bulutu, 2013.

U višedimenzionalnim međučestičnim modelima, mjerni instrument sadrži nekoliko subskala koje mjere povezane, ali različite latentne dimenzije. Ovaj tip MIRT modela u literaturi je još poznat i kao više-jednodimenzionalni model, jer se svaka subskala tretira kao jednodimenzionalna, dok cjelokupna struktura testa postaje višedimenzionalna (Sheng i Wikle, 2007).

Najčešći način korištenja IRT-a za analizu više-jednodimenzionalnih struktura je procjena parametara čestica i sposobnosti ispitanika za svaku subskalu zasebno putem jednodimenzionalnih IRT modela ili tretiranje cijelog mjernog instrumenta kao jednodimenzionalnog. Oba ova pristupa imaju određene slabosti koje ih čine manje poželjnima u odnosu na provođenje višedimenzionalne kalibracije (Adams i sur., 1997; Zhang, 2012).

Kada pak mjerni instrument mjeri više latentnih dimenzija a neke od čestica zahtijevaju kombinaciju više sposobnosti da bi se na njih ispravno odgovorilo, tada test pokazuje unutarčestičnu višedimenzionalnost. Modeli koji uključuju unutarčestičnu višedimenzionalnost prikladni su za modeliranje interakcija između različitih sposobnosti i zahtjeva čestice. U ovom slučaju se vjerojatnost rješavanja zadatka može modelirati kao funkcija kombinacije različitih dimenzija sposobnosti.

1.2. Diferencijalno funkcioniranje čestica (DIF)

1.2.1. Diferencijalno funkcioniranje čestica, valjanost i pravednost

Diferencijalno funkcioniranje čestica (engl. *differential item functioning, DIF*) je pojava u kojoj ispitanici koji imaju istu razinu latentne osobine, ali pripadaju različitim skupinama, na istoj čestici instrumenta postižu različite rezultate (Holland i Wainer, 1993). Ovaj fenomen ukazuje na potencijalnu pristranost čestica, koja može utjecati na valjanost i pravednost psihološkog i obrazovnog testiranja.

Naime, postojanje pristranosti čestica u instrumentima dovodi u pitanje ispravnost zaključaka koji su doneseni temeljem tih instrumenata, odnosno njihovu valjanost. Osim toga, upitna je i pravednost, jer pristrane čestice mogu dovesti do diskriminacije određenih (demografskih) skupina.

Pravednost testiranja (engl. *test fairness*) je jedan od aspekata metrijske karakteristike valjanosti, a aktualni ga *Standardi za obrazovno i psihologijsko testiranje* (AERA, APA i NCME, 2014) definiraju kao prilagodbu instrumenta i uvjeta testiranja individualnim osobitostima pojedinca kako bi dobiveni rezultati bili temelj za valjane zaključke. Pravedan mjerni instrument mjeri konstrukt koji ima jednaku mentalnu reprezentaciju i psihološko značenje za sve pojedince iz testirane populacije.

Ponekad se događa da na neekvivalentnost bodova ispitanika utječu neke osobine mjernog instrumenta nepovezane sa samim konstruktom (ili načinom primjene instrumenta). U tim slučajevima kažemo da čestica diferencijalno funkcionira.

Kako bi se osigurala valjanost zaključivanja, važno je uzeti u obzir sve individualne karakteristike različitih pojedinaca iz ciljne populacije (kao što su rasa, spol, nacionalnost, dob, socioekonomski status, jezično ili kulturno porijeklo) i tijekom razvoja instrumenta, ali i tijekom njegove primjene, bodovanja i interpretacije. Prilagodba uključuje pažljivu razradu

čestica, osiguranje adekvatnih uputa i uvjeta testiranja, te promišljenu interpretaciju rezultata. Pritom treba voditi računa da se samom prilagodbom ne naruši valjanost mjernog instrumenta. Naime, ponekad prilagođavanje standardnih procedura kako bi se omogućila pristupačnost određenim ispitanicima može utjecati na sam mjerni konstrukt i tako dovesti do smanjenja valjanosti zaključaka te stvaranja nepravedne prednosti za pojedince kojima je materijal prilagođen (poznato kao „pozitivna diskriminacija“). S druge strane, u kontekstu dijagnostičkog testiranja, prilagodba mjernog instrumenta radi povećanja pristupačnosti može povećati točnost dijagnoze i valjanost zaključivanja.

No, bez obzira na svrhu testiranja, primarni zadatak pravednosti jest omogućiti svim ispitanicima priliku da u potpunosti iskažu svoje sposobnosti i točno predstave razinu konstrukta koji se mjeri. Stoga prilagodba mjernih instrumenata treba biti pažljivo promišljena i usklađena sa specifičnim potrebama i kontekstom testiranja kako bi se postigla ravnoteža između pravednosti i valjanosti zaključivanja.

1.2.2. Mjerna invarijantnost i pristranost

Važan aspekt pravednosti je ekvivalentnost mjerenog konstrukta odnosno mjerna invarijantnost (Mellenbergh, 1989). S druge strane, mjerna pristranost (engl. *measurement bias*) jedna je od glavnih prijetnji pravednosti.

Ekvivalentnost je temeljni preduvjet za grupne usporedbe rezultata mjerenja različitih skupina ispitanika. Uobičajeno se ističu tri razine ekvivalentnosti: konstruktna (strukturalna), metrička i skalarna (potpuna) ekvivalentnost (Vandenberg i Lance, 2000).

Konstruktna (strukturalna, funkcionalna) ekvivalentnost podrazumijeva jednako psihološko značenje konstrukta u različitim skupinama. To osim iste interpretacije konstrukta predstavlja i istu važnost konstrukta za sve ispitanike, bez obzira na grupnu pripadnost, te osigurava da mjerni instrument zaista mjeri isti konstrukt u različitim skupinama. Ova razina mjerne ekvivalentnosti implicira postojanje univerzalnog psihološkog funkcioniranja izraženog putem zajedničkih latentnih dimenzija ili faktora. Drugim riječima, konstrukt je isti i funkcionira na isti način kod svih ispitanika bez obzira na njihovu grupnu pripadnost.

Metrička ekvivalentnost, također poznata kao ekvivalentnost mjernih jedinica, odnosi se na prisutnost istih mjernih skala s različitim ishodištima. Drugim riječima, mjerni instrumenti koriste iste jedinice mjere, ali mogu imati različito početno mjesto na skali (konstantni pomak). Instrumenti koji postižu ovu razinu ekvivalentnosti omogućavaju uspoređivanje razlika unutar jedne grupe s razlikama unutar druge grupe, ali direktne usporedbe među skupinama nisu moguće.

Najviša razina ekvivalentnosti je skalarna ekvivalentnost, koja omogućava direktne usporedbe između skupina. Skalarna ekvivalentnost postiže se kada mjerni instrumenti imaju iste mjerne jedinice i isto ishodište. To znači da se rezultati mogu uspoređivati između skupina bez ikakvih preinaka ili prilagodbi.

Sve ove razine ekvivalentnosti važne su za osiguravanje pravednosti testiranja, jer omogućuju objektivno i usporedivo mjerenje konstrukta među različitim skupinama ispitanika.

Pristranost predstavlja izvor sistematskih pogrešaka koje mogu nastati u istraživanjima na razini konstrukta, metode ili čestice. Ekvivalentnost i pristranost su međusobno povezani, te odsustvo ili eliminacija pristranosti doprinosi mjernoj ekvivalentnosti instrumenata (van de Vijver i Leung, 1997). Također se uobičajeno razlikuju tri vrste pristranosti: konstruktna pristranost, pristranost metode i pristranost čestica.

Konstruktna pristranost odnosi se na odstupanja u razumijevanju određenog konstrukta među različitim skupinama. Kod ove vrste pristranosti psihološko značenje konstrukta nije jednako u svim skupinama koje se uspoređuju.

Pristranost metode može biti povezana s uzorkovanjem, odabirom instrumenata ili načinom njegove primjene. Pristranost uzorkovanja nastaje kada uzorak ispitanika nije reprezentativan za populaciju koju istraživanje namjerava istražiti. To može dovesti do neopravdanih generalizacija i zaključaka koji ne odražavaju stvarnu distribuciju karakteristika ili stavova unutar ciljane populacije. Primjerice, ako se online anketa o političkim stavovima distribuira preko platformi koje posjećuju pretežno mladi, rezultati možda neće točno predstavljati političke stavove cijele populacije. Pristranost instrumenta javlja se kada korišteni instrumenti ne mjere ono što im je namjena, ili kada su instrumenti prilagođeni specifičnim karakteristikama određene skupine ispitanika na način koji isključuje ili marginalizira druge skupine. Na primjer, test inteligencije koji naglašava verbalne sposobnosti možda neće biti objektivan prema onima čiji prvi jezik nije jezik na kojem je test napisan, potencijalno umanjujući njegovu sposobnost da nepristrano mjeri inteligenciju ispitanika koji koriste različite jezike. Način primjene instrumenata također može uvesti pristranost. To uključuje uvjete pod kojima se testiranje odvija, kao što su vrijeme, lokacija, atmosfera, kao i način interakcije ispitivača s ispitanicima. Na primjer, test koji se primjenjuje u školskom okruženju gdje su neki učenici pod stresom zbog akademskog pritiska može dati drugačije rezultate nego isti test primijenjen u opuštenijem okruženju. Slično tome, ako ispitivač daje dodatne upute ili poticaje određenim ispitanicima, to može utjecati na njihove rezultate, čime se uvodi pristranost u proces mjerenja.

Pristranost čestica (engl. *item bias*) prvotno se poistovjećivala s diferencijalnim funkcioniranjem čestica. Međutim, s vremenom je došlo do razgraničenja između ta dva pojma. Diferencijalno funkcioniranje čestica je neutralniji termin koji se odnosi na statističku činjenicu da ispitanici iz različitih skupina, koji imaju jednaku razinu latentne osobine, dobivaju različite rezultate na tim česticama. No, kako bi se čestice smatrale pristranima, nije dovoljna samo postojanje te činjenice. Potrebna je i logička analiza uzroka diferencijalnog funkcioniranja čestica te dokaz o postojanju diskriminirajućih posljedica za jednu od tih skupina (Camilli i Shepard, 1994).

Pitanje konstruktne valjanosti mjernih instrumenata i pitanje pristranosti mjerenja je međusobno povezano, a prema Ackermanu im je u osnovi broj konstrukata koji mjerni instrument odnosno čestica mjere (Ackerman, 1992). Ako je konstruktna valjanost mjernog

instrumenta niska, to znači da takav mjerni instrument sadrži čestice koje mjere i druge konstrukte osim onog za koji je namijenjen, što upućuje i na moguću pristranost mjernog instrumenta prema određenoj skupini ispitanika.

Kako bismo bolje razumjeli što točno mjerna pristranost znači, važno je razumjeti razliku između prave i ometajuće (eng. *nuisance*) sposobnosti. Ackerman definira pravu sposobnost kao sposobnost koju želimo izmjeriti mjernim instrumentom, dok je ometajuća sposobnost ona sposobnost koju su ispitanici koristili prilikom odgovaranja na čestice, a koju mjerni instrument nema namjeru izmjeriti. Ako se količina ometajuće sposobnosti razlikuje među grupama koje uspoređujemo, te ako čestice nenamjerno mjere više od jedne sposobnosti (a boduju samo jednu), to može dovesti do pristranosti mjerenja.

Pristranost čestice nužno je razlikovati od drugih srodnih koncepata. Jedan od njih je koncept kojeg nazivamo „učinak čestice“ (engl. *item impact*), a koji predstavlja stvarnu razliku u latentnim osobinama među skupinama. „Učinak čestice“ se odnosi na stvarne razlike u latentnim osobinama među skupinama koje se očituju u različitim vjerojatnostima točnog odgovaranja na određenu česticu (Wainer, 1993). Ako sve čestice testa mjere samo željeni konstrukt, tada se razlike u sposobnostima među grupama ne javljaju zbog pristranosti mjerenja, već odražavaju stvarne razlike u konstrukt. Međutim, ako vjerojatnost točnog odgovaranja na česticu varira među ispitanicima s istom razinom sposobnosti, tada govorimo o diferencijalnom funkcioniranju čestice, a razlike u ukupnom rezultatu testa koje proizlaze iz toga su posljedica ometajućih sposobnosti.

Važno je znati da samo zato što jedna grupa postiže niže rezultate na testu, to ne znači da su čestice instrumenta pristrane. Pristranost mjerenja se odnosi na nepravilnosti u testiranju koje rezultiraju sistematskom pogreškom u procjeni određenih skupina ispitanika. Cilj istraživanja pristranosti je utvrditi jesu li razlike među grupama stvarne (temelje se na faktorima koji nisu povezani s testom) ili su artificijelne (uzrokovane postupcima mjerenja). Na primjer, žene mogu postizati niže rezultate na matematičkim testovima, ali to može biti posljedica činjenice da žene češće odabiru srednje škole i fakultete koji imaju manje sati matematike ili manje matematičkih kolegija. Stoga bi takve razlike u rezultatima testiranja predstavljale stečene razlike u matematičkom znanju, a ne pristranost testa. Zaključak o pristranosti testa mogao bi se donijeti samo ako bi se takve razlike pojavile između muškaraca i žena koji imaju jednaka prethodna obrazovna iskustva (isti smjer srednje škole i fakulteta te iste ocjene).

Osim toga, pristranost čestice potrebno je razlikovati od diferencijalnog funkcioniranja čestice, jer je DIF, kako je već rečeno, samo statistički pokazatelj različitog funkcioniranja čestice u različitim grupama, te kao takav nužan ali ne i dovoljan preduvjet za proglašenje čestice pristranom (Clauser i Mazor, 1998; Zumbo, 1999). Kako bi se čestica proglasila pristranom, bitno je nakon što se utvrdi da određena čestica različito funkcionira u dvije skupine ispitanika razmotriti je li ta razlika u funkcioniranju čestice imala diskriminirajuće posljedice za neku od skupina.

Identificiranje uzroka diferencijalnog funkcioniranja čestica također je važan korak u otkrivanju pristranosti. Potrebno je istražiti razloge zbog kojih se različite skupine ispitanika razlikuju u odgovorima na određenu česticu. To može uključivati analizu jezičnih ili kulturnih razlika, razlika u obrazovanju ili iskustvu, ili druge faktore koji mogu utjecati na različito funkcioniranje čestica.

Također je važno naglasiti da pristranost ne proizlazi iz slučajnih pogreški mjerenja. Nijedan test ne mjeri savršeno pouzdano i bez pogreške, no dok god pogreška mjerenja jednako pogađa različite skupine ispitanika, test neće biti pristran. Pristranost se odnosi na sistematsku pogrešku mjerenja koja utječe samo na određenu skupinu ispitanika, dok slučajne pogreške mjerenja utječu na sve ispitanike jednako. Kada test nije pristran, pogreška mjerenja se ravnomjerno raspodjeljuje među različitim skupinama ispitanika. Pristranost se javlja tek kada postoji sistematska pogreška u mjerenju koja utječe na rezultate za određenu skupinu, stvarajući netočne ili iskrivljene rezultate za tu skupinu (možemo to usporediti sa situacijom u kojoj se vrijeme trčanja jednog skupini mjeri ispravnom štopericom, a drugoj neispravnom). U takvoj situaciji bi rangiranje ispitanika unutar pojedine skupine ispitanika bilo moguće i ispravno, ali usporedbe među skupinama bile bi netočne i pristrane.

Pristranost testova najčešće je povezana s istraživanjem razlika u uspješnosti rasnih, spolnih i etničkih skupina, no isti koncepti i pokazatelji mogu se primijeniti i na skupine koje se razlikuju po socijalnom statusu, dobi, regiji, urbanoj sredini i drugim čimbenicima.

1.2.3. Definicija DIF-a

No što je zapravo DIF? U objašnjenju će pomoći primjer koji su koristili Millsap i Everson (1993), Penfield i Camilli (2006), te Osterlind i Everson (2009).

U njemu se kreće od jednostavne, idealizirane situacije u kojoj na to koji će odgovor (Y) na određenoj čestici testa biti odabran utječe samo ciljna sposobnost odnosno latentna osobina koja se mjeri (θ), te slučajna pogreška. Odgovor na čestici Y je, dakle, indikator latentne osobine θ . Distribuciju uvjetne vjerojatnosti odgovora Y ovisno o razini latentne osobine θ možemo izraziti kao $f(Y|\theta)$.

Recimo da želimo usporediti uvjetne distribucije vjerojatnosti odgovora Y u dvije skupine ispitanika (simbolom G označava se grupirajuća varijabla): fokalnoj ($G=F$) i referentnoj ($G=R$). U ovoj idealnoj situaciji u kojoj je uvjetna vjerojatnost odgovora Y ovisna isključivo o θ -i, a distribucije pogrešaka su identične u obje grupe, vrijedi sljedeće:

$$f(Y|\theta, G = R) = f(Y|\theta, G = F)$$

(12)

Gornjom formulom prikazuje se situacija u kojoj nema DIF-a, odnosno distribucija uvjetne vjerojatnosti odgovora Y ne ovisi o pripadnosti grupi.

Na primjeru jedne dihotomne čestice na kojoj $Y=1$ označava točan, a $Y=0$ netočan odgovor, to je situacija u kojoj je vjerojatnost točnog odgovora jednaka u obje grupe za pojedince s istom količinom latentne osobine θ . Takva čestica ne daje prednost niti jednoj od skupina, pa stoga nije niti pristrana niti pokazuje DIF.

Suprotno tome, u slučaju u kojem su uvjetne vjerojatnosti odgovora Y različite za referentnu i fokalnu skupinu, pojedinci s istom razinom latentne osobine a koji su pripadnici različitih grupa imaju različite distribucije vjerojatnosti odgovora Y . To znači da postoji uvjetna ovisnost funkcioniranja čestice o pripadnosti grupi i nakon što su ispitanici izjednačeni po latentnoj osobini (ponovno uz pretpostavku jednakih distribucija pogrešaka).

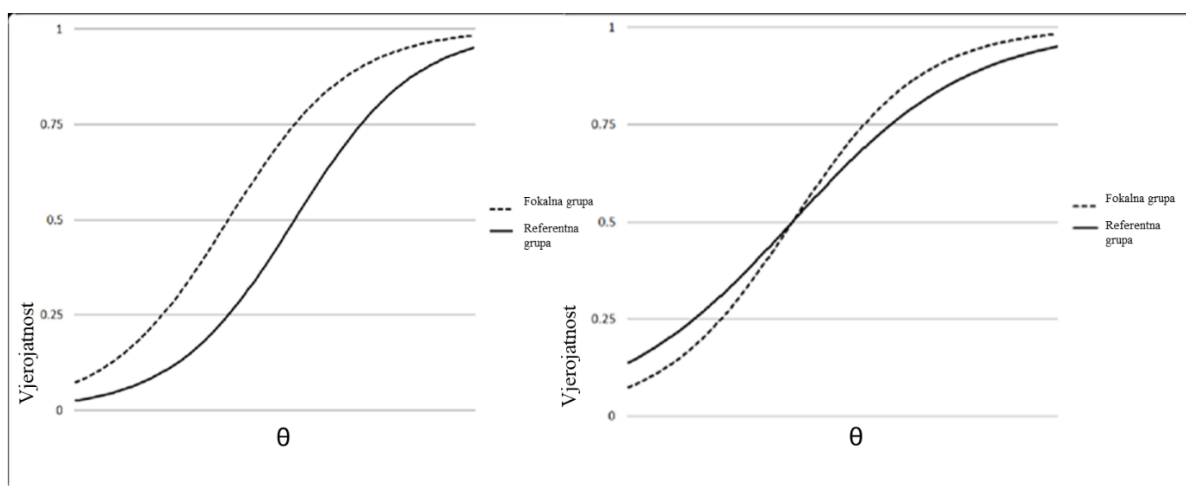
Na primjeru dihotomne čestice, ova uvjetna ovisnost implicira da pripadnici referentne i fokalne skupine s jednakom razinom latentne osobine θ nemaju jednaku vjerojatnost točnog odgovora. Skupina s nižom uvjetnom vjerojatnošću točnog odgovora je u tom slučaju u nepovoljnijem položaju. Tada možemo reći da čestica funkcionira različito za dvije skupine, odnosno da je prisutan DIF. Drugim riječima, pripadnici različitih skupina imaju različite očekivane rezultate na čestici i nakon što se kontroliraju stvarne razlike u sposobnosti (engl. impact) (Drasgow, 1987; Stark i sur., 2004).

1.2.4. Jednoliki i nejednoliki DIF

S obzirom da se DIF definira kao uvjetna ovisnost funkcioniranja čestice o grupnoj pripadnosti, činjenica da priroda te ovisnosti može biti različita na različitim razinama kontinuuma latentne osobine/sposobnosti dovodi do definicije dvije glavne vrste DIF-a. U najjednostavnijem slučaju, uvjetna ovisnost je jednake relativne veličine na svim razinama latentne osobine θ . To znači da čestica pruža konstantnu relativnu prednost istoj skupini, neovisno o razini θ -e. Ovo se naziva *jednolikim DIF-om* (engl. *uniform DIF*), jer je učinak DIF-a jednak na svim razinama θ -kontinuumu.

Druga vrsta DIF-a je *nejednoliki DIF* (engl. *nonuniform DIF*) koji se javlja kada uvjetna ovisnost nije jednaka na različitim razinama latentne osobine, i to ili po veličini ili po smjeru. Ova vrsta DIF-a izaziva prednost koja varira u jačini i/ili smjeru ovisno o razini latentne osobine pojedinaca (Mellenbergh, 1982). Primjerice, referentna skupina može imati malu relativnu prednost nad fokalnom skupinom na nižim razinama θ -e, ali veliku relativnu prednost na visokim razinama θ -e. Kod jednolikog DIF-a ne postoji interakcija između razine sposobnosti i grupne pripadnosti, dok kod nejednolikog postoji.

Slika 16. Prikaz čestice s jednolikim (lijevo) i nejednolikim DIF-om (desno)



Teorija odgovora na zadatak na vrstu DIF-a gleda iz rakursa parametara čestice, pa tako jednoliki DIF definira kao grupne razlike u b parametrima odnosno težini čestice, a nejednoliki kao grupne razlike u a parametru odnosno diskriminativnosti čestice (iako zapravo može biti riječ i o kombinaciji razlike u oba parametra).

Kod jednolikog DIF-a, karakteristične krivulje čestice dvije skupine su paralelne i ne presijecaju se, jer čestica konzistentno favorizira jednu skupinu ispitanika u odnosu na drugu, na svim razinama sposobnosti. To znači da je ta čestica sustavno teža za ispitanike fokalne skupine u odnosu na referentnu skupinu (ili obrnuto), i to neovisno o njihovoj stvarnoj razini sposobnosti. Primjerice, ako dva ispitanika, jedan iz referentne i drugi iz fokalne skupine, imaju istu razinu matematičke sposobnosti, ispitanik iz fokalne skupine će imati manju vjerojatnost davanja točnog odgovora na pristranu česticu, što ukazuje na diferencijalnu težinu čestice. Isto će vrijediti za sve ispitanike iz fokalne skupine u odnosu na sve ispitanike iz referentne skupine, neovisno o razini njihovog znanja matematike.

Nejednoliki DIF, s druge strane, pokazuje interakciju između grupne pripadnosti ispitanika i njihove razine latentne osobine, rezultirajući karakterističnim krivuljama koje se presijecaju. To znači da se diskriminativnost čestice mijenja ovisno o razini sposobnosti ispitanika, što može rezultirati situacijom gdje čestica favorizira jednu skupinu na niskim razinama sposobnosti, ali drugu skupinu na visokim razinama sposobnosti. Nejednoliki DIF ukazuje na to da čestica ne samo da različito mjeri latentnu osobinu između skupina, već i da se njezina diskriminativnost mijenja ovisno o razini sposobnosti ispitanika, što ju čini manje pouzdanom u procjeni stvarne sposobnosti ispitanika unutar fokalne skupine u usporedbi s referentnom skupinom. Dakle, općenito možemo reći da su čestice s nejednolikim DIF-om manje diskriminativne u fokalnoj skupini, dok su čestice s jednolikim DIF-om teže za fokalnu skupinu (Wanichtanom, 2001; Woods, 2008).

1.3. Metode detekcije DIF-a

Otkrivanje DIF-a je jedan od ključnih koraka u validaciji instrumenata i utvrđivanju njihove pravednosti u procjenjivanju osobina različitih skupina ispitanika. Razvoj statističkih metoda za identificiranje potencijalno pristranih čestica započeo je 70.-ih godina prošlog stoljeća, no tek je potkraj 80.-ih diferencijalno funkcioniranje čestica razdvojeno od analize pristranosti čestica (Penfield i Camilli, 2006). Sam termin diferencijalno funkcioniranje čestica (DIF) skovali su Paul Holland i Dorothy Thayer 1988. godine u poglavlju knjige „Test Validity“ (Holland i Thayer, 1988). Naredna su desetljeća svjedočila ogromnom interesu za istraživanje i mjerenje ovog fenomena, a iz tog su se interesa izrodile brojne metode za njegovo otkrivanje, čiji ekstenzivan pregled se može naći u radovima Millsapa i Eversona (1993), Camillija i Sheparda (1994), Clausera i Mazora (1998), Potenze i Doransa (1995), Penfielda i Lama (2000), Teresija i Fleishmana (2007), Penfielda i Camillija (2006), Mapurange, Doransa i Middletona (2008), te Osterlinda i Eversona (2009).

1.3.1. Zajednička obilježja metoda detekcije DIF-a

Unatoč mnogim razlikama među metodama detekcije DIF-a, neka su im obilježja svima zajednička.

Kao prvo, u istraživanjima DIF-a uobičajeno je uspoređivati dvije skupine ispitanika: **fokalnu i referentnu skupinu** (iako je ponekad moguće imati i više fokalnih skupina). Fokalna skupina je ona skupina koja je u fokusu istraživanja (obično manjinska skupina), i za koju se vjeruje da postoji nepravedan tretman instrumenta, a referentna je ona koja se pritom koristi za usporedbu (obično većinska skupina). Dakle, prvo zajedničko obilježje svih metoda detekcije DIF-a je da je nužno postojanje grupirajuće varijable.

Osim toga, kako bi se DIF uspio statistički odvojiti od stvarne razlike među grupama, potrebna je **varijabla za izjednačavanje/uparivanje** (engl. *matching variable*) pomoću koje ispitanike izjednačujemo s obzirom na njihovu sposobnost.

Dvije su vrste varijabli koje pritom možemo koristiti: interne ili eksterne. Kada kao varijablu za uparivanje koristimo uspješnost na testu čije čestice provjeravamo, onda je riječ o internoj varijabli za uparivanje (pristup internog kriterija). S druge strane, ako se kao varijabla za uparivanje koristi rezultat na nekom drugom testu ili neki drugi vanjski kriterij, onda je riječ o eksternoj varijabli za uparivanje (pristup prediktivne valjanosti). Metode diferencijalnog funkcioniranja čestica koriste se unutar pristupa internog kriterija (Camilli i Shepard, 1994).

S obzirom na vrstu internog kriterija koji se koristi za uparivanje ispitanika (tj. način na koji se postiže ekvivalentnost ispitanika u skupinama), razlikujemo dvije skupine metoda DIF-a. Prve se temelje na već spomenutom ukupnom rezultatu na testu (koji je vrlo praktičan za upotrebu zbog svoje lake dostupnosti i relativne točnosti), a druge na procjeni latentne sposobnosti. Točnije, razlikujemo metode koje su temeljene na teoriji odgovora na zadatak (procjena latentne sposobnosti) i one koje nisu temeljene na teoriji odgovora na zadatak. Neke

od ovih metoda koje ne proizlaze iz pristupa teorije odgovora na zadatak utemeljene su na klasičnoj teoriji testova, neke su kombinacija IRT-a i CTT-a, a neke nemaju teorijsku osnovu unutar teorija testova (Mapuranga i sur., 2008).

Za preciznost detekcije DIF-a, važno je osigurati da varijable koje koristimo za uparivanje nisu kontaminirane DIF česticama, jer u suprotnom raste broj lažnih detekcija DIF-a, tj. stopa pogrešaka tipa I. Međutim, u praksi se suočavamo s izazovom već u početnoj fazi - prepoznavanju čestica koje se diferencijalno ponašaju. Postupci za detekciju DIF-a primarno su namijenjeni identifikaciji tih specifičnih čestica. Stoga, očekivati da već prije same analize možemo sa sigurnošću odrediti koje čestice će pokazati DIF je nelogično i cirkularno.

Tom problemu se nastoji doskočiti procedurom koja se naziva purifikacija tj. pročišćavanje čestica (engl. *item purification*) ili purifikacija varijable za uparivanje (engl. *matching criterion purification*). Ovaj postupak uključuje izvođenje preliminarnih analiza DIF-a kako bi se identificirane DIF čestice eliminirale iz skupa podataka prije formiranja varijable za uparivanje. Tim postupkom se nastoji osigurati da varijabla za uparivanje sadrži samo invarijantne čestice koje se zatim koriste u glavnoj DIF analizi (Clauser i Mazor, 1998; French i Maller, 2007; Holland i Thayer, 1988).

Pritom se uobičajeno razlikuju dva pristupa. Prvi je dvostupanjski postupak (engl. *two-step purification*), koji su predložili Holland i Thayer (1988), a drugi je iterativni postupak. Razlika između ovih pristupa leži u broju provedenih preliminarnih DIF analiza. Dvostupanjski postupak koristi jednu preliminarnu analizu, dok iterativni pristup nastavlja s ponavljanjem analiza DIF-a sve dok se u preostalom skupu više nijedna čestica ne detektira kao DIF čestica (French i Maller, 2007).

Procesom pročišćavanja čestica dolazi do određivanja skupa tzv. sidrišnih čestica (engl. *anchor items*), odnosno čestica za koje se pretpostavlja da su invarijantne. Ove čestice služe kao temelj za kreiranje ekvivalentne skale, omogućavajući usporedbu latentnih varijabli između različitih grupa. Taj proces je ključan za formiranje nepristrane varijable za uparivanje za koju se smatra da jednako mjeri latentnu osobinu u obje skupine, te posljedično i za smislenu i pouzdanu analizu DIF čestica.

1.3.2. Različite klasifikacije metoda detekcije DIF-a

Metode za detekciju diferencijalnog funkcioniranja čestica primarno se klasificiraju na temelju dva ključna kriterija: kako definiraju **odsustvo DIF-a** i koju **varijablu za uparivanje** koriste. Prema ova dva kriterija razlikujemo dvije glavne vrste: metode opaženog rezultata i metode latentnih osobina. Metode opaženog rezultata pretpostavljaju da odsustvo DIF-a postoji kada svi ispitanici s identičnim ukupnim rezultatom na testu imaju jednake šanse za točan odgovor na određenu česticu, neovisno o pripadnosti referentnoj ili fokalnoj skupini. S druge strane, metode latentnih osobina definiraju odsustvo DIF-a kao situaciju gdje ispitanici s istom

razinom latentne osobine imaju istu vjerojatnost točnog odgovora na česticu, također bez obzira na grupnu pripadnost.

Dok metode opaženog rezultata koriste ukupni rezultat kao varijablu za uparivanje, metode latentnih osobina se oslanjaju na procjene latentne varijable, dobivene ili temeljem teorije odgovora na zadatak (IRT) ili klasične teorije testova.

Metode opaženog rezultata, kao tradicionalniji pristup, dugo su bile metodologija izbora za detekciju DIF-a. Međutim, s razvojem računalnih tehnologija i dostupnosti softvera, metode latentnih osobina postale su sve popularnije zahvaljujući svojoj sposobnosti sofisticiranije analize.

Neke od korištenijih metoda latentnih osobina su: test omjera izglednosti iz područja teorije odgovora na zadatak (engl. *IRT-LR test*) (Thissen i sur., 1986), Lordov (1977, 1980) Wald (1943) χ^2 test, poboljšani Waldov test (engl. *improved Wald test*) (Langer, 2008; Woods i sur., 2013; Tay i sur., 2015; Cao i sur., 2017), MIMIC model (Jöreskog i Goldberger, 1975; Muthén, 1989; Woods, 2009), MIMIC model s interakcijom (engl. *MIMIC-interaction model*) (Woods i Grimm, 2011), te MIMIC model sa sekvencijalno slobodnim osnovnim modelom (engl. *MIMIC with sequential-free baseline*) (Chun i sur., 2016).

Metode latentnih osobina mogu se podvući pod dva glavna pristupa: strukturalno modeliranje (engl. *Structural Equation Modeling*, SEM) i teoriju odgovora na zadatak (IRT). Oba ova pristupa omogućavaju otkrivanje pristranosti čestica i usporedbu ekvivalentnosti mjernih instrumenata. No, pritom koriste različitu teorijsku osnovu, različite procedure za ispitivanje odnosa među česticama i skalama, kao i potpuno različitu terminologiju. Zbog ovog raskola među pristupima, istraživači se obično opredjeljuju za jedan, no ima i onih koji esperimentiraju s metodama iz oba pristupa kako bi dobili sveobuhvatniji uvid u fenomen pristranosti mjernih instrumenata (Zimbra, 2018).

Također, metode se mogu klasificirati i s obzirom na **vrstu čestica** za koju su primjenjive. Prve DIF metode su nastale prvenstveno za upotrebu na dihotomnim česticama, a tek desetljeće kasnije neke od njih su adaptirane i za politomne čestice. No, u usporedbi s metodama dihotomnih čestica, ove druge sadrže određene probleme koji proizlaze iz samih osobina politomnih čestica. Prema preglednom radu Penfielda i Lama (2000), to su sljedeće karakteristike i poteškoće: 1) niža pouzdanost politomnih čestica; 2) teže definiranje varijable za uparivanje kod politomnih čestica i 3) problem stvaranja jedinstvene mjere funkcioniranja svake pojedine kategorije politomnih čestica (ekvivalentne težini čestica kod dihotomnih čestica).

Niža pouzdanost politomnih čestica često proizlazi iz specifičnih karakteristika ovog tipa čestica, koje uključuju tipično kraće skale, veću nedosljednost u procjenama ispitanika i veću raznolikost sadržajnih domena. Ove karakteristike su izraženije kod politomnih nego kod dihotomnih čestica. Takva niža pouzdanost može dovesti do nepreciznosti u procjenama latentnih osobina, što zauzvrat može rezultirati neopravdanim identificiranjem čestica koje navodno pokazuju diferencijalno funkcioniranje (DIF), odnosno pogreškom tipa I.

Drugi nedostatak povezan je s težim definiranjem varijable za uparivanje kod politomnih čestica. Naime, varijabla za uparivanje trebala bi biti dostatna za dobru procjenu latentne osobine, odnosno ona bi ju trebala valjano i pouzdano mjeriti. No, ukoliko varijabla za uparivanje nije dobra mjera latentne osobine, u tom slučaju također može doći do povišenih stopa pogrešaka tipa I. Međutim, zbog već spomenutih karakteristika instrumenata s politomnim česticama, a to su manji broj čestica, niža pouzdanost te potencijalna višedimenzionalnost, definiranje dobre varijable za uparivanje nije tako jednostavno (Zwick i sur., 1993). Neka od mogućih rješenja ovog problema su upotreba vanjskog kriterija (koji mora biti visoko koreliran s latentnom osobinom koja se mjeri, što je pak drugi problem koji se u tom slučaju javlja), purifikacija čestica ili upotreba procjena latentne osobine umjesto ukupnog rezultata.

Treće, procjena karakteristika čestice je složenija za politomne čestice. Kod dihotomnih čestica, karakteristike (parametri) čestice mogu se lako definirati procjenom vjerojatnosti točnog odgovora. Međutim, za politomne čestice s više kategorija odgovora, ne postoji jedinstvena mjera, već više stupnjeva točnog odgovora; osim toga, postoji i više potencijalnih razlika među grupama za svaku kategoriju odgovora.

Iz prethodno navedenog razloga, neke metode za detekciju DIF-a uspješne su samo u identificiranju DIF-a kod dihotomnih čestica, dok su druge prilagođene ili konstruirane za upotrebu i s politomnim česticama.

Osim toga, metode se razlikuju i s obzirom na **broj grupa** koje mogu simultano analizirati. Iako je vjerojatno bolje analizirati DIF koristeći sve grupe simultano (zbog statističke efikasnosti, te mogućnosti analize interakcija), većina DIF metoda uspoređuje samo po dvije grupe u jednoj analizi (Mapuranga i sur., 2008; Zhang i sur., 2005).

Još jedna klasifikacija metode DIF-a razvrstava s obzirom na **teoretsku osnovu** (Mapuranga i sur., 2008). Tako se razlikuju metode iz okvira klasične testne teorije, metode iz okvira teorije odgovora na zadatak, miješane metode te metode koje nemaju teorijsku podlogu. Među metodama koje se temelje na klasičnoj testnoj teoriji najpoznatije su Mantel-Haenszelov postupak (Holland i Thayer, 1988) – metoda izbora organizacije ETS; DIF disekcija (Zhang i sur., 2005) i standardizacija (Clauser i Mazor, 1998). U okviru teorije odgovora na zadatak nastao je niz metoda, no sve one mogu se razvrstati u tri skupine: metode koje se usmjeravaju na pronalazak razlika među parametrima čestica, metode koje izračunavaju površinu između karakterističnih krivulja čestice u dvije grupe i metode koje se koriste omjerima izglednosti za usporedbu promjene u slaganju niza ugniježđenih modela (npr. IRT-LR, Rajueva metoda površine (Raju, 1988), Lordov Wald test (Lord, 1980)). Među miješanim metodama ističe se SIBTEST (Shealy i Stout, 1993), a među metodama koje nemaju teorijsku osnovu logistička regresija (Swaminathan i Rogers, 1990; Miller i Spray, 1993; French i Miller, 1996; Zumbo, 1999).

Najnoviji interes istraživača je usmjeren na metode koje imaju sposobnost detekcije DIF-a u instrumentima s namjerno višedimenzionalnom strukturom. S obzirom na to, još jedna je moguća klasifikacija s obzirom na **dimenzionalnost** podataka koja metode dijeli na metode

primjerene za jednodimenzionalne podatke i one koje mogu otkrivati DIF i u višedimenzionalnim podacima. Naime, kada instrument namjerno mjeri više od jedne dimenzije, upotreba metoda koje su namijenjene jednodimenzionalnim instrumentima može dovesti do značajnih pogrešaka u procjeni parametara (Kirisci i sur., 2001; Dorans i Kingston, 1958) i krivom identificiranju DIF čestica (Snow i Oshima, 2009). Iz tog razloga je bitno prije provedbe DIF analiza odrediti točan broj latentnih dimenzija, te ako se utvrdi da je instrument višedimenzionalan, trebalo bi koristiti metode za višedimenzionalne instrumente.

Do danas je predloženo samo nekoliko pristupa istraživanju DIF-a u višedimenzionalnim podacima. Uglavnom su to višedimenzionalne nadgradnje uobičajenih DIF metoda za jednodimenzionalne instrumente, kao što je recimo višedimenzionalni SIBTEST, diferencijalno funkcioniranje čestica i testova (DFIT) (Raju i sur., 1995), logistička regresija (Mazor i sur., 1998), IRT-LR test (Suh i Cho, 2014) i MIMIC model (engl. *multiple indicators multiple causes*) (Lee i sur., 2016).

Među njima, najčešće se koriste MIMIC, IRT-LR i logistička regresija, jer su najjednostavniji, najlakše dostupni, te imaju jasnu vezu s IRT modelima (parametrijske metode) (Bulut i Suh, 2017).

S obzirom na velik broj dostupnih metoda, te njihovih varijacija, u nastavku će detaljnije biti opisane samo metode za višedimenzionalne instrumente s politomnim česticama koje su odabrane za upotrebu u ovom istraživanju.

1.3.3. Metode za višedimenzionalne instrumente s politomnim česticama

1.3.3.1. Omjer izglednosti iz okvira teorije odgovora na zadatak (IRT LR)

Metoda omjera izglednosti je za detekciju DIF-a prvi puta upotrijebljena u radu Thissena, Steinberga i Gerarda 1986. godine. Ovaj pristup koristi pokazatelj omjera izglednosti, označen kao G^2 , kako bi testirao nultu hipotezu koja pretpostavlja jednakost parametara čestica među grupama (Thissen, 2001). Njime se za svaku pojedinu česticu uspoređuje nekoliko ugniježđenih IRT modela. Pritom se mogu koristiti različiti IRT modeli, ovisno o vrsti čestica u instrumentu. Vrijednosti pokazatelja omjera izglednosti distribuiraju se približno kao vrijednosti hi-kvadrata, pri čemu je broj stupnjeva slobode jednak razlici u broju parametara između dva modela (Thissen, 1991). Pritom je za točnost rezultata ove metode nužno definiranje skupa sidrišnih čestica koje se koriste za procjenu latentne varijable, te za izjednačavanje njezine skale u obje skupine.

Metodu su za korištenje s politomnim česticama adaptirali Thissen i sur.(1993), a Suh i Cho su ju 2014. godine prvi put primijenili na višedimenzionalnim podacima. Postupak započinje s onim što se naziva „omnibus“ testom, koji simultano uspoređuje a i b parametre čestice, služeći kao početna provjera postojanja bilo kakvog DIF-a. U ovom koraku,

osnovni model (engl. *baseline or constrained model*), gdje su parametri a i b izjednačeni u obje grupe, stavlja se nasuprot proširenom modelu (engl. *augmented model*), koji dozvoljava individualne procjene tih parametara za svaku grupu. Statistička značajnost u „omnibus“ testu ukazuje na razlike među grupama, ali ne specificira da li te razlike proizlaze iz parametra a , b , ili oboje.

Da bi se preciziralo koji parametri čestica izazivaju DIF, slijede dvije dodatne usporedbe. Prva od njih usmjerena je na detekciju nejednolikog DIF-a, uspoređujući model gdje je parametar a fiksiran, a parametar b slobodan (drugi kompaktni model) s prethodno spomenutim proširenim modelom. Ako omjer izglednosti pokazuje statističku značajnost, to upućuje na prisutnost nejednolikog DIF-a.

Ako pak prva dodatna usporedba ne pokaže statistički značajne rezultate, provodi se druga dodatna usporedba koja istražuje postojanje jednolikog DIF-a. Ova usporedba odvija se između drugog kompaktnog modela, gdje je b parametar slobodan, i osnovnog modela, gdje su oba parametra, a i b , jednaka među grupama. Statistički značajan omjer izglednosti ukazuje na jednoliki DIF.

Bitan element ove analize je izjednačenost parametara sidrišnih čestica; oni ostaju nepromijenjeni kroz sve tri usporedbe, osiguravajući stabilnost i pouzdanost modela u identifikaciji DIF-a.

Za procjenu parametara koristi se metoda marginalnog maksimalnog izgleda prema Bocku i Aitkinu (1981) s algoritmom maksimizacije očekivanja (EM MML algoritam). Veličina učinka može se odrediti pomoću vrijednosti G^2 , te se na taj način može procijeniti praktična važnost DIF-a, no ovo je zasad moguće samo kod jednodimenzionalnih modela (Bulut i Suh, 2017).

Detaljni postupci koji se koriste u ovoj metodi mogu se pronaći u radovima Thissena, Steinberga i Gerarda (1986), Thissena, Steinberga i Weinera (1988; 1993), Thissena (2001) i Woods (2008).

Simulacijska istraživanja su IRT-LR metodu za otkrivanje DIF-a provjeravala u različitim realističnim uvjetima, u kojima je pogreška tipa I za omnibus test bila blizu nominalnoj vrijednosti, što znači da je stopa lažno pozitivnih rezultata pod uvjetima koji su testirani bila prihvatljiva (Ankenmann i sur., 1999; Bolt, 2002; Cohen i sur., 1996; Kim i Cohen, 1998; Stark i sur., 2006; Sweeney, 1997; Wang i Yeh, 2003; Woods, 2009).

Cohen i suradnici su pak 1996. godine kod testa s 50 čestica prema 3PL modelu i s uzorcima od 250, ali i 1000 ispitanika, otkrili da je LR pokazatelj imao pogrešku tipa I višu od nominalne razine od 0.05. Wanichtanom je 2001. godine proširio Cohenovo istraživanje dodavši u njega i statističku snagu kao zavisnu varijablu, te je otkrio da je prosječna stopa detekcije jednolikog DIF-a (veličine 0.2, 0.5 i 0.8) bila 0.97.

Od mogućih čimbenika koji utječu na statističku snagu, tj. osjetljivost IRT-LR metode, provjeravan je utjecaj veličine uzorka, diskriminativnosti čestice, broja sidrišnih čestica, te

količina DIF čestica u instrumentu. Učinkovitost je provjeravana s različitim brojem čestica u instrumentu, s različitim modelima u podlozi, različitim omjerom grupa, različitim brojem sidrišnih čestica, te s prisustvom odnosno odsustvom grupnih razlika u latentnoj osobini.

Ta istraživanja pokazala su da povećanje veličine uzorka, diskriminativnosti čestice, broja sidrišnih čestica ili veličine DIF-a utječe na povećanje osjetljivosti IRT-LR metode (Ankenmann i sur., 1999; Wang i Yeh, 2003; Woods, 2009). Osim toga, IRT-LR je pokazala veću snagu od ostalih uspoređivanih metoda (Thissen i sur., 1993; Teresi i sur., 2000; Wainer, 1995; Wanichthanom, 2001), te je dobro kontrolirala pogrešku tipa I (Cohen i sur., 1996; Stark i sur., 2006; Woods, 2009).

No, u detekciji jednolikog DIF-a, IRTLR je pokazao nešto povišene stope pogreške tipa I u slučaju modela s tri parametra (3PL) (Cohen i sur., 1996) i kada je sidrišni skup čestica kontaminiran s DIF-om (Finch, 2005; Wang i Yeh, 2003). Bez kontaminacije, stope pogrešaka bile su nominalne (tj. na razini od .05), a snaga testa bila je znatno veća od .90 u slučaju modela s dva parametra (2PL) i iznad .80 u slučaju 3PL modela (Bolt, 2002; Finch, 2005; Kim i Cohen, 1998). Međutim, snaga testa bila je smanjena kad su korišteni manji uzorci ili kad je bila prisutna grupna razlika u latentnoj osobini, a u Finchevu istraživanju iz 2007. godine IRT-LR metoda pokazala je iznimno niske stope točne detekcije nejednolikog DIF-a (u razini uobičajeno očekivanih stopi pogrešaka tipa I) u uvjetima u kojima je postojala kontaminacija sidrišnog skupa DIF-om.

Sve u svemu, uz niz prednosti koja ova metoda ima, jedan od njenih glavnih nedostataka metode je računaska i posljedično vremenska zahtjevnost, a s obzirom da modeliranje u okviru teorije odgovora na zadatak zahtjeva velik uzorak, to ograničenje vrijedi i za ovu metodu detekcije DIF-a.

1.3.3.2. MIMIC model s interakcijom

Metodu modela s višestrukim indikatorima i višestrukim uzrocima (engl. *multiple indicators multiple causes*), tj. MIMIC metodu, prvobitno su kao poseban oblik strukturalnog modeliranja (SEM) predstavili Jöreskog i Goldberger (1975), a zatim ju je Muthén (1985) prvi put upotrijebio za otkrivanje DIF-a. U MIMIC modelu je standardni model konfirmatorne faktorske analize (CFA) proširen uključivanjem egzogenih varijabli koje uzrokuju latentne faktore (tj. kovarijata).

Metoda je prvobitno bila upotrebljiva samo za otkrivanje jednolikog DIF-a, ali su ju 2011. godine Woods i Grimm prilagodili za korištenje s nejednolikim DIF-om, te svoju verziju nazvali MIMIC modelom s interakcijom. Nju su 2016. godine Lee i suradnici adaptirali za otkrivanje DIF-a u višedimenzionalnim podacima (Lee i sur., 2016).

Osnovna logika MIMIC modela za otkrivanje jednolikog DIF-a je u provjeri da li istovremeno dodavanje izravnih putanja iz varijable pripadnosti grupi (grupirajuće varijable) prema česticama za koje se pretpostavlja da diferencijalno funkcioniraju (tzv. indikator varijable), te neizravnih putanja iz varijable pripadnosti grupi prema latentnoj osobini mijenja prikladnost

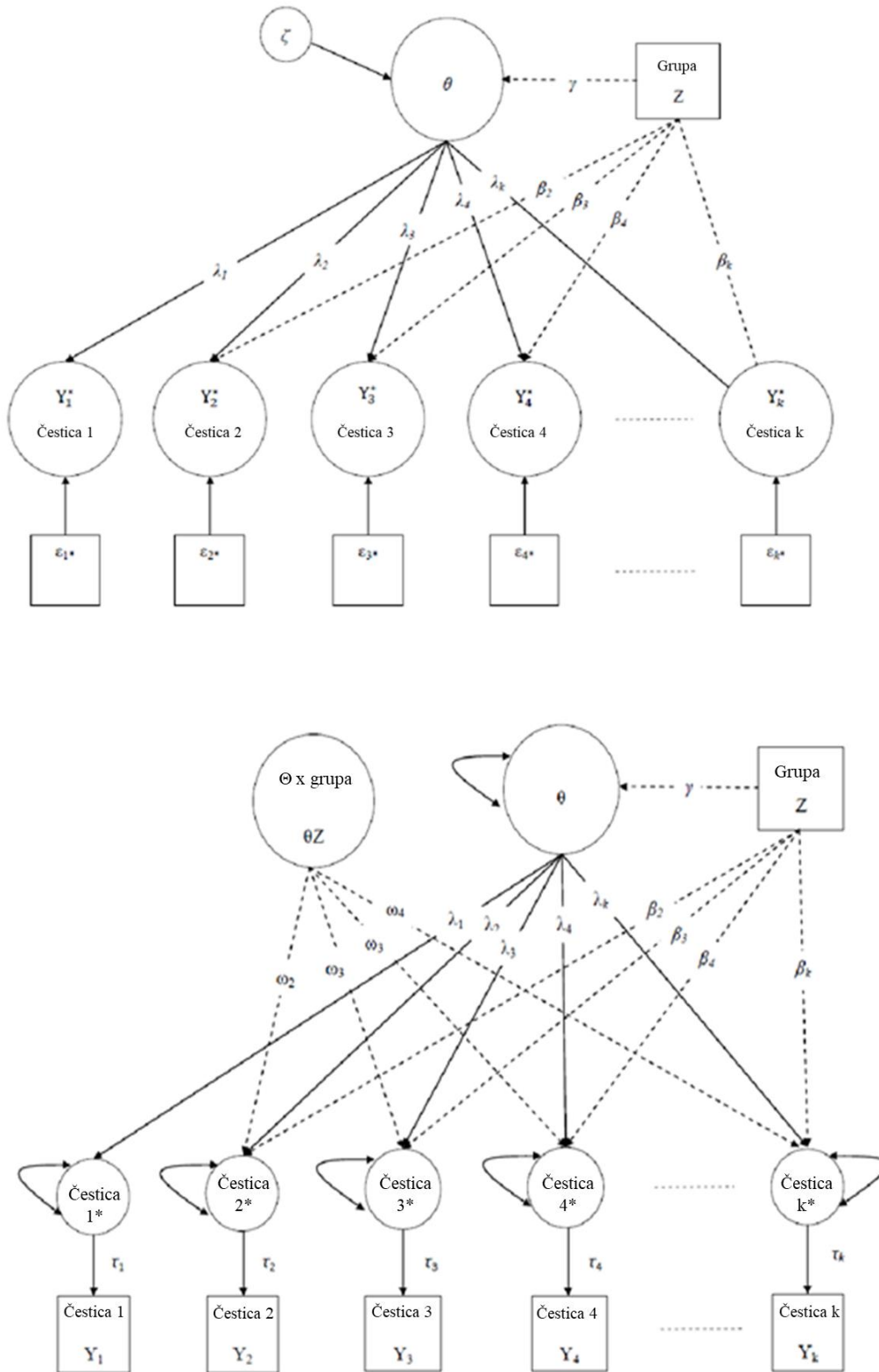
modela. Pri tome se jednoliki DIF pojedine čestice provjerava dodavanjem izravne putanje od grupirajuće varijable do indikatora, a nejednoliki DIF dodavanjem izravne putanje od interakcijskog termina (između latentne osobine i pripadnosti grupi) do indikatora. Značajne izravne putanje ukazuju na prisutnost DIF-a (jednolikog ili nejednolikog), a regresija latentne osobine na grupirajuću varijablu se koristi kako bi se kontrolirale grupne razlike u srednjoj vrijednosti latentne osobine (odnosno „impact“).

MIMIC metoda je konceptualno vrlo slična logističkoj regresiji jer procjenjuje uspješnost na čestici ovisno o pripadnosti grupi i o interakciji između grupirajuće varijable i latentne osobine, dok pritom istovremeno kontrolira razlike u srednjim vrijednostima grupa.

Finch (2005) je usporedio učinkovitost četiri metode u otkrivanju jednolikog DIF-a (Mantel-Haenszel (MH) postupak, SIBTEST, MIMIC i IRT-LR metodu) te ustanovio sljedeće: 1) MIMIC je imao veću stopu pogreške tipa I od svih ostalih metoda kada je korišten kratki test sastavljen od 20 čestica za koje su korišteni podaci generirani iz 3PL IRT modela; 2) kada su simulirani podaci iz 2PL IRT modela za duži test sastavljen od 50 čestica, MIMIC je imao stopu pogreške tipa I usporedivu s ostalim metodama, što se stope točnih identifikacija tiče, ona je također bila usporediva, a u nekim situacijama i viša nego kod ostalih metoda. Osim toga, na ostale tri metode negativno je utjecala količina DIF čestica u varijabli za uspoređivanje, dok je MIMIC po tom pitanju bio relativno robustan. Iz tog bi se razloga MIMIC mogao preporučiti kao metoda izbora u situacijama kada se očekuje velik broj DIF čestica u instrumentu. Osim toga, čini se da na ostale tri metode značajnije utječe veličina fokalne skupine, a pogotovo u prisustvu DIF-a u varijabli za uspoređivanje.

Wang i sur. (2009) su usporedili standardnu MIMIC metodu (*AOAA; all-others-as-anchors* ili *AOI; all-other-item* metodu) s MIMIC metodom s postupkom purifikacije skale koristeći simulirane podatke. Postupak purifikacije skale sastoji se od četiri koraka: (a) inicijalna procjena DIF-a za sve čestice, (b) uklanjanje čestica kod kojih je otkriven DIF iz skupa referentnih čestica (tzv. sidrišne čestice), (c) ponovna procjena DIF-a za sve čestice i (d) ponavljanje koraka b i c sve dok se u dvije uzastopne iteracije ne dobije isti skup DIF čestica (Wang i sur., 2009). U tom istraživanju MIMIC metoda s purifikacijom pokazala se učinkovitijom od standardne MIMIC metode.

Slika 17. Prikaz MIMIC modela s interakcijom za detekciju jednolikog (gore) i nejednolikog DIF-a



Napomena: Prilagođeno prema Zimbri, 2018.

Zatim su Wang i Shih (2010) proširili prethodno istraživanje koristeći politomne čestice i upotrijebivši još jednu verzije purifikacije MIMIC metode, koju su nazvali MIMIC sa čistim sidrišnim skupom (engl. *MIMIC pure anchor*, *MIMIC PA*). U ovoj novoj metodi purifikacije, čisti sidrišni skup odabire se prethodno opisanim iterativnim postupkom, a zatim se taj skup koristi u novoj DIF analizi u kojoj se parametri sidrišnih čestica izjednače u obje grupe, a parametri ispitivanih čestica oslobode. Ovo istraživanje je prvo koristilo tzv. *DIF-free-than-DIF* strategiju na politomnim česticama, te se pokazalo da je ona točnija u odnosu na druge metode, te da može zadržati visoku snagu i dobru kontrolu stope pogrešaka tipa I čak i kada mjerni instrument sadrži do 40% DIF čestica.

Stope pogrešaka tipa I i snagu metoda MIMIC i IRT-LR za otkrivanje DIF-a kod malih fokalnih grupa usporedila je Woods u svom simulacijskom istraživanju. Rezultati su pokazali da je u tom uvjetu MIMIC pristup otkrivao jednoliki DIF bolje od IRT-LR metode. Osim toga, zaključila je da, uz držanje drugih varijabli, kao što su broj čestica i postotak čestica s DIF-om, konstantnima, MIMIC model treba manji uzorak za dobivanje adekvatne snage i preciznu procjenu parametara (Woods, 2009).

U sličnom simulacijskom istraživanju u kojem su također uspoređivali MIMIC i IRT-LR metodu, Woods i Grimm (2011) su prvi put upotrijebili interakcijsku MIMIC metodu za otkrivanje nejednolikog DIF-a. To su učinili tako da su u MIMIC model dodali interakciju latentne varijable i grupne pripadnosti. No, ovakva upotreba pristupa strukturalnog modeliranja s latentnom moderator varijablom za otkrivanje nejednolikog DIF-a dovela do inflacije pogreški tipa I. Njihovo objašnjenje ovog nalaza bilo je da je uzrok visokim stopama pogrešaka tipa I, metoda koju Mplus koristi za procjenu interakcija latentnih varijabli, a to je pristup latentnih moderiranih strukturalnih jednadžbi, tj. LMS (engl. *Latent Moderated Structural Equations*; Klein i Moosbrugger, 2000). U njihovom istraživanju navodi se da LMS pretpostavlja normalnu distribuciju obje latentne varijable između kojih se formira interakcija (pomoću XWITH naredbe), te da ukoliko ova pretpostavka nije zadovoljena, dolazi do povećanje pogrešaka tipa I (Klein i Moosbrugger, 2000). S obzirom da će u interakciji kontinuirane latentne varijable i kategorijalnih opaženih kovarijata pretpostavka o multivarijantnoj normalnosti najvjerojatnije biti narušena. Međutim, u međuvremenu je došlo do adaptacije ove metode u okviru MPlus programa, tako da LMS metoda sad koristi procjenu maksimalne vjerojatnosti s potpunim informacijama (engl. *full-information maximum likelihood estimation*) za analizu interakcijskih učinaka koji nisu normalno distribuirani između latentne varijable i opažene kategorijalne varijable, te se na taj način uklanja problem odstupanja od pretpostavke normalnosti (Barendse i sur., 2012).

Navedeno istraživanje Woods i Grimma proširili su Lee i sur. (2016) ispitivanjem učinkovitosti višedimenzionalnog MIMIC-interakcijskog modela u različitim simulacijskim uvjetima u pogledu kontrole stopa pogreške tipa I i statističke snage, tj. osjetljivosti, te zaključili da su stope snage bile veće u uvjetima jednolikog DIF-a nego u uvjetima nejednolikog DIF-a. Osim toga, snaga otkrivanja DIF-a bila je viša u uvjetima s većim uzorcima i s više sidrišnih čestica. Općenito, višedimenzionalni MIMIC-interakcijski model

učinkovit je za otkrivanje jednolikog DIF-a, ali njegova sposobnost otkrivanja nejednolikog DIF-a i dalje je upitna.

Chun i sur. (2016) su u simulacijskom istraživanju istraživali učinkovitost MIMIC metoda za otkrivanje jednolikog i nejednolikog DIF-a simultano na više grupa. Najprije su simulirali dvije pozadinske varijable koje su modelirali kao uzrok DIF-u (npr. spol i etnička pripadnost), a zatim su usporedili tri različite verzije MIMIC modela: ograničeni osnovni model (engl. *constrained baseline model*), slobodni osnovni model (engl. *free baseline model*) i sekvencijalno-slobodni osnovni model (engl. *sequential-free model*).

Istraživanje Buluta i Suh, objavljeno 2017. godine, usporedilo je učinkovitost interakcijskog MIMIC modela, IRT-LR metode i logističke regresije u detekciji DIF-a u namjerno višedimenzionalnim instrumentima koji nemaju jednostavnu strukturu. U istraživanju su korišteni dihotomni podaci. U njihovom istraživanju najučinkovitijim se pokazala IRT-LR metoda, no i MIMIC model je funkcionirao dobro po pitanju snage i stopi pogrešaka tipa I, a u usporedbi s druge dvije metode pogotovo je dobro detektirao DIF u kraćim instrumentima. Logistička regresija bila je najlošija po pitanju statističke snage, a imala je i najveću inflaciju pogrešaka tipa I (Bulut i Suh, 2017).

Montoya i Jeon (2020) dali su novi pogled na MIMIC model, jer su ga konceptualizirali kao konfirmatorni faktorski model u kojem je jednoliki DIF predstavljen kao medijacija, a nejednoliki DIF kao moderirana medijacija. Prednost ove konceptualizacije DIF-a je u tome što se odmiče od tradicionalnog pogleda na DIF koji je isključivo eksplorativan, te vodi prema konfirmacijskom istraživanju DIF-a u kojem se unaprijed postavljaju hipoteze o funkcioniranju pojedinačnih čestica i mogućim uzrocima njihovog različitog funkcioniranja.

Tradicionalno se MIMIC metoda za detekciju DIF-a koristila metodom ograničenog osnovnog modela. U njemu se prisustvo DIF-a provjerava tako da se rade usporedbe niza reduciranih modela s punim modelom, u kojima se dodaju putanje od pojedine grupne varijable do pojedinačnih čestica. Ako puni model pokaže statistički značajno bolje slaganje s podacima u odnosu na reducirani osnovni model, tada je DIF prisutan (Kim i sur., 2012). Iako je metoda ograničenog osnovnog modela praktična jer omogućuje evaluaciju svake čestice, ona često dovodi do visokih stopi pogreške tipa I. Uzrok tome je potencijalno netočna specifikacija osnovnog modela kada je u česticama prisutan DIF (Stark i sur., 2006).

Slobodni osnovni model počiva na formiranju modela s minimalnim ograničenjima potrebnima za njegovu identifikaciju. To se čini izjednačavanjem faktorskih opterećenja i limena (engl. *threshold*) jedne čestice u svim grupama. Ovaj model se zatim uspoređuje s reduciranim modelima, u kojima se parametri svake čestice ograničavaju pojedinačno, te se zatim analiziraju promjene u indeksima slaganja.

Sekvencijalno-slobodni osnovni model prvo koristi ograničeni pristup za identifikaciju čestica bez DIF-a. Iz ove skupine se bira čestica s najvećom diskriminativnošću kao sidrišna čestica za drugi korak koji koristi slobodni osnovni model za analizu svih ostalih čestica.

Ovo istraživanje je pokazalo da je sekvencijalno-slobodni pristup bolji od ostalih modela, jer ima visoke stope statističke snage i niske stope pogreške tipa I, čime se između ova tri ističe kao preferirani model za detekciju DIF-a.

Dosadašnja istraživanja MIMIC modela otkrila su da on ima nekoliko prednosti. Kao prvo, ova je metoda fleksibilna jer omogućava da grupirajuća varijabla bude bilo koja vrsta varijable (binarna, kategorijalna ili kontinuirana) koja ima dvije ili više razina. S obzirom na to, MIMIC omogućava simultanu usporedbu više od dviju grupa, te slojevito kvantitativno istraživanje uzroka DIF-a (Chun i sur., 2016; Muthén, 1988).

Za razliku od drugih metoda iz okvira konfirmatorne faktorske analize (multigrupna konfirmatorna faktorska analiza, MG-CFA (Pae i Park, 2006) i metoda modifikacijskih indeksa (Chan, 2000) koje zahtijevaju dva odvojena skupa parametara modela za dvije grupe koje se uspoređuju, prednost MIMIC-a je da se koristi samo jednim skupom parametara modela za cijeli uzorak, jer pretpostavlja da je populacijska matrica faktorskih opterećenja u obje grupe jednaka. To zapravo znači da se za otkrivanje DIF-a pomoću MIMIC metode mogu koristiti dvostruko manji uzorci. Osim toga, u situacijama u kojima se uspoređuje više od dviju grupa, uzorak se ne mora povećavati. No, ova pretpostavka se s druge strane može gledati i kao jedno od ograničenja metode.

Među metodama otkrivanja DIF-a, MIMIC je također najmanje pogođen slučajnim uključivanjem DIF-a u sidrišne čestice (Finch, 2005).

Sve u svemu, MIMIC se pokazao učinkovitim u dihotomnim (Finch, 2005; Stark i sur., 2006) i politomnim česticama (Chun i sur., 2016; Hong i sur., 2008; Wang i Shih, 2010) u jednodimenzionalnim instrumentima, a nedavno istraživanje Buluta i Suha (2017) je pokazalo i njegovu učinkovitost u višedimenzionalnim instrumentima s dihotomnim česticama.

1.3.3.3. Logistička regresija

Logističku regresiju su za otkrivanje jednolikog i nejednolikog DIF-a kod dihotomnih čestica prvi put upotrijebili Swaminathan i Rogers (1990). Kasnije je ova metoda prilagođena i za upotrebu s politomnim česticama (ordinalna logistička regresija). Učinili su to Miller i Spray (1993) te French i Miller (1996), a Zumbo (1999) je doprinio poboljšanju metode predloživši nekoliko pseudo R^2 pokazatelja kao mjera veličine efekta. Logističku regresiju možemo konceptualizirati kao poveznicu između metoda kontingencijskih tablica (tj. metoda omjera izgleda) i IRT metoda (Clauser i Mazor, 1998).

Prilikom otkrivanja DIF-a, logistička regresija nastoji modelirati učinak pojedinca na čestici kao funkciju njegovog ukupnog rezultata (ili procijenjene razine latentne osobine) i pripadnosti grupi, te zatim uspoređuje slaganje tog modela s podacima. Proces modeliranja slijedi tri koraka, koji su slični koracima u IRT-LR metodi.

Prvi korak koristi ukupni rezultat ili procjenu latentne osobine (Mazor i sur., 1998) za predviđanje vjerojatnosti ispravnog odgovora na čestici (ili, u slučaju politomnih čestica, zaokruživanje određene kategorije odgovora). U drugom koraku, u predikciju se dodaju informacije o pripadnosti grupi, a u posljednjem koraku uključuje se i interakcijski termin. Pritom se za modeliranje koristi logistička funkcija.

Ako ukupni rezultat sam po sebi može predvidjeti uspješnost na proučavanoj čestici, može se zaključiti da je čestica invarijantna. No, ako uključivanje pripadnosti grupi poboljšava točnost predviđanja, tada se zaključuje da je u čestici prisutan jednoliki DIF. I konačno, ako je osim ukupnog rezultata i pripadnosti grupi za točnije predviđanje potrebno uključiti i interakciju između te dvije varijable, tada je u čestici prisutan nejednoliki DIF (Zumbo, 1999).

Ovaj postupak se matematički može prikazati kao:

$$\text{Ln} \left(\frac{P_{mi}}{1 - P_{mi}} \right) = b_0 + b_1 \text{tot} + b_2 \text{group} + b_3 (\text{tot} * \text{group}) \quad (13)$$

U ovoj formuli b_0 je sjecište, $b_1 \text{tot}$ je utjecaj varijable za uparivanje (ukupni rezultat ili procjena latentne osobine), $b_2 \text{group}$ je grupirajuća varijabla, a $b_3 (\text{tot} \times \text{group})$ je interakcijski efekt latentne osobine i pripadnosti grupi. Osim toga, formula se temelji na logističkoj funkciji $\text{Ln} \left(\frac{P_{mi}}{1 - P_{mi}} \right)$ u kojoj je P_{mi} vjerojatnost da će ispitanik m točno odgovoriti na česticu i , a $1 - P_{mi}$ je vjerojatnost njegovog netočnog odgovaranja. Drugim riječima, radi se o prirodnom logaritmu omjera izgleda za uspjeh naspram izgledima za neuspjeh.

Identifikacija DIF-a pomoću logističke regresije nalikuje hijerarhijskoj regresiji, jer se kroz nekoliko koraka grade sukcesivni modeli, te se u svakom koraku dodaje nova varijabla a zatim se provjerava da li ta nova varijabla poboljšava model u odnosu na prethodni.

Kao test značajnosti DIF-a, od hi-kvadrat vrijednosti iz trećeg koraka oduzme se hi-kvadrat vrijednost iz prvog koraka. Tako se dobiva ukupni indeks značajnosti DIF-a. Hi-kvadrat vrijednost iz drugog koraka može se oduzeti od hi-kvadrata iz trećeg koraka kao mjera nejednolikog DIF-a. Usporedba hi-kvadrata iz prvog i drugog koraka je dobar indikator jednolikog DIF-a.

Za postupak logističke regresije, Zumbo (1999) je predložio upotrebu usporedbe R^2 vrijednosti za model iz prvog koraka i R^2 vrijednosti modela iz trećeg koraka kako bi se procijenila veličina DIF-a (tzv. $R^2\Delta$). No, ukoliko nas zanimaju veličine efekta za jednoliki odnosno nejednoliki DIF, onda ćemo se koristiti razlikom u R^2 vrijednostima između modela iz drugog i prvog koraka, odnosno između modela iz trećeg i drugog koraka. Slijedeći konvencije Cohena (1992) za male, srednje i velike efekte, Zumbo i Thomas (1996) su predložili smjernice za identificiranje zanemarivog, umjerenog i velikog DIF-a koristeći $R^2\Delta$. Prema njima, za vrijednosti manje od 0.13, količina DIF-a smatrala se zanemarivom, za

vrijednosti između 0.13 i 0.26, količina DIF-a smatrala se umjerenom, a za vrijednosti veće od 0.26, količina DIF-a smatrala se velikom. No, naknadna istraživanja Jodoina i Gierla (2001) i Kima i sur. (2007) su Zumbo-Thomasove veličine efekta dovele u pitanje, te navedeni autori smatraju kako je nužno daljnje istraživanje mjera veličine efekata.

Mazor i sur. (1998) su prvi koristili procjenu latentne osobine kao varijablu za uparivanje u logističkoj regresiji. Osim toga su i prilagodili ovu metodu za upotrebu s dvije ili više latentnih osobina, odnosno za upotrebu na višedimenzionalnim podacima.

S obzirom da se logistička regresija i IRT-LR metode smatraju učinkovitijima u otkrivanju nejednolikog DIF-a, Le i sur. (2006) predložili su načine prilagodbe tih metoda za računalno adaptivno testiranje (engl. *computerized adaptive testing*, CAT) i procijenili njihovu učinkovitost u otkrivanju jednolikih i nejednolikih DIF čestica. Rezultati simulacija pokazali su da su obje metode podjednako učinkovite u otkrivanju i jednolikih i nejednolikih DIF čestica. Međutim, utvrđeno je povećanje pogreške tipa I kod logističke regresije u uvjetima kada postoji razlika u prosječnim vrijednostima osobine između grupa

Osim toga, Wiberg (2007) je primijetila da se logistička regresija, kao i MH pokazatelji, vrlo često upotrebljavaju zbog činjenice da se mogu koristiti za otkrivanje DIF-a i kod uzoraka male veličine. Uz to, Zumbo (1999) je istaknuo kako je za LR potrebno 200 ispitanika po grupi što nije tako velik uzorak ukoliko znamo da drugi modeli, primjerice IRT s tri parametra, zahtijevaju više od 1000 ispitanika po skupini.

McNamara i Roever (2006) smatraju LR korisnom jer je riječ o neparametrijskoj metodi koja omogućava modeliranje i jednolikog i nejednolikog DIF-a, a jednostavnija je za upotrebu od analiza temeljenih na teoriji odgovora na zadatak. Također, postoji i mnoštvo softverskih programa za DIF analizu pomoću logističke regresije.

Jodoin i Gierl (2001) pak navode kako logistička regresija ima poteškoću s povišenim stopama pogrešaka tipa I kad su prisutni veliki uzorci, te to objašnjava time što su relativno male razlike u uzorkovanju označene kao statistički značajne, te posljedično proglašene DIF-om. No, problematika zapravo leži u hi-kvadrat pokazatelju kojeg ova metoda koristi, a koji je osjetljiv na veličinu uzorka. Iz tog razloga, veliki uzorci mogu dovesti do proglašenja čestica DIF česticama, zbog značajnog, ali malog i praktički nevažnog DIF-a. Ovo je potvrđeno u nekoliko istraživanja koja su pokazala da i snaga LR metode, ali i stopa pogrešaka tipa I rastu kako se poveća broj ispitanika (Narayanan & Swaminathan, 1996; Rogers & Swaminathan, 1993; Swaminathan & Rogers, 1990).

2. CILJ ISTRAŽIVANJA

2.1. Problemi i hipoteze istraživanja

S obzirom na sve veću kompleksnost današnjih psihologijskih procjena, svrha ovog istraživanja bila je usporediti učinkovitost IRT-LR metode, MIMIC modela i logističke regresije u detekciji jednolikog i nejednolikog DIF-a kod višedimenzionalnih instrumenata s politomnim česticama. Učinkovitost navedenih metoda se određivala s obzirom na stopu pogrešaka tipa I (lažnih detekcija) i stopu odbacivanja nulte hipoteze (ispravnih detekcija), i to upotrebom simuliranih i stvarnih podataka.

U simulacijskom dijelu istraživanja provjeravani su sljedeći čimbenici koji mogu utjecati na učinkovitost metoda:

- 1) vrsta DIF-a,
- 2) odnos veličina referentne i fokalne skupine,
- 3) razlike u distribucijama sposobnosti između fokalne i referentne skupine,
- 4) veličina DIF-a i
- 5) korelacije između latentnih osobina.

Kako bi se omogućila generalizacija rezultata dobivenih na simuliranim podacima, korišteni su i empirijski dobiveni podaci.

S obzirom na preporuke iz prethodnih istraživanja o tome koji čimbenici utječu na otkrivanje DIF-a, postavljena su sljedeća istraživačka pitanja:

- 1) Koja od ove 3 metode (IRT-LR, MIMIC ili LogR) najbolje funkcionira u različitim simuliranim scenarijima po pitanju pogreške tipa I koja predstavlja stupanj neispravnih otkrivanja DIF-a?
- 2) Koja od ove 3 metode najbolje funkcionira u različitim simuliranim scenarijima po pitanju statističke snage koja predstavlja točnost detekcije DIF-a?
- 3) Koja od ove 3 metode najbolje funkcionira u realnim uvjetima, te postoji li slaganje između učinkovitosti metoda na simuliranim i realnim podacima?

Postavljene su sljedeće hipoteze:

H1: Postoji razlika među metodama s obzirom na pogrešku tipa I u različitim simuliranim scenarijima.

H2: Postoji razlika među metodama s obzirom na statističku snagu u različitim simuliranim scenarijima.

H2: Postoji slaganje u učinkovitosti dobivenoj u simuliranim i stvarnim uvjetima, odnosno, metoda koja je u simuliranim uvjetima bila najučinkovitija, bit će najučinkovitija i u realnim uvjetima.

3. METODA

3.1. Simulirani podaci

3.1.1. Monte Carlo simulacije

Kako bi se istražili učinci nekoliko složenih uvjeta povezanih s učinkovitošću metoda za detekciju višedimenzionalnog politomnog DIF-a, u prvom dijelu istraživanja korištena je metoda Monte Carlo simulacija pomoću koje su generirani računalno simulirani podaci.

Znanstvenici i psihometričari najčešće navode sljedeće razloge za upotrebu simuliranih podataka umjesto prikupljanja empirijskih podataka:

- 1) nepraktičnost i visoka cijena prikupljanja empirijskih podataka kada se manipulira s više uvjeta, odnosno nezavisnih varijabli (kao što su veličina uzorka, dužina instrumenta ili karakteristike čestica),
- 2) nemogućnost istraživanja stvarnog utjecaja nezavisnih varijabli bez poznavanja pravih osobina čestica i ispitanika (npr. parametri čestica i raspodjela latentne osobine ispitanika), te
- 3) empirijski podaci su vrlo često nepotpuni, što može utjecati na ishod istraživanja, pogotovo kada je mnogo podataka koji nedostaju, a obrazac nepotpunih podataka nije slučajaj (Brown, 2006; Bulut & Sunbul, 2017).

Osim toga, u empirijskim istraživanjima je često nemoguće isključiti utjecaj nepoželjnih varijabli na ispitanike (npr. spol, stavovi i motivacija) i čestice instrumenta (npr. sadržaj, lingvistička i kognitivna složenost) (Bulut & Sunbul, 2017).

Za razliku od realnih podataka, kod simuliranih podataka unaprijed sami određujemo raspodjelu sposobnosti/latentne osobine ispitanika i karakteristike čestica, zbog čega je moguće provjeriti stvarni utjecaj nezavisnih varijabli na zavisne varijable, te je zapravo riječ o svojevrsnom računalnom eksperimentu.

Simulacijska istraživanja se vrlo često koriste za provjeru osobina metoda za detekciju diferencijalnog funkcioniranja, s obzirom da je tim pristupom moguće modelirati realistične uvjete u kojima se mogu usporediti različite metodologije i pokazatelji na način koji empirijski podaci, u kojima parametri populacije nisu poznati, ne omogućavaju.

Monte Carlo simulacijska istraživanja obično se sastoje od sljedećih koraka:

1. Istraživač određuje skup simulacijskih faktora za koje se očekuje da će utjecati na funkcioniranje određenog statističkog postupka tj. metode. Simulacijski faktori, kao i nezavisne varijable u eksperimentu, mogu biti u potpunosti ukriženi ili djelomično ukriženi. Ako su simulacijski faktori u potpunosti ukriženi, tada je potrebno generirati skup podataka za svaku moguću kombinaciju simulacijskih faktora. Ako su, pak, djelomično ukriženi, pretpostavlja se da su samo neki simulacijski faktori u interakciji, te je potreban manji broj skupova podataka.

2. Istraživač donosi niz pretpostavki o prirodi podataka koji će biti generirani. Te pretpostavke mogu uključivati, na primjer, to da će podaci slijediti normalnu distribuciju, ili da će određeni parametri imati specifične vrijednosti, te su one ključne za autentičnost simulacijskog istraživanja jer kvaliteta rezultata simulacije ovisi o tome koliko su odabrane pretpostavke realistične.
3. Računalni program generira više skupova podataka na temelju simulacijskih faktora i pretpostavki o prirodi podataka. Proces generiranja više skupova podataka naziva se replikacija. Monte Carlo simulacije često uključuju višestruke replikacije (a) kako bi se dobila distribucija uzorkovanja procjena parametara, (b) kako bi se smanjila šansa za dobivanje nevjerojatnih rezultata iz jednog skupa podataka i (c) kako bi se imala opcija ponovnog uzorkovanja pravih parametara na temelju pretpostavke napravljene u 2. koraku.
4. Računalni program provodi statističke analize na simuliranim skupovima podataka, te izvještava procjene parametara koji su u fokusu analiza. Procjene parametara mogu biti p -vrijednosti, koeficijenti korelacije ili određeni elementi statističkog modela koje je istraživač zadao.
5. Konačno, procijenjeni parametri se evaluiraju na temelju jednog ili više kriterija - kao što su pogreška tipa I, statistička snaga, korelacija, pristranost i srednja kvadratna greška (RMSE).

Valja napomenuti da gornjih pet koraka može varirati ovisno o istraživačkim pitanjima na koje se nastoji odgovoriti.

Osim toga, prilikom dizajniranja i provođenja Monte Carlo simulacijskih istraživanja, istraživači moraju uzeti u obzir tri ključna načela: autentičnost, izvedivost i reproducibilnost tj. ponovljivost.

Autentičnost se odnosi na stupanj u kojem simulacijsko istraživanje odražava stvarne uvjete, pri čemu se naglašava važnost izbora scenarija koji zaista mogu postojati u stvarnom svijetu. Izvedivost implicira balans između ciljeva i opsega istraživanja, pri čemu se u obzir treba uzeti složenost kombinacija simulacijskih faktora i broj replikacija, kako bi istraživanje bilo dovršeno u razumnom vremenskom roku. Reproducibilnost odnosno ponovljivost se odnosi na mogućnost ponavljanja rezultata istraživanju u kasnijem trenutku od strane istog ili drugih istraživača koji imaju pristup parametrima simulacije.

Za osiguravanje reproducibilnosti, istraživač treba specificirati "sjemenku" (engl. *seed*) prije generiranja podataka i pohraniti zapis o odabranim sjemenkama. Sjemenka je početna vrijednost koja se koristi za inicijalizaciju generatora slučajnih brojeva; određivanjem sjemenke osigurava se da generirani niz slučajnih brojeva može biti ponovljen. Međutim, treba napomenuti da čak i upotreba istih sjemenki možda neće garantirati dobivanje identičnih rezultata simulacije zbog mogućih razlika u mehanizmu generatora slučajnih brojeva između različitih računala ili verzija istog softverskog programa.

3.1.2. Generiranje podataka

Simuliran je dvodimenzionalni mjerni instrument s jednostavnom strukturom i politomnim česticama. U skladu s jednostavnom strukturom, svaka čestica je bila povezana samo s jednom latentnom varijablom. Prvih 14 čestica bilo je povezano samo s prvom latentnom dimenzijom, dok je drugih 14 čestica bilo povezano samo s drugom latentnom dimenzijom.

Parametri čestica su bili slični onima korištenim u istraživanjima Jianga i sur. (2016) te Buluta i Sunbula (2017). Kao i u istraživanju Buluta i Sunbula, parametri diskriminativnosti su nasumično generirani iz uniformne distribucije $a \sim U(1.1, 2.8)$, parametar težine prve kategorije odgovora nasumično je generiran iz uniformne distribucije $b_1 \sim U(0.67, 2)$, dok su parametri težine ostale tri kategorije odgovora formirani tako da se nasumično generirana vrijednost oduzela od parametra težine prethodne kategorije ($b_2 = b_1 - U(0.67, 1.34)$, $b_3 = b_2 - U(0.67, 1.34)$, $b_4 = b_3 - U(0.67, 1.34)$).

U situacijama u kojima je simuliran jednoliki DIF, vrijednost koja je predstavljala veličinu DIF-a dodala se parametrima težine pojedinih kategorija odgovora, a u situacijama u kojima je simuliran nejednoliki DIF, vrijednost veličine DIF-a dodavana je parametru diskriminativnosti. S obzirom da je pretpostavljeno da sve DIF čestice predstavljaju prednost za referentnu grupu, ova vrijednost veličine DIF-a dodavana je samo parametrima čestica fokalne grupe.

3.1.3. Simulacijski faktori

Što se vrste DIF-a tiče, prethodna istraživanja su pokazala da neke metode bolje detektiraju određene vrste DIF-a. U vezi s ovim čimbenikom, simulirane su tri situacije: nepostojanje DIF-a, jednoliki DIF i nejednoliki DIF. U uvjetima u kojima DIF nije bio simuliran, parametri čestica bili su isti za obje skupine. U uvjetima kad je simuliran jednoliki DIF, veličina DIF-a (operacionalizirana kao razlika između parametara težine ili diskriminativnosti fokalne i referentne grupe) dodana je parametrima težine, dok je u situacijama u kojima je simuliran nejednoliki DIF veličina DIF-a dodana parametrima diskriminativnosti.

Veličina DIF-a pokazala se kao jedan od važnijih faktora koji utječu na statističku snagu metoda, kao i na pojavu pogreške tipa I (Berrio i sur., 2020). Naime, što je DIF veći, veća je i vjerojatnost njegovog otkrivanja, a manja je vjerojatnost pogreške tipa I. U ovom istraživanju varirane su tri veličine DIF-a: 0, 0.3 i 0.6 logita razlike između fokalne i referentne grupe. Ove tri veličine DIF-a predstavljale su nulti (0 logita razlike), mali (0.3 logita razlike) i srednji DIF (0.6 logita razlike). Iste veličine DIF-a korištene su za čestice s jednolikim i s nejednolikim DIF-om.

Važnost veličine uzorka za snagu otkrivanja DIF-a je očita. Kao i kod svake druge vrste istraživanja, ako uzorak za analizu nije dovoljno velik, čak i veliki efekt koji je u podacima prisutan (u ovom slučaju DIF) može proći nezamijećeno. Osim toga, čak i ako je ukupni uzorak dovoljno velik za otkrivanje efekta, ako je u grupama za usporedbu različit broj ispitanika, na nekim razinama uparujuće varijable bit će manje podataka, a posljedica toga bit će manje pouzdano uparivanje (Kristjansson i sur., 2005), koje onda može dovesti do izostanka detekcije DIF-a.

Kako bi se ovo istražilo, simulirane su dvije situacije za omjer fokalne i referentne grupe: R1500/F500 i R1000/F1000. Prvi uvjet (R1500/F500) odabran je jer je u empirijskim istraživanjima fokalna grupa obično manja od referentne grupe, dok je drugi uvjet (R1000/F1000) korišten kako bi se usporedio utjecaj uravnoteženih i neuravnoteženih veličina grupa na učinkovitost detekcije DIF-a.

U neuravnoteženom uvjetu, u fokalnoj grupi bilo je 500 simulanata, što se pokazalo kao minimalna veličina uzorka za adekvatnu procjenu parametara modela u višedimenzionalnim modelima stupnjevitih odgovora (Jiang i sur., 2016). Iste ove veličine uzorka kao i omjeri veličina grupa često su korištene i u prethodnim simulacijskim istraživanjima (Finch, 2005; Jin i sur., 2012; Lee i sur., 2016).

Kako bi se istražio utjecaj stvarnih razlika u sposobnostima, simulirana su dva različita uvjeta. U uvjetu "bez stvarnih razlika", referentna i fokalna grupa se nisu razlikovale po latentnim osobinama. U obje grupe latentne osobine su simulirane korištenjem dvodimenzionalne normalne distribucije, a pritom su obje latentne osobine imale aritmetičku sredinu 0 i varijancu 1. U drugom uvjetu, uvjetu „s prisustvom stvarne razlike u osobinama", fokalna skupina bila je inferiorna u odnosu na referentnu skupinu na obje latentne varijable, koje su sada simulirane s aritmetičkom sredinom od -0.5 u fokalnoj skupini, te 0 u referentnoj skupini. Varijanca je u obje skupine bila ista i iznosila 1.

Utjecaj korelacije između latentnih varijabli ranije nije previše istraživan. U ovom istraživanju, simulirane su dvije različite korelacije između latentnih osobina, $\rho = 0.0$ i $\rho = 0.5$, kako bi se usporedili uvjeti sa ili bez korelacije između latentnih varijabli.

Navedene nezavisne varijable su bile u potpunosti ukrižene, osim kod uvjeta bez DIF-a koji je bio isti kao i uvjet vrste DIF-a: bez DIF-a i uvjet veličine DIF-a: 0. To je rezultiralo s 40 različitih simulacijskih scenarija, koji su replicirani 100 puta za svaku uspoređivanu metodu. Ovaj broj replikacija je odabran kao najčešće korišten broj replikacija u simulacijskim istraživanjima metoda za otkrivanje DIF-a, u razdoblju između 1990. i 2018. godine (korišten u 37.5% studija prema meta-analizi Berrija i suradnika iz 2020. godine).

Sve simulacijske situacije odnosno simulacijski scenariji prikazani su u nastavku tabelarno:

Tablica 1. Prikaz simulacijskih scenarija u kojima su testirane metode detekcije DIF-a

Scen.	Vrsta DIF-a	Odnos referentne i fokalne skupine	Veličina DIF-a	Prisustvo stvarne razlike	Korelacije latentnih osobina	Evaluacijski kriterij
1	bez DIF-a	R1500/F500	0	stvarna razlika	0	pogreška tipa I
2	bez DIF-a	R1000/F1000	0	stvarna razlika	0	pogreška tipa I
3	bez DIF-a	R1500/F500	0	bez stvarne razlike	0	pogreška tipa I
4	bez DIF-a	R1000/F1000	0	bez stvarne razlike	0	pogreška tipa I
5	bez DIF-a	R1500/F500	0	stvarna razlika	0.5	pogreška tipa I
6	bez DIF-a	R1000/F1000	0	stvarna razlika	0.5	pogreška tipa I
7	bez DIF-a	R1500/F500	0	bez stvarne razlike	0.5	pogreška tipa I
8	bez DIF-a	R1000/F1000	0	bez stvarne razlike	0.5	pogreška tipa I
9	jednoliki DIF	R1500/F500	0.3	stvarna razlika	0	snaga
10	jednoliki DIF	R1000/F1000	0.3	stvarna razlika	0	snaga
11	jednoliki DIF	R1500/F500	0.3	bez stvarne razlike	0	snaga
12	jednoliki DIF	R1000/F1000	0.3	bez stvarne razlike	0	snaga
13	jednoliki DIF	R1500/F500	0.3	stvarna razlika	0.5	snaga
14	jednoliki DIF	R1000/F1000	0.3	stvarna razlika	0.5	snaga
15	jednoliki DIF	R1500/F500	0.3	bez stvarne razlike	0.5	snaga
16	jednoliki DIF	R1000/F1000	0.3	bez stvarne razlike	0.5	snaga
17	jednoliki DIF	R1500/F500	0.6	stvarna razlika	0	snaga
18	jednoliki DIF	R1000/F1000	0.6	stvarna razlika	0	snaga
19	jednoliki DIF	R1500/F500	0.6	bez stvarne razlike	0	snaga
20	jednoliki DIF	R1000/F1000	0.6	bez stvarne razlike	0	snaga
21	jednoliki DIF	R1500/F500	0.6	stvarna razlika	0.5	snaga
22	jednoliki DIF	R1000/F1000	0.6	stvarna razlika	0.5	snaga
23	jednoliki DIF	R1500/F500	0.6	bez stvarne razlike	0.5	snaga
24	jednoliki DIF	R1000/F1000	0.6	bez stvarne razlike	0.5	snaga
25	nejednoliki DIF	R1500/F500	0.3	stvarna razlika	0	snaga
26	nejednoliki DIF	R1000/F1000	0.3	stvarna razlika	0	snaga
27	nejednoliki DIF	R1500/F500	0.3	bez stvarne razlike	0	snaga
28	nejednoliki DIF	R1000/F1000	0.3	bez stvarne razlike	0	snaga
29	nejednoliki DIF	R1500/F500	0.3	stvarna razlika	0.5	snaga
30	nejednoliki DIF	R1000/F1000	0.3	stvarna razlika	0.5	snaga
31	nejednoliki DIF	R1500/F500	0.3	bez stvarne razlike	0.5	snaga
32	nejednoliki DIF	R1000/F1000	0.3	bez stvarne razlike	0.5	snaga
33	nejednoliki DIF	R1500/F500	0.6	stvarna razlika	0	snaga
34	nejednoliki DIF	R1000/F1000	0.6	stvarna razlika	0	snaga
35	nejednoliki DIF	R1500/F500	0.6	bez stvarne razlike	0	snaga
36	nejednoliki DIF	R1000/F1000	0.6	bez stvarne razlike	0	snaga
37	nejednoliki DIF	R1500/F500	0.6	stvarna razlika	0.5	snaga
38	nejednoliki DIF	R1000/F1000	0.6	stvarna razlika	0.5	snaga
39	nejednoliki DIF	R1500/F500	0.6	bez stvarne razlike	0.5	snaga
40	nejednoliki DIF	R1000/F1000	0.6	bez stvarne razlike	0.5	snaga

Kako bi se olakšala interpretacija rezultata, konstantni faktori bili su duljina instrumenta, broj kategorija odgovora i broj čestica s DIF-om. Broj čestica u svim simulacijskim scenarijima bio je 28 (paran broj čestica odabran je kako bi obje dimenzije imale jednak broj čestica) a broj kategorija odgovora 5. Slične duljine instrumenta korištene su i u prethodnim istraživanjima učinkovitosti metoda otkrivanja DIF-a (Bolt, 2002; Chun i sur., 2016; Finch, 2005). Broj čestica s DIF-om bio je 2, kao i u istraživanju koje su proveli Bulut i Suh (2017).

3.1.4. Evaluacijski kriteriji

Učinkovitost metode mjerena je pomoću dvije zavisne varijable: snage testa i stope pogreški tipa I, što su, prema već spomenutoj meta analizi, i inače najčešće korištene mjere učinkovitosti metoda u simulacijskim istraživanjima. Mjerenjem ovih dviju varijabli omogućena je usporedba kvalitete upotrijebljenih metoda u detekciji jednolikog i nejednolikog DIF-a.

Statistička snaga ili osjetljivost testa mjeri stopu točnih prepoznavanja čestica kod kojih je simuliran DIF, dok je stopa pogreške tipa I mjera pogrešnih identifikacija čestica kod kojih nije simuliran DIF, te njihovo svrstavanje među DIF čestice. Oba indikatora izračunavaju se kao broj odgovarajućih (točnih ili lažnih, ovisno o vrsti indikatora) identifikacija podijeljen s ukupnim brojem replikacija. S nominalnom stopom značajnosti od .05 očekuje se do 5% pogrešnih identifikacija, što se smatra prihvatljivim, ali kada se koristi Bradleyjev (Bradley, 1978) liberalni kriterij robusnosti, raspon prihvatljivih stopa grešaka tipa I proširuje se na vrijednosti od .025 do .075 za istu nominalnu razinu alfe. U tom slučaju, ako je stopa pogrešaka tipa I viša od .075, statistički test smatra se liberalnim, a ako je ona ispod .025, test se smatra konzervativnim. Povišena stopa pogrešaka tipa I predstavlja veći problem, jer su zaključci do kojih se putem takvih testova dolazi pogrešni i nisu valjani, dok snižena stopa pogrešaka tipa I uzrokuje samo smanjenje statističke snage otkrivanja efekata (Nordstokke i sur., 2011).

Temeljem toga, i u ovom se istraživanju stopa pogrešaka tipa I u rasponu od .025–.075 smatrala dobro kontroliranom, te je metoda koja je greške zadržala u tom rasponu smatrana robusnom u određenim uvjetima. Što se pak stopa snage tiče, korištene su granične vrijednosti od .80 odnosno .70 kao i u prethodnim simulacijskim studijama. Stope snage iznad .80 smatrane su izvrsnima, dok su stope snage u rasponu .70 - .80 smatrane umjerenima (Cohen, 1992). Također je važno imati na umu odnos između pogrešaka tipa I i stopa statističke snage. Povećane stope pogrešaka tipa I umjetno povećavaju i stope snage, pa u tim situacijama treba biti oprezan s interpretacijama (Woods, 2008).

3.1.3.1. Pogreška tipa I

Pogreška tipa I definira se kao postotni omjer slučajeva u kojima je neka čestica proglašena DIF česticom u odnosu na ukupni broj replikacija u kojima **nije** simulirano različito funkcioniranje čestica među grupama. To u stvari znači da je došlo do pogrešne detekcije, te je čestica krivo proglašena DIF česticom. Na taj način stopa pogreški tipa I predstavlja vjerojatnost odbacivanja točne nul-hipoteze o nepostojanju DIF-a.

Nekoliko istraživanja pokazalo je da je kod metoda detekcije DIF-a stopa pogrešaka tipa I visoka kada se grupe razlikuju u svojoj prosječnoj sposobnosti i kad je diskriminativnost ispitivane čestice visoka (Chang i sur., 1996; Tian, 1999; Zwick i sur., 1997).

Primjerice, Chang i sur. (1996), Spray i Miller (1994), Tian (1999), kao i Zwick i sur. (1997) otkrili su da metode detekcije DIF-a u politomnim česticama imaju vrlo niske stope pogreške tipa I u situacijama u kojima nema grupnih razlika u sposobnosti, a čak i u nekim slučajevima u kojima postoji razlika u sposobnosti.

U simulacijskim istraživanjima čestica bez DIF-a, pogreška tipa I je kod IRT-LR metode bila kontrolirana u različitim situacijama, uz uvjet da je većina sidrišnih čestica dobro odabrana, te zaista ne sadrži DIF. Stope pogrešaka tipa I IRT-LR metode bile su u blizini nominalne razine za 2PL, 3PL i stupnjevite modele (Ankenmann i sur., 1999; Cohen i sur., 1996; Kim i Cohen, 1998; Sweeney, 1997; Wang i Yeh, 2003).

Finch je pak 2005. godine proveo simulacijsko istraživanje kako bi usporedio MIMIC metodu s IRT-LR-om. U tom istraživanju, otkrio je da je MIMIC metoda vrlo dobro kontrolirala pogrešku tipa I kod duljih testova (s 50 čestica), te 2PL modela, i to neovisno o veličini fokalne grupe, prisustvu razlika u grupnim sposobnostima i količini DIF-a među sidrišnim česticama. S druge strane, imala je vrlo visoku stopu pogrešaka tipa I kod kraćih testova s 3PL modelom. U isto vrijeme, na IRT-LR je znatno utjecala količina DIF čestica među sidrišnim česticama. Finch je zaključio da je, ukoliko se očekuje velik broj DIF čestica u instrumentu, preporučljivo koristiti MIMIC metodu. Uz to, u Woodsinom istraživanju iz 2009. godine MIMIC pristup se pokazao boljim od IRT-LR metode po pitanju kontroliranja pogreške tipa I.

Općenito govoreći, pogreška tipa I je značajna odrednica učinkovitosti metode i može se koristiti za preporuku o prednosti jedne metode nad drugom u određenim eksperimentalnim (i realnim) uvjetima.

3.1.3.2. Statistička snaga

Kao i stope pogreške tipa I, tako se i stope statističke snage razlikuju ovisno o upotrijebljenoj metodi i simuliranim uvjetima. Stopa snage definira se kao proporcija DIF čestica koje su točno detektirane u odnosu na ukupan broj replikacija. Važno je pritom imati na umu da je interpretacija teoretskih stopa snage ovisna o stopama pogreški tipa I za određeni

nivo značajnosti. Razlog tome je činjenica da stopa pogreške tipa I može umjetno povećati statističku snagu. Drugim riječima, visoka snaga može biti uzrokovana inflacijom stope pogrešaka tipa I (Woods, 2008), pa se iz tog razloga ona mora interpretirati s oprezom.

U prethodnim simulacijskim istraživanjima, za izvrsnu stopu statističke snage korištena je granična vrijednost od 0.8, a proporcije između 0.7 i 0.8 smatrane su umjerenima (Cohen, 1992).

Kod IRT-LR-a, statistička snaga za detekciju jednolikog i nejednolikog DIF-a povećavala se s povećanjem veličine uzorka, diskriminativnosti čestice, broja sidrišnih čestica, i količinom DIF čestica (Ankenmann i sur., 1999; Wang i Yeh, 2003; Woods, 2009; Woods i sur., 2013). Primjerice, u istraživanju Ankenmanna i suradnika (1999), optimalni uvjeti za snagu detekcije bili su visoki parametar diskriminativnosti, veliki uzorak i iste grupne sposobnosti.

Snaga DIF metoda izravno je povezana s veličinom uzorka. Kod vrlo malih uzoraka bilo samo referentne, samo fokalne ili obje grupe, čak i čestice kod kojih postoji vrlo velika razlika u funkcioniranju mogu ostati neidentificirane. Obično se sugerira upotreba većih uzoraka kad se koriste IRT metode s dva ili tri parametra (Clauser i Mazor, 1998).

Što se pak MIMIC metode tiče, kako bi se postigli točni rezultati fokalna skupina ne bi smjela imati manje od 100 ispitanika. Iako, s obzirom da snaga raste s porastom diskriminativnosti čestica (Ankenmann i sur., 1999), manje fokalne skupine mogu biti prihvatljive ukoliko su čestice visoko diskriminativne (kao primjerice kod instrumenata iz područja psihopatologije (Rodebaugh i sur., 2004).

Prema Finchu (2005), veličina fokalne skupine značajnije utječe na snagu IRT-LR metode nego na MIMIC. Pogotovo je tome tako kada je prisutan DIF među sidrišnim česticama.

Što se omjera između referentne i fokalne skupine tiče, snaga je manja kad su uzorci nejednaki (Kristjansson i sur., 2005). To nas ne bi trebalo čuditi, s obzirom da velike razlike u veličini grupa znače da na svakoj razini ukupnog rezultata postoji relativno manje ispitanika u fokalnoj nego u referentnoj skupini koje je moguće izravno usporediti. S obzirom na to, efektivni uzorak za usporedbu je manji, a onda je i snaga niža. U istraživanju Houa i suradnika (2014), stope snage su se povećavale s porastom veličine uzorka, neovisno o drugim čimbenicima. Woods je 2009. otkrila da je potrebna veličina uzorka za adekvatnu snagu i relativno točne procjene većine parametara čestica manja kod MIMIC modela nego kod IRT-LR metode. Unatoč tome, IRT-LR metoda je uvijek imala veću snagu u detekciji nejednolikog DIF-a.

Duljina instrumenta vjerojatno nema veliki utjecaj na statističku snagu metoda ili točnost procjene parametara čestice, no duži instrumenti mogu dati točnije procjene razlika u grupnim sposobnostima (Woods, 2009). U nekim se pak situacijama MIMIC metoda pokazala snažnijom od IRT-LR metode kod duljih testova (Finch, 2005). Slično su potvrdili i Wang i

Shih (2010). Naime, oni su ustanovili da MIMIC ima jednaku ili veću snagu otkrivanja DIF-a kod duljih instrumenata (50 čestica), ali nije tako učinkovit kod kraćih instrumenata (20 čestica).

Veličina DIF-a također ima utjecaj na stope snage. U već spomenutom istraživanju Houa i suradnika (2014), veći DIF odgovarao je većim stopama snage otkrivanja jednolikog DIF-a, neovisno o veličini uzorka.

3.2. Empirijski podaci

3.2.1. Izvor podataka

U dijelu istraživanja u kojem su korišteni empirijski podaci riječ je bilo o sekundarnim podacima prikupljenih Upitnikom emocionalne kompetentnosti, UEK-45, koji su dobiveni od samog autora instrumenta, Vladimira Takšića.

Tablica 2. Broj ispitanika u pojedinim državama

Država	M	Ž	NA	N	%
Hrvatska	235	522	1	758	14.9
Portugal	157	192	0	349	6.9
Finska	54	174	0	228	4.5
Švedska	79	133	0	212	4.2
Slovenija	171	389	0	560	11.0
Španjolska	13	106	0	119	2.3
Japan	132	483	0	615	12.1
Kina	247	637	0	884	17.4
Indija	181	223	0	404	8.0
Argentina	95	208	0	303	6.0
Srbija	64	162	0	226	4.4
SAD	37	55	1	93	1.8
Turska	229	101	0	330	6.5
Ukupno	1694	3385	2	5081	100

Baza podataka objedinjavala je odgovore ispitanika dobivene u različitim istraživanjima u kojima je korišten upitnik UEK-45, provedenih u 13 zemalja: Hrvatskoj, Portugalu, Finskoj, Švedskoj, Sloveniji, Španjolskoj, Srbiji, Japanu, Kini, Indiji, Argentini, Turskoj i Sjedinjenim Američkim Državama. Upitnik je ispunio 5081 ispitanik (3385 ženskog i 1694 muškog spola). Nacionalna i spolna raspodjela prikazana je u Tablici 1. Ženske ispitanice činile su dvije trećine, a muški ispitanici jednu trećinu uzorka. Najmanji uzorak bio je prikupljen u SAD-u (93 ispitanika), a najveći u Kini (615 ispitanika).

3.2.2. Instrument

Upitnik emocionalne kompetentnosti (UEK-45) (Takšić, 2002) mjera je emocionalne inteligencije kao osobine ličnosti. Konstruiran je na uzorku hrvatskih srednjoškolaca produkcijom čestica u okviru Mayer-Salovey modela (Mayer i Salovey, 1997) s četiri grane i četiri razvojna stupnja seta emocionalnih sposobnosti. Tako je u prvoj verziji dobiveno 16 različitih skala s ukupno 136 čestica (Takšić, 1998). Kasnije je Upitnik skraćen na tri skale s ukupno 45 čestica (Takšić, 2002).

Skala sposobnosti uočavanja i razumijevanja emocija (UR) sastoji se od 15 čestica (primjer čestice: „Kada sretnem poznanika, odmah shvatim kako je raspoložen.“), Skala sposobnosti izražavanja i imenovanja emocija (II) sastoji se od 14 čestica (primjer čestice: „Gotovo uvijek mogu riječima opisati svoje osjećaje i emocije.“), a Skala sposobnosti regulacije i upravljanja emocijama (RU) sastoji se od 16 čestica (primjer čestice: „Dobro raspoloženje mogu zadržati i ako mi se nešto loše dogodi.“).

UEK-45 preveden je prema načelima dvostrukog prijevoda (engl. *back-translation*) (van de Vijver i Hambleton, 1996) i primijenjen u različitim državama. Zadatak ispitanika prilikom ispunjavanja upitnika tipa papir-olovka je procijeniti koliko se pojedina tvrdnja sadržana u upitniku odnosi na njega odgovarajući na skali od 1 („uopće NE“) do 5 („u potpunosti DA“).

Skale UEK-45 pokazale su zadovoljavajuću pouzdanost u različitim uzorcima ispitanika (Takšić i sur., 2009). Korelacije među skalama UEK se kreću u rasponu od .35 do .51 i omogućuju formiranje ukupnog rezultata, kao mjere opće emocionalne kompetentnosti.

Kraća verzija upitnika koja sadrži 27 čestica odabrana je radi pojednostavljivanja nacrtu istraživanja, te s obzirom da je u istraživanju Takšića i suradnika (2013) imala najveću kroskulturalnu ekvivalentnost.

4. REZULTATI

4.1. Rezultati simulacijskog istraživanja

4.1.1. Rezultati pogreške tipa I

U ovom dijelu bit će opisani rezultati analiza pomoću kojih se pokušao dati odgovor na prvo istraživačko pitanje, koje je glasilo:

Koja od ove 3 metode (IRT-LR, MIMIC ili LogR) najbolje funkcionira u različitim simuliranim scenarijima po pitanju pogreške tipa I koja predstavlja stupanj neispravnih otkrivanja DIF-a?

S obzirom da su u ovim analizama promatrane samo situacije bez simuliranog DIF-a, vrsta DIF-a i veličina DIF-a su bile konstantne varijable, a manipulirane su razine samo preostale 3 nezavisne varijable: prisustvo/odsustvo stvarnih grupnih razlika, odnos veličina referentne i fokalne skupine i korelacija među latentnim osobinama. S obzirom na to, postojalo je 8 različitih situacija (2*2*2).

Tablica 3. Stope lažnih detekcija (pogreška tipa I) u svim simuliranim uvjetima

Uvjet	ρ	R1500/F500				R1000/F1000			
		IRT-LR	LR-R	LR-T	MIMIC	IRT-LR	LR-R	LR-T	MIMIC
Grupna razlika	.00	.040	.060	.070	.170	.065	.075	.055	.200
	.50	.045	.085	.070	.140	.035	.090	.045	.275
	M	.043	.073	.070	.155	.050	.083	.050	.238
Bez grupne razlike	.00	.020	.040	.035	.070	.085	.040	.040	.045
	.50	.045	.045	.040	.060	.045	.040	.070	.045
	M	.033	.043	.038	.065	.065	.040	.055	.045
Prosjek		.038	.058	.054	.110	.058	.061	.053	.141

Napomena. R1500/F500 = uvjet s neuravnoteženim veličinama grupa; R1000/F1000 = uvjet s uravnoteženim veličinama grupa

Radi lakše usporedbe, grafički je prikazan prosjek rezultata stopi pogreški tipa I na obje analizirane čestice (čestica 1 i čestica 28). Na grafičkom prikazu vidljive su i granične vrijednosti Bradleyevog kriterija liberalne robusnosti ($0.025 < \text{pogreška tipa I} < 0.075$).

MIMIC metoda je u svim kombinacijama u kojima je bila prisutna stvarna razlika među osobinama, neovisno o druge dvije nezavisne varijable, imala stopu pogreški tipa I iznad gornje vrijednosti Bradleyevog kriterija liberalne robusnosti, te je dobro kontrolirala pogrešku tipa I samo u situacijama odsustva stvarne razlike među grupama.

Osim analize deskriptivnih pokazatelja i vizualne inspekcije grafova, napravljena je analiza varijance s mješovitim nacrtom u kojoj je metoda bila faktor unutar subjekata, a prethodno navedene tri nezavisne varijable bile su faktori između subjekata. Mauchlyjev test nije bio značajan, pa je s obzirom na to pretpostavka sferičnosti bila ispunjena.

Četverosmjerna interakcija nije bila značajna, ali dobivene su dvije značajne trosmjerne interakcije, trosmjerna interakcija metode, prisustva stvarne razlike među skupinama i odnosa veličina skupina ($F(1.938, 15.502)=10.715, p<.01, np^2=.573$), te trosmjerna interakcija metode, odnosa veličina skupina i korelacije među latentnim osobinama ($F(1.938, 15.502)=5.271, p<.05, np^2=.397$).

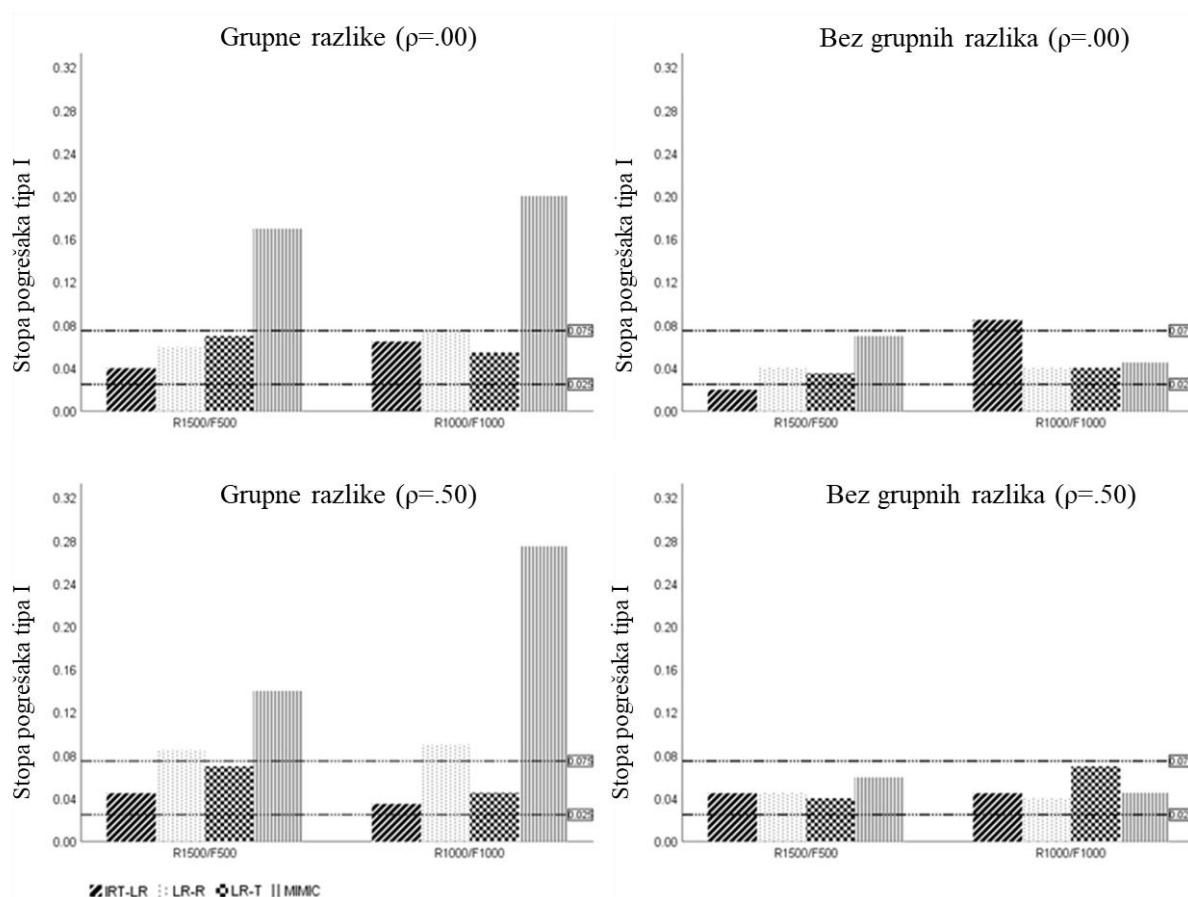
I u situaciji s neuravnoteženim i u situaciji s uravnoteženim odnosom veličina referentne i fokalne skupine dvosmjerna interakcija metode i prisustva stvarne razlike među skupinama bila je statistički značajna ($F(1.901, 11.405)=8.115, p<.01, np^2=.575$, odnosno $F(3, 18)=22.688, p<.001, np^2=.791$). Stoga su dalje provjeravani jednostavni glavni efekti obje varijable.

I u situaciji s neuravnoteženim i u situaciji s uravnoteženim odnosom veličina referentne i fokalne skupine, uz prisustvo stvarne razlike među grupama, MIMIC metoda imala je statistički značajno više pogrešaka tipa I od svih ostalih metoda, a stopa pogrešaka tipa I nije se značajno razlikovala između preostale tri metode. U odsustvu pak grupnih razlika metode se međusobno nisu značajno razlikovale po pitanju broja pogrešaka tipa I.

Što se jednostavnog glavnog efekta prisustva/odsustva stvarne razlike među grupama tiče, u situaciji s neravnotežom veličina grupa samo je MIMIC metoda imala značajno više pogrešaka tipa I u prisustvu stvarne razlike među grupama nego u njenom odsustvu. U situaciji uravnoteženih grupa, osim MIMIC metode i logistička regresija s bruto rezultatima imala je značajno više pogrešaka u prisustvu grupnih razlika.

Također, MIMIC metoda je jedina metoda kod koje se broj pogrešaka tipa I značajno razlikovao ovisno o tome jesu li grupe bile jednake ili različite veličine. Kad su grupe bile jednake veličine MIMIC metoda je imala značajno više pogrešaka tipa I, nego kad su grupe bile različite veličine. Isto tako, MIMIC metoda je jedina kod koje je broj pogrešaka tipa I ovisio o korelaciji među latentnim osobinama, ali samo u situaciji s uravnoteženim omjerom veličina grupa, i to na način da je više pogrešaka tipa I bilo u situaciji u kojoj nije bilo korelacije među latentnim osobinama.

Slika 18. Prikaz stopa pogrešaka tipa I



Napomena: Referentne linije upućuju na Bradleyjev kriterij liberalne robusnosti.

Općenito govoreći, obje vrste logističke regresije i IRT-LR metoda dobro su kontrolirale pogrešku tipa I, a interakcijska MIMIC metoda imala je povišene stope lažnih detekcija u situacijama u kojima je postojala stvarna grupna razlika u sposobnostima. Logistička regresija s bruto rezultatima imala je značajno više pogrešaka tipa I u prisustvu grupnih razlika, ali samo u uvjetu s uravnoteženim grupama. Interakcijska MIMIC metoda imala je značajno više pogrešaka tipa I u uvjetu s uravnoteženim grupama i nekoreliranim latentnim osobinama.

4.1.2. Rezultati statističke snage

U ovom dijelu navedeni su rezultati analiza pomoću kojih se pokušalo odgovoriti na drugo istraživačko pitanje, a koje je glasilo:

Koja od ove 3 metode najbolje funkcionira u različitim simuliranim scenarijima po pitanju statističke snage koja predstavlja točnost detekcije DIF-a?

U donjim tablicama navedene su stope odbacivanja nulte hipoteze odnosno stope točne detekcije DIF-a za sve četiri uspoređivane metode, i to zasebno za jednoliki i zasebno za nejednoliki DIF. Stope točne detekcije su prosječne stope dobivene na obje ispitivane čestice u kojima je simuliran DIF (čestica 1 i čestica 28).

Tablica 4. Stope snage (točnih identifikacija) u uvjetima s jednolikim DIF-om

JEDNOLIKI DIF	ρ	R1500/F500				R1000/F1000			
		IRT-LR	LR-R	LR-T	MIMIC	IRT-LR	LR-R	LR-T	MIMIC
niski DIF, bez grupne razlike	.00	.575	.780	.770	.730	.740	.860	.880	.865
	.50	.580	.765	.780	.780	.720	.865	.875	.870
	M	.578	.773	.775	.755	.730	.863	.878	.868
niski DIF, grupna razlika	.00	.535	.690	.760	.590	.680	.780	.855	.725
	.50	.520	.675	.750	.575	.690	.850	.910	.760
	M	.528	.683	.755	.583	.685	.815	.883	.743
NISKI DIF		.553	.728	.765	.669	.708	.839	.880	.805
umjereni DIF, bez grupne razlike	.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	.50	.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	M	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
umjereni DIF, grupna razlika	.00	.990	.995	.995	.995	1.000	1.000	1.000	1.000
	.50	.995	1.000	1.000	.990	1.000	1.000	1.000	1.000
	M	.993	.998	.998	.993	1.000	1.000	1.000	1.000
UMJERENI DIF		.995	.999	.999	.996	1.000	1.000	1.000	1.000

Napomena. R1500/F500 = uvjet s neuravnoteženim veličinama grupa; R1000/F1000 = uvjet s uravnoteženim veličinama grupa

Jednoliki DIF

Kao granična vrijednost za stopu točnih detekcija korištena je vrijednost od 0.80 za izvrsnu statističku snagu i 0.70 za umjerenu. S obzirom na to, najboljom u detekciji jednolikog DIF-a se pokazala logistička regresija s procjenom latentne vrijednosti, koja je u svim situacijama s uravnoteženim odnosom veličina skupina imala stopu točnih identifikacija DIF-a višu od granične vrijednosti. Vrlo slično funkcionirala je i logistička regresija s bruto rezultatima.

Sve su metode bile učinkovite u detekciji umjerenog jednolikog DIF-a, uz (gotovo) 100%-tnu uspješnost. Otkrivanje niskog jednolikog DIF-a je bilo prilično lošije, posebno u situacijama s neuravnoteženom veličinom skupina, u kojima nijedna metoda nije dosegla zadovoljavajuću razinu detekcije.

Kod uravnoteženih veličina skupina najslabije je funkcionirala IRT-LR metoda sa stopama detekcije od 0.68 do 0.74. I inače je ova metoda imala niske stope snage gotovo u svim situacijama s nejednolikim DIF-om, a funkcionirala je nešto bolje, ali ipak nedovoljno dobro, samo u situacijama s uravnoteženim grupama i bez prisustva grupne razlike u sposobnostima.

Logistička regresija s bruto rezultatima imala je odlične stope snage gotovo u svim uvjetima s uravnoteženim omjerom grupa. S druge strane, kad su veličine grupa bile različite, na snagu je utjecalo prisustvo grupnih razlika u sposobnosti, u čijem prisustvu je ova metoda imala neprihvatljive stope snage.

Interakcijska MIMIC metoda imala je isti obrazac stopa snage, samo uz nešto niže vrijednosti. U uvjetima jednakih grupa funkcionirala je izvrsno kada nije bila simulirana razlika u grupnim sposobnostima, a umjereno dobro kada je ona simulirana. U uvjetima nejednakih grupa funkcionirala je umjereno dobro bez prisustva razlika u sposobnosti, ali je imala neprihvatljive stope snage kad su razlike bile prisutne.

Logistička regresija s latentnom procjenom nadmašila je ostale metode u svim simuliranim uvjetima. Imala je izvrsne stope snage u uvjetima s jednakim grupama, a funkcionirala je umjereno dobro kada su grupe bile nejednake veličine.

Korelacija među latentnim osobinama nije pokazala jednoobrazan utjecaj na učinkovitost metoda po pitanju stopa snage detekcije DIF-a.

Nejednoliki DIF

Što se detekcije nejednolikog DIF-a tiče, stope točne identifikacije su generalno niže nego stope detekcije jednolikog DIF-a. Nijedna metoda nije postigla zadovoljavajuću razinu snage otkrivanja niskog nejednolikog DIF-a.

Kod te vrste DIF-a najvišu stopu detekcije postigla je logistička regresija s bruto rezultatima u situaciji s uravnoteženim odnosom grupa, prisutnom grupnom razlikom te

korelacijom među latentnim osobinama (0.65). No, s obzirom da je u ovim uvjetima logistička regresija s bruto rezultatima imala i povišene stope pogrešaka tipa I, ovaj nalaz treba uzeti s rezervom.

Najnižu stopa detekcije imala je IRT-LR metoda u uvjetu s nejednakim grupama, korelacijom među latentnim osobinama i grupnom razlikom u sposobnostima (0.365).

Tablica 5. Stope snage (točnih identifikacija) u uvjetima s nejednolikim DIF-om

NEJEDNOLIKI DIF	ρ	R1500/F500				R1000/F1000			
		IRT-LR	LR-R	LR-T	MIMIC	IRT-LR	LR-R	LR-T	MIMIC
niski DIF, bez grupne razlike	.00	.380	.400	.435	.540	.465	.515	.545	.430
	.50	.370	.355	.385	.455	.420	.475	.525	.545
	M	.375	.378	.410	.498	.443	.495	.535	.488
niski DIF, grupna razlika	.00	.375	.530	.455	.430	.485	.635	.535	.430
	.50	.365	.515	.430	.410	.445	.650	.545	.485
	M	.370	.523	.443	.420	.465	.643	.540	.458
NISKI DIF		.373	.450	.426	.459	.454	.569	.538	.473
umjereni DIF, bez grupne razlike	.00	.835	.830	.855	.895	.895	.940	.950	.910
	.50	.820	.845	.865	.895	.870	.890	.895	.930
	M	.828	.838	.860	.895	.883	.915	.923	.920
umjereni DIF, grupna razlika	.00	.810	.880	.885	.695	.920	.925	.860	.745
	.50	.780	.910	.855	.705	.920	.955	.950	.740
	M	.795	.895	.870	.700	.920	.940	.905	.743
UMJERENI DIF		.811	.866	.865	.798	.901	.928	.914	.831

Napomena. R1500/F500 = uvjet s neuravnoteženim veličinama grupa; R1000/F1000 = uvjet s uravnoteženim veličinama grupa

Što se pak umjerenog nejednolikog DIF-a tiče, sve metode su funkcionirale značajno bolje, s tim što je najviše detekcija ponovno imala logistička regresija s procjenom latentne osobine, a najslabije je funkcionirala MIMIC metoda i to u situacijama s prisutnom stvarnom razlikom među grupama. No, i u tim situacijama stope su bile umjereno dobre.

Generalno govoreći, s povećanjem veličine DIF-a od niskog do umjerenog došlo je do znatnog povećanja učinkovitosti svih metoda.

Nisu uočene velike razlike u stopi detekcije s obzirom na prisustvo ili odsustvo korelacije među latentnim osobinama.

Učinci simulacijskih čimbenika

Statistička značajnost uočenih razlika na deskriptivnoj razini provjerena je pomoću šestosmjerne analize varijance s mješovitim nacrtom. Parcijalna kvadrirana eta korištena je kao mjera veličine učinka svake varijable.

S obzirom da je Mauchly's W ($\chi^2(5)=110.435$, $p<.001$) bio statistički značajan, zbog narušene pretpostavke sferičnosti u interpretaciji je upotrijebljena p -vrijednost dobivena pomoću Greenhouse-Geiserove korekcije stupnjeva slobode.

Dobivene su dvije značajne četverosmjerne interakcije velikog efekta između sljedećih nezavisnih varijabli: 1) metode, vrste DIF-a, veličine DIF-a i prisustva/odsustva prave razlike među skupinama ($F(2.059, 65.899)=12.150$, $p<.001$, $np^2=.275$); te 2) metode, vrste DIF-a, veličine DIF-a i uravnoteženosti veličina referentne i fokalne skupine ($F(2.059,65.899)=3.277$, $p<.05$, $np^2=.093$). Korelacija među latentnim osobinama nije se pokazala značajnim faktorom pri razlikovanju učinkovitosti pojedinih metoda, a nije utjecala niti na broj točnih identifikacija općenito.

Neovisno o veličini DIF-a i prisustvu grupnih razlika u sposobnosti, sve su metode bolje detektirale jednoliki nego nejednoliki DIF. Također, sve metode su bile konzistentno bolje u detekciji umjerenog nego niskog DIF-a.

Stvarne razlike u sposobnosti

U Tablici 6. sažeti su rezultati prve četverosmjerne interakcije (metoda X vrsta DIF-a X veličina DIF-a X stvarne razlike u sposobnosti).

Kada su istovremeno bile prisutne grupne razlike u sposobnosti i niski DIF, logistička regresija s latentnom procjenom bila je statistički značajno bolja od ostalih metoda u identificiranju jednolikog DIF-a, a logistička regresija s bruto rezultatima u identificiranju nejednolikog DIF-a. Najniže stope snage i s jednolikim i nejednolikim DIF-om imala je IRT-LR metoda, čak i u odsustvu grupnih razlika u sposobnosti.

U uvjetima umjerenog DIF-a, metode su bile podjednako dobre u identificiranju jednolikog DIF-a, i to neovisno o prisustvu stvarnih razlika u sposobnosti. S druge strane, kad je simuliran nejednoliki DIF, interakcijska MIMIC metoda imala je poteškoće u otkrivanju DIF-a u situacijama s prisutnom stvarnom razlikom u sposobnostima. U tim je uvjetima najbolje po pitanju snage funkcionirala logistička regresija s bruto rezultatima.

Tablica 6. Usporedba učinkovitosti metoda detekcije DIF-a u ovisnosti o stvarnim grupnim razlikama

Vrsta i veličina DIF-a	Stvarna grupna razlika	Bez stvarne grupne razlike
niski jednoliki DIF	LR-T > LR-R > MIMIC > IRT-LR	IRT-LR < (LR-R = LR-T = MIMIC)
umjereni jednoliki DIF	nema razlike u učinkovitosti metoda	nema razlike u učinkovitosti metoda
niski nejednoliki DIF	LR-R > LR-T, LR-T=MIMIC, MIMIC > IRT-LR	IRT-LR < (LR-R = LR-T = MIMIC)
umjereni nejednoliki DIF	LR-T > MIMIC, MIMIC = LR-R, LR-R > IRT-LR	IRT-LR < (LR-R = LR-T = MIMIC)

Napomena. Tablica predstavlja sažetak jednostavnih glavnih efekata metode u četverosmjernoj ANOVA-i između metode, vrste DIF-a, veličine DIF-a i grupne razlike. Učinkovitost metoda je uspoređena upotrebom simbola <, > i =.

Dakle, prisustvo grupnih razlika u sposobnostima najviše je utjecalo na MIMIC metodu, koja je u prisustvu stvarnih razlika imala statistički značajno manje točnih identifikacija. Ostale metode su bile relativno robusne po pitanju ovog faktora.

Omjer veličina grupa

Tablica 7. sažima rezultate druge četverosmjerne interakcije (metoda X vrsta DIF-a X veličina DIF-a X omjer veličina grupa).

Sve metode su imale podjednako dobre stope snage kad su veličine grupa bile jednake. U uvjetima s nejednakim grupama, učinkovitost je ovisila o ostala dva čimbenika. U otkrivanju niskog jednolikog DIF-a najbolje je funkcionirala logistička regresija s latentnom procjenom, a najlošije IRT-LR metoda. Kad umjerenog jednolikog DIF-a nije bilo razlike među metodama. Kod nejednolikog DIF-a najlošije je funkcionirala IRT-LR metoda i to neovisno o njegovoj veličini.

Tablica 7. Usporedba učinkovitosti metoda detekcije DIF-a u ovisnosti o omjeru veličina grupa

Vrsta i veličina DIF-a	R1000/F1000	R1500/F500
niski jednoliki DIF	nema razlike u učinkovitosti metoda	IRT-LR < LR-R, LR-R = MIMIC, MIMIC < LR-T
umjereni jednoliki DIF	nema razlike u učinkovitosti metoda	nema razlike u učinkovitosti metoda
niski nejednoliki DIF	nema razlike u učinkovitosti metoda	IRT-LR < LR-R, LR-R = MIMIC, MIMIC < LR-T
umjereni nejednoliki DIF	nema razlike u učinkovitosti metoda	IRT-LR < (LR-R = MIMIC = LR-T)

Napomena. Tablica predstavlja sažetak jednostavnih glavnih efekata metode u četverosmjernoj ANOVA-i između metode, vrste DIF-a, veličine DIF-a i omjera veličine grupa. Učinkovitost metoda je uspoređena upotrebom simbola <, > i =.

Odabir najučinkovitije metode

S obzirom na prethodno opisane analize stopi pogrešaka tipa I i stopi točnih identifikacija DIF-a, najučinkovitijom se po pitanju oba kriterija pokazala logistička regresija s latentnom procjenom osobina.

4.2. Rezultati DIF-analiza na realnim podacima

U ovom dijelu istraživanja provjeravala se učinkovitost ispitivanih metoda u otkrivanju različitog funkcioniranja čestica s obzirom na varijablu spola, kako bi se pokušalo odgovoriti na treće istraživačko pitanje, a koje je glasilo:

Koja od ove 3 metode najbolje funkcionira u realnim uvjetima, te postoji li slaganje između učinkovitosti metoda u simuliranim i realnim uvjetima?

4.2.1. Mjerna invarijantnost Upitnika emocionalne kompetentnosti UEK-27

Preduvjet za provođenje analiza dimenzionalnosti Upitnika emocionalne kompetentnosti UEK-27, kao i analiza diferencijalnog funkcioniranja čestica za muške i ženske ispitanike na kros-kulturalnom uzorku bilo je pronalazak mjerno invarijantnog uzorka s obzirom na državu.

Za provjeru mjerne invarijantnosti upitnika korištena je nova metoda iz područja mrežne psihometrije – eksploratorna grafička analiza (EGA) koja omogućava provjeru dimenzionalnosti psihologijskih konstrukata (Golino & Epskamp, 2017). Za ovu analizu korišten je paket *EGAnet* (Golino & Christensen, 2019), verzija 2.0.0. Eksploratorna grafička analiza provedena je na svih 13 zemalja zasebno, a rezultati su prikazani na Slici 7. u Pravitku 4.

Kada se primjenjuje EGA, smatra se da je razina konfiguralne invarijantnosti postignuta ukoliko se isti čvorovi (čestice) nalaze u istim zajednicama u svim skupinama koje se uspoređuju. Naime, za razliku od provjere mjerne invarijantnosti u okviru konfirmatorne faktorske analize, ovaj pristup ne dopušta postojanje bilo koje razine neslaganja u konfiguracijama (tj. ne dozvoljava se parcijalna konfiguralna invarijantnost) prije nego što se prijeđe na provjeru metričke invarijantnosti (Jamison i sur., 2022). Jamison i sur. (2022) stoga preporučuju provođenje bootEGA analize na cjelokupnom uzorku, provjeru razine stabilnosti čestica, te eliminaciju onih koji imaju tu vrijednost nižu od .70. Zatim se provodi zasebna EGA analiza na svakoj od komparativnih grupa.

S obzirom da u ovom slučaju eliminacija čestica nije bila opcija, nego je cilj bio zadržati što veći broj originalnih čestica, a smanjenje veličine ukupnog uzorka, s obzirom na početnu veličinu, jest bila opcija, bootEGA i analiza stabilnosti čestica nisu provedene, nego se odmah pristupilo usporedbi EGA rješenja pojedinih zemalja. Zatim su eliminirane one zemlje u kojima nije dobivena teorijski očekivana konfiguracija.

Riječ je o sljedećim zemljama: Hrvatska, Portugal, Finska, Švedska, Japan, Indija, Srbija, SAD i Turska. Dakle, konfiguralnu invarijantnost iskazale su samo Slovenija, Španjolska, Kina i Argentina. Metrička invarijantnost je stoga provjerena samo na uzorcima iz te četiri zemlje. Metrička invarijantnost se u okviru eksploratorne grafičke analize dokazuje tako da se ekvivalentnost mrežnih opterećenja provjerava pomoću permutacijskih testova. Glavna prednost testiranja permutacija u odnosu na tradicionalno testiranje hipoteza je znatno manje pretpostavki o distribuciji varijable u populaciji (Chihara i Hesterberg, 2018).

S obzirom da se višestrukim usporedbama povećava vjerojatnost javljanja pogreške tipa I, preporuka je da se dobivene p -vrijednosti prilagode pomoću Benjamini-Hochbergovog postupka (BH-postupak) (Benjamini i Hochberg, 1995). Prije provedenog BH postupka metričku neekvivalentnost iskazivale su tri čestice sa Subskale sposobnosti izražavanja i imenovanja emocija, čestica II2 („Kada me netko pohvali (nagradi), radim s više elana.“), čestica II4 („Kada sam dobrog raspoloženja svaki mi se problem čini lako rješiv.“) i čestica II7 („Većinu svojih osjećaja mogu prepoznati.“).

Tablica 8. Rezultati eksplorativne grafičke analize UEK-27

Čvor	Zajednica	Razlika	<i>p</i> - vrijednost	B-H prilagođena <i>p</i> -vrijednost
Kada vidim kako se netko osjeća, obično znam što mu se dogodilo. (UR1)	1	-.014	.840	.986
Kod prijatelja mogu razlikovati kada je tužan, a kada razočaran. (UR2)	1	.045	.296	.666
Mogu točno odrediti emocije neke osobe ako je promatram u društvu. (UR3)	1	.010	.840	.986
Mogu zapaziti kada se netko osjeća bespomoćno. (UR4)	1	-.099	.054	.364
Prema izrazu lica mogu prepoznati nečija osjećanja. (UR5)	1	-.057	.372	.688
Prepoznajem prikrivenu ljubomoru kod svojih prijatelja. (UR6)	1	-.067	.138	.621
Primijetim kada netko nastoji prikriti loše raspoloženje. (UR7)	1	-.012	.834	.986
Primijetim kada netko osjeća krivnju. (UR8)	1	-.004	.956	.998
Primijetim kada netko prikriva svoje prave osjećaje. (UR9)	1	-.05	.382	.688
Primijetim kada se netko osjeća potišteno. (UR10)	1	.091	.108	.583
Uočim kada se netko ponaša različito od onoga kako je raspoložen. (UR11)	1	-.025	.624	.986
Iz neugodnih iskustava učim kako se ubuduće ne treba ponašati. (RU1)	2	-.006	.948	.998
Kada me netko pohvali (nagradi), radim s više elana. (RU2)	2	.119	.162	.624
Kada sam dobro raspoložen, teško me je oneraspoložiti. (RU3)	2	.042	.522	.880
Kada sam dobrog raspoloženja svaki mi se problem čini lako rješiv. (RU4)	2	-.001	.998	.998
Kada se osjećam sretno i raspoloženo, tada najbolje učim i pamtim. (RU5)	2	.017	.816	.986
Nastojim ublažiti neugodne emocije, a pojačati pozitivne. (RU6)	2	.024	.720	.986

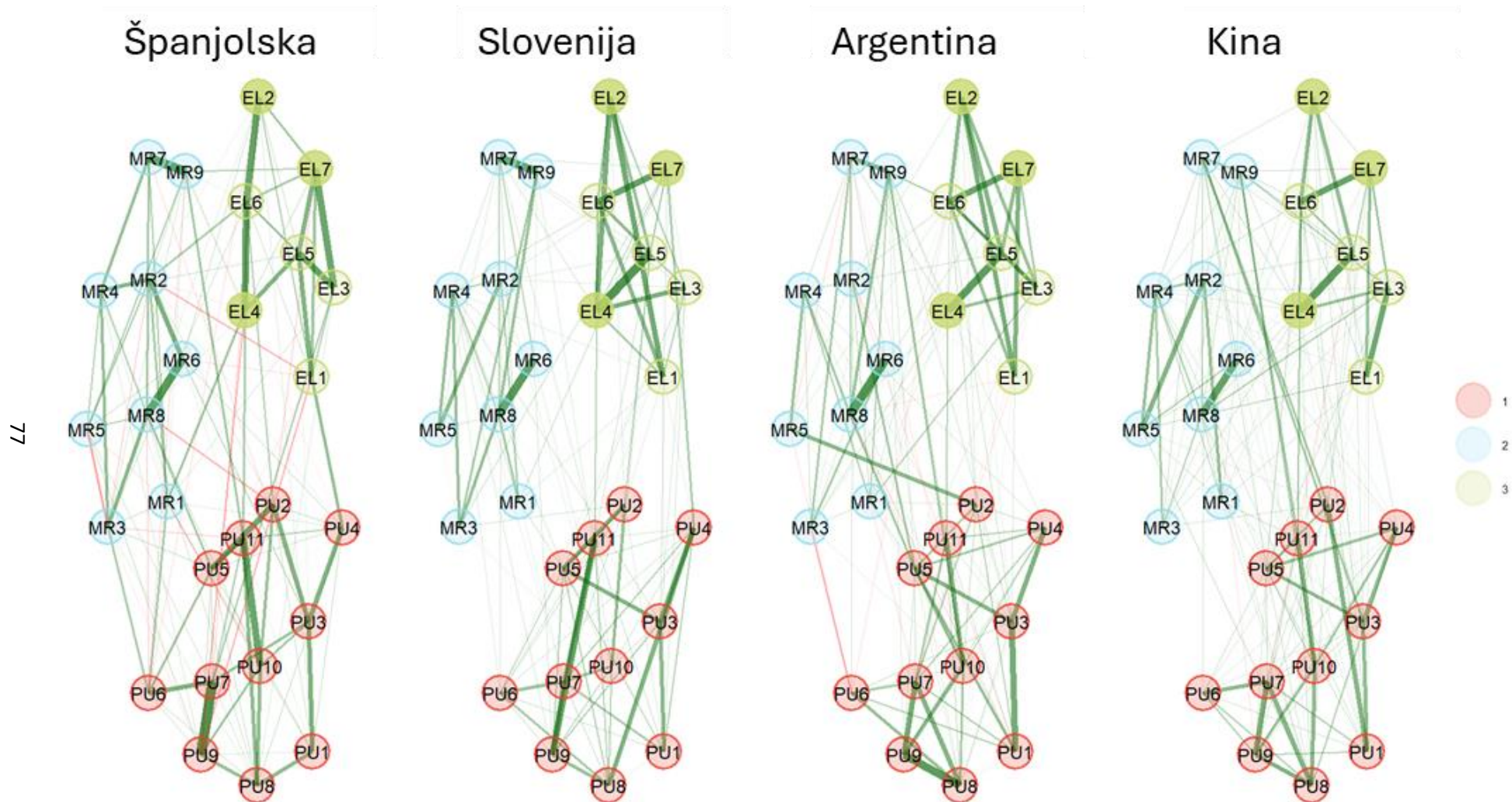
Nema ništa loše u tome kako se obično osjećam. (RU7)	2	.019	.826	.986
Pokušavam održati dobro raspoloženje. (RU8)	2	-.108	.254	.666
Što se mene tiče, potpuno je u redu ovako se osjećati. (RU9)	2	-.071	.290	.666
Gotovo uvijek mogu riječima opisati svoje osjećaje i emocije. (II1)	3	-.046	.380	.688
Lako mogu nabrojiti emocije koje trenutno doživim. (II2)	3	-.126	.016*	.324
Mogu dobro izraziti svoje emocije. (II3)	3	.087	.204	.666
Mogu opisati kako se osjećam. (II4)	3	-.241	.002**	.054
Mogu opisati svoje sadašnje stanje. (II5)	3	.076	.244	.666
Većinu svojih osjećaja lako mogu imenovati. (II6)	3	-.001	.992	.998
Većinu svojih osjećaja mogu prepoznati. (II7)	3	.160	.006**	.081

*Napomena: * $p < .05$; ** $p < .01$*

Usporedba mrežnih opterećenja u navedene četiri zemlje uz primjenu BH postupka pokazala je da su sve čestice UEK-27 metrički invarijantne (Slika 20.).

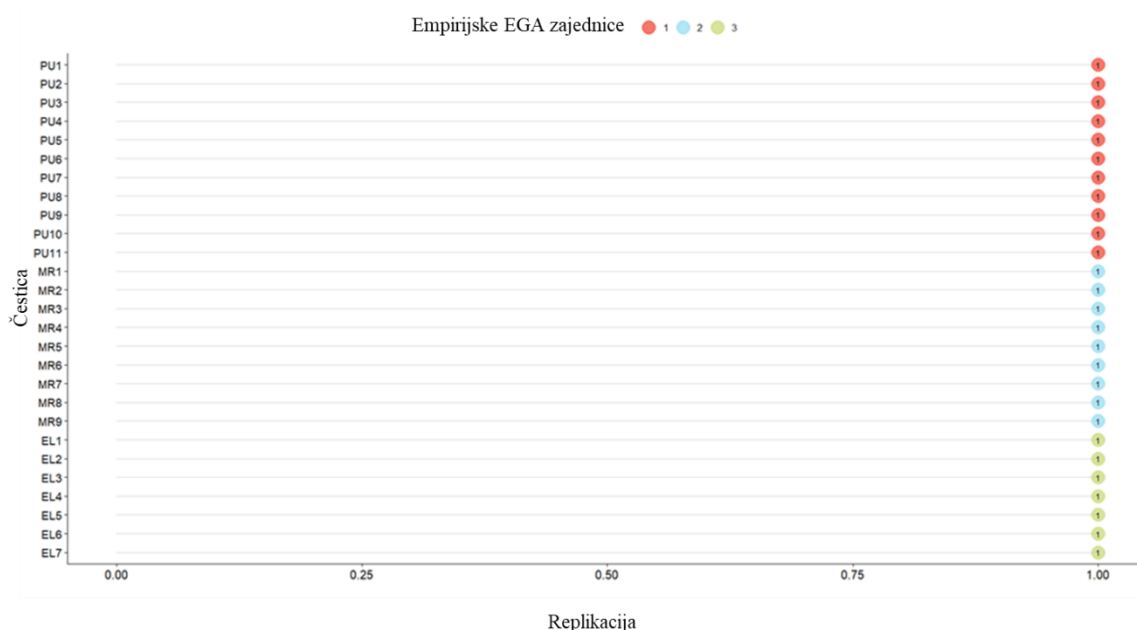
Provjerena je i stabilnost trofaktorskog rješenja na ovom mjerno-invarijantnom uzorku pomoću bootEGA analize s 500 replikacija. Dobiveno je savršeno stabilno rješenje u svih 500 replikacija (Slika 21.).

Slika 19. Prikaz metrički invarijantnih uzoraka dobiven metodom eksploratorne grafičke analize



Napomena: 1 = Skala sposobnosti uočavanja i razumijevanja emocija (UR); 2 = Skala sposobnosti regulacija i upravljanja emocijama (RU); 3= Skala sposobnosti izražavanja i imenovanja emocija (II)

Slika 20. Rezultati bootEGA analize stabilnosti tro-faktorskog rješenja na mjerno-invarijantnom uzorku



Napomena: 1 = Skala sposobnosti uočavanja i razumijevanja emocija (UR); 2 = Skala sposobnosti regulacija i upravljanja emocijama (RU); 3= Skala sposobnosti izražavanja i imenovanja emocija (II)

4.2.2. Dimenzionalnost Upitnika emocionalne kompetentnosti UEK-27

Zatim je provjeravana dimenzionalnost Upitnika emocionalne kompetentnosti, verzije s 27 čestica. Prethodna istraživanja UEK, u okviru klasične testne teorije uz upotrebu konfirmatorne faktorske analize, pokazala su da postoje 3 važna faktora koji odgovaraju trima subskalama upitnika, te da su oni međusobno dovoljno korelirani da se može pretpostaviti da je u pozadini generalni faktor emocionalne inteligencije.

Dimenzionalnost se sada željela provjeriti u okviru teorije odgovora na zadatak. Pritom je korišten *mirt* paket u programu RStudio.

Uspoređena su četiri jednodimenzionalna modela: model stupnjevitih odgovora (GRM), generalizirani model djelomičnog bodovanja (GPCM), model skala ocijenjivanja (RSM), model djelomičnog bodovanja (PCM), te jedan višedimenzionalni model – višedimenzionalni model stupnjevitih odgovora (eng. *multidimensional graded response model, mGRM*).

Pretpostavku jednodimenzionalnosti provjerili smo upotrebom eksplorativne faktorske analize. Prikladnost polihorične korelacijske matrice za provođenje faktorske analize provjerena je Bartlettovim, te Kaiser-Meyer-Olkinovim testom. Determinanta korelacijske matrice bila je adekvatna ($1.6e^{-05}$). Vrijednost Kaiser-Meyer-Olkinovog testa od.92 upućuje na

činjenicu da je obrazac korelacija relativno kompaktan (parcijalne korelacije među česticama niske), te da bi se faktorskom analizom trebali dobiti jasni i pouzdani faktori. Bartlettov test, koji provjerava da li se dobivena korelacijska matrica razlikuje od matrice identiteta, odnosno matrice nultih korelacija, je statistički značajan, što znači da je faktorska analiza prikladan postupak.

Faktorska analiza provedena je metodom glavnih osi uz forsirano izlučivanje jednog faktora, a zatim je pretpostavka jednodimenzionalnosti provjerena kriterijem omjera prve i druge svojstvene vrijednosti, Velicerovim MAP testom (engl. *minimum average partial test*) i Cattellovim scree testom. Iako ne postoji „zlatni standard“ u pogledu omjera prve i druge svojstvene vrijednosti koji upućuje na jednodimenzionalnost, ipak, obično se omjer veći od 3 smatra dovoljnim (Hattie, 1985; Morizot i sur. , 2007; Slocum-Gori i Zumbo, 2011).

Prvih četiri svojstvenih vrijednosti bilo je veće od 1, a prva i druga iznosile su 7.89 odnosno 3.09, što znači da im je omjer bio 2.55. S obzirom da je ovaj omjer bio manji od 3, to je upućivalo na višedimenzionalnost UEK-27. No, prva svojstvena vrijednost objasnila je 29% varijance, a još jedan od kriterija kaže da ukoliko prva svojstvena vrijednost objašnjava više od 20% varijance može se smatrati da je riječ o jednodimenzionalnom rješenju (Hattie, 1985). S druge strane, Velicerov test minimalne prosječne parcijalnosti sugerirao je rješenje sa tri faktora (Velicer, 1976).

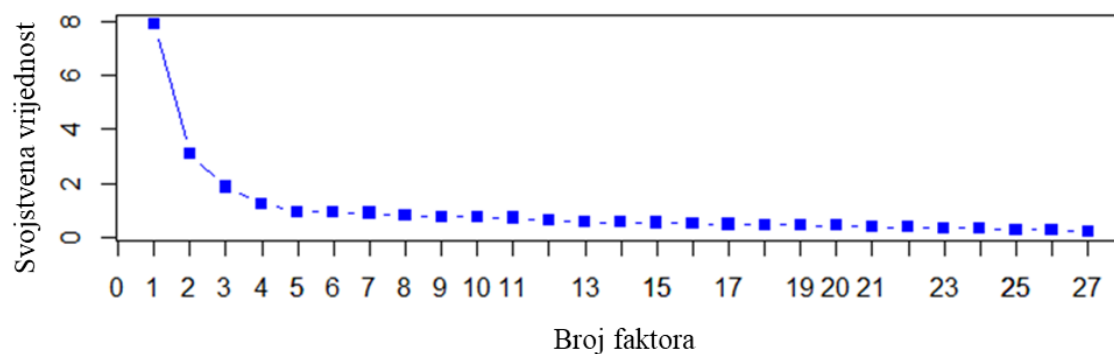
Tablica 9. Testiranje primjerenosti podataka za faktorsku analizu - KMO mjera prikladnosti uzorkovanja i Bartlettov test

Kaiser-Meyer-Olkinova mjera prikladnosti uzorkovanja		.92
Bartlettov test sfericiteta	Hi-kvadrat	20485.02
	Stupnjevi slobode	351
	Značajnost	.000

Vizualni pregled Cattellovog *scree* testa upućivao je ponovno na četverofaktorsko rješenje. Sve u svemu, eksploratorna faktorska analiza je dala proturječne rezultate, ukazujući na mogućnost postojanja i jednodimenzionalnosti i višedimenzionalnosti u podacima. Ipak, više kriterija išlo je u prilog višedimenzionalnosti.

Zatim je provjeravana pretpostavka lokalne neovisnosti, prema kojoj bi, nakon izjednačavanja ispitanika po latentnoj osobini (procjeni *thete*), odgovori na pojedinačne čestice trebali biti međusobno neovisni. Kršenje ove pretpostavke može iskriviti procjenu parametara čestica, njihove standardne pogreške, kao i pokazatelje prikladnosti modela za podatke (de Ayala, 1993).

Slika 21. Cattellov (scree) dijagram



Tablica 10. Rezultati faktorske analize

Faktor	Početne svojstvene vrijednosti			Izlučene sume kvadriranih faktorskih opterećenja		
	Ukupno	Proporcija varijance	Kumulativna proporcija varijance	Ukupno	Proporcija varijance	Kumulativna proporcija varijance
1	7.89	.29	.29	7.21	.27	.27
2	3.09	.11	.41			
3	1.87	.07	.48			
4	1.26	.05	.52			
<i>Omjer prve i druge svojstvene vrijednosti</i>						2.553
<i>Velicerov MAP test</i>						
<i>Velicerov minimum</i>						
			<i>Minimum</i>	<i>Broj faktora</i>		
MAP ²			0.00972	3		
MAP ⁴			0.00039	3		

Za ovu LD analizu (engl. *local dependance*) korišteni su Yenovi Q3 pokazatelji koji se temelje na rezidualima između opaženih i očekivanih korelacija među parovima česticama a izračunati su pomoću funkcije **residuals** u paketu *mirt*.

Korištena je granična vrijednost od .2 (Christensen i sur., 2017) – rezidualne korelacije više od ove vrijednosti ukazuju na mogućnost da u podacima postoji dodatna dimenzija odnosno latentna osobina pored one koja se mjeri. Ove rezidualne korelacije izračunate su u odnosu na svih pet uspoređivanih IRT modela.

Svi jednodimenzionalni modeli imali su dosta parova čestica koje su odstupale od zadanog modela (između 46 i 75). Najmanje pak parova čestica koje su odstupale imao je višedimenzionalni model stupnjevitih odgovora, njih 9.

Dakle, i ova analiza sugerirala je upotrebu višedimenzionalnog modela kao primjerenijeg.

Slaganje čestica s modelom

Kod teorije odgovora na zadatak slaganje modela s podacima provjerava se i na razini testa i na razini čestice. Slaganje na razini čestice provjereno je upotrebom Orlando-Thissen-Bjornerovog pokazatelja $S-\chi^2$ (Bjorner i sur., 2007; Orlando i Thissen, 2000, 2003). Vrijednosti koje nisu značajne ukazuju da se čestica slaže s predloženim IRT modelom. Kako bi se izbjegla inflacija pogreški tipa I zbog višestrukih usporedbi korišten je Benjamini-Hochberg postupak za korekciju (Benjamini i Hochberg, 1995).

Predzadnji redak u Tablici 9. prikazuje koji broj čestica je pokazao slaganje s pojedinim modelom. Analiza je pokazala da je svim česticama instrumenta od jednodimenzionalnih modela jednako odgovarao i model stupnjevitih odgovora (GRM) kao i generalizirani model djelomičnih odgovora (GPCM). No, i višedimenzionalni model stupnjevitih odgovora odgovarao je svim česticama modela. Ostala dva modela pokazala su slaganje sa znatno manjim brojem čestica, model djelomičnih odgovora slagao se sa 16, a model skala ocjenjivanja slagao se samo s 8 čestica.

Globalno slaganje modela s podacima i usporedba modela

Za razliku od konfirmatorne faktorske analize, usporedba različitih modela i odabir najprikladnijeg u okviru IRT-a mogu se provesti tek nakon što se dokazalo postojanje pretpostavki modela i nakon što je ustanovljeno da postoji pristajanje modela česticama, odnosno prikladnost modela na razini čestice (de Ayala, 2009), zbog čega se pretpostavlja da je za svaku česticu identificiran odgovarajući IRT model.

Prikladnost modela provjerena je pomoću C_2 (Cai i Monroe, 2014) pokazatelja i s njime povezanim indeksima slaganja. C_2 pokazatelj je nastao specifično za potrebe procjene prikladnosti modela iz teorije odgovora na zadatak koji se temelje na ordinalnim podacima, a

njegova distribucija je slična hi-kvadrat distribuciji. Odabran je jer se pokazalo da funkcionira bolje od drugih sličnih pokazatelja, M_2 (Maydeu-Olivares i Joe, 2006) i M_2^* (Cai i Hansen, 2013). Kao primarni indeks slaganja korištena je $RMSEA_{C_2}$ (engl. *Root Mean Square Error of Approximation*), pogreška aproksimacije C_2 pokazatelja, za koju su izračunate i granice pouzdanosti. Tradicionalna granična vrijednost za RMSEA-u je ≤ 0.08 i upućuje na primjereno slaganje, no u slučaju IRT modela treba biti oprezan sa donošenjem zaključaka, jer se ova granična vrijednost temelji na modelima koji pretpostavljaju linearan odnos između odgovora na česticu i latentne osobine (Toland i sur., 2017).

Tablica 11. Usporedba prikladnosti različitih IRT modela za podatke dobivene UEK-27 upitnikom

	PCM	GRM	GPCM	RSM	mGRM
LL	-57378.62	-56594.02	-56750.54	-57805.32	-54246.92
Broj parametara	109	135	135	135	138
BIC	115578.2	114204.8	114517.8	115844.1	109533.2
AIC	114975.2	113458	113771.1	115672.6	108769.8
C_2 (df)	11872.13*** (350)	11923.84*** (324)	11598.67*** (324)	12553.04*** (428)	1010.86*** (240)
$RMSEA_{C_2}$ (90% CI)	0.133 (0.130-0.135)	0.138 (0.136-0.140)	0.136 (0.134-0.139)	0.123 (0.121-0.125)	0.041 (0.039-0.044)
SRMSR	0.127	0.098	0.099	0.121	0.046
Broj adekvatnih čestica	16	27	27	8	27
Broj značajnih Q3 parova	46	75	74	51	9

Napomena. PCM = model djelomičnih odgovora; GRM = model stupnjevitih odgovora; GPCM = generalizirani model djelomični odgovora; RSM = model skala ocijenjivanja; mGRM = multidimenzionalni model stupnjevitih odgovora

Također je korišten i standardizirani prosječni rezidual SRMSR (engl. *Standardized Root Mean Square Residual*), koji se preporuča za upotrebu s velikim modelima, a za usporedbu različitih modela korišteni su Akaikeov (engl. *Akaike information criterion, AIC*) i Bayesov informacijski kriterij (engl. *Bayesian information criterion, BIC*). Kod ovih kriterija odabire se kao najprikladniji onaj model koji ima najmanji BIC ili AIC.

Svi pokazatelji slaganja modela s podacima bili su u korist višedimenzionalnog modela stupnjevutih odgovora. Stoga su za višedimenzionalni model stupnjevutih odgovora provjereni još i pokazatelji apsolutnog slaganja modela s podacima iz okvira strukturalnog modeliranja, TFI i CFI, te su oba pokazala adekvatno slaganje (.93 odnosno .94).

Dakle, i pokazatelji slaganja modela s podacima i rezultati LD analiza išli su u korist višedimenzionalnog stupnjevutog modela.

Sve u svemu, rezultati provjere dimenzionalnosti sugerirali su da je UEK-27 najvjerojatnije višedimenzionalan instrument. Stoga je daljnja analiza pojedinačnih čestica napravljena temeljem višedimenzionalnog modela stupnjevutih odgovora.

No, s obzirom da je više od dvije latentne osobine gotovo nemoguće grafički prikazati u dvodimenzionalnom sustavu, a osim toga ih je vrlo teško za interpretirati, grafički prikazi čestica su temeljeni na jednodimenzionalnom modelu stupnjevutih odgovora. Također, uspoređeni su parametri čestica jednodimenzionalnog GRM s parametrima čestica GRM kako bi se analizirali efekti krivog modeliranja.

Parametri čestica

Parametri čestica izračunati su temeljem višedimenzionalnog modela stupnjevutih odgovora i prikazani su u Tablicama 10., 11. i 12. Za usporedbu su ondje navedeni i parametri dobiveni temeljem jednodimenzionalnog modela stupnjevutih odgovora.

Parametri nagiba (a) ili parametri diskriminativnosti mjere koliko dobro čestica razlikuje ispitanike s različitim razinama latentne osobine. Što su više vrijednosti tog parametra strmiji su nagibi, a čestice su bolje u razlikovanju ispitanika. Nagib može još služiti i kao indikator snage odnosa između čestice i latentne osobine, pa više vrijednosti parametra odgovaraju jačoj povezanosti.

Prema vrijednosti parametra nagiba, diskriminativnost čestice možemo klasificirati na sljedeći način (Baker, 2001):

- nulta: 0,
- vrlo niska: (0.01, 0.34),
- niska: (0.35, 0.64),
- umjerena: (0.65, 1.34),
- visoka: (1.35, 1.69),
- vrlo visoka: (1.70, ∞).

Parametri nagiba UEK-27 variraju od 0.68 do 2.70, s tim da je čestica 11 najmanje diskriminativna, a čestica 23 najdiskriminativnija čestica upitnika. Sve čestice su ili umjereno ili visoko ili vrlo visoko diskriminativne.

Osim parametra nagiba, za svaku česticu su izračunata i 4 parametra lokacije (b). Parametri lokacije u okviru modela stupnjevutih odgovora interpretiraju se kao vrijednost *thete*

koja odgovara 50%-tnoj vjerojatnosti da će ispitanik na česticu odgovoriti tom kategorijom ili kategorijama koje se nalaze iznad nje. Uobičajeno kod modela stupnjevitih odgovora postoji $m-1$ parametara lokacije, pri čemu m odgovara broju kategorija odgovora na skali čestice.

Prema nekim autorima (Baker, 2001; Hambleton i sur., 1991; Hasmy, 2014), težina čestice može se klasificirati u sljedeće kategorije:

- vrlo lagana čestica: $(-\infty, -2.00)$,
- lagana čestica: $(-2.00, -0.05)$,
- srednje teška čestica: $(-0.05, 0.05)$
- teška čestica: $(0.05, 2.00)$,
- vrlo teška čestica: $(2.00, \infty)$.

Parametri lokacije (b) UEK-27 upućuju na to da odgovori na čestice pokrivaju širok raspon mjerenih latentnih osobina, no i na to da neke čestice to čine na nerealno visokim i niskim razinama. Parametar lokacije b_1 je parametar težine za odabir kategorije „uopće NE“ ili više; b_2 je parametar težine za odabir srednje („uglavnom NE“) kategorije ili više; b_3 je parametar težine za odabir kategorije „kako kada“ ili više; a b_4 je parametar težine za odabir kategorije „u potpunosti DA“.

Tablica 12. Usporedba parametara čestica Skale sposobnosti uočavanja i razumijevanja emocija prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora

Čestica	Višedimenzionalni GRM					Jednodimenzionalni GRM				
	A ₁	b ₁	b ₂	b ₃	b ₄	a	b ₁	b ₂	b ₃	b ₄
UR9	2.20 ⁺	-3.08	-1.39	0.19	1.89	1.36***	-3.95	-1.72	0.26	2.37
UR8	2.08 ⁺	-3.34	-1.53	0.14	2.00	1.38***	-4.17	-1.84	0.19	2.43
UR7	2.07 ⁺	-3.20	-1.51	0.09	1.82	1.29**	-4.16	-1.90	0.13	2.31
UR5	1.81 ⁺	-3.90	-2.21	-0.33	1.73	1.58***	-4.16	-2.33	-0.32	1.84
UR10	1.75 ⁺	-4.27	-2.59	-0.77	1.39	1.35***	-5.00	-2.97	-0.85	1.60
UR3	1.54***	-3.59	-1.64	0.27	2.49	1.33**	-3.89	-1.75	0.31	2.70
UR4	1.51***	-4.03	-1.91	-0.12	2.03	1.22**	-4.61	-2.14	-0.11	2.29
UR11	1.44***	-4.14	-2.46	-0.45	1.80	1.38***	-4.24	-2.50	-0.45	1.83
UR6	1.22**	-3.30	-1.36	0.32	2.28	0.88**	-4.19	-1.68	0.43	2.88
UR1	1.19**	-3.62	-1.32	0.94	3.22	1.06**	-3.88	-1.39	1.02	3.47
UR2	1.08**	-5.15	-2.81	-0.96	1.31	1.08**	-5.08	-2.77	-0.94	1.30

Napomena: značenje oznaka parametara diskriminativnosti: ⁺ vrlo visoka; *** visoka; ** umjerena;

* niska; ⁺ vrlo niska

Sve čestice u Skali sposobnosti uočavanja i razumijevanja emocija imaju visoku ili vrlo visoku diskriminativnost. Parametri lokacija su također dobri, te većina čestica dobro mjeri i ispodprosječne i iznadprosječne razine ove grane emocionalne kompetentnosti. Ono što se pak uočava je da je odabir srednje kategorije odgovora („kako kada“) ili viših pomaknut prema ispodprosječnim vrijednostima, na način da će, za većinu čestica, 50%-tna vjerojatnost odabira

ove kategorije biti kod ispitanika čija razina ove sposobnosti je u intervalu od -2.81 do -1.36 *thete*. Kategorija odgovora „uglavnom DA“ pokriva razinu *thete* u intervalu oko 0, odnosno između -0.96 i 0.94. To bi značilo da je 50%-vjerojatnost odabira ove kategorije ili više prisutna za ispitanike prosječne sposobnosti uočavanja i razumijevanja emocija.

Čestice na skali također imaju visoku ili vrlo visoku diskriminativnost. Što se parametara lokacije tiče, uočava se da su dva središnja parametra lokacije, *b2* i *b3* još više pomaknuti prema negativnim vrijednostima *thete*. To bi značilo da ova skala, još više nego prethodna, bolje razlikuje ispitanike na ispodprosječnim razinama ove grane emocionalne kompetentnosti.

Tablica 13. *Usporedba parametara čestica Skale sposobnosti izražavanja i imenovanja emocija prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora*

Čestica	Višedimenzionalni GRM						Jednodimenzionalni GRM			
	A ₂	b ₁	b ₂	b ₃	b ₄	a	b ₁	b ₂	b ₃	b ₄
II5	2.70 ⁺	-3.14	-1.87	-0.62	1.08	1.59***	-3.99	-2.29	-0.74	1.32
II6	2.50 ⁺	-2.94	-1.54	-0.19	1.46	1.65***	-3.54	-1.78	-0.21	1.70
II4	2.36 ⁺	-3.36	-1.90	-0.71	1.02	1.49***	-4.24	-2.32	-0.84	1.24
II7	1.99 ⁺	-3.55	-1.95	-0.51	1.36	1.48***	-4.17	-2.24	-0.57	1.55
II3	1.97 ⁺	-3.21	-1.60	-0.18	1.65	1.43***	-3.85	-1.84	-0.18	1.92
II2	1.91 ⁺	-3.10	-1.41	-0.11	1.51	1.32**	-3.80	-1.64	-0.09	1.80
III1	1.55***	-3.21	-1.39	-0.14	1.66	1.09**	-4.04	-1.69	-0.16	2.03

Napomena: značenje oznaka parametara diskriminativnosti: ⁺ vrlo visoka; *** visoka; ** umjerena; * niska; ⁺ vrlo niska

Skala sposobnosti regulacije i upravljanja emocijama imala je najlošije parametre diskriminativnosti, pri čemu je većina čestica imala umjerenu diskriminativnost, a samo jedna je imala visoku. Što se parametara lokacije tiče, pomaknutost u negativne vrijednosti latentne osobine je na ovoj skali još uočljivija.

Osim toga, neke čestice imale su nerealno široke raspone parametara težine za kategorije odgovora. Ovakvi nerealno široki rasponi upućuju na to da se ove kategorije odgovora odabiru na nerealnim razinama emocionalne kompetentnosti, što znači da čestice nisu funkcionalne pri mjerenju uobičajenog i očekivanog raspona emocionalne kompetentnosti.

Tablica 14. Usporedba parametara čestica Skale sposobnosti regulacije i upravljanja emocijama prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora

Čestica	Višedimenzionalni GRM					Jednodimenzionalni GRM				
	A ₃	b ₁	b ₂	b ₃	b ₄	a	b ₁	b ₂	b ₃	b ₄
RU8	1.43***	-4.94	-3.27	-1.36	0.81	0.72**	-8.58	-5.50	-2.20	1.31
RU7	1.20**	-4.49	-2.69	-0.52	1.82	0.86**	-5.82	-3.42	-0.64	2.33
RU9	1.17**	-4.34	-2.80	-1.14	1.17	0.84**	-5.65	-3.59	-1.43	1.49
RU4	1.12**	-5.34	-3.13	-0.93	1.60	0.73**	-7.61	-4.35	-1.24	2.23
RU6	1.10**	-5.06	-2.62	-0.67	1.71	0.59*	-8.64	-4.30	-1.05	2.80
RU1	0.77**	-6.50	-3.93	-1.57	1.61	0.57*	-8.49	-5.07	-2.00	2.08
RU3	0.76**	-6.07	-2.56	-0.29	2.85	0.45*	-9.79	-4.00	-0.40	4.55
RU2	0.70**	-7.60	-5.29	-2.95	0.42	0.51*	-10.15	-7.03	-3.87	0.57
RU5	0.68**	-8.41	-4.69	-1.70	1.14	0.48*	-11.56	-6.38	-2.27	1.55

Napomena: Značenje oznaka parametara diskriminativnosti: + vrlo visoka; *** visoka; ** umjerena;

* niska; † vrlo niska

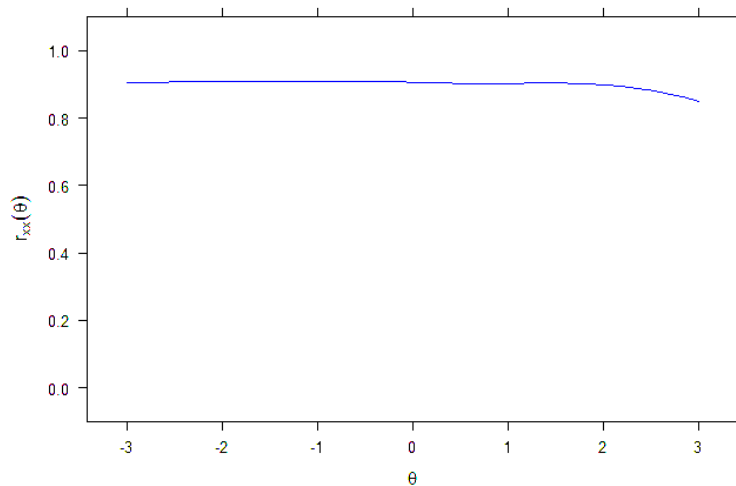
Kada usporedimo parametre dobivene temeljem jednodimenzionalnog u odnosu na višedimenzionalni model stupnjevitih odgovora, možemo vidjeti da je kod svih čestica došlo do poboljšanja mjernih svojstava – diskriminativnost čestice je porasla, a raspon težina čestica se uravnotežio.

Provjerene su i korelacije među latentnim osobinama, te one iznose: korelacija između sposobnosti uočavanja i razumijevanja emocija i sposobnosti izražavanja i imenovanja emocija .40, korelacija između sposobnosti izražavanja i imenovanja emocija i sposobnosti regulacije i upravljanja emocijama .53, a korelacija između sposobnosti uočavanja i razumijevanja emocija i sposobnosti regulacije i upravljanja emocijama .44.

Za razliku od klasične testne teorije, IRT nudi uvjetnu mjeru pouzdanosti koja ovisi o razini latentne osobine. Ova uvjetna pouzdanost UEK-27 grafički je prikazana na Slici 24. Iz nje je vidljivo da je UEK-27 visoko pouzdan u intervalu od -3 do +2 emocionalne kompetentnosti, a zatim mu pouzdanost ponešto opada za vrijednosti *thete* iznad +2. Indeks marginalne (opće) pouzdanosti (ekvivalentne Cronbachovoj α) iznosi .90.

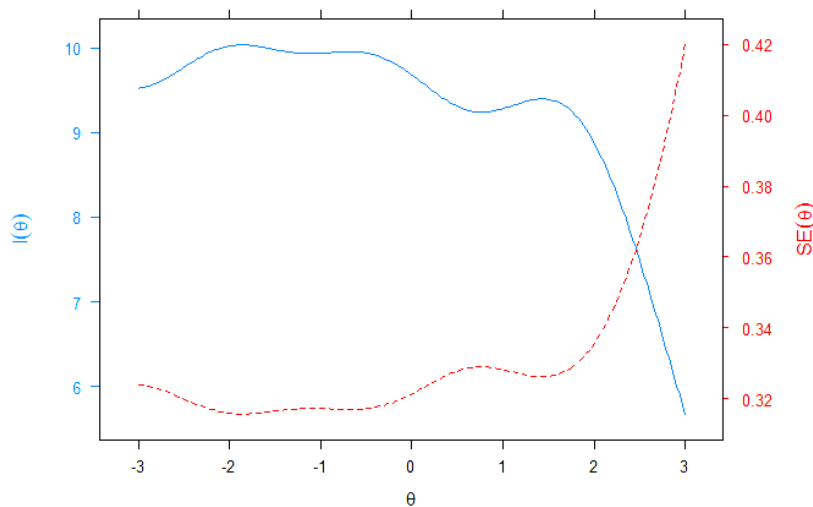
Grafički su prikazane karakteristične krivulje kategorija, zajedno s informacijskim funkcijama za svaku pojedinu česticu. Osim toga, prikazane su i kumulativne funkcije kategorija odgovora za svaku česticu (Privitak 3.).

Slika 22. Uvjetna pouzdanost UEK-27



Temeljem informacijskih krivulja pojedinih čestica izračunata je informacijska krivulja upitnika (Slika 24.), koja pokazuje da je UEK-27 najinformativniji za ispitanike čija je emocionalna kompetentnost u intervalu od -3 do 0, nešto manje informativna za ispitanike s emocionalnom kompetentnošću između 0 i +2, a nakon te vrijednosti *thete* informativnost rapidno opada. To znači da je UEK-27 najprilagođeniji osobama prosječne i ispodprosječne emocionalne kompetentnosti, dok za osobe iznadprosječne emocionalne kompetentnosti nije informativan.

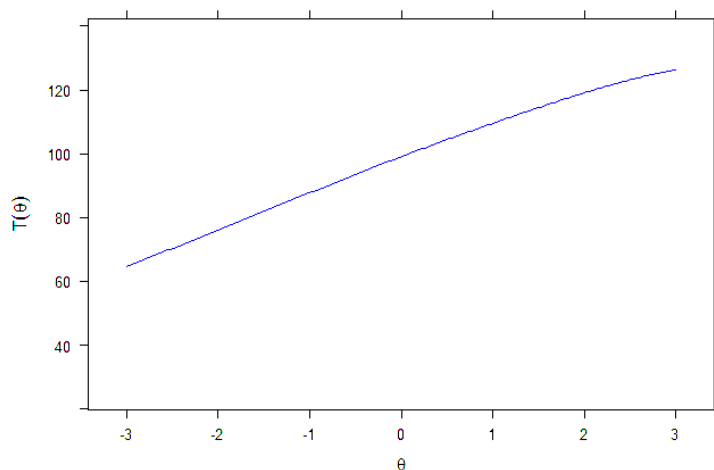
Slika 23. Informacijska krivulja UEK-27



Na kraju je prikazana i karakteristična krivulja upitnika (Slika 25.). Iz nje su vidljivi očekivani ukupni rezultati osoba s određenom razinom emocionalne kompetentnosti. Ako pogledamo interval između -3 i -1 (ispodprosječna emocionalna kompetentnost) možemo

vidjeti da je očekivani raspon ukupnih rezultata ispitanika između 60 i 90 bodova. Kod ispitanika prosječne emocionalne kompetentnosti (interval od -1 do +1 *theta*), očekivani raspon bodova je između 90 i 110. Kod ispitanika pak iznadprosječne emocionalne kompetentnosti (+1 do +3 *thete*) očekuje se raspon ukupnih rezultata od 110 do 120. I iz ovog je također vidljivo kako je diskriminativnost čestica, informativnost upitnika i pouzdanost dobivenih rezultata niža s porastom latentne varijable.

Slika 24. Karakteristična krivulja UEK-27



4.2.3. Diferencijalno funkcioniranje čestica

Nakon što je identificiran metrički invarijantan uzorak moglo se pristupiti analizi diferencijalnog funkcioniranja čestica po pitanju spola. Ukupni broj ispitanika u sva četiri metrički invarijantna uzorka bio je 1866, od čega je 1340 bilo ženskog, a 526 muškog spola.

Kako bi se ispitala učinkovitost metoda za otkrivanje DIF-a na realnim podacima i zatim napravila usporedba učinkovitosti dobivene na simuliranim podacima, iste četiri metode za detekciju DIF-a primijenjene su na metrički invarijantnom uzorku, a različito funkcioniranje provjeravalo su u odnosu na varijablu spola.

Interakcijska MIMIC metoda testirala je simultano jednoliki i nejednoliki DIF pomoću *MplusAutomation* paketa (Hallquist i Wiley, 2018) u sprezi s *Mplus 7* programom (Muthén i Muthén, 1998-2017). Kao metoda procjene parametara korištena je procjena najveće vjerojatnosti s robusnim standardnim pogreškama (engl. *maximum likelihood estimation with robust standard errors*). Specificiran je trofaktorski konfirmatorni model s kovarijatama. Faktori su specificirani u skladu s teorijskom strukturom subskala UEK-27, odnosno prvih 11 čestica pripadalo je prvom faktoru (Uočavanje i razumijevanje emocija), drugih 7 drugom

faktoru (Izražavanje i imenovanje emocija), a zadnjih 9 trećem faktoru (Razumijevanje i upravljanje emocijama). Zatim je svaki faktor zasebno regresiran na varijablu spola. U model su dodane i tri interakcijske latentne varijable kao kovarijate pomoću kojih je provjeravan nejednoliki DIF. Riječ je o interakciji varijable spola i svakog od pojedinačna tri faktora.

Zatim je svaka pojedina čestica testirana na jednoliki DIF tako što je regresirana na varijablu spola, te na nejednoliki DIF tako što je regresirana na interakcijski termin između varijable spola i odgovarajuće latentne varijable/faktora koji mjeri. U svakoj subskali po jedna varijabla koja je preliminarnim analizama identificirana kao najdiskriminativnija služila je kao sidrišna čestica, što je značilo da je bila fiksirana kao jednaka u obje grupe ispitanika (tj. i za muške i za ženske ispitanike). Konkretno, to su bile čestice UR3, II5 i RU8. Uloga sidrišnih čestica je da se temeljem njih odredi zajednička skala latentne varijable u drugoj analizi.

Kod IRT-LR metode detekcije DIF-a uspoređivalo se nekoliko modela s različitim ograničenjima, za svaku pojedinu česticu. Pritom se analiza provela u dva koraka, prema proceduri Meade i Wright (2012). U prvom koraku koristi se tzv. AOAA pristup (engl. *all others as anchors*) u kojem sve čestice osim ispitivane čestice ulaze u sidrišni skup za koji se smatra da ne sadrži DIF čestice. Zatim se temeljem rezultata analize iz prvog koraka odabire jedna ili nekoliko invarijantnih čestica koje imaju visoke parametre diskriminativnosti i one ulaze u sidrišni skup čestica te se ponovno provodi DIF analiza. U ovom istraživanju u sidrišni skup čestica ušle je po jedna čestica sa svake subskale, i to ona koja je u prvom koraku detektirana kao invarijantna i imala je najviši parametar diskriminativnosti među invarijantnim česticama sa svoje subskale. Drugim riječima, parametri ovih čestica se izjednače u obje grupe, te ih se ne ispituje na DIF u drugom koraku. To su kod IRT-LR metode bile čestice UR9, II5 i RU8.

U glavnoj analizi DIF-a provedena su sva tri testa za svaku pojedinačnu česticu: omnibus test i testovi za jednoliki i nejednoliki DIF.

Logistička regresija s bruto rezultatima je uz pomoć logističke funkcije modelirala rezultat na pojedinoj čestici kao funkciju ukupnog rezultata ispitanika i pripadnosti grupi kroz tri sukcesivna koraka. U prvom je rezultat čestice modeliran isključivo temeljem ukupnog rezultata. U drugom koraku, predikcijska jednadžba proširena je dodavanjem učinka grupa koji se koristio kao test za provjeru postojanja jednolikog DIF-a. U trećem koraku, u model je uključena i interakcija između ukupnog rezultata i pripadnosti grupi, kao osnova za provjeru postojanja nejednogolikog DIF-a. Zatim su provedene usporedbe slaganja omnibus modela i pojedinačnih modela za jednoliki i nejednoliki DIF s podacima, tj. ANOVA testom je provjereno koji je model bolji za predviđanje rezultata na čestici.

Logistička regresija s procjenom latentne osobine slijedila je istu proceduru kao i logistička regresija s bruto rezultatima, samo što su umjesto ukupnog rezultata korišteni faktorski bodovi. I jedna i druga vrsta logističke regresije provedene su dvaput. U drugom pokušaju iz sidrišnog skupa čestica izbačene su čestice koje je prva analiza koja je koristila AOAA pristup identificirala kao DIF čestice. To su bile čestice UR2 i UR7, II7 i RU2. To je značilo da te čestice kod logističke regresije s bruto rezultatom nisu ulazile u izračun ukupnog

rezultata (*matching* varijable), a kod logističke regresije s procjenom latentne osobine nisu ulazile u izračun faktorskih bodova (tj. nisu definirale pojedini faktor subskale kojoj inače pripadaju).

Tablica 15. Rezultati dvostupanjskih analiza diferencijalnog funkcioniranja čestica UEK-27

Čestica	LR-T (G ²)	LR-R (G ²)	IRT- LR (X ²)	MIMIC
UR1	5.02	4.31	11.75*	0.085*
UR2	8.39**	8.26**	19.45**	0.12**
UR5	4.31	4.31	8.27*	-0.08*
UR7	18.94**	23.41**	10.43*	0.07*
UR8	6.44	6.87	4.40	-0.07*
UR9	6.89**	8.35**	2.49	-0.10**
UR10	4.86	4.46	11.99*	0.06
II3	7.95	8.69	8.42	-0.11*
II4	7.73	8.05**	9.01	-0.09*
II7	25.43**	26.65**	9.31*	0.11**
RU1	7.31	4.80	2.62	-0.39**
RU2	19.34**	20.37**	31.13**	0.21**
RU3	6.85**	5.80*	6.72	-0.12**
RU5	4.17*	9.47	9.84	0.08
RU9	13.81**	9.39*	12.92	-0.17**

Napomena: * p < .05; ** p < .001

U Tablici 15. prikazane su čestice koje je jedna ili više metoda označila kao DIF čestice. Metode su se u potpunosti složile oko određivanja je li čestica DIF ili nije u otprilike 60 % slučajeva. S obzirom da kod realnih, za razliku od simuliranih podataka, nije moguće unaprijed znati je li neka čestica DIF ili nije DIF, a metode nisu u potpunosti učinkovite, kako bismo provjerili koja od njih je najučinkovitija za procjenu je li čestica DIF ili nije upotrijebljeno je konsenzualno rješenje.

Naime, pretpostavljeno je da ukoliko se većina metoda slaže da je neka čestica DIF, ona to zaista i jest. Isto je vrijedilo i u suprotnom slučaju, kod čestica koje su se pokazale invarijantnima. U slučaju da su se sve četiri ili tri od četiri metode složile bilo je jednostavno donijeti zaključak o tome kakva je čestica.

Pritom je u obzir uzeta i činjenica da su LR-T i LR-R međusobno inherentno sličnije (po osnovnoj logici i matematičkoj podlozi) od ostale dvije metode, stoga, ukoliko su ove dvije metode govorile jedno, a druge dvije suprotno, prednost se davala suprotnom zaključku.

U slučaju da se identifikacija u potpunosti razlikovala, pa se tako jedna od logističkih regresija slagala s jednom od preostale dvije, a druga logistička regresija s drugom, tada nije donesena odluka o tome je li čestica DIF ili nije i ta čestica nije ulazila u daljnju kalkulaciju postotka točnih identifikacija. Ovakva situacija dogodila se samo s česticom II4, kod koje su DIF identificirale logistička regresija s bruto rezultatima i MIMIC metoda.

Što se vrste DIF-a tiče, kod čestica kod kojih je postojalo konsenzualno rješenje, metode su se složile i oko vrste DIF-a. Tako su jednoliki DIF sadržavale čestice UR1, UR2, UR9, RU2 i RU3, a nejednoliki čestice UR5, UR7, II7 i RU9.

Tablica 16. *Usporedba slaganja pojedinih metoda s konsenzualnim rješenjem*

Čestica	Konsenzualno rješenje	LR-T	LR-R	IRT- LR	MIMIC
UR1	1	0	0	1	1
UR2	1	1	1	1	1
UR3	0	0	0	0	0
UR4	0	0	0	0	0
UR5	1	0	0	1	1
UR6	0	0	0	0	0
UR7	1	1	1	1	1
UR8	0	0	0	0	1
UR9	1	1	1	0	1
UR10	0	0	0	1	0
UR11	0	0	0	0	0
II1	0	0	0	0	0
II2	0	0	0	0	0

II3	0	0	0	0	1
II4	/	/	/	/	/
II5	0	0	0	0	0
II6	0	0	0	0	0
II7	1	1	1	1	1
RU1	0	0	0	0	1
RU2	1	1	1	1	1
RU3	1	1	1	0	1
RU4	0	0	0	0	0
RU5	0	1	0	0	0
RU6	0	0	0	0	0
RU7	0	0	0	0	0
RU8	0	0	0	0	0
RU9	1	1	1	0	1
Broj ukupnih točnih identifikacija	23	24	22	23	
Broj ukupnih netočnih identifikacija	3	2	4	3	
Broj točno identificiranih DIF čestica	7	7	6	9	
Broj točno identificiranih invarijantnih čestica	16	17	16	14	
Broj lažno pozitivnih detekcija	1	0	1	3	
Broj lažno negativnih detekcija	2	2	3	0	
Stopa točnih identifikacija (TPR)	77.78%	77.78%	66.67%	100%	
Stopa pogreški tipa I (FPR)	5.88%	0%	5.88%	17.65%	

Napomena: / = metode nisu došle do konsenzualnog rješenja

Veličina DIF-a nije mogla biti provjerena s obzirom na činjenicu da trenutno ne postoje razvijeni pokazatelji veličine efekta za metode koje DIF otkrivaju u višedimenzionalnim politomnim podacima.

Kako bi se odredila stopa točnih identifikacija (engl. *true positive rate*, *TPR*) i stopa pogreški tipa I odnosno stopa lažno pozitivnih klasifikacija (engl. *false positive rate*, *FPR*) napravljena je matrica konfuzije (engl. *confusion matrix*) za svaku metodu (Cheng i sur., 2023).

Stopa točnih identifikacija se izračunava kao omjer broja ispravnih klasifikacija (TP) i ukupnog broja čestica u kojima postoji DIF (P), odnosno $TPR = TP/P$.

Stopa pogreški tipa I se izračunava kao omjer lažno pozitivnih klasifikacija (FP) i ukupnog broja čestica koje su invarijantne (N), odnosno $FPR = FP/N$.

Ukupni broj čestica oko kojih su se metode složile da diferencijalno funkcioniraju bio je 9, a ukupni broj čestica oko kojih su se metode složile da su invarijantne bio je 17.

Iako je najveći broj točnih identifikacija imala logistička regresija s bruto rezultatima (24), najvišu stopu točnih identifikacija imala je MIMIC metoda. Ona je točno detektirala svih 9 DIF čestica, no pritom je imala i najviši broj lažnih detekcija (3). Logistička regresija s bruto rezultatima, pak, nije imala niti jednu lažnu detekciju.

Najnižu stopu točnih detekcija imala je IRT-LR metoda (66.67%). Obje vrste logističke regresije imale su jednaku stopu točnih identifikacija (77.78%), ali je logistička regresija s procjenom latentne osobine imala nekoliko lažnih detekcija. Sve tri pak metode imale su stopu pogreški tipa I unutar Bradleyevog kriterija liberalne robusnosti, dok je MIMIC model imao povišenu stopu.

4.2.4. Usporedba rezultata simulacijskog i empirijskog istraživanja

Među nalazima dobivenim simulacijskim istraživanjem dva nalaza potvrđena su i na realnim podacima. Naime, interakcijska MIMIC metoda u oba je slučaja imala povišenu stopu pogreški tipa I, a IRT-LR metoda funkcionirala je najlošije po pitanju snage otkrivanja DIF-a.

No, rang učinkovitosti metoda se promijenio, pa je tako u simulacijskom istraživanju najučinkovitija bila logistička metoda s latentnom procjenom, dok je DIF u realnim podacima najučinkovitije otkrivala logistička metoda s bruto rezultatima. Obje vrste logističke regresije imale su umjerenu statističku snagu otkrivanja DIF-a u realnim podacima, dok je MIMIC model imao izvrsnu, dapače, stopostotnu snagu otkrivanja DIF-a, ali praćenu povišenom stopom pogrešaka tipa I. Samo IRT-LR metoda imala je neadekvatnu snagu otkrivanja DIF-a.

Važno je, doduše, na umu imati to da su rezultati simulacijskog istraživanja dobiveni u različitim, unaprijed poznatim, uvjetima u kojima su manipulirane različite vrste i veličine DIF-a, različit omjer veličina grupa, prisustvo odnosno odsustvo stvarnih grupnih razlika i prisustvo odnosno odsustvo korelacija među latentnim osobinama. Osim toga, svi ovi različiti uvjeti su replicirani 100 puta, što je računalni ekvivalent provedbe svakog eksperimenta (pojedinačne kombinacije nezavisnih varijabli koje su definirale svaki pojedini simulacijski scenarij) na 100

različitih uzoraka. S obzirom na kompleksnost simulacijskih uvjeta pod kojima su se metode uspoređivale ograničena je mogućnost generalizacije nalaza na stvarne testne situacije.

U empirijskom istraživanju smo, s druge strane, imali samo jedan uzorak a od uvjeta je u početku bilo poznato samo to da su referentna i fokalna skupina nejednakih veličina, te da postoji korelacija među latentnim osobinama. Vrsta DIF-a određena je nakon provedenih metoda, a veličina DIF-a nije mogla biti izračunata, jer za višedimenzionalne podatke još nisu razvijene mjere veličine efekata.

Imajući sve navedeno na umu, treba biti oprezan u usporedbi nalaza dobivenih u simulacijskim i stvarnim uvjetima.

5. RASPRAVA

Metode za otkrivanje DIF-a tradicionalno su se primjenjivale na jednodimenzionalne mjerne instrumente ili uz pretpostavku jednodimenzionalnosti podataka. Ovakav pristup, iako koristan u određenim kontekstima, ne odražava složenost mnogih psihologijskih instrumenata koji su inherentno višedimenzionalni. Takvi instrumenti su konstruirani za procjenu složenih konstrukata te se obično sastoje od više skupina čestica koje mjere različite latentne osobine (odnosno subskele). Pretpostavka jednodimenzionalnosti, stoga, često ne odgovara stvarnoj strukturi podataka, što može dovesti do netočnih zaključaka o prisutnosti DIF-a (Ackerman i sur., 2003).

Naime, jednodimenzionalna teorija odgovora na zadatak smatra da je već samo prisustvo višedimenzionalnosti u česticama indikator postojanja DIF-a, odnosno da je druga dimenzija izvor te pristranosti (tzv. ometajući faktor ili dimenzija). Važno je stoga razlikovati namjernu od nenamjerne višedimenzionalnosti u mjernim instrumentima (Oshima i sur., 1997). Kada je instrument nenamjerno višedimenzionalan, on je zapravo konstruiran u svrhu mjerenja samo jedne latentne osobine, no pojedine njegove čestice su slučajno osjetljive i na druge ometajuće dimenzije, koje su onda izvor DIF-a u tim česticama. Namjerno višedimenzionalni instrumenti konstruirani su u svrhu mjerenja različitih latentnih osobina odnosno više od jedne ciljane dimenzije. Primjena jednodimenzionalnih modela kod namjerno višedimenzionalnih instrumenata može biti izvor pogrešaka u procjenama parametara i otkrivanju DIF-a (Snow i Oshima, 2009).

Višedimenzionalna teorija odgovora na zadatak (MIRT), iako relativno nov pristup, smatra se adekvatnim načinom za modeliranje kompleksnih instrumenata u kojima čestice mjere više od jedne latentne osobine. Glavna razlika MIRT modela u odnosu na jednodimenzionalne IRT modele je u broju latentnih dimenzija kojima se objašnjava funkcioniranje odgovora na česticu.

U ovom istraživanju korištena je višedimenzionalna verzija modela stupnjevitih odgovora (Samejima, 1969) koji je jedan od nekoliko IRT modela koji se koriste za politomne čestice. Zapravo je riječ o adaptaciji dvoparametarskog logističkog IRT modela za dihotomne čestice za upotrebu s politomnim česticama. Kao i kod 2-PL modela, karakteristike čestice koje se ovim modelom uključuju u objašnjenje vjerojatnosti određenog odgovora na čestici su diskriminativnost čestice (a -parametar), te parametri limena (engl. *threshold parameters*) odnosno lokacije. Dok u 2-PL modelu postoji samo jedan parametar lokacije, b -parametar (koji se u tom slučaju naziva težinom čestice), u GRM-u, kao i u ostalim politomnim modelima, postoji $k-1$ parametara (k označava broj kategorija odgovora). Ovaj model dopušta da b -parametri variraju među česticama, te ih ne fiksira na jednu vrijednost identičnu za sve čestice, ali, za razliku od nekih drugih modela, u njega nije uključen c -parametar, parametar pogađanja. Samejimin GRM i sve njegove ekstenzije koriste se u analizi instrumenata kod kojih se odgovori formiraju u obliku ordinalnih politomnih čestica, kao što su primjerice skale procjene Likertovog tipa (Samejima, 1969; Samejima, 1972).

Imajući u vidu prethodno navedeno, upotreba metoda za analizu DIF-a namijenjenih jednoj latentnoj osobini (kao što su Mantel-Haenszelova metoda, simultani test pristranosti čestica (SIBTEST) i Lordova metoda hi-kvadrata) na višedimenzionalnim podacima može dovesti do krivog proglašavanja čestice DIF česticom, dok ustvari ona samo namjerno mjeri više od jedne dimenzije (Bulut i Suh, 2017). Zato je neophodno pronaći adekvatnu metodu detekcije DIF-a pomoću koje se ispitanici uparaju temeljem zajedničke distribucije latentnih osobina, a kako bi bili usporedivi na svim primarnim latentnim osobinama (Bulut i Suh, 2017; Mazor i sur., 1998).

Dosadašnja istraživanja o detekciji DIF-a u višedimenzionalnim kontekstima su rijetka, ali ključna za razumijevanje kako različite metode funkcioniraju u složenijim mjernim okvirima. Većina dosadašnjih istraživanja usmjerila se na detekciju DIF-a u jednodimenzionalnim instrumentima, a tek je nekoliko metoda adaptirano za detekciju DIF-a u višedimenzionalnim podacima. Višedimenzionalne adaptacije tradicionalnih DIF metoda, kao što su višedimenzionalni SIBTEST, DFIT, logistička regresija, test omjera vjerojatnosti i MIMIC modeli, predstavljaju značajan napredak u ovoj oblasti.

Među njima IRT-LR, logistička regresija i MIMIC model smatraju se najlakšima i najjednostavnijima za primjenu i na dihotomnim i na politomnim podacima (Bulut i Suh, 2017). Bulut i Suh su 2017. godine usporedili učinkovitost ovih metoda u višedimenzionalnom latentnom prostoru na dihotomnim podacima. Svrha ove disertacije bila je proširiti usporedbu višedimenzionalnih verzija MIMIC modela, IRT-LR i logističke regresije na politomne podatke. Pri tome je logistička regresija upotrijebljena u oba svoja pojavna oblika, koristeći kao varijablu za uparivanje i bruto rezultate i procjenu latentne osobine.

U simulacijskom dijelu istraživanja komparirane su ove četiri metode za otkrivanje DIF-a u višedimenzionalnim politomnim podacima u nizu simulacijskih scenarija. Koristeći Monte Carlo simulacije, istraživanje je usmjereno na dubinsko razumijevanje kako različiti kompleksni uvjeti utječu na sposobnost svake od ovih metoda da precizno identificira diferencijalno funkcioniranje čestica. Glavni razlozi za odabir simuliranih podataka uključuju izazove u prikupljanju empirijskih podataka, kao što su visoki troškovi i praktične poteškoće pri manipulaciji s brojnim nezavisnim varijablama (npr. veličina uzorka, dužina testa, karakteristike čestica), te ograničenja u istraživanju stvarnog utjecaja ovih varijabli bez točnog poznavanja parametara čestica i raspodjele latentnih osobina ispitanika. Dodatno, empirijski podaci često sadrže nepotpune informacije, što može narušiti valjanost rezultata, posebno kada postoji sistematsko nedostajanje podataka.

U ovom dijelu istraživanja, simulirani su podaci koji odražavaju dvodimenzionalni mjerni instrument s politomnim česticama i jednostavnom strukturom, gdje svaka čestica pripada samo jednoj latentnoj dimenziji. Parametri čestica generirani su na način sličan metodama korištenim u prethodnim istraživanjima, s diskriminativnošću čestica nasumično određenom iz uniformne distribucije i parametrima težine kategorija odgovora formiranim tako da održavaju ordinalnu strukturu.

Simulacijski faktori uključuju različite vrste DIF-a (bez DIF-a, jednoliki DIF, nejednoliki DIF), različite veličine DIF-a, veličinu uzorka i omjer veličina referentne i fokalne grupe, te prisutnost ili odsutnost stvarnih razlika u sposobnostima između grupa. Osim toga, ispitan je i utjecaj korelacije između latentnih osobina, s ciljem razumijevanja kako ovi čimbenici utječu na učinkovitost metoda za detekciju DIF-a.

Simulacijski faktori odnosno nezavisne varijable bili su kombinirani u ukupno 40 različitih scenarija. Svaki od ovih scenarija repliciran je 100 puta za svaku od metoda koje su se uspoređivale, pri čemu je ovaj broj replikacija odabran na temelju uobičajene prakse u sličnim istraživanjima provedenim u razdoblju od 1990. do 2018. godine (Berrio i sur., 2020).

Da bi se osigurala jasnoća i jednostavnost interpretacije rezultata, nekoliko faktora bilo je konstantno kroz sve scenarije. Duljina instrumenta, broj kategorija odgovora, i broj čestica s DIF-om držani su fiksima. Instrument je obuhvatio ukupno 28 čestica, raspoređenih jednako među dvije latentne dimenzije, sa 5 kategorija odgovora po čestici. Ovaj nacrt odražava strukturu i dužinu instrumenta sličnu onoj korištenoj u prethodnim istraživanjima efikasnosti metoda za detekciju DIF-a, a s druge strane omogućava direktnu usporedbu s instrumentom koji je korišten u drugom dijelu istraživanja. Konkretno, u istraživanju je izabran broj od dvije čestice s DIF-om, kako bi se olakšala usporedba s pristupom koji su koristili Bulut i Suh (2017).

Evaluacija metoda temeljila se na statističkoj snazi (ili osjetljivosti) testa i stopi pogrešaka tipa I. Statistička snaga odnosi se na sposobnost metode da točno identificira DIF čestice, dok stopa pogreške tipa I pokazuje koliko često metoda neopravdano označava česticu kao DIF česticu kada DIF nije prisutan. Ovi kriteriji su ključni za procjenu kvalitete i pouzdanosti metoda detekcije DIF-a.

Kroz Monte Carlo simulacije, istraživanje je nastojalo pružiti dublje razumijevanje kako različiti simulacijski uvjeti utječu na učinkovitost metoda za detekciju DIF-a. Time se omogućila usporedba metodologija u kontekstima koji bi bili teško dostupni ili neizvedivi u tradicionalnim empirijskim studijama, čime se doprinosi boljem razumijevanju mogućnosti i ograničenja pojedinih pristupa u detekciji DIF-a.

Analiza pogreške tipa I u simulacijskom istraživanju usmjerila se na određivanje koja od tri metode (IRT-LR, MIMIC, LR) najučinkovitije smanjuje neispravne detekcije DIF-a u različitim simuliranim scenarijima. Analize su se fokusirale na situacije bez simuliranog DIF-a, gdje su varijable kao što su prisutnost stvarnih grupnih razlika, omjer veličina referentne i fokalne skupine te korelacija među latentnim osobinama bili ključni faktori u ocjenjivanju metoda.

Rezultati su pokazali da su obje vrste logističke regresije i IRT-LR metoda generalno dobro kontrolirale pogrešku tipa I, što implicira njihovu pouzdanost u izbjegavanju lažnih detekcija DIF-a. IRT-LR metoda se istaknula s najmanjim brojem nepravilno detektiranih DIF čestica u svim simuliranim uvjetima. Ovaj konzistentno nizak broj pogrešaka tipa I ukazuje na visoku robusnost IRT-LR metode, čineći je pouzdanim izborom za istraživače koji teže smanjenju rizika od lažnih pozitivnih nalaza. Osim toga, logistička regresija, kako u varijanti s bruto rezultatima tako i s procjenom latentne osobine, je također pokazala zadovoljavajuće

rezultate. Obje varijante logističke regresije bile su učinkovite u detekciji DIF-a, što implicira da se mogu koristiti kao efikasne alternative IRT-LR metodi, osobito u situacijama gdje je potrebna fleksibilnost u obradi podataka ili kada su dostupni različiti tipovi podataka.

Međutim, MIMIC metoda pokazala je povišene stope pogrešaka tipa I u uvjetima s prisutnom stvarnom razlikom u sposobnostima među grupama. To ukazuje na manju robusnost MIMIC metode u takvim scenarijima, što je bitan nalaz s obzirom na čestu prisutnost grupnih razlika u stvarnim podacima. Povećana stopa pogrešaka tipa I u ovim scenarijima sugerira da interakcijska MIMIC metoda može nepravilno interpretirati strukturalne razlike između grupa kao DIF, što može dovesti do pogrešnih zaključaka o pristranosti čestica. Mješovita ANOVA pokazala je kako je interakcijska MIMIC metoda u tim uvjetima imala znatno više pogrešaka tipa I od ostale tri metode, i to neovisno o omjeru veličina grupa. Uz to, u situacijama u kojima nije bilo korelacije među latentnim osobinama, te je omjer grupa bio uravnotežen, samo je ova metoda imala znatno više pogrešaka tipa I. Oba ova nalaza upućuju da je potreban dodatni oprez prilikom korištenja interakcijske MIMIC metode u praksi, naglašavajući važnost detaljne provjere modela i prilagodbe analitičkih strategija specifičnostima podataka prije donošenja konačnih zaključaka.

Povećanu pogrešku tipa I kod MIMIC metode primijetili su i Gelin i Zumbo (2007) u svom simulacijskom istraživanju. Naime, oni su MIMIC koristili kod kratkih instrumenata, skala s 10 i 20 ordinalnih čestica, te ustanovili da je stopa lažnih detekcija u svim simuliranim uvjetima i neovisno o metodi procjene parametara (ML ili WLS), bila povećana. S obzirom na to, preporučili su oprez u interpretaciji rezultata o identificiranom DIF-u kada se koristi ova metoda. Iako je u Finchevom istraživanju (2005) MIMIC dobro kontrolirao stopu pogrešaka tipa I i kod kraćih i kod duljih instrumenata kojima je u podlozi 2PL model, i Finch je dobio slične nalaze kada je simuliran kratki instrument (20 čestica) s 3PL modelom. Razlika oba prethodna istraživanja u odnosu na ovo je u tome što se u ovom istraživanju koristio MIMIC model s interakcijom.

Uključivanje interakcija u MIMIC model ne samo da unosi dodatnu složenost već i stavlja povećan naglasak na važnost precizne specifikacije i identifikacije modela. Kada se u MIMIC model dodaju interakcijski termini, precizna specifikacija modela postaje zahtjevnija. Interakcije mogu značajno promijeniti kako model interpretira veze među varijablama, što zahtijeva pažljivo razmatranje koje interakcije uključiti, kako bi se izbjegla pretjerana složenost ili pogrešna interpretacija uzročno-posljedičnih veza. Osim toga, problemi sa specifikacijom mogu dovesti do poteškoća u identifikaciji modela, gdje model možda neće moći jednoznačno procijeniti parametre zbog nedostatnih informacija ili zbog prekomjerne prilagodbe podacima.

Također i identifikacija modela, proces kojim se osigurava da su parametri modela jedinstveno određeni, može postati izazovnija s interakcijama. Ovo je zato što dodavanje interakcija povećava broj parametara koje model treba procijeniti, što može dovesti do situacija gdje model ima više parametara nego što podaci mogu pouzdano podržati. To može rezultirati nestabilnim procjenama i posljedično većom stopom pogrešaka tipa I.

Dakle, interakcije u model unose dodatnu kompleksnost, a ona povećava osjetljivost metode na razlike koje inače ne bi bile statistički značajne. Povećana osjetljivost MIMIC metode zamijećena je i u istraživanju Carol Woods i Kevina Grimma (2011)(Woods & Grimm, 2011) u kojem je prvi puta upotrijebljen MIMIC model s interakcijom kao metoda za otkrivanje nejednolikog DIF-a. Naime, i u tom istraživanju stope pogrešaka tipa I bile su iznad dozvoljene granične vrijednosti Bradleyevog liberalnog kriterija u svim simuliranim uvjetima, za razliku od klasične MIMIC metode i IRT-LR metode koje su ih dobro kontrolirale. U istraživanju Lee i sur.(2016) MIMIC-interakcijska metoda je pak dobro kontrolirala stopu pogrešaka tipa I. No, za razliku od ovog i istraživanja Woods i Grimma, oni su koristili prilagodbe p -vrijednosti radi višestrukih usporedbi (konkretno BH postupak).

Ponešto neočekivani rezultati vezani su uz pojavu povišenih stopi pogrešaka tipa I u nekim simulacijskim situacijama koje su imale uravnotežen odnos grupa. Tako je recimo jedina povišena vrijednost kod IRT-LR metode vezana uz situaciju u kojoj nije bilo korelacije među latentnim varijablama, nisu postojale razlike u grupnim sposobnostima i odnos grupa je bio uravnotežen. Logistička regresija s bruto rezultatima je imala povišene vrijednosti pogreške u uvjetu s grupnim razlikama u sposobnosti i uravnoteženim odnosom veličine grupa, a interakcijska MIMIC metoda u uvjetu s uravnoteženim grupama i bez korelacije među latentnim osobinama. Iako bi ovakvi rezultati mogli biti simulacijski i metodološki artefakt, jedno od mogućih objašnjenja je i to da je povećanju stopi pogrešaka uzrok povećanje ukupnog usporedivog uzorka. Dok je u situaciji R1500/F500 ukupni broj usporedbi 500, kod R1000/F1000 on je dvostruko veći. Iako mali uzorak općenito utječe na lošije procjene, što onda rezultira time da se čestice koje različito funkcioniraju ne uspijevaju otkriti, a veći uzorak povećava preciznost njihovog otkrivanja, on također povećava i vjerojatnost lažne identifikacije čestica bez DIF-a ili čestica s malim, praktički nevažnim DIF-om. Drugim riječima, povećanje uzorka utječe na povećanje statističke snage otkrivanja efekta, ali i na povećanje osjetljivosti metoda na proglašavanje malih, praktički neznačajnih razlika, DIF-om (Jodoin i Gierl, 2001; Kim, 2010; Zumbo, 1999).

Što se snage tiče, rezultati su pokazali da su sve četiri metode imale odličnu statističku snagu otkrivanja umjerenog jednolikog DIF-a. No, kada je jednoliki DIF bio nizak, sve su metode bile lošije, a pogotovo u uvjetima s neuravnoteženim veličinama grupa.

Nijedna metoda nije imala zadovoljavajuću razinu snage otkrivanja niskog nejednolikog DIF-a, a sve su metode funkcionirale bolje kad je nejednoliki DIF bio umjerene veličine. Logistička regresija s procjenom latentne osobine je i ovdje imala najbolje stope detekcije, a MIMIC metoda je bila najneuspješnija.

Logistička regresija s procjenom latentne osobine bila je najbolja u svim simuliranim uvjetima, dok je IRT-LR imala najlošije stope snage. Ovi rezultati su u suprotnosti s istraživanjem Buluta i Suh (Bulut i Suh, 2017), koji su uspoređivali iste četiri metode na višedimenzionalnim podacima, ali s dihotomnim česticama. Naime, u njihovom je istraživanju IRT-LR metoda pokazala više stope točne identifikacije DIF-a od obje vrste logističke regresije i MIMIC modela s interakcijom, te je njihov zaključak bio da je IRT-LR preferirana metoda u detekciji jednolikog i nejednolikog DIF-a. S druge strane, Zimbra (2018) je dobio rezultate

koje su slični ovom istraživanju. I u njegovom istraživanju IRT-LR je pokazao povećanje stope snage s porastom veličine DIF-a, neovisno o njegovoj vrsti. No, u tom istraživanju uspješnost detekcije IRT-LR bila je adekvatna samo u uvjetima velikog DIF-a (0.7 logita razlike između referentne i fokalne skupine), dok su stope snage u otkrivanju niskog i umjerenog DIF-a bile neadekvatne. Također, u istraživanju Starka i suradnika (2006), IRT-LR je imala niske stope točnih identifikacija u uvjetima s niskim DIF-om. Niske stope snage IRT-LR metode istraživanju Lopez Rivas i sur.(2009) objašnjene su karakteristikama skupa sidrišnih čestica. Naime, u uvjetima malog DIF-a snaga metode se značajno poboljšala kada su u sidrišnom skupu korištene čestice visoke diskriminativnosti.

Istraživanje nije otkrilo utjecaj korelacije među latentnim osobinama na učinkovitost metoda. Iako se pretpostavljalo da će veća korelacija među latentnim osobinama izazvati manji broj točnih identifikacija, s obzirom na to da kod veće korelacije među osobinama postoji i veće preklapanje u distribucijama latentnih osobina, taj efekt, kao ni u istraživanju Buluta i Suha (2017) nije dobiven. Naime, visoka korelacija među latentnim osobinama znači da će ispitanici koji su na jednoj osobini imali visok ukupni rezultat imati visok ukupni rezultat i na drugoj, zbog čega će obrazac bodova biti sličan u obje latentne osobine. Zbog te sličnosti, metode mogu imati poteškoća u razlučivanju da li je grupna razlika u uspješnosti na čestici uzrokovana stvarnim razlikama u sposobnosti, DIF-om ili je samo artefakt visoke korelacije. No, dosadašnja istraživanja utjecaja korelacije među latentnim osobinama dala su suprotstavljene rezultate (Kunina-Habenicht i sur., 2012; Svetina i sur., 2018), a u ovom se istraživanju korelacija nije pokazala značajnim faktorom.

Na učinkovitost metoda su utjecali i veličina i vrsta DIF-a, te su ti rezultati u skladu s onima iz prethodnih istraživanja, koja su također pokazala da metode generalno nemaju poteškoća s detekcijom velikog DIF-a, pogotovo kad je on jednolik.

Kada je riječ o jednolikom DIF-u, sve su metode pokazale relativno dobre stope snage u uvjetima umjerenog DIF-a, pritom gotovo dosežući 100% uspješnost detekcije. Međutim, kod niskog DIF-a, stope detekcije su bile niže, a posebno u situacijama s neuravnoteženim veličinama grupa. Logistička regresija s procjenom latentne osobine pokazala je najbolje rezultate u detekciji jednolikog DIF-a u većini situacija, posebno u uvjetima s uravnoteženim veličinama grupa. Ova metoda je pokazala dosljedno visoku točnost u detekciji DIF-a, što sugerira da je učinkovita u različitim uvjetima simulacije.

Kada je u pitanju nejednoliki DIF, stope detekcije su bile generalno niže nego kod jednolikog DIF-a, a niti jedna od analiziranih metoda nije postigla zadovoljavajuću razinu snage detekcije niskog nejednolikog DIF-a. Najviše stope detekcije postigla je logistička regresija s bruto rezultatima u situaciji s uravnoteženim veličinama grupa, no, ovaj nalaz je potrebno uzeti s rezervom, s obzirom da je u tom uvjetu ova metoda imala i povišene stope pogrešaka tipa I.

Prisustvo stvarnih razlika u grupnim sposobnostima utjecalo je najviše na učinkovitost interakcijske MIMIC metode, koja je u tim uvjetima imala značajno manje točnih

identifikacija. Na ostale metode ova varijabla nije utjecala, pa se može reći da su one po tom pitanju robusnije.

U situacijama u kojima su grupe bile jednake veličine, sve su metode bile jednako uspješne, no u situacijama s nejednakim grupama, učinkovitost metoda ovisila je o ostalim varijablama. Kod niskog jednolikog DIF-a, logistička regresija s procjenom latentne osobine bila je najučinkovitija metoda, a IRT-LR najneučinkovitija. Kod umjerenog jednolikog DIF-a, nije bilo značajne razlike u funkcioniranju metoda. Obje su vrste logističke regresije bile podjednako dobre u otkrivanju nejednolikog DIF-a, neovisno o njegovoj veličini, a IRT-LR je i u ovom slučaju bila inferiornija. Nalaz da su stope detekcije manje u uvjetima grupa nejednake veličine ne iznenađuje, s obzirom da, ukoliko postoje velike razlike u veličinama grupa, ukupno je manji uzorak usporedivih ispitanika u obje grupe na svakoj razini ukupnog rezultata (Kristjansson i sur., 2005).

Nalazi ovog istraživanja idu u prilog logističke regresije s procjenom latentne varijable kao najučinkovitije metode za otkrivanje DIF-a u politomnim višedimenzionalnim podacima s jednostavnom strukturom. Ova metoda bila je uspješna u otkrivanju i jednolikog i nejednolikog DIF-a, a u usporedbi s drugim metoda bila je posebno učinkovita u otkrivanju niskog DIF-a. Osim toga, pokazala se relativno otpornom na stvarne razlike u grupnim sposobnostima, a bila je i najbolji izbor kad su veličine grupa bile nejednake.

S obzirom da veličina i vrsta DIF-a često nisu unaprijed poznate, čini se da bi ova vrsta logističke regresije mogla biti najpouzdaniji izbor kada imamo višedimenzionalne instrumente s jednostavnom strukturom. Osim toga, IRT-LR se, bez obzira na nisku stopu lažno pozitivnih nalaza, čini kao najmanje učinkovit izbor, iako treba biti oprezan pri tumačenju snage interakcijske MIMIC metode, jer ona može biti umjetno povećana zbog utjecaja visoke stope pogreški tipa I.

U drugom dijelu istraživanja korišteni su empirijski podaci slične strukture. Podaci su prikupljeni Upitnikom emocionalne kompetentnosti, koji sadrži politomne čestice s 5 kategorija (Likertova skala), te je iz prethodnih istraživanja bilo poznato da mu je struktura višedimenzionalna (Takšić i sur., 2013). S obzirom da su prethodni rezultati koji su govorili o višedimenzionalnosti dobiveni pomoću metoda klasične testne teorije, najprije se isto htjelo provjeriti metodologijom iz okvira teorije odgovora na zadatak.

S obzirom da su podaci prikupljeni tijekom niza istraživanja provedenih u različitim zemljama, kako bi se pristupilo provjeri dimenzionalnosti i diferencijalnog funkcioniranja čestica, najprije je bilo potrebno pronaći mjerno invarijantan uzorak. To se učinilo upotrebom napredne metode iz područja mrežne psihometrije, eksploratorne grafičke analize (EGA). Mrežni modeli (EGA) i modeli latentnih varijabli (CFA), iako različiti, vrlo su usko povezani te često proizvode slične parametre modela (Christensen i Golino, 2021; Golino i Epskamp, 2017; Golino i sur., 2020). EGA je u nekoliko prethodnih istraživanja pokazala superiornost u odnosu na tradicionalne tehnike faktorske analize (Cosemans i sur., 2022; Golino i sur., 2020; Golino i Epskamp, 2017; Jiménez i sur., 2023), a posebno u slučajevima s većim faktorskim strukturama i značajnim korelacijama između faktora, kao i sa srednjim do visokim

faktorskim opterećenjima. Njezina jednostavnost, odsustvo potrebe za odlukama o rotaciji faktora, te intuitivna priroda ističu se kao njezine jedinstvene prednosti.

Analiza mjerne invarijantnosti s *EGAnet* paketom imala je za cilj pronaći kros-kulturalno invarijantan uzorak kako bi se na njemu mogla provesti provjera dimenzionalnosti i kasnije provjere diferencijalnog funkcioniranja čestica s obzirom na varijablu spola.

Eksplozivna grafička analiza pokazala je konfiguralnu invarijantnost UEK-27 među četiri kros-kulturalna uzorka (Slovenija, Španjolska, Kina, Argentina), dok su druge zemlje eliminirane zbog nedostatka očekivane konfiguracije. Metrička invarijantnost, testirana permutacijskim testovima, potvrđena je za sve čestice u navedenim uzorcima, nakon korekcije *p*-vrijednosti Benjamini-Hochbergovim postupkom. Osim toga, trofaktorsko je rješenje pokazalo visoku stabilnost u bootEGA analizi.

Daljnja provjera dimenzionalnosti upitnika pomoću IRT-a otkrila je višedimenzionalnost instrumenta, te tako potvrdila prethodne nalaze klasične testne teorije koji su već ukazivali na postojanje više faktora. Naime, klasična testna teorija, uz upotrebu konfirmatorne faktorske analize, prethodno je identificirala tri važna faktora koji odgovaraju trima subskalama upitnika. Ova tri faktora su međusobno korelirana, što je sugeriralo postojanje generalnog faktora emocionalne kompetentnosti. I IRT je potvrdio postojanje korelacije među tri faktora, zbog čega bi u daljnjim istraživanjima bilo dobro provjeriti mogućnost da je struktura UEK-27 bifaktorska.

Višedimenzionalni model stupnjevitih odgovora (mGRM) pokazao se primjerenijim od jednodimenzionalnih modela, prema pokazateljima modeliranja kao što su C2, RMSEA_{C2}, i SRMSR. Osim toga, analiza je pokazala da UEK-27 najbolje diferencira ispitanike s ispodprosječnim i prosječnim razinama emocionalne kompetentnosti, dok je za one s iznadprosječnom razinom manje informativan. To znači da upitnik pruža više informacija o razlikama u emocionalnoj kompetentnosti u ovom rasponu, što je posebno korisno za identificiranje ispitanika koji možda trebaju podršku ili intervencije za poboljšanje svojih emocionalnih vještina.

Za ispitanike s iznadprosječnom razinom emocionalne kompetentnosti, upitnik postaje manje informativan. To bi moglo značiti da su čestice upitnika konstruirane na način koji bolje obuhvaća niže i srednje razine emocionalne kompetentnosti, dok možda nedostaju čestice koje bi mogle diferencirati visoke razine kompetentnosti. U kontekstu emocionalne inteligencije, ovo može sugerirati potrebu za razvojem dodatnih čestica koje ciljaju na više aspekte emocionalne kompetentnosti ili revidiranje postojećih čestica kako bi se poboljšala njihova sposobnost diferencijacije na cijelom spektru emocionalne kompetentnosti.

Isto je vidljivo i uvidom u parametre čestica. Naime, parametri čestica dobiveni višedimenzionalnim modelom upućuju na visoku do vrlo visoku diskriminativnost većine čestica, uz dobru pokrivenost širokog spektra mjerenih latentnih osobina. Međutim, parametri lokacije sugeriraju da neke čestice, posebno one sa Skale sposobnosti regulacije i upravljanja emocijama, nisu dovoljno funkcionalne za diferenciranje ispitanika s višim razinama emocionalne kompetentnosti. To se događa kada čestice ne omogućavaju adekvatno

diferenciranje među ispitanicima na gornjem kraju skale emocionalne kompetentnosti, jer većina ispitanika bira više kategorije odgovora, implicirajući da niže kategorije odgovora za te čestice ne pružaju dodatne informacije.

Dobiveni nalazi iz simulacijskog dijela istraživanja i njihova potvrda na realnim podacima ukazuju na važnost konteksta i prirode podataka prilikom odabira metodologije za detekciju DIF-a. Ovo je posebno bitno jer različite metode imaju svoje specifične prednosti i ograničenja koje mogu varirati ovisno o karakteristikama podataka.

Inflacija stope pogrešaka tipa I kod interakcijske MIMIC metode, koja je zabilježena i u simulacijskom i u realnom kontekstu, može se djelomično objasniti njezinom visokom osjetljivošću na specifikaciju modela. MIMIC modeli su složeni i mogu biti osjetljivi na prisutnost neprepoznatih višedimenzionalnih struktura unutar podataka, što može rezultirati lažno pozitivnim nalazima. S obzirom na to, važno je detaljno procijeniti i testirati model prije njegove primjene na specifičan skup podataka, posebno ako on uključuje više latentnih dimenzija koje nisu direktno povezane s konstruktima koji se ispituju.

Zanimljivo je promatrati kako se rang učinkovitosti metoda mijenja između simulacijskog i realnog konteksta. Naime, logistička metoda s bruto rezultatima upotrijebljena na realnim podacima nadmašila je ostale metode, dok je s druge strane simulacija favorizirala logističku metodu s latentnom procjenom. Ova promjena može biti rezultat specifičnih karakteristika realnih podataka, poput distribucije odgovora ili prisutnosti nekontroliranih ometajućih varijabli, koje nisu bile prisutne ili su bile kontrolirane u simulacijskom okruženju. Stoga, ovi nalazi naglašavaju važnost validacije metodoloških pristupa na širokom spektru realnih podataka prije donošenja zaključaka o njihovoj generalnoj primjenjivosti.

Dok simulacijske studije omogućuju kontrolu nad mnogim aspektima istraživačkog nacrtu i mogu replicirati eksperimentalne uvjete s visokom točnošću, one također imaju određena ograničenja. Konkretno, simulacije se temelje na pretpostavkama koje možda u potpunosti ne odražavaju stvarnost. Primjerice, pretpostavke o raspodjeli latentnih osobina, veličini DIF-a, i slično, mogu biti prilično pojednostavnjene ili jednostavno neprikladne za stvarne situacije koje su mnogo složenije. Zbog toga, iako su simulacije vrijedne za teoretsku validaciju metodoloških pristupa, empirijsko istraživanje pruža neophodan test njihove primjenjivosti u praksi.

Empirijsko istraživanje, s druge strane, suočava se s izazovima poput ograničenih informacija o podacima i složenosti realnih skupova podataka. Međutim, pruža neprocjenjiv uvid u to kako metode funkcioniraju kada se primjenjuju na stvarne probleme, uključujući interakcije koje možda nisu bile evidentne u kontroliranom okruženju simulacije. Empirijski nalazi mogu otkriti potrebu za prilagodbom ili rafiniranjem metodoloških pristupa kako bi se poboljšala njihova točnost i pouzdanost u specifičnim kontekstima primjene.

Na kraju, razlike u nalazima između simulacijskih i stvarnih uvjeta naglašavaju važnost kombiniranja teoretskih, simulacijskih, i empirijskih pristupa u metodološkom istraživanju. Kritičko vrednovanje i prilagodba metodoloških pristupa, uzimajući u obzir specifičnosti pojedinog istraživanja, ključno je za pouzdanu detekciju DIF-a.

6. DOPRINOSI I OGRANIČENJA ISTRAŽIVANJA

U simulacijskom dijelu istraživanja uspoređivana je efikasnost različitih metoda za identifikaciju DIF-a u politomnim višedimenzionalnim podacima s jednostavnom strukturom, u kojima su sve čestice instrumenta pripadale samo jednoj od dvije simulirane latentne osobine. No, u stvarnosti, višedimenzionalni politomni podaci mogu imati kompleksnu strukturu, s više od dvije latentne osobine, što može utjecati na učinkovitost metoda. Ova složenost može utjecati na sposobnost metoda da precizno detektira DIF, budući da interakcije među više latentnih dimenzija mogu otežati identifikaciju čestica koje nisu invarijantne. Buduća istraživanja trebala bi istražiti kako različite strukture latentnih prostora utječu na detekciju DIF-a, što bi moglo dovesti do razvoja sofisticiranijih metoda prilagođenih složenijim podacima.

Osim toga, isto istraživanje se isključivo bavilo grupnim razlikama u težini čestice (jednoliki DIF) ili diskriminativnosti čestice (nejednoliki DIF), no, nejednoliki se DIF može manifestirati i kao razlika u oba parametra. Razlikovanje između jednolikog i nejednolikog DIF-a ključno je za razumijevanje kako se pristranost manifestira unutar čestica mjernog instrumenta. Dok je jednoliki DIF odraz konzistentne pristranosti na svim razinama sposobnosti ispitanika, nejednoliki DIF ukazuje na pristranost koja se mijenja ovisno o razini sposobnosti. Pristup koji uzima u obzir samo jednu vrstu pristranosti može propustiti složenije obrasce DIF-a, posebno kada čestice pokazuju pristranost u oba parametra (težina i diskriminativnost). Istraživanje ovih složenijih oblika DIF-a također bi moglo dovesti do razvoja preciznijih metoda detekcije.

Uz to, distribucije latentnih osobina mogu imati različite obrasce, koje mogu biti nesimetrične ili višemodalne, a postoje i različiti modeli za generiranje podataka koji bi se mogli koristiti u budućim istraživanjima kako bismo dobili bolju sliku o ovom fenomenu. S obzirom da različiti modeli za generiranje simuliranih podataka mogu imati različite pretpostavke o distribucijama i interakcijama među česticama i latentnim dimenzijama, razmatranje ovih varijabli može obogatiti razumijevanje učinkovitosti metoda za detekciju DIF-a pod raznovrsnim uvjetima.

Još jedan nedostatak simulacijskog istraživanja je i upotreba isključivo statističke značajnosti za definiranje DIF-a, no, iz praktičnih bi razloga upotreba pokazatelja veličine efekta bila korisnija. Pokazatelji veličine efekta mogu pomoći u procjeni koliko je DIF značajan u kontekstu primjene testa. Uključivanje ovih pokazatelja u analizu omogućava informiranje odluke o tome treba li i kako modificirati čestice mjernog instrumenta. S obzirom da za ovu vrstu podataka pokazatelji veličine efekta ili nisu razvijeni ili još uvijek nisu lako dostupni i primjenjivi, buduća istraživanja mogla bi se pozabaviti njihovim razvijanjem i upotrebom.

Također, u istraživanju su korištena samo dva omjera veličina grupa (1:1 i 3:1), dok se u stvarnim uvjetima testiranja javljaju mnogi različiti omjeri fokalne i referentne grupe. Različiti omjeri veličina fokalne i referentne grupe mogu značajno utjecati na snagu detekcije DIF-a. Većina istraživanja ograničava se na usporedbu nekoliko omjera, što ne odražava širok

raspon omjera koji se mogu pojaviti u stvarnosti, stoga bi neko buduće istraživanje utjecaja različitih omjera na sposobnost detekcije DIF-a moglo pružiti uvide u to kako optimalno konfigurirati istraživanja kojima je cilj detekcija DIF-a u mjernim instrumentima.

Radi pojednostavljenja nacрта broj čestica (dužina testa), broj kategorija odgovora kod politomnih čestica, i broj DIF čestica držani su konstantnima, no to su također čimbenici koji mogu utjecati na učinkovitost metoda.

Osim toga, prilikom izračunavanja stopa statističke snage i pogrešaka tipa I nije korištena neka od korekcija za višestruke usporedbe (npr. Bonferroni ili B-H postupak). S obzirom da u analizama u kojima se provode mnoge statističke usporedbe postoji povećan rizik od pogrešaka tipa I, primjena korektivnih postupaka, kao što su Bonferroni ili Benjamini-Hochberg, može pomoći u kontroli ovog rizika. Buduća istraživanja trebala bi razmotriti uključivanje takvih korektivnih mjera za rigorozniju statističku analizu.

S obzirom da je u fokusu istraživanja bila usporedba metoda i njihove učinkovitosti za detekciju DIF-a u politomnim višedimenzionalnim instrumentima, ovo istraživanje nije koristilo tehnike purifikacije skale za odabir sidrišnih čestica koje ne sadrže DIF, nego je pretpostavljeno da su ispravne sidrišne čestice unaprijed poznate. No, kod empirijski prikupljenih podataka, ne postoji način da se unaprijed zna da li su sidrišne čestice zapravo bez DIF-a i upravo stoga se koristi purifikacijski proces. Buduća istraživanja mogu se usmjeriti na metode za purifikaciju skala u kontekstu politomnih višedimenzionalnih podataka. Osim toga, bilo bi korisno saznati kako različiti omjeri veličina grupa utječu na proces purifikacije skale, s obzirom da bi to moglo imati značajan učinak na učinkovitost otkrivanja DIF-a.

Što se pak drugog dijela istraživanja tiče, njegov je doprinos, osim naravno provjere nalaza dobivenih simulacijskim istraživanjem na empirijskim podacima, u tome što je Upitnik emocionalne kompetentnosti ovim istraživanjem prvi puta analiziran u okviru teorije odgovora na zadatak. Ovaj pristup omogućava detaljniju i matematički rigorozniju analizu svojstava čestica upitnika nego što je to slučaj s tradicionalnim pristupima poput klasične teorije testova. Teorija odgovora na zadatak omogućava procjenu kako čestice razlikuju ispitanike s različitim razinama latentne osobine (u ovom slučaju, emocionalne kompetentnosti), što daje dublji uvid u diskriminativnost i težinu svake čestice, kao i njezinu informativnost, tj. doprinos cjelokupnom upitniku. Ovo je posebno korisno za daljnje fino podešavanje upitnika s ciljem poboljšanja njegove preciznosti i pouzdanosti, što su također i preduvjeti za bolju valjanost kao najvažniju metrijsku osobinu.

Osim toga, korištenjem metodologije iz okvira teorije odgovora na zadatak potvrđena je višedimenzionalnost upitnika, koja je dosad proučavana samo u okviru klasične teorije testova. Na ovaj način je dan doprinos razumijevanju strukture konstrukta emocionalne kompetentnosti, te je potvrđeno postojanje više latentnih faktora (dimenzija) koji su međusobno povezani, ali i različito doprinose emocionalnoj kompetentnosti.

Eksploratorna grafička analiza dala je dodatnu potvrdu teorijskoj strukturi upitnika UEK-27, te je osim toga pokazala postojanje konfiguralne i metričke invarijantnosti upitnika u 4 od 13 zemalja (Slovenija, Španjolska, Kina i Argentina). Uzroci nedostatka konfiguralne

invarijantnosti u ostalih 9 zemalja nisu bili u fokusu ovog istraživanja, ali bi ih bilo dobro provjeriti u budućim istraživanjima.

Međutim, empirijsko istraživanje se suočava s nekoliko ograničenja koja utječu na interpretaciju i generalizaciju nalaza. Korištenje samo jednog uzorka u empirijskom istraživanju može ograničiti mogućnost generalizacije rezultata na različite uvjete testiranja i različite populacije. Ovo ograničenje zahtijeva oprez u interpretaciji rezultata i njihovoj primjeni u različitim kontekstima.

Nadalje, nedostatak razvijenih i lako dostupnih mjera veličine efekata za višedimenzionalne podatke otežava precizno kvantificiranje praktičnog utjecaja diferencijalnog funkcioniranja čestica (DIF), što je ključno za točno određivanje stvarnog efekta DIF-a.

Kompleksnost i nepredvidivost stvarnih podataka, u usporedbi s kontroliranim simulacijskim uvjetima, dodatno kompliciraju analizu i interpretaciju nalaza, budući da stvarni podaci često sadrže nepredvidive varijable i složene interakcije koje nisu prisutne u simulacijama. Konačno, neadekvatna kontrola uvjeta u empirijskim istraživanjima također otežava usporedbu nalaza dobivenih u različitim testnim situacijama.

Buduća istraživanja trebala bi nastaviti s istraživanjem i razvojem metodologija za detekciju DIF-a, s posebnim naglaskom na rješavanje ovih ograničenja i proširenje primjenjivosti nalaza u različitim testnim situacijama.

7. ZAKLJUČAK

Suvremena obrazovna i psihologijska testiranja sve više prepoznaju višedimenzionalnost inherentnu ljudskim kognitivnim sposobnostima, osobinama ličnosti i drugim konstruktima povezanim s ljudskim funkcioniranjem. Tradicionalni jednodimenzionalni pristupi testiranju i analizi skloni su pretjeranom pojednostavljivanju konstrukata koje mjere, što može dovesti do netočnih ili nepotpunih tumačenja rezultata testova. Višedimenzionalni modeli mjernih instrumenata prepoznaju ovu složenost dopuštajući procjenu više osobina ili sposobnosti istovremeno, što rezultira puno jasnijim razumijevanjem pojedinca koji se procjenjuje.

Globalna i multikulturalna priroda suvremenog društva znači da su populacije ispitanika koje pristupaju različitim vrstama testiranja danas raznovrsnije nego ikada prije. Ova raznolikost uključuje, ali nije ograničena na, jezične, kulturalne, obrazovne i socioekonomske razlike. Takva raznolikost povećava rizik od različitog funkcioniranja testova za različite grupe ispitanika - fenomen poznat kao diferencijalno funkcioniranje čestica (DIF). Detekcija DIF-a tako postaje jedan od ključnih aspekata nužnih za osiguranje pravednosti i valjanosti procjena.

Tradicionalne metode detekcije DIF-a nastale su u okviru jednodimenzionalnog pristupa testiranju, te su primarno bile primjenjive za dihotomne čestice. Osim toga, svaku registriranu višedimenzionalnost u česticama pripisivale su DIF-u. Kao takve one često nisu adekvatne za identifikaciju DIF-a u višedimenzionalnim politomnim česticama. Ovakve čestice pak mogu otkriti suptilnije razlike među ispitanicima i nude veću količinu informacija o konstruktima koji se mjeri, no s obzirom na veći broj parametara koji sadrže, postoji i veći broj mogućih uzroka diferencijalnom funkcioniranju takvih čestica.

S obzirom na vrijednost koju donose višedimenzionalni instrumenti koji koriste politomne čestice, bilo je potrebno razviti i metode koje mogu otkrivati različito funkcioniranje čestica u takvim kontekstima. Prilagođavanjem višedimenzionalnoj prirodi konstrukata i nijansiranim odgovorima koje politomne čestice mogu registrirati, ove metode mogu točnije identificirati gdje se DIF pojavljuje. To zatim vodi do boljih odluka o prilagodbi ili uklanjanju čestica kako bi se osigurala pravednost i valjanost testa.

Ukratko, potreba za višedimenzionalnim politomnim metodama detekcije DIF-a vođena je povećanom složenošću mjernih konstrukata, raznolikošću populacija ispitanika i ograničenjima tradicionalnih metoda detekcije DIF-a. Usvajanjem ovih novih metoda, istraživači i praktičari mogu poboljšati točnost, pravednost i valjanost procjena, i pritom osigurati da one nisu pristrane.

U simulacijskom dijelu ovog istraživanja analizirane su i međusobno uspoređene četiri metode detekcije DIF-a prilagođene za upotrebu na višedimenzionalnim instrumentima s politomnim česticama. Koristeći Monte Carlo simulacije, istraživanje je pokušalo razumjeti kako različiti složeni uvjeti utječu na sposobnost svake od ovih metoda da precizno identificira čestice koje diferencijalno funkcioniraju, tj. koje nisu invarijantne. Ove metode, pažljivo odabrane zbog njihove mogućnosti i jednostavnosti primjene u analizi politomnih podataka i

potencijala u višedimenzionalnom kontekstu, provjeravane su u nizu rigorozno dizajniranih simulacijskih scenarija.

Simulacijski scenariji su kreirani tako da odražavaju širok spektar stvarnih i teoretskih situacija, uključujući različite kombinacije vrste i veličine DIF-a, omjera veličina referentne i fokalne skupine, prisustva stvarnih razlika među grupama i prisustva korelacije među latentnim osobinama. Ovi scenariji su omogućili detaljan uvid u funkcioniranje metoda u različitim uvjetima, pritom nudeći objektivnu procjenu njihove učinkovitosti i pouzdanosti.

Osim toga, Monte Carlo simulacije su, kao metoda izbora za ovaj dio istraživanja, omogućile generiranje širokog raspona podataka pod kontroliranim uvjetima, što je značajno doprinijelo razumijevanju dinamike i izazova povezanih s detekcijom DIF-a u višedimenzionalnim politomnim kontekstima. To je omogućilo preciznu manipulaciju ključnim varijablama, te analizu utjecaja svakog pojedinog faktora na točnost i pouzdanost izabranih metoda detekcije DIF-a.

Ovaj pristup ne samo da je pružio vrijedne uvide u sposobnosti i ograničenja pojedinih metoda u različitim scenarijima, već je također naglasio važnost rigoroznog testiranja i validacije metoda detekcije DIF-a prije njihove primjene u empirijskim istraživanjima. Konačni cilj ovog dijela istraživanja bio je identificirati metode koje su najpouzdanije i najučinkovitije u identificiranju DIF-a u višedimenzionalnim politomnim podacima, te tako pružiti smjernice za buduća istraživanja i praksu u psihometriji.

Stoga su simulacijskim istraživanjem uspoređene četiri trenutno dostupne metode detekcije DIF-a u višedimenzionalnim politomnim podacima. Riječ je o dvije verzije logističke regresije, konkretno logističkoj regresiji s bruto rezultatima i logističkoj regresiji s procjenom latentne osobine, MIMIC modelu s interakcijom i omjeru izglednosti iz okvira teorije odgovora na zadatak. Njihova učinkovitost ocijenjena je upotrebom stope statističke snage odnosno točnih identifikacija i stope pogrešaka tipa I odnosno lažnih detekcija u različitim simuliranim scenarijima. Nezavisne varijable čiji utjecaj na učinkovitost metoda se provjeravao bile su: 1) vrsta i veličina DIF-a, 2) omjer veličina referentne i fokalne skupine, 3) prisustvo stvarnih razlika među grupama; i 4) prisustvo korelacije među latentnim osobinama.

Rezultati su pokazali da je logistička regresija s procjenom latentnih osobina imala najvišu stopu točnih identifikacija DIF-a, čime se istakla kao metoda izbora u psihometrijskoj praksi.

Empirijski dio istraživanja, temeljen na podacima prikupljenim upitnikom UEK-45 u 13 zemalja, omogućio je stvarni uvid u primjenjivost metoda i istaknuo potrebu za njihovim prilagođavanjem specifičnim kontekstima.

Analiza UEK-27 pokazala je visoku pouzdanost i informativnost upitnika za ispitanike s ispodprosječnim i prosječnim nivoima emocionalne kompetentnosti, dok se informativnost upitnika smanjuje za ispitanike s iznadprosječnom emocionalnom kompetentnošću, konkretno za vrijednosti θ iznad +2. Ovo ukazuje na potrebu za daljnjim prilagodbama čestica

upitnika, posebno u smislu poboljšanja diferencijacije na višim razinama emocionalne kompetentnosti.

Također, korištenjem metodologije iz okvira teorije odgovora na zadatak potvrđena je višedimenzionalnost upitnika, koja je dosad proučavana samo u okviru klasične teorije testova. Upotrebom eksploratorne grafičke analize ustanovljena je metrička invarijantnost upitnika u 4 od 13 zemalja.

Provjera metoda detekcije diferencijalnog funkcioniranja čestica na empirijskim podacima nije u potpunosti potvrdila nalaze dobivene simulacijskim istraživanjem. Naime, na podacima metrijski invarijantnog uzorka UEK-27 najefikasnijom metodom detekcije DIF-a pokazala se logistička regresija s bruto rezultatima, dok su interakcijska MIMIC metoda i IRT-LR metoda pokazale manju efikasnost i pouzdanost.

S druge strane, nalazi simulacijskog istraživanja, koji su pokazali povišenu stopu pogrešaka tipa I za interakcijsku MIMIC metodu i slabu snagu otkrivanja DIF-a za IRT-LR metodu, potvrđeni su i na realnim podacima. Međutim, dok je u simulaciji najučinkovitija bila logistička regresija s latentnom procjenom, empirijski podaci su pokazali veću učinkovitost logističke regresije s bruto rezultatima.

Obje vrste logističke regresije pokazale su umjerenu statističku snagu u empirijskim uvjetima, dok je MIMIC model imao izvrsnu, ali prekomjerno liberalnu snagu otkrivanja DIF-a, praćenu visokom stopom pogrešaka tipa I. IRT-LR metoda se istaknula kao najmanje učinkovita zbog svoje neadekvatne snage otkrivanja DIF-a.

Važno je naglasiti da su simulacijski uvjeti bili konstruirani s precizno kontroliranim varijablama i replikacijama, što otežava generalizaciju nalaza na stvarne uvjete testiranja. Empirijsko istraživanje je provedeno u stvarnim uvjetima s neizvjesnim parametrima, što predstavlja izazov u interpretaciji i primjeni ovih metoda. Stoga treba biti oprezan pri usporedbi nalaza dobivenih u kontroliranim simulacijskim uvjetima i onih iz realnih testnih okruženja.

Ovim istraživanjem znatno se doprinijelo razumijevanju kako različite metode funkcioniraju u teorijskim i praktičnim okvirima, što ima važne implikacije za praksu u obrazovnim i psihometrijskim testiranjima, posebno u kontekstu postizanja veće valjanosti i pravednosti donesenih zaključaka.

Ovo istraživanje također naglašava važnost adekvatnog odabira metode ovisno o specifičnim karakteristikama podataka i uvjetima testiranja. S obzirom na složenost i različite efekte koje svaka metoda pokazuje u različitim uvjetima, ključno je izabrati metodu koja je najprikladnija za specifične zahtjeve istraživanja, kako bi se osigurala visoka pouzdanost i valjanost rezultata mjerenja.

8. LITERATURA

- Ackerman, T. A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective. *Journal of Educational Measurement*, 29(1), 67–91. Dohvaćeno iz <http://www.jstor.org/stable/1434777>
- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 20, 309–310.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*, 22, 37-53.
- Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23.
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. (A. E. Association, A. P. Association, & N. C. Education, Ur.) American Educational Research Association.
- Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594.
- Ankenmann, R., Witt, E., & Dunbar, S. (1999). An investigation of the power of the likelihood ratio goodness-of-fit statistics in detecting differential item functioning. *Journal of Educational Measurement*, 4, str. 277-300. doi:<https://doi.org/10.1111/j>.
- Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 37-48.
- Baker, F. B. (1961). Empirical comparison of item parameters based on the logistic and normal functions. *Psychometrika*, 3, str. 239–246.
- Baker, F. B. (2001). *The basics of item responses theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Barendse, M. T., Oort, F. J., Werner, C. S., Ligtoet, R., & Schermelleh-Engel, K. (2012). Measurement bias detection through factor analysis. *Structural Equation Modeling*, 4, str. 561–579. doi:<https://doi.org/10.1080/10705511.2012.713261>
- Barton, M. A., & Lord, F. M. (1981). *An upper asymptote for the three-parameter logistic item-response model*. Princeton, NJ: Educational Testing Service.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*.

Series B (Methodological), 1, str. 289–300. Dohvaćeno iz
<http://www.jstor.org/stable/2346101>

- Berrio, M., Gomez-Benito, J., & Arias, V. B. (2020). Developments and trends in research on methods of detecting differential item functioning. *Educational Research Review*, 31. doi:<https://doi.org/10.1016/j.edurev.2020.100340>
- Birnbaum, A. (1957). *Efficient design and use of tests of a mental ability for various decision-making problems*. Randolph Air Force Base, TX: USAF School of Aviation Medicine.
- Birnbaum, A. (1958a). *On the Estimation of Mental Ability*. Randolph Air Force Base, Texas: USAF School of Aviation Medicine.
- Birnbaum, A. (1958b). *Further Considerations of Efficiency in Tests of a Mental Ability*. Randolph Air Force Base, Texas: USAF School of Aviation Medicine.
- Bjorner, J. B., Smith, K. J., Edelen, M. O., Stone, C., Thissen, D., & Sun, X. (2007). *IRTFIT: A macro for item fit and local dependence tests under IRT models*. QualityMetric Incorporated.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443–459.
- Bock, R. D., & Liberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 283–319.
- Bolt, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education*, 2, str. 113-141. doi:https://doi.org/10.1207/S15324818AME1502_01
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, 27(6), 395-414.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, str. 144-152. doi:10.1111/j.2044-8317.1978.tb00581.x
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Bulut, O. (2013). Between-person and within-person subscore reliability: comparison of unidimensional and multidimensional IRT models. *Unpublished doctoral dissertation*. University of Minnesota Digital Conservancy. Dohvaćeno iz <https://hdl.handle.net/11299/155592>.

- Bulut, O., & Suh, Y. (2017). Detecting Multidimensional Differential Item Functioning with the Multiple Indicators Multiple Causes Model, the Item Response Theory Likelihood Ratio Test, and Logistic Regression. *Frontiers in Education*, 51. doi:<https://doi.org/10.3389/feduc.2017.00051>
- Bulut, O., & Sunbul, Ö. (2017). Monte Carlo simulation studies in item response theory with the R programming language. *Journal of Measurement and Evaluation in Education and Psychology*, 3, str. 266–287.
- Cai, L., & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *The British Journal of Mathematical and Statistical Psychology*, 2, str. 245-276. doi:<https://doi.org/10.1111/j.2044-8317.2012.02050.x>
- Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESST Report 839)*. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Camilli, G., & Shepard, L. (1994). *Methods for Identifying Biased Test Items*. SAGE Publications.
- Cao, M., Tay, L., & Liu, Y. (2017). A Monte Carlo Study of an Iterative Wald Test Procedure for DIF Analysis. *Educational and Psychological Measurement*, 1, str. 104-118. doi:<https://doi.org/10.1177/0013164416637104>
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 6, str. 1-29. doi:<https://doi.org/10.18637/jss.v048.i06>
- Chan, D. (2000). Detection of differential item functioning on the Kirton Adaptation-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research*, str. 169-199. doi:10.1207/S15327906MBR3502_2
- Chang, C., & Reeves, B. (2005). Item Response Theory and its applications to patient-reported outcomes measurement. *Evaluation and the Health Professions*, 28, 264-282.
- Chang, H.-H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for Polytomously Scored Items: An Adaptation of the SIBTEST Procedure. *Journal of Educational Measurement*, 3(33), str. 333-353. Dohvačeno iz <http://www.jstor.org/stable/1435376>
- Chang, Y. W. (1992). A comparison of unidimensional and multidimensional IRT approaches to test information in a test battery. *Unpublished Doctoral Dissertation*. University of Minnesota ProQuest Dissertations Publishing.
- Cheng, Y., Pérez-Díaz, P. A., Petrides, K. V., & Li, J. (2023). Monte Carlo simulation with confusion matrix paradigm - A sample of internal consistency indices. *Frontiers in psychology*, 14, 1298534. <https://doi.org/10.3389/fpsyg.2023.1298534>

- Chihara, L. M., & Hesterberg, T. C. (2018). *Mathematical statistics with resampling and R*. Wiley.
- Christensen, A. P., & Golino, H. (2021). On the equivalency of factor and network loadings. *Behavior Research Methods*, 4, str. 1563–1580. doi:<https://doi.org/10.3758/s13428-020-01500-6>.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied psychological measurement*, 41(3), 178–194. doi:<https://doi.org/10.1177/014662161666>
- Chun, S., Stark, S., Kim, E. S., & Chernyshenko, O. S. (2016). MIMIC methods for detecting DIF among multiple groups: Exploring a new sequential-free baseline procedure. *Applied psychological measurement*, 7, str. 486-499.
- Chun, S., Stark, S., Kim, E. S., & Chernyshenko, O. S. (2016). MIMIC Methods for Detecting DIF Among Multiple Groups: Exploring a New Sequential-Free Baseline Procedure. *Applied Psychological Measurement*, 7, str. 486-499.
- Clauser, B. E., & Mazor, K. M. (1998). Using Statistical Procedures To Identify Differentially Functioning Test Items. An NCME Instructional Module. *Educational Measurement: issues and practice*, 17(1), 31-44.
- Cohen, A. S., Kim, S.-H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*, 1, str. 15-26. doi:<https://doi.org/10.1177/014662169602000102>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 1, str. 155.
- Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement*, 5, str. 880-910. doi:<https://doi.org/10.1177/00131644211059089>.
- Crocker, L. M., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.
- de Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, 25, 172-189.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- de Ayala, R. J., & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, 23, 3-19.

- de Boeck, P., & Wilson, M. (2004). *Explanatory Item Response Models: A Generalized and Nonlinear Approach*. New York: Springer-Verlag.
- de la Tore, J., & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item response theory test scoring. *Journal of Educational and Behavioral Statistics*, 30(3), 295-311.
- deMars, C. (2010). *Item Response Theory*. Oxford University Press.
- Dorans, N. J., & Kingston, N. M. (1958). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. *Journal of Educational Measurement*, 4, str. 249–262.
- Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied psychology*, 72(1), 19-29.
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 3, str. 357-381. doi:<https://doi.org/10.1177/0013164498058003001>
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement*, 4, str. 278–295.
- Fox, J.-P. (2007). Multilevel IRT Modeling in Practice with the Package mlirt. *Journal of Statistical Software*, 5, str. 1-16. doi:<https://doi.org/10.18637/jss.v020.i05>
- French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, str. 315-332.
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, str. 373-393.
- Gelin, M. N., & Zumbo, B. D. (2007). Operating characteristics of the DIF MIMIC approach using Jöreskog's covariance matrix with ML and WLS estimation for short scales. *Journal of Modern Applied Statistical Methods*, 2. doi:<https://doi.org/10.56801/10.56801>

- Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*, *6*. doi:<https://doi.org/10.1371/journal.pone.0174035>
- Golino, H., & Christensen, A. P. (2019). EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics. Dohvaćeno iz <https://CRAN.R-project.org/package=EGAnet>
- Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., . . . Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *3*, str. 292-320.
- Hallquist, M. N., & Wiley, J. F. (2018). Mplus Automation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, *4*, str. 621-638. doi:<https://doi.org/10.1080/10705511.2017.1402334>
- Hambleton, R. K. (1989). *Principles and selected applications of item response theory* (Svez. Educational Measurement). (R. L. Linn, Ur.) New York: Macmillan.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *3*, str. 38-47. doi:<https://doi.org/10.1111/j.1745-3992.199>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, Inc.
- Hasmy, A. (2014). Compare unidimensional & multidimensional Rasch model for test with multidimensional construct and items local dependence. *Journal of Education and Learning*, *3*, str. 187-194.
- Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *2*, str. 139-164. doi:<https://doi.org/10.1177/014662168500900204>
- Holland, P. W., & Thayer, D. T. (1985). *An Alternate Definition of the ETS Delta Scale of Item Difficulty*. Technical Report. doi:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2330-8516.1985.tb00128.x>
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and Mantel-Haenszel procedure. U H. Wainer, & H. I. Braun (Ur.), *Test validity* (str. 129-145). Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and Mantel-Haenszel procedure. U H. Wainer, & H. I. Braun, *Test validity* (str. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
- Hong, T., Wu, N., Maller, S. J., & Pei, L. (2008). Assessing DIF in polytomous items using the MIMIC modeling approach. *Paper presented at the annual meeting of the National Council on Measurement in Education*. New York.
- Hou, L., la Torre, J. D., & Nandakumar, R. (2014). Differential item functioning assessment in cognitive diagnostic modeling: Application of the Wald test to investigate DIF in the DINA model. *Journal of Educational Measurement, 1*, str. 98–125.
- Jamison, L., Golino, H., & Christensen, A. P. (26. May 2022). Metric Invariance in Exploratory Graph Analysis via Permutation Testing. *PsycArxiv*.
doi:<https://doi.org/10.31234/osf.io/j4rx9>
- Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology, 109*. doi:<https://doi.org/10.3389/fpsyg.2016.00109>
- Jiménez, M., Abad, F. J., Garcia-Garzon, E., Golino, H., Christensen, A. P., & Garrido, L. E. (2023). Dimensionality assessment in bifactor structures with multiple general factors: A network psychometrics approach. *Psychological Methods*.
- Jin, Y., Myers, M. D., Ahn, S., & Penfield, R. D. (2012). A comparison of uniform DIF effect size estimators under the MIMIC and Rasch models. *Applied Psychological Measurement, 73*, str. 339-358.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education, 4*, str. 329-349.
doi:<https://doi.org/10.1207/S15324818AME1404>
- Johnson, M. S. (2007). Marginal Maximum Likelihood Estimation of Item Response Models in R. *Journal of Statistical Software, 10*, str. 1-24.
doi:<https://doi.org/10.18637/jss.v020.i1>
- Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 351a*, str. 631–639.
- Kim, E. S., Yoon, M., & Lee, T. (2012). Testing measurement invariance using MIMIC likelihood ratio test with a critical value adjustment. *Educational and Psychological Measurement, 3*, str. 469-492.
- Kim, J. (2010). Controlling Type 1 Error Rate in Evaluating Differential Item Functioning for Four DIF Methods: Use of Three Procedures for Adjustment of Multiple Item Testing. *Doctoral dissertation, Georgia State University*. Georgia State University ScholarWorks. Dohvaćeno iz https://scholarworks.gsu.edu/eps_diss/67/

- Kim, S.-H., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement, 4*, str. 345-355. doi:<https://doi.org/10.1177/014662169802200403>
- Kim, S.-H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement, 2*, str. 93-116. doi:<https://doi.org/10.1111/j.1745-3984.2007.00029.x>
- Kirisci, L., Hsu, T.-c., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*(2), 2, str. 146–162.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 4*, str. 457–474.
- Kristjansson, E., Aylesworth, R., Mcdowell, I., & Zumbo, B. D. (2005). A comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement, 65*(6), 935-953.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. *Journal of Educational Measurement, 1*, str. 59-81. doi:[doi:10.1111/j.1745-3984.2011.00160.x](https://doi.org/10.1111/j.1745-3984.2011.00160.x)
- Langer, M. (2008). A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation. *Unpublished doctoral dissertation*. Chapel Hill: University of North Carolina.
- Le, L. T. (2006). Investigating gender differential item functioning across countries and test languages for PISA science items . *Paper Presented at the International Test Commission Conference, Brussels, Belgium*. Dohvačeno iz https://www.acer.org/files/gendif__pisa_trial_2006_itc_ver2.pdf
- Lee, S., Bulut, O., & Suh, Y. (2016). Multidimensional extension of multiple indicators multiple causes models to detect DIF. *Educational Psychological Measurement*. doi:[doi:10.1177/0013164416651116](https://doi.org/10.1177/0013164416651116)
- Lee, Y.-H., Jiao, H., & Wang, S. (2019). The Effects of Q-Matrix Misspecification on Parameter Estimation and Classification Accuracy in the DINA Model. *Frontiers in Psychology, str. 1137*. doi:<https://doi.org/10.3389/fpsyg.2019.01137>
- Lopez Rivas, G. E., Stark, S., & Chernyshenko, O. S. (2009). The effects of referent item parameters on differential item functioning detection using the free baseline likelihood ratio test. *Applied Psychological Measurement, 4*, str. 251–265.
- Lord, F. M. (1953). The Relation of Test Score to the Trait Underlying the Test. *Educational and Psychological Measurement, 13*(4), 517-549. doi:<https://doi.org/10.1177/001316445301300401>

- Lord, F. M. (1977). A study of item bias, using item characteristic curve theory. U Y. H. Poortinga, *Basic problems in cross-cultural psychology* (str. 19-29). Amsterdam.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 9.
- Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). *A review of recent developments in differential item functioning*. ETS Research Report Series.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187–212.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, 9, str. 1-21.
doi:<https://doi.org/10.18637/jss.v042.i09>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 149-174.
- Maydeu-Olivares, A., & Joe, H. (2006). Limited Information Goodness-of-fit Testing in Multidimensional Contingency Tables. *Psychometrika*, str. 713-732.
doi:<https://doi.org/10.1007/s11336-005-1295-9>
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? U P. Salovey, & D. Sluyter (Ur.), *Emotional development and emotional intelligence: Educational implications* (str. 3-31). New York: Basic Books.
- Mazor, K. M., Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF analyses: the effects of matching on unidimensional subtest scores. *Applied Psychological Measurement*, str. 357-367. doi:10.1177/014662169802200404
- McDonald, R. P. (1982). Unidimensional and multidimensional models for item response theory. U D. J. Weiss (Ur.), *Proceedings of the 1982 Computerized Adaptive Testing Conference* (str. 127-148). Minneapolis: University of Minnesota, Department of Psychology.
- McNamara, T., & Roever, C. (2006). Psychometric Approaches to Fairness: Bias and DIF. *Language Learning*, Suppl 2, str. 81–128. doi:<https://doi.org/10.1111/j.1467-9922.2006.00381.x>
- Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, 5, str. 1016-1031.
doi:<https://doi.org/10.1037/a0027934>

- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of educational statistics*, 7(2), 105-118.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International journal of educational research*, 13(2), 127-143.
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, str. 107-122.
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. NY: Taylor & Francis.
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied psychological measurement*, 17(4), 297-334.
- Montoya, A. K., & Jeon, M. (2020). MIMIC Models for Uniform and Nonuniform DIF as Moderated Mediation Models. *Applied psychological measurement*, 2, str. 118–136. doi:<https://doi.org/10.1177/014662161983549>
- Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. U R. C. R. W. Robins (Ur.), *Handbook of Research Methods in Personality Psychology* (str. 407-423). New York: Guilford.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14, 59-71.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational and Behavioral Statistics*, 2, str. 121-132.
- Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. U In H. Wainer & H. I. Braun (Eds.) (Ur.), *Test validity* (str. 213-238). Lawrence Erlbaum Associates, Inc.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, str. 557-585.
- Muthén, L. K., & Muthén, B. (1998-2017). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 3, str. 257–274. doi:<https://doi.org/10.1177/014662169602000306>

- Nordstokke, D. W., Zumbo, B. D., & Saklofske, D. H. (2011). The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data. *Practical Assessment, Research, and Evaluation, 1*.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, str.* 50-64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, str.* 289–298.
- Oshima, T. C., & Morris, S. B. (2008). An NCME instructional module on Raju's differential functioning of items and tests (DFIT). *Educational Measurement: Issues and Practice, 27*(3), 43-50.
- Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and Demonstration of Multidimensional IRT-Based Internal Measures of Differential Functioning of Items and Tests. *Journal of Educational Measurement, 3*, str. 253–272. Dohvačeno iz <http://www.jstor.org/stable/1435445>
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning (No. 7)*. Sage.
- Pae, T.-I., & Park, G.-P. (2006). Examining the relationship between differential item functioning and differential test functioning. *Language Testing, 4*, str. 475-496.
- Penfield, R. D., & Camilli, G. (2006). *Differential item functioning and item bias. Handbook of statistics (Svez. 26)*. 125-167.
- Penfield, R. D., & Lam, T. C. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice, 19*(3), 5-15.
- Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied psychological measurement, 19*(1), 23-37.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (Ur.). (2004). *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*. doi:doi:10.1007/BF02294403
- Raju, N. S., Van der Linden, W., & Fleer, P. F. (1995). IRT-Based Internal Measures of Differential Functioning of Items and Tests. *Applied Psychological Measurement, 4*, str. 353-368. doi:10.1177/014662169501900405.

- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* ((Expanded edition with foreword and afterword by B.D. Wright, Chicago: The University of Chicago Press) izd.). Danish Institute for Educational Research.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4), 401-412.
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21, 25-36.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York: Springer.
- Reckase, M. D., & McKinley, R. L. (1982). The feasibility of a multidimensional latent trait model. *aper presented at the annual meeting of the American Psychological Association*. Washington.
- Reckase, M., & Hirsh, T. (1991). Interpretation of number correct scores when the true number of dimensions assessed by a test is greater than two. *Annual meeting of the National Council on Measurement in Education*. Chicago, IL.
- Reise, S. P., & Revicki, D. A. (Ur.). (2014). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. Routledge.
- Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Analysis. *Journal of Statistical Software*, 5, str. 1-25.
doi:<https://doi.org/10.18637/jss.v017.i05>
- Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: the factor structure and item properties of the original and brief fear of negative evaluation scale. *Psychological assessment*, 2, str. 169.
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 2, str. 105–116. doi:
<https://doi.org/10.1177/014662169301700201>
- Rost, J. (1988). Measuring Attitudes With a Threshold Model Drawing on a Traditional Scaling Concept. *Applied Psychological Measurement*, 12(4), 397-409.
doi:<https://doi.org/10.1177/014662168801200408>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 1, str. 139.

- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph Supplement, 1, Pt. 2*, str. 68.
- Samejima, F. (1988). Comprehensive Latent Trait Theory. *Behaviormetrika, 15*, 1-24. doi:https://doi.org/10.2333/bhmk.15.24_1
- Samejima, F. (1996). Evaluation of Mathematical Models for Ordered Polychotomous Responses. *Behaviormetrika, 23*, 17-35. doi: <https://doi.org/10.2333/bhmk.23.17>
- Shealy, R. T., & Stout, W. F. (1991). *A procedure to detect test bias present simultaneously in several items*. Washington, DC: Office of Naval Research.
- Shealy, R. T., & Stout, W. F. (1993). A model-based standardization approach that true bias/DIF from group ability differences and detects test bias/DIF as well bias/DIF. *Psychometrika*, str. 159-194.
- Sheng, Y., & Wikle, C. K. (2007). Comparing Multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement, 67*(6), 899–919.
- Sijtsma, K., & Meijer, R. R. (2007). Nonparametric item response theory and special topics. U C. R. Rao, & S. Sinharary (Ur.), *Handbook of Statistics Vol. 26: Psychometrics* (str. 719-746). Amsterdam: Elsevier.
- Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research, 3*, str. 443–461. doi:<https://doi.org/10.1007/s11205-010-9682-8>
- Snow, T., & Oshima, T. (2009). A Comparison of Unidimensional and Three-Dimensional Differential Item Functioning Analysis Using Two-Dimensional Data. *Educational and Psychological Measurement, 5*, str. 732-747.
- Spray, J., & Miller, T. (1994). Identifying Nonuniform DIF in Polytomously Scored Test Items. *ACT Research Report Series 94-1*.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology, 89*(3), 497.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *The Journal of Applied Psychology, 6*, str. 1292-1306. doi:10.1037/0021-9010.91.6.1292
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, str. 589–617.

- Suh, Y., & Cho, S.-J. (2014). Chi-square difference tests for detecting functioning in a multidimensional IRT model: a Monte Carlo study. *Applied Psychological Measurement*, str. 359–375. doi:10.1177/0146621614523116
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement*, 27(4), 361-370.
- Sweeney, K. P. (1997). A Monte Carlo investigation of the likelihood-ratio procedure in the detection of differential item functioning. *Doctoral dissertation*. ETD Collection for Fordham University.
- Sympson, J. B. (1978). A model for testing with multidimensional items. U D. J. Weiss (Ur.), *Proceedings of the 1977 Computerized Adaptive Testing Conference Minneapolis* (str. 82-98). University of Minnesota, Department of Psychology, Psychometrics Methods Program.
- Takšić, V. (1998). Validacija konstrukta emocionalne inteligencije. *Neobjavljena doktorska disertacija*. Sveučilište u Zagrebu.
- Takšić, V. (2002). The importance of emotional intelligence (competence) in positive psychology. *Rad prezentiran na the First International Positive Psychology Summit, Washington, DC, 4-6.10.2002*.
- Takšić, V., Bradić, S., & Žauhar, V. (2013). Prikaz metoda za procjenu kroskulturalne ekvivalentnosti na primjeru upitnika emocionalne kompetentnosti. *Primenjena psihologija*, 2, str. 101-120.
- Takšić, V., Mohorić, T., & Duran, M. (2009). Emotional skills and competence questionnaire (ESCQ) as a self-report measure of emotional intelligence. *Psihološka obzorja (Horizons of psychology)*, 3, str. 7-21.
- Tay, L., Meade, A. W., & Cao, M. (2015). An Overview and Practical Guide to IRT Measurement Equivalence Analysis. *Organizational Research Methods*, 18(1), 3-46. doi:https://doi.org/10.1177/1094428114553062
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, 16, 33-42.
- Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in medicine*, 11-12, str. 1651–1683.
- Thissen, D. (1991). *MULTILOG™ User's Guide: Multiple, Categorical Item Analysis and Test Scoring Using Item Response Theory*. Chicago: Scientific Software Inc.
- Thissen, D. (2001). IRTLDRDIF v 2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning

- [Computer software]. *Documentation for computer program*. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. U D. Thissen, & H. Wainer (Ur.), *Test scoring* (str. 73-140). Lawrence Erlbaum Associates Publishers.
- Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika*, 49(4), 501-519. doi:<https://doi.org/10.1007/BF02302588>
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: the concept of item bias. *Psychological Bulletin*, str. 118-128. doi:10.1037/0033-2909.99.1.118
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of Item Response Theory in the Study of Group Differences in Trace Lines IRT DIF. U H. Wainer, & H. I. Braun (Ur.), *Test Validity* (str. 147-169). Lawrence Erlbaum Associates.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models IRT DIF. U P. W. Holland, & H. Wainer (Ur.), *Differential Item Functioning* (str. 67-113). Lawrence Erlbaum Associates,.
- Tian, F. (1999). Detecting DIF in polytomous item responses. [*Doctoral dissertation (Publication No. AAT 9902072) University of Ottawa*], 61, 02. Dissertation Abstracts International.
- Toland, M. D., Sulis, I., Giambona, F., Porcu, M., & Campbell, J. M. (2017). Introduction to bifactor polytomous item response theory analysis. *Journal of School Psychology*, str. 41-63. doi:<https://doi.org/10.1016/j.jsp.2016.11.001>
- van de Vijver, F. L., & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks, CA: Sage.
- van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 2, str. 89-99. doi:<https://doi.org/10.1027/1016-9040.1.2.89>
- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer.
- Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4-70. doi:<https://doi.org/10.1177/109442>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327.

- Wainer, H. (1993). Model-based standardized measurement of an item's differential impact. U P. W. Holland, & H. Wainer (Ur.), *Differential item functioning* (str. 123-135). Hillsdale NJ: Erlbaum.
- Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admissions Test as an example. *Applied Measurement in Education*, 2, str. 157-186.
- Wainer, H., & Kiely, G. L. (1987). Item Clusters and Computerized Adaptive Testing: A Case for Testlets. *Journal of Educational Measurement*, 185-201. doi:<http://www.jstor.org/stable/1434630>
- Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 3, str. 426-482. doi:<https://doi.org/10.2307/1990256>
- Walker, C. M., & Beretvas, S. N. (2000). Using multidimensional versus unidimensional ability estimates to determine student proficiency in mathematics. *Paper presented at the annual meeting of the American Educational Research Association*. New Orleans, LA.
- Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, 40, 255-275.
- Wang, W. C., & Shih, C. L. (2010). MIMIC methods for assessing differential item functioning in polytomous items. *Applied Psychological Measurement*, 3, str. 166-180.
- Wang, W. C., & Yeh, Y. L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, 6, str. 479-498.
- Wang, W. C., Shih, C. L., & Yang, C. C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 5, str. 713-731. doi:<https://doi.org/10.1177/0013164409332218>
- Wang, W., Chen, P., & Cheng, Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9, 116-136.
- Wanichtanom, R. (2001). Methods of detecting differential item functioning: A comparison of item response theory and confirmatory factor analysis . *Unpublished doctoral dissertation*. Old Dominion University.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimation. *Applied Psychological Measurement*, 12, 239-252.

- Wiberg, M. (2007). Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods. *Dissertation, Umea University*. Umea University Libraries EM No 60.
- Wilson, M. (1992). The ordered partition model: An extension of the partial credit model. *Applied Psychological Measurement, 16*, 309-325.
doi:doi:10.1177/014662169201600401
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Woods, C. M. (2008). Likelihood-Ratio DIF Testing: Effects of Nonnormality. *Applied Psychological Measurement, 32*(7), 511-526.
doi:https://doi.org/10.1177/0146621607310402
- Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate behavioral research, 1*, str. 1-27.
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied psychological measurement, 5*, str. 339-361.
- Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-Improved Wald Test for DIF Testing With Multiple Groups: Evaluation and Comparison to Two-Group IRT. *Educational and Psychological Measurement, 3*, str. 532-547.
doi:https://doi.org/10.1177/0013164412464875
- Wright, B. D. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on Testing Problems* (str. 85-101). Princeton, NI: Educational Test Service.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 2*, str. 125-145.
- Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement, 36*(5), 375–398.
- Zhang, Y., Dorans, N. J., & Matthews-López, J. L. (2005). *Using DIF dissection method to assess effects of item deletion*. ETS Research Report Series.
- Zimbra, D. J. (2018). An Examination of the MIMIC Method for Detecting DIF and Comparison to the IRT Likelihood Ratio and Wald Tests. *Unpublished doctoral dissertation*. Honolulu, HI: University of Hawai‘i at Mānoa.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary*

and Likert-type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, str. 233-251.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education*, 4, str. 321-344.

9. POPIS SLIKA

Slika 1. Karakteristične krivulje odgovora na česticu „Jeste li u protekla 4 tjedna zbog utjecaja bilo kakvih emocionalnih problema (npr. osjećaj depresije ili tjeskobe) na poslu ili pri obavljanju nekih drugih svakodnevnih aktivnosti obavili manje nego ste željeli?“	6
Slika 2. Prikaz čestica različite težine	8
Slika 3. Prikaz čestica različite diskriminativnosti	9
Slika 4. Informacijska funkcija	11
Slika 5. Karakteristična krivulja testa	12
Slika 6. Prikaz čestica s različitim parametrom pseudo-pogađanja	15
Slika 7. Usporedba dvoparametarskog (lijevo) i jednoparametarskog modela (desno).....	17
Slika 8. Grafički prikaz odnosa broja točnih odgovora (CTT) i latentne osobine.....	19
Slika 9. Usporedba 1PL/Raschovog, 2PL i 3PL modela kada se vjerojatnost točnog odgovora prikazuje u logitima	20
Slika 10. Prikaz politomne čestice modelirane pomoću modela stupnjevitih odgovora	21
Slika 11. Prikaz vjerojatnosti odabira određene kategorije politomne čestice prema modelu stupnjevitih odgovora.....	22
Slika 12. Prikaz politomne čestice modelirane pomoću generaliziranog modela djelomičnih odgovora	24
Slika 13. Plošni grafik (lijevo) i konturni grafik (desno) za vjerojatnost točnog odgovora na dvodimenzionalnoj kompenzacijskoj čestici s parametrima $a1=1.5$, $a2=.5$, $d=.7$	29
Slika 14. Plošni grafik (lijevo) i konturni grafik (desno) za vjerojatnost točnog odgovora na dvodimenzionalnoj nekompenzacijskoj čestici s parametrima $a1=1.5$, $a2=.5$, $b1=-1.0$, $b2=0.0$	30
Slika 15. Primjer međučestičnih i unutarčestičnih modela.....	31
Slika 16. Prikaz čestice s jednolikim (lijevo) i nejednolikim DIF-om (desno).....	38
Slika 17. Prikaz MIMIC modela s interakcijom za detekciju jednolikog (gore) i nejednolikog DIF-a.....	47
Slika 18. Prikaz stopa pogrešaka tipa I	67
Slika 19. Prikaz metrički invarijantnih uzoraka dobiven metodom eksploratorne grafičke analize	77
Slika 20. Rezultati bootEGA analize stabilnosti tro-faktorskog rješenja na mjerno-invarijantnom uzorku	78
Slika 21. Cattellov (scree) dijagram	80
Slika 22. Uvjetna pouzdanost UEK-27.....	87
Slika 23. Informacijska krivulja UEK-27	87
Slika 24. Karakteristična krivulja UEK-27	88

10. POPIS TABLICA

Tablica 1. <i>Prikaz simulacijskih scenarija u kojima su testirane metode detekcije DIF-a</i>	59
Tablica 2. <i>Broj ispitanika u pojedinim državama</i>	63
Tablica 3. <i>Stope lažnih detekcija (pogrešaka tipa I) u svim simuliranim uvjetima</i>	65
Tablica 4. <i>Stope snage (točnih identifikacija) u uvjetima s jednolikim DIF-om</i>	68
Tablica 5. <i>Stope snage (točnih identifikacija) u uvjetima s nejednolikim DIF-om</i>	70
Tablica 6. <i>Usporedba učinkovitosti metoda detekcije DIF-a u ovisnosti o stvarnim grupnim razlikama</i>	72
Tablica 7. <i>Usporedba učinkovitosti metoda detekcije DIF-a u ovisnosti o omjeru veličina grupa</i>	73
Tablica 8. <i>Rezultati eksplorativne grafičke analize UEK-27</i>	75-76
Tablica 9. <i>Testiranje primjerenosti podataka za faktorsku analizu - KMO mjera prikladnosti uzorkovanja i Bartlettov test</i>	77-79
Tablica 10. <i>Rezultati faktorske analize</i>	80
Tablica 11. <i>Usporedba prikladnosti različitih IRT modela za podatke dobivene UEK-27 upitnikom</i>	82
Tablica 12. <i>Usporedba parametara čestica Skale sposobnosti uočavanja i razumijevanja emocija prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora</i>	84
Tablica 13. <i>Usporedba parametara čestica Skale sposobnosti izražavanja i imenovanja emocija prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora</i>	85
Tablica 14. <i>Usporedba parametara čestica Skale sposobnosti regulacije i upravljanja emocijama prema jednodimenzionalnom i prema višedimenzionalnom modelu stupnjevitih odgovora</i>	86
Tablica 15. <i>Rezultati dvostupanjskih analiza diferencijalnog funkcioniranja čestica UEK-27</i>	90
Tablica 16. <i>Usporedba slaganja pojedinih metoda s konsenzualnim rješenjem</i>	91-92

11. POPIS JEDNADŽBI

(1) Dvoparametrijski logistični model vjerojatnosti odgovora „ne“ na dihotomnoj čestici	7
(2) Dvoparametrijski logistični model vjerojatnosti odgovora „da“ na dihotomnoj čestici	7
(3) Standardna pogreška mjerenja	11
(4) Četveroparametrijski logistični model za dihotomne čestice.....	16
(5) Troparametrijski logistični model za dihotomne čestice.....	16
(6) Dvoparametrijski logistični model za dihotomne čestice	16
(7) Jednparametrijski logistični model za dihotomne čestice	17
(8) Raschov model	18
(9) Raschov model u obliku logaritama izgleda	18
(10) Model stupnjevitih odgovora	23
(11) Generalizirani model djelomičnih odgovora.....	24
(12) Jednakost uvjetnih distribucija odgovora u referentnoj i fokalnoj skupini	36
(13) Logistička regresija	51

12. PRIVITCI

- Privitak 1. Kodovi za simulaciju podataka i provedene analize
- Privitak 2. Upitnik emocionalne kompetentnosti UEK-45
- Privitak 3. Slika 1. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti uočavanja i razumijevanja emocija
- Slika 2. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti izražavanja i imenovanja emocija
- Slika 3. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti regulacije i upravljanja emocijama
- Slika 4. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti uočavanja i razumijevanja emocija
- Slika 5. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti izražavanja i imenovanja emocija
- Slika 6. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti regulacije i upravljanja emocijama
- Privitak 4. Usporedni prikaz EGA rješenja svih 13 zemalja

Privitak 1: Kodovi za simulaciju podataka i provedene analize

```
###Detecting Differential Item Functioning in Multidimensional IRT
## IRT-LR method, conditions with impact:
library("doParallel")
library("mirt")
library("MASS")
library("readr")

detectDIF <- function(sample.size, DIF.size, cor, seed) {

  require("mirt")
  require("MASS")
  set.seed(seed)

  #Define multidimensional abilities for reference and focal groups
  theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
  theta.foc <- mvrnorm(n = sample.size[2], rep(-0.5, 2),
matrix(c(1,cor,cor,1),2,2))

  #Generate item parameters for reference and focal groups
  a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
  a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
  a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
  b1 <- runif(n = 28, min = 0.67, max = 2)
  b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
  b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
  b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
  b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)

  #Uniform and nonuniform DIF for items 1 and 28
  b.foc <- b.ref
  b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
  a.foc <- a.ref
  a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
  a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]
```

```

#Generate item responses according to MGRM
ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
dat <- rbind(ref, foc)

#Define the group variable (0=reference; 1=focal) and test DIF using mirt
group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))
itemnames <- colnames(dat)
model <- 'f1 = 1-14
          f2 = 15-28
          COV = F1*F2'
model.mgrm <- mirt.model(model)

#Test uniform DIF
if(DIF.size[1]>0 & DIF.size[2]==0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)],
                                       'free_means', 'free_var'), verbose = FALSE)

  mod1 <- DIF(mod0, c('d1','d2','d3','d4' ), items2test = c(1,28))

  result <- data.frame(item1_DIF=mod1[1,8],
                      item28_DIF=mod2[1,8])
} else

#Test nonuniform DIF
if(DIF.size[1]==0 & DIF.size[2]>0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)],
                                       'free_means', 'free_var'), verbose =
FALSE)

  mod1 <- DIF(mod0, c('a1'), items2test = c(1))
  mod2 <- DIF(mod0, c('a2'), items2test = c(28))

  result <- data.frame(item1_DIF=mod1[1,8],
                      item28_DIF=mod2[1,8])
} else

```

```

#Test type I error
if(DIF.size[1]==0 & DIF.size[2]==0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)],
                                       'free_means', 'free_var'), verbose =
FALSE)
  mod1 <- DIF(mod0, c('a1','d1','d2','d3','d4' ), items2test = c(1))
  mod2 <- DIF(mod0, c('a2','d1','d2','d3','d4'), items2test = c(28))

  result <- data.frame(item1_DIF=mod1[1,8],
                      item28_DIF=mod2[1,8])
}
return(result)
}

# Set the seed
myseed <- sample.int(n=10000, size = 100)
detectCores()
registerDoParallel(4)

### UNIFORM DIF WITH IMPACT

### SCENARIO 1

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```

```

results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsimsцен1.csv")

### SCENARIO 2

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```

write_csv(sum0128, file = "IRTLRsims scen2.csv")

### SCENARIO 5
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1", "item28")

item <- c(1,28)
type <- c("DIF", "DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <- data.frame(item, type, signif01, signif05, hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)), digits = 2)

write_csv(sum0128, file = "IRTLRsims scen5.csv")

### SCENARIO 6
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1", "item28")

item <- c(1,28)

```



```

type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen6.csv")

### SCENARIO 9
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1

```

```

    if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen9.csv")

### SCENARIO 10
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen10.csv")

### SCENARIO 13
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}

```

```

}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen13.csv")

### SCENARIO 14
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen14.csv")

### SCENARIO 17
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen17.csv")

```

SCENARIO 18

Run the simulation using parallel computing

```
result <- foreach (i = 1:100, .combine=rbind) %dopar% {  
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])  
}
```

```
colnames(result) <- c("item1","item28")
```

```
item <- c(1,28)
```

```
type <- c("DIF","DIF")
```

```
signif01 <- c(0,0)
```

```
signif05 <- c(0,0)
```

```
hitrate.05 <- c(0,0)
```

```
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(result)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "IRTLRsims scen18.csv")
```

SCENARIO 21

Run the simulation using parallel computing

```
result <- foreach (i = 1:100, .combine=rbind) %dopar% {  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])  
}
```

```
colnames(result) <- c("item1","item28")
```

```
item <- c(1,28)
```

```
type <- c("DIF","DIF")
```

```
signif01 <- c(0,0)
```

```
signif05 <- c(0,0)
```

```

hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen21.csv")

### SCENARIO 22
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen22.csv")

### NONUNIFORM DIF WITH IMPACT
### SCENARIO 25
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen25.csv")

### SCENARIO 26
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

```

```

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen26.csv")

```

SCENARIO 29

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

```

```

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1

```



```

    if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
    if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
    if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen29.csv")

### SCENARIO 30
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen30.csv")

### SCENARIO 33
# Run the simulation using parallel computing

```

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen33.csv")

```

SCENARIO 34

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```

```

results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen34.csv")

```

SCENARIO 37

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```

write_csv(sum0128, file = "IRTLRsims scen37.csv")

### SCENARIO 38
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen38.csv")

## IRT-LR method, conditions without impact:
detectDIF <- function(sample.size, DIF.size, cor, seed) {

  require("mirt")
  require("MASS")
  set.seed(seed)

  #Define multidimensional abilities for reference and focal groups
  theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))

```

```

theta.foc <- mvrnorm(n = sample.size[2], rep(0, 2), matrix(c(1,cor,cor,1),2,2))

#Generate item parameters for reference and focal groups
a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
b1 <- runif(n = 28, min = 0.67, max = 2)
b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)

#Uniform and nonuniform DIF for items 1 and 28
b.foc <- b.ref
b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
a.foc <- a.ref
a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]

#Generate item responses according to MGRM
ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
dat <- rbind(ref, foc)

#Define the group variable (0=reference; 1=focal) and test DIF using mirt
group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))
itemnames <- colnames(dat)
model <- 'f1 = 1-14
          f2 = 15-28
          COV = F1*F2'
model.mgrm <- mirt.model(model)

#Test uniform DIF
if(DIF.size[1]>0 & DIF.size[2]==0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)]),

```

```

FALSE)
                                'free_means', 'free_var'), verbose =

mod1 <- DIF(mod0, c('d1','d2','d3','d4' ), items2test = c(1,28))

result <- data.frame(item1_DIF=mod1[1,8],
                    item28_DIF=mod2[1,8])
} else

#Test nonuniform DIF
if(DIF.size[1]==0 & DIF.size[2]>0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)],
                                'free_means', 'free_var'), verbose =
FALSE)
  mod1 <- DIF(mod0, c('a1'), items2test = c(1))
  mod2 <- DIF(mod0, c('a2'), items2test = c(28))

  result <- data.frame(item1_DIF=mod1[1,8],
                      item28_DIF=mod2[1,8])
} else

#Test type I error
if(DIF.size[1]==0 & DIF.size[2]==0) {
  mod0 <- multipleGroup(data = dat, model = model.mgrm, group = group,
                        invariance = c(itemnames[-c(1,28)],
                                'free_means', 'free_var'), verbose =
FALSE)
  mod1 <- DIF(mod0, c('a1','d1','d2','d3','d4' ), items2test = c(1))
  mod2 <- DIF(mod0, c('a2','d1','d2','d3','d4'), items2test = c(28))

  result <- data.frame(item1_DIF=mod1[1,8],
                      item28_DIF=mod2[1,8])
}
return(result)
}

```

```

# Set the seed
myseed <- sample.int(n=10000, size = 100)
detectCores()
registerDoParallel(4)

### UNIFORM DIF WITHOUT IMPACT
# SCENARIO 3

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen3.csv")

### SCENARIO 4

```

```

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen4.csv")

### SCENARIO 7
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)

```



```

sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen7.csv")

### SCENARIO 8
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```

write_csv(sum0128, file = "IRTLRsims scen8.csv")

### SCENARIO 11
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1", "item28")

item <- c(1,28)
type <- c("DIF", "DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <- data.frame(item, type, signif01, signif05, hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)), digits = 2)

write_csv(sum0128, file = "IRTLRsims scen11.csv")

### SCENARIO 12
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1", "item28")

item <- c(1,28)

```

```

type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsimsцен12.csv")

```

SCENARIO 15

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

```

```

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1

```

```

    if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen15.csv")

### SCENARIO 16
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen16.csv")

### SCENARIO 19
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}

```

```

}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen19.csv")

### SCENARIO 20
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```
write_csv(sum0128, file = "IRTLRsims scen20.csv")
```

SCENARIO 23

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}

```

```
colnames(result) <- c("item1","item28")
```

```
item <- c(1,28)
```

```
type <- c("DIF","DIF")
```

```
signif01 <- c(0,0)
```

```
signif05 <- c(0,0)
```

```
hitrate.05 <- c(0,0)
```

```
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(result)
```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "IRTLRsims scen23.csv")
```

SCENARIO 24

Run the simulation using parallel computing

```
result <- foreach (i = 1:100, .combine=rbind) %dopar% {  
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])  
}
```

```
colnames(result) <- c("item1","item28")
```

```
item <- c(1,28)
```

```
type <- c("DIF","DIF")
```

```
signif01 <- c(0,0)
```

```
signif05 <- c(0,0)
```

```
hitrate.05 <- c(0,0)
```

```
sum0128 <- data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(result)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "IRTLRsims scen24.csv")
```

NONUNIFORM DIF WITH IMPACT

SCENARIO 27

Run the simulation using parallel computing

```
result <- foreach (i = 1:100, .combine=rbind) %dopar% {  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])  
}
```

```
colnames(result) <- c("item1","item28")
```

```
item <- c(1,28)
```

```
type <- c("DIF","DIF")
```

```
signif01 <- c(0,0)
```

```

signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen27.csv")

### SCENARIO 28
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```



```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen28.csv")

### SCENARIO 31
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen31.csv")

### SCENARIO 32
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

```

```

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen32.csv")

### SCENARIO 35
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1

```

```

    if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
    if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
    if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen35.csv")

### SCENARIO 36
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen36.csv")

### SCENARIO 39
# Run the simulation using parallel computing

```

```

result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen39.csv")

### SCENARIO 40
# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind) %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```

```

results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "IRTLRsims scen40.csv")

## LR method, conditions with impact:
#Detecting Differential Item Functioning in Multidimensional IRT with Logistic
Regression

library("doParallel")
library("mirt")
library("MASS")
library("readr")

detectDIF <- function(sample.size, DIF.size, cor, seed) {

  require("mirt")
  require("MASS")
  set.seed(seed)

  #Define multidimensional abilities for reference and focal groups
  theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
  theta.foc <- mvrnorm(n = sample.size[2], rep(-0.5, 2),
matrix(c(1,cor,cor,1),2,2))

  #Generate item parameters for reference and focal groups
  a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
  a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
  a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
  b1 <- runif(n = 28, min = 0.67, max = 2)

```

```

b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)

#Uniform and nonuniform DIF for items 1 and 28
b.foc <- b.ref
b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
a.foc <- a.ref
a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]

#Generate item responses according to MGRM
ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
dat <- rbind(ref, foc)

#Define the group variable (0=reference; 1=focal) and test DIF using mirt
group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))
itemnames <- colnames(dat)
model <- 'f1 = 1-14
          f2 = 15-28
          COV = F1*F2'
mod <- mirt(dat[,1:28], model, verbose = FALSE)
theta <- fscores(mod, full.scores = TRUE, method = "EAP")
theta <- as.data.frame(theta)

#Save raw scores

raw <- data.frame(raw1=rowSums(dat[,c(1:14)]), raw2=rowSums(dat[,c(15:28)]))

data <- cbind(dat,theta,raw, group)

#Test uniform DIF
if(DIF.size[1]>0 & DIF.size[2]==0) {
  data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"), ordered
= TRUE)
}

```

```

data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)

data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
#raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
raw.unifDIF.Item_1 <- anova(raw1model1, raw1model2, test = "Chisq")

raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
#raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
raw.unifDIF.Item_28 <- anova(raw28model1, raw28model2, test = "Chisq")

theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)
theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)
#theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.unifDIF.Item_1 <- anova(theta1model1, theta1model2, test = "Chisq")

theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
#theta1model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.unifDIF.Item_28 <- anova(theta28model1, theta28model2, test = "Chisq")

result <- data.frame(item1_rawunif=raw.unifDIF.Item_1[2,7],
                    item28_rawunif=raw.unifDIF.Item_28[2,7],
                    item1_thetaunif=theta.unifDIF.Item_1[2,7],
                    item28_thetaunif=theta.unifDIF.Item_28[2,7])

return(result)
} else

#Test nonuniform DIF

if(DIF.size[1]==0 & DIF.size[2]>0) {
  data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)

```

```

    data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)

    data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
    #raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
    raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
    raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
    raw.nonunifDIF.Item_1 <- anova(raw1model2, raw1model3, test = "Chisq")

    #raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
    raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess =
TRUE)
    raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
    raw.nonunifDIF.Item_28 <- anova(raw28model2, raw28model3, test = "Chisq")

    #theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)
    theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)
    theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
    theta.nonunifDIF.Item_1 <- anova(theta1model2, theta1model3, test = "Chisq")

    #theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
    theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
    theta28model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess =
TRUE)
    theta.nonunifDIF.Item_28 <- anova(theta28model2, theta28model3, test =
"Chisq")

    result <- data.frame(item1_rawnonunif=raw.nonunifDIF.Item_1[2,7],
                        item28_rawnonunif=raw.nonunifDIF.Item_28[2,7],
                        item1_thetanonunif=theta.nonunifDIF.Item_1[2,7],
                        item28_thetanonunif=theta.nonunifDIF.Item_28[2,7])

    return(result)
}
#Test type I error
if(DIF.size[1]==0 & DIF.size[2]==0) {
    data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"), ordered
= TRUE)
    data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)

```



```

data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
#raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess = TRUE)
raw.DIF.Item_1 <- anova(raw1model1, raw1model3, test = "Chisq")

raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
#raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess =
TRUE)
raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
raw.DIF.Item_28 <- anova(raw28model1, raw28model3, test = "Chisq")

theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)
#theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)
theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.DIF.Item_1 <- anova(theta1model1, theta1model3, test = "Chisq")

theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
#theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
theta28model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.DIF.Item_28 <- anova(theta28model1, theta28model3, test = "Chisq")

result <- data.frame(item1_rawDIF=raw.DIF.Item_1[2,7],
                    item28_rawDIF=raw.DIF.Item_28[2,7],
                    item1_thetaDIF=theta.DIF.Item_1[2,7],
                    item28_thetaDIF=theta.DIF.Item_28[2,7])
}
return(result)
}

# Set the seed
myseed <- sample.int(n=10000, size = 100)
detectCores()
registerDoParallel(4)

```

UNIFORM DIF WITH IMPACT

SCENARIO 1

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")
```

```
item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)
```

```
for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "LRsimscen1.csv")
```

SCENARIO 2

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
```

```

}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen2.csv")

### SCENARIO 5

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)

```

```

signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen5.csv")

```

SCENARIO 6

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {

```

```

if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen6.csv")

```

SCENARIO 9

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1

```

```

    if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
    if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
    if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen9.csv")

### SCENARIO 10

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```
write_csv(sum0128, file = "LRsimscen10.csv")
```

SCENARIO 13

```
# Run the simulation using parallel computing
```

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])  
}
```

```
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")
```

```
item <- c(1,28,1,28)
```

```
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
```

```
signif01 <- c(0,0,0,0)
```

```
signif05 <- c(0,0,0,0)
```

```
hitrate.05 <- c(0,0,0,0)
```

```
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(results)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
```

```
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
```

```
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
```

```
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "LRsimscen13.csv")
```

SCENARIO 14

```
# Run the simulation using parallel computing
```

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
```

```

    detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
  }
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen14.csv")

```

SCENARIO 17

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")

```



```

signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen17.csv")

```

SCENARIO 18

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "Lrsimscen18.csv")

```

SCENARIO 21

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

```

```

    if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
    if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
    if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
    if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen21.csv")

```

SCENARIO 22

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen22.csv")

### NONUNIFORM DIF WITH IMPACT
### SCENARIO 25

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen25.csv")

### SCENARIO 26

```

```

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen26.csv")

# SCENARIO 29

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}

```

```

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen29.csv")

# SCENARIO 30

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)

```

```

signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen30.csv")

# SCENARIO 33

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen33.csv")

# SCENARIO 34

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```



```

if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen34.csv")

# SCENARIO 37

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

```

```

}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen37.csv")

# SCENARIO 38

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen38.csv")

```

```
#Detecting Differential Item Functioning in Multidimensional IRT with Logistic Regression
```

```
library("doParallel")
```

```
library("mirt")
```

```
library("MASS")
```

```
library("readr")
```

```
detectDIF <- function(sample.size, DIF.size, cor, seed) {
```

```
  require("mirt")
```

```
  require("MASS")
```

```
  set.seed(seed)
```

```
  #Define multidimensional abilities for reference and focal groups
```

```
  theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
```

```
  theta.foc <- mvrnorm(n = sample.size[2], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
```

```
  #Generate item parameters for reference and focal groups
```

```
  a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
```

```
  a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
```

```
  a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
```

```
  b1 <- runif(n = 28, min = 0.67, max = 2)
```

```
  b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
```

```
  b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
```

```
  b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
```

```
  b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)
```

```
  #Uniform and nonuniform DIF for items 1 and 28
```

```
  b.foc <- b.ref
```

```
  b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
```

```
  a.foc <- a.ref
```

```
  a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
```

```
  a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]
```

```
  #Generate item responses according to MGRM
```

```
  ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
```

```

foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
dat <- rbind(ref, foc)

#Define the group variable (0=reference; 1=focal) and test DIF using mirt
group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))
itemnames <- colnames(dat)
model <- 'f1 = 1-14
          f2 = 15-28
          COV = F1*F2'
mod <- mirt(dat[,1:28], model, verbose = FALSE)
theta <- fscores(mod, full.scores = TRUE, method = "EAP")
theta <- as.data.frame(theta)

#Save raw scores

raw <- data.frame(raw1=rowSums(dat[,c(1:14)]), raw2=rowSums(dat[,c(15:28)]))

data <- cbind(dat,theta,raw, group)

#Test uniform DIF
if(DIF.size[1]>0 & DIF.size[2]==0) {
  data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"), ordered
= TRUE)
  data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)
  data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
  raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
  raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
  #raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
  raw.unifDIF.Item_1 <- anova(raw1model1, raw1model2, test = "Chisq")

  raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
  raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
  #raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
  raw.unifDIF.Item_28 <- anova(raw28model1, raw28model2, test = "Chisq")

  theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)

```

```

theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)
#theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.unifDIF.Item_1 <- anova(theta1model1, theta1model2, test = "Chisq")

theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
#theta1model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.unifDIF.Item_28 <- anova(theta28model1, theta28model2, test = "Chisq")

result <- data.frame(item1_rawunif=raw.unifDIF.Item_1[2,7],
                    item28_rawunif=raw.unifDIF.Item_28[2,7],
                    item1_thetaunif=theta.unifDIF.Item_1[2,7],
                    item28_thetaunif=theta.unifDIF.Item_28[2,7])

return(result)
} else

#Test nonuniform DIF

if(DIF.size[1]==0 & DIF.size[2]>0) {
  data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)
  data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)
  data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
  #raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
  raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
  raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
  raw.nonunifDIF.Item_1 <- anova(raw1model2, raw1model3, test = "Chisq")

  #raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
  raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess =
TRUE)
  raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
  raw.nonunifDIF.Item_28 <- anova(raw28model2, raw28model3, test = "Chisq")

```

```

#theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)
theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)
theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.nonunifDIF.Item_1 <- anova(theta1model2, theta1model3, test = "Chisq")

#theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
theta28model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess =
TRUE)
theta.nonunifDIF.Item_28 <- anova(theta28model2, theta28model3, test =
"Chisq")

result <- data.frame(item1_rawnonunif=raw.nonunifDIF.Item_1[2,7],
                    item28_rawnonunif=raw.nonunifDIF.Item_28[2,7],
                    item1_thetanonunif=theta.nonunifDIF.Item_1[2,7],
                    item28_thetanonunif=theta.nonunifDIF.Item_28[2,7])

return(result)
}
#Test type I error
if(DIF.size[1]==0 & DIF.size[2]==0) {
  data$Item_1 = factor(data$Item_1, levels = c("0", "1", "2", "3", "4"), ordered
= TRUE)
  data$Item_28 = factor(data$Item_28, levels = c("0", "1", "2", "3", "4"),
ordered = TRUE)
  data$group = factor(data$group, levels = c("0", "1"), ordered = TRUE)
  raw1model1= polr(data$Item_1 ~ raw1 + raw2, data = data, Hess = TRUE)
#raw1model2= polr(data$Item_1 ~ raw1 + raw2 + group, data = data, Hess = TRUE)
raw1model3= polr(data$Item_1 ~ (raw1 + raw2) * group, data = data, Hess = TRUE)
raw.DIF.Item_1 <- anova(raw1model1, raw1model3, test = "Chisq")

raw28model1= polr(data$Item_28 ~ raw1 + raw2, data = data, Hess = TRUE)
#raw28model2= polr(data$Item_28 ~ raw1 + raw2 + group, data = data, Hess =
TRUE)
raw28model3= polr(data$Item_28 ~ (raw1 + raw2) * group, data = data, Hess =
TRUE)
raw.DIF.Item_28 <- anova(raw28model1, raw28model3, test = "Chisq")

theta1model1= polr(data$Item_1 ~ f1 + f2, data = data, Hess = TRUE)
#theta1model2= polr(data$Item_1 ~ f1 + f2 + group, data = data, Hess = TRUE)

```

```

theta1model3= polr(data$Item_1 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.DIF.Item_1 <- anova(theta1model1, theta1model3, test = "Chisq")

theta28model1= polr(data$Item_28 ~ f1 + f2, data = data, Hess = TRUE)
#theta28model2= polr(data$Item_28 ~ f1 + f2 + group, data = data, Hess = TRUE)
theta28model3= polr(data$Item_28 ~ (f1 + f2) * group, data = data, Hess = TRUE)
theta.DIF.Item_28 <- anova(theta28model1, theta28model3, test = "Chisq")

result <- data.frame(item1_rawDIF=raw.DIF.Item_1[2,7],
                    item28_rawDIF=raw.DIF.Item_28[2,7],
                    item1_thetaDIF=theta.DIF.Item_1[2,7],
                    item28_thetaDIF=theta.DIF.Item_28[2,7])
}
return(result)
}

# Set the seed
myseed <- sample.int(n=10000, size = 100)
detectCores()
registerDoParallel(4)

### UNIFORM DIF WITH IMPACT
### SCENARIO 3

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw", "item28_raw", "item1_theta", "item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw", "DIF_raw", "DIF_theta", "DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```

```

results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen3.csv")

### SCENARIO 4

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1

```



```

if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen4.csv")

### SCENARIO 7

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
}

```

```

    if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen7.csv")

### SCENARIO 8

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen8.csv")

```

SCENARIO 11

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])  
}  
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")
```

```
item <- c(1,28,1,28)
```

```
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
```

```
signif01 <- c(0,0,0,0)
```

```
signif05 <- c(0,0,0,0)
```

```
hitrate.05 <- c(0,0,0,0)
```

```
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(results)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
```

```
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
```

```
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
```

```
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "LRsimscen11.csv")
```

SCENARIO 12

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {  
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])  
}
```

```

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen12.csv")

```

SCENARIO 15

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)

```

```

hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen15.csv")

```

SCENARIO 16

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

```

```

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1

```

```

if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen16.csv")

```

SCENARIO 19

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

```

```

    if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
    if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen19.csv")

### SCENARIO 20

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen20.csv")

```

SCENARIO 23

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])  
}  
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")
```

```
item <- c(1,28,1,28)
```

```
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
```

```
signif01 <- c(0,0,0,0)
```

```
signif05 <- c(0,0,0,0)
```

```
hitrate.05 <- c(0,0,0,0)
```

```
sum0128 <- data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(results)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
```

```
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
```

```
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
```

```
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "LRsimscen23.csv")
```

SCENARIO 24

Run the simulation using parallel computing

```
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {  
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
```



```

}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen24.csv")

### NONUNIFORM DIF WITH IMPACT
### SCENARIO 27

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)

```

```

type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
write_csv(sum0128, file = "LRsimscen27.csv")

```

SCENARIO 28

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000,1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen28.csv")

```

SCENARIO 31

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}

colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1

```

```

if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen31.csv")

# SCENARIO 32

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
}

```

```

    if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen32.csv")

# SCENARIO 35

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```

write_csv(sum0128, file = "LRsimscen35.csv")

# SCENARIO 36

# Run the simulation using parallel computing
results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen36.csv")

```

```

# SCENARIO 39

```

```

# Run the simulation using parallel computing

```

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

```

```

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```

write_csv(sum0128, file = "LRsimscen39.csv")

```

SCENARIO 40

Run the simulation using parallel computing

```

results <- foreach (i = 1:100, .combine=rbind, .packages='mirt') %dopar% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(results) <- c("item1_raw","item28_raw","item1_theta","item28_theta")

```

```

item <- c(1,28,1,28)
type <- c("DIF_raw","DIF_raw","DIF_theta","DIF_theta")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <- data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(results)

for (i in 1:nrow(results)) {
  if(results$item1_raw[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1_raw[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28_raw[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28_raw[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

  if(results$item1_theta[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item1_theta[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
  if(results$item28_theta[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28_theta[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "LRsimscen40.csv")

#Detecting Differential Item Functioning in Multidimensional IRT with MIMIC -
interaction Model

library("MplusAutomation")
library("doParallel")
library("mirt")
library("MASS")
library("readr")

detectDIF <- function(sample.size, DIF.size, cor, seed) {

  require("mirt")
  require("MASS")
  set.seed(seed)

```



```

#Define multidimensional abilities for reference and focal groups
theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
theta.foc <- mvrnorm(n = sample.size[2], rep(-0.5, 2),
matrix(c(1,cor,cor,1),2,2))

#Generate item parameters for reference and focal groups
a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
b1 <- runif(n = 28, min = 0.67, max = 2)
b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)

#Uniform and nonuniform DIF for items 1 and 28
b.foc <- b.ref
b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
a.foc <- a.ref
a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]

#Generate item responses according to MGRM
ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
dat <- rbind(ref, foc)

#Define the group variable (0=reference; 1=focal) and test DIF using mirt
group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))
itemnames <- colnames(dat)
data <- cbind(dat, group)
data <- as.data.frame(data)
names(data) <- sub("Item_", "item", names(data))

#Test omnibus DIF

mimic_omnibus1 <- mplusObject(

```

```

TITLE= "MIMIC multidimensional omnibus test Item1;",
VARIABLE=
  "usevar = item1-group;
  CATEGORICAL ARE item1-item28;",
MODEL=
  "!LATENT TRAIT 1;
  [factor_1@0];
  [item1$4-item14$4];

  factor_1 BY item1* item2-item14;
  factor_1@1;

  factor_1 on group;

  !UNIFORM DIF;
  item1 on group;

  !NONUNIFORM DIF;
  zxfact1 | group XWITH factor_1;
  item1 ON zxfact1;

  !LATENT TRAIT 2;
  [factor_2@0];
  [item15$4-item28$4];

  factor_2 BY item28* item15-item27;
  factor_2@1;

  factor_2 on group;",

ANALYSIS=
  "estimator = mlr;type= random;",
  usevariables = colnames(data),
  rdata=data
)
mimic_omnibus1_fit <- mplusModeler(mimic_omnibus1,
                                  modelout="mimic_omnibus1.inp",

```

```

check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus28 <- mplusObject(
  TITLE= "MIMIC multidimensional omnibus test Item28;",
  VARIABLE=
    "usevar = item1-group;
    CATEGORICAL ARE item1-item28;",
  MODEL=
    "!LATENT TRAIT 1;
    [factor_1@0];
    [item1$4-item14$4];

    factor_1 BY item1* item2-item14;

    factor_1@1;

    factor_1 on group;

    !LATENT TRAIT 2;
    [factor_2@0];
    [item15$4-item28$4];

    factor_2 BY item28* item15-item27;
    factor_2@1;

    factor_2 on group;

    !UNIFORM DIF;
    item28 on group;

    !NONUNIFORM DIF;
    zxfact2 | group XWITH factor_2;
    item28 ON zxfact2;",
  ANALYSIS=
    "estimator = mlr;type= random;",
  usevariables = colnames(data),
  rdata=data
)

```

```

mimic_omnibus28_fit <- mplusModeler(mimic_omnibus28,
                                   modelout="mimic_omnibus28.inp",
                                   check=TRUE, run = TRUE, hashfilename = FALSE)

#Return uniform DIF
  if(DIF.size[1]>0 & DIF.size[2]==0) {
    result <-
data.frame(item1_uniform=mimic_omnibus1_fit[["results"]][["parameters"]][["unstanda
rdized"]][32,6],

item28_uniform=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]
[32,6])

    return(result)

  } else

#Return nonuniform DIF

  if(DIF.size[1]==0 & DIF.size[2]>0) {

    result <-
data.frame(item1_nonunif=mimic_omnibus1_fit[["results"]][["parameters"]][["unstanda
rdized"]][31,6],

item28_nonunif=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]
[31,6])

    return(result)

  }

#Return type I error
  if(DIF.size[1]==0 & DIF.size[2]==0) {

    result <-
data.frame(item1_uniformDIF=mimic_omnibus1_fit[["results"]][["parameters"]][["unsta
ndardized"]][32,6],

item1_nonuniformDIF=mimic_omnibus1_fit[["results"]][["parameters"]][["unstandardize
d"]][31,6],

item28_uniformDIF=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized
"]][32,6],

```

```

item28_nonuniformDIF=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardi
zed"]][31,6])
}
return(result)
}

# Set the seed
myseed <- sample.int(n=10000, size = 100)

### UNIFORM DIF WITH IMPACT
### SCENARIO 1

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
}

```

```

    if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen1.csv")

### SCENARIO 2

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

```

```
write_csv(sum0128, file = "MIMICsimscen2.csv")
```

SCENARIO 5

Run the simulation using parallel computing

```
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%  
{  
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])  
}
```

```
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",  
"item28nonuniform")
```

```
item <- c(1,1,28,28)
```

```
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
```

```
signif01 <- c(0,0,0,0)
```

```
signif05 <- c(0,0,0,0)
```

```
hitrate.05 <- c(0,0,0,0)
```

```
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
```

```
results <- na.omit(result)
```

```
for (i in 1:nrow(results)) {
```

```
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
```

```
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
```

```
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
```

```
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
```

```
  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
```

```
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
```

```
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
```

```
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
```

```
}
```

```
sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)
```

```
write_csv(sum0128, file = "MIMICsimscen5.csv")
```

SCENARIO 6

```

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen6.csv")

### SCENARIO 9

# Run the simulation
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{

```



```

    detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
  }
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen9.csv")

### SCENARIO 10

# Run the simulation
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)

```

```

sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen10.csv")

```

SCENARIO 13

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen13.csv")

### SCENARIO 14

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen14.csv")

### SCENARIO 17

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{

```

```

    detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
  }
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen17.csv")

```

SCENARIO 18

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)

```

```

sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen18.csv")

```

SCENARIO 21

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen21.csv")

### SCENARIO 22

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen22.csv")

### NONUNIFORM DIF WITH IMPACT
### SCENARIO 25

# Run the simulation using parallel computing

```

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen25.csv")

```

SCENARIO 26

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)

```

```

hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen26.csv")

```

SCENARIO 29

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```



```

}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen29.csv")

### SCENARIO 30

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen30.csv")

### SCENARIO 33

# Run the simulation using parallel computing

```

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen33.csv")

```

SCENARIO 34

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)

```

```

hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen34.csv")

```

SCENARIO 37

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```

}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen37.csv")

### SCENARIO 38

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %dopar%
{
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1","item28")

item <- c(1,28)
type <- c("DIF","DIF")
signif01 <- c(0,0)
signif05 <- c(0,0)
hitrate.05 <- c(0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item28[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen38.csv")

#Detecting Differential Item Functioning in Multidimensional IRT with MIMIC -
interaction Model

library("MplusAutomation")
library("doParallel")
library("mirt")

```

```

library("MASS")
library("readr")

detectDIF <- function(sample.size, DIF.size, cor, seed) {

  require("mirt")
  require("MASS")
  set.seed(seed)

  #Define multidimensional abilities for reference and focal groups
  theta.ref <- mvrnorm(n = sample.size[1], rep(0, 2), matrix(c(1,cor,cor,1),2,2))
  theta.foc <- mvrnorm(n = sample.size[2], rep(0, 2), matrix(c(1,cor,cor,1),2,2))

  #Generate item parameters for reference and focal groups
  a1 <- c(runif(n = 14, min = 1.1, max = 2.8), rep(0,14))
  a2 <- c(rep(0,14), runif(n = 14, min = 1.1, max = 2.8))
  a.ref <- as.matrix(cbind(a1, a2), ncol = 2)
  b1 <- runif(n = 28, min = 0.67, max = 2)
  b2 <- b1 - runif(n = 28, min = 0.67, max = 1.34)
  b3 <- b2 - runif(n = 28, min = 0.67, max = 1.34)
  b4 <- b3 - runif(n = 28, min = 0.67, max = 1.34)
  b.ref <- as.matrix(cbind(b1, b2, b3, b4), ncol = 4)

  #Uniform and nonuniform DIF for items 1 and 28
  b.foc <- b.ref
  b.foc[c(1,28),] <- b.foc[c(1,28),]+DIF.size[1]
  a.foc <- a.ref
  a.foc[c(1),1] <- a.foc[c(1),1]+DIF.size[2]
  a.foc[c(28),2] <- a.foc[c(28),2]+DIF.size[2]

  #Generate item responses according to MGRM
  ref <- simdata(a = a.ref, d = b.ref, itemtype = 'graded', Theta = theta.ref)
  foc <- simdata(a = a.foc, d = b.foc, itemtype = 'graded', Theta = theta.foc)
  dat <- rbind(ref, foc)

  #Define the group variable (0=reference; 1=focal) and test DIF using mirt
  group <- c(rep("0", sample.size[1]), rep("1", sample.size[2]))

```

```

itemnames <- colnames(dat)
data <- cbind(dat, group)
data <- as.data.frame(data)
names(data) <- sub("Item_", "item", names(data))

#Test omnibus DIF

mimic_omnibus1 <- mplusObject(
  TITLE= "MIMIC multidimensional omnibus test Item1;",
  VARIABLE=
    "usevar = item1-group;
    CATEGORICAL ARE item1-item28;",
  MODEL=
    "!LATENT TRAIT 1;
    [factor_1@0];
    [item1$4-item14$4];

    factor_1 BY item1* item2-item14;
    factor_1@1;

    factor_1 on group;

    !UNIFORM DIF;
    item1 on group;

    !NONUNIFORM DIF;
    zxfact1 | group XWITH factor_1;
    item1 ON zxfact1;

    !LATENT TRAIT 2;
    [factor_2@0];
    [item15$4-item28$4];

    factor_2 BY item28* item15-item27;
    factor_2@1;

    factor_2 on group;",

```

```

ANALYSIS=
  "estimator = mlr;type= random;",
usevariables = colnames(data),
rdata=data
)
mimic_omnibus1_fit <- mplusModeler(mimic_omnibus1,
                                  modelout="mimic_omnibus1.inp",
                                  check=TRUE, run = TRUE, hashfilename = FALSE)
mimic_omnibus28 <- mplusObject(
  TITLE= "MIMIC multidimensional omnibus test Item28;",
  VARIABLE=
    "usevar = item1-group;
    CATEGORICAL ARE item1-item28;",
  MODEL=
    "!LATENT TRAIT 1;
    [factor_1@0];
    [item1$4-item14$4];

    factor_1 BY item1* item2-item14;

    factor_1@1;

    factor_1 on group;

    !LATENT TRAIT 2;
    [factor_2@0];
    [item15$4-item28$4];

    factor_2 BY item28* item15-item27;
    factor_2@1;

    factor_2 on group;

    !UNIFORM DIF;
    item28 on group;

```

```

!NONUNIFORM DIF;
zxfact2 | group XWITH factor_2;
item28 ON zxfact2;",
ANALYSIS=
  "estimator = mlr;type= random;",
usevariables = colnames(data),
rdata=data
)
mimic_omnibus28_fit <- mplusModeler(mimic_omnibus28,
                                   modelout="mimic_omnibus28.inp",
                                   check=TRUE, run = TRUE, hashfilename = FALSE)

#Return uniform DIF
if(DIF.size[1]>0 & DIF.size[2]==0) {
  result <-
data.frame(item1_uniform=mimic_omnibus1_fit[["results"]][["parameters"]][["unstandardized"]][32,6],

item28_uniform=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]][32,6])

  return(result)

} else

#Return nonuniform DIF

if(DIF.size[1]==0 & DIF.size[2]>0) {

  result <-
data.frame(item1_nonunif=mimic_omnibus1_fit[["results"]][["parameters"]][["unstandardized"]][31,6],

item28_nonunif=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]][31,6])

  return(result)

}

#Return type I error
if(DIF.size[1]==0 & DIF.size[2]==0) {

```



```

    result <-
data.frame(item1_uniformDIF=mimic_omnibus1_fit[["results"]][["parameters"]][["unstandardized"]][32,6],

item1_nonuniformDIF=mimic_omnibus1_fit[["results"]][["parameters"]][["unstandardized"]][31,6],

item28_uniformDIF=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]][32,6],

item28_nonuniformDIF=mimic_omnibus28_fit[["results"]][["parameters"]][["unstandardized"]][31,6])
  }
  return(result)
}

# Set the seed
myseed <- sample.int(n=10000, size = 100)

### UNIFORM DIF WITHOUT IMPACT
### SCENARIO 3

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen3.csv")

```

SCENARIO 4

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1

```

```

if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen4.csv")

### SCENARIO 7

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
}

```

```

    if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
    if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
  }

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen7.csv")

```

SCENARIO 8

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform","item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen8.csv")

### SCENARIO 11

# Run the simulation
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen11.csv")

```

```

### SCENARIO 12

# Run the simulation
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen12.csv")

### SCENARIO 15

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {

```

```

    detectDIF(sample.size=c(1500, 500), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
  }
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")
item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen15.csv")

### SCENARIO 16

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.3, 0), cor=0.5, seed=myseed[i])
}
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")
item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")

```

```

signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen16.csv")

```

SCENARIO 19

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500,500), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```



```

results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen19.csv")

```

SCENARIO 20

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1

```

```

if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen20.csv")

### SCENARIO 23

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
}

```

```

if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen23.csv")

```

SCENARIO 24

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0.6, 0), cor=0.5, seed=myseed[i])
}

```

```

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

```

```

item <- c(1,1,28,28)

```

```

type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")

```

```

signif01 <- c(0,0,0,0)

```

```

signif05 <- c(0,0,0,0)

```

```

hitrate.05 <- c(0,0,0,0)

```

```

sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)

```

```

results <- na.omit(result)

```

```

for (i in 1:nrow(results)) {

```

```

  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1

```

```

  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1

```

```

  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1

```

```

  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

```

```

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1

```

```

  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1

```

```

  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1

```

```

  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1

```

```

}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICscen24.csv")

### NONUNIFORM DIF WITHOUT IMPACT
### SCENARIO 27

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform","item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

```

```

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen27.csv")

### SCENARIO 28
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen28.csv")

### SCENARIO 31

```

```

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen31.csv")

```

SCENARIO 32

```

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.3), cor=0.5, seed=myseed[i])
}

```

```

}
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen32.csv")

### SCENARIO 35

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}
colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)

```

```

type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen35.csv")

```

SCENARIO 36

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)

```



```

sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen36.csv")

```

SCENARIO 39

Run the simulation using parallel computing

```

result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1500, 500), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {

```

```

if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1

if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen39.csv")

### SCENARIO 40

# Run the simulation using parallel computing
result <- foreach (i = 1:100, .combine=rbind, .packages='MplusAutomation') %do% {
  detectDIF(sample.size=c(1000, 1000), DIF.size=c(0, 0.6), cor=0.5, seed=myseed[i])
}

colnames(result) <- c("item1uniform","item1nonuniform", "item28uniform",
"item28nonuniform")

item <- c(1,1,28,28)
type <- c("uniform DIF","nonuniform DIF","uniform DIF","nonuniform DIF")
signif01 <- c(0,0,0,0)
signif05 <- c(0,0,0,0)
hitrate.05 <- c(0,0,0,0)
sum0128 <-data.frame(item,type,signif01,signif05,hitrate.05)
results <- na.omit(result)

for (i in 1:nrow(results)) {
  if(results$item1uniform[i]<0.01) sum0128[1,3]=sum0128[1,3]+1
  if(results$item1uniform[i]<0.05) sum0128[1,4]=sum0128[1,4]+1
  if(results$item28uniform[i]<0.01) sum0128[3,3]=sum0128[3,3]+1
  if(results$item28uniform[i]<0.05) sum0128[3,4]=sum0128[3,4]+1
}

```

```

  if(results$item1nonuniform[i]<0.01) sum0128[2,3]=sum0128[2,3]+1
  if(results$item1nonuniform[i]<0.05) sum0128[2,4]=sum0128[2,4]+1
  if(results$item28nonuniform[i]<0.01) sum0128[4,3]=sum0128[4,3]+1
  if(results$item28nonuniform[i]<0.05) sum0128[4,4]=sum0128[4,4]+1
}

sum0128$hitrate.05 <- round((sum0128$signif05/nrow(results)),digits = 2)

write_csv(sum0128, file = "MIMICsimscen40.csv")

```

###Real data study

##EGA measurement invariance

```

library(knitr)
library(dplyr)
library(psych)
library(EGAnet)

```

#Load data:

```

ESCQ_27 <- read.csv(file.choose(),fileEncoding="UTF-8-BOM", header=T)
options(max.print=999999)
ESCQ <- na.omit(ESCQ_27)

ESCQ <- ESCQ %>% rename(PU1=escq14, PU2=escq15, PU3= escq25, PU4=escq26,
PU5=escq34, PU6=escq35, PU7=escq36, PU8=escq37, PU9=escq38, PU10=escq39,
PU11=escq42,
                        EL1=escq2, EL2=escq17,EL3=escq21, EL4=escq22, EL5=escq23,
EL6=escq43, EL7=escq44,
                        MR1=escq4, MR2=escq5, MR3=escq8, MR4=escq9, MR5=escq11,
MR6=escq29, MR7=escq30, MR8=escq33, MR9=escq40)

scale <-(ESCQ[,1:27])
head(scale, 3)

```

#Subset data according to country:

```

CroData <- subset(ESCQ, country == "1",select=c(1:27))
PortData <- subset(ESCQ, country == "2", select=c(1:27))
FinData <- subset(ESCQ, country == "3", select=c(1:27))
SweData <- subset(ESCQ, country == "4", select=c(1:27))
SloData <- subset(ESCQ, country == "5", select=c(1:27))
SpainData <- subset(ESCQ, country == "6", select=c(1:27))

```

```

JapData <- subset(ESCQ, country == "7", select=c(1:27))
ChiData <- subset(ESCQ, country == "8", select=c(1:27))
IndData <- subset(ESCQ, country == "9", select=c(1:27))
ArgData <- subset(ESCQ, country == "10", select=c(1:27))
SerbData <- subset(ESCQ, country == "11", select=c(1:27))
USAData <- subset(ESCQ, country == "12", select=c(1:27))
TurkData <- subset(ESCQ, country == "13", select=c(1:27))

```

Run exploratory graph analyses for each country:

```

ega.Cro <- EGA(
  CroData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)

```

```

ega.Port <- EGA(
  PortData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)

```

```

ega.Fin <- EGA(
  FinData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)

```

```

ega.Swe <- EGA(

```

```

SweData,
uni.method = "expand",
corr = "cor_auto",
model = "glasso",
algorithm = "walktrap",
plot.EGA = TRUE,
plot.type = "GGally"
)
ega.Slo <- EGA(
  SloData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "glasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.Spain <- EGA(
  SpainData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "glasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.Jap <- EGA(
  JapData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "glasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.Chi <- EGA(
  ChiData,

```

```

uni.method = "expand",
corr = "cor_auto",
model = "lasso",
algorithm = "walktrap",
plot.EGA = TRUE,
plot.type = "GGally"
)
ega.Ind <- EGA(
  IndData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.Arg <- EGA(
  ArgData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.Serb <- EGA(
  SerbData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "lasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
ega.USA <- EGA(
  USAData,
  uni.method = "expand",

```

```

    corr = "cor_auto",
    model = "glasso",
    algorithm = "walktrap",
    plot.EGA = TRUE,
    plot.type = "GGally"
)
ega.Turk <- EGA(
  TurkData,
  uni.method = "expand",
  corr = "cor_auto",
  model = "glasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally"
)
groups <- as.vector(ESCQ$country)
results <- invariance(scale, groups)
set.seed(1)
boot.escq <- bootEGA(scale,
  iter = 500)
escq.stability <- dimensionStability(boot.escq)

plot.all <- compare.EGA.plots(ega.Cro, ega.Port, ega.Fin, ega.Swe, ega.Slo,
  ega.Spain, ega.Jap, ega.Chi, ega.Ind, ega.Arg, ega.Serb, ega.USA, ega.Turk,
  base.plot = 6,
  labels = c("Hrvatska", "Portugal", "Finska",
    "Švedska", "Slovenija", "Španjolska", "Japan", "Kina", "Indija", "Argentina",
    "Srbija", "SAD", "Turska"),
  rows = 4, columns = 4)

MIData <- ESCQ[!(ESCQ$country=="1" |ESCQ$country=="2" |ESCQ$country=="3" |
  ESCQ$country=="4" |ESCQ$country=="7" |ESCQ$country=="9" |ESCQ$country=="11" |
  ESCQ$country=="12" | ESCQ$country=="13"), ]
MIScale <- (MIData[,1:27])

##Checking metric invariance:
set.seed(1)
results <- invariance(data = MIScale,
  groups = MIData$country)

```

```

plot(results)

# Applying BH-procedure
adjusted.p <- p.adjust(
  results$results$p, method = "BH",
  n = length(results$results$p)
)
# Uncorrected p
results$results[results$results$p < .05,]
# Corrected p
results$results[adjusted.p < .05,]

ega.MI <- EGA(
  MIScale,
  uni.method = "LE",
  corr = "cor_auto",
  model = "glasso",
  algorithm = "walktrap",
  plot.EGA = TRUE,
  plot.type = "GGally")
net.loads(ega.MI)
network.descriptives(ega.MI)

# Run bootEGA analysis

set.seed(1)
boot.escq.MI <- bootEGA(MIScale,
  iter = 500)
boot.escq.MI$summary.table
boot.escq.MI$frequency
escq.MI.stability <- dimensionStability(boot.escq.MI)

###ESCQ-27 dimensionality
library(EFA.dimensions)
ESCQ_27 <- read.csv(file.choose(),fileEncoding="UTF-8-BOM", header=T)
options(max.print=999999)
ESCQ_27 <- na.omit(ESCQ_27)

```



```

MIData <- ESCQ_27[!(ESCQ_27$country=="1" |ESCQ_27$country=="2"
|ESCQ_27$country=="3" | ESCQ_27$country=="4" |ESCQ_27$country=="7"
|ESCQ_27$country=="9" |ESCQ_27$country=="11" | ESCQ_27$country=="12" |
ESCQ_27$country=="13"), ]

scale <-(MIData[,1:27])

head(scale, 3)

FACTORABILITY(scale,corkind='polychoric',Ncases=NULL,verbose=TRUE)

PCA(scale,corkind='polychoric',Nfactors=1,Ncases=NULL,ppower=3,verbose=TRUE)

PA_FA(scale,corkind='polychoric',Nfactors=1,Ncases=NULL,ppower=3,verbose=TRUE)

SCREE_PLOT(scale,corkind='polychoric',verbose=TRUE)

MAP(scale,corkind='polychoric',verbose=TRUE)

```

Model comparison

Load the necessary libraries:

```

library(mirt)
library(knitr)
library(dplyr)

```

#Load data:

```

ESCQ_27 <- read.csv(file.choose(),fileEncoding="UTF-8-BOM", header=T)
options(max.print=999999)
ESCQ_27 <- na.omit(ESCQ_27)

MIData <- ESCQ_27[!(ESCQ_27$country=="1" |ESCQ_27$country=="2"
|ESCQ_27$country=="3" | ESCQ_27$country=="4" |ESCQ_27$country=="7"
|ESCQ_27$country=="9" |ESCQ_27$country=="11" | ESCQ_27$country=="12" |
ESCQ_27$country=="13"), ]

```

Define scale variable (containing all 27 items of ESCQ-27) and check that it is done properly:

```

scale <-(MIData[,1:27])

head(scale, 3)

```

#Fit a graded response model (GRM):

```

mod1 <- (mirt(scale, 1, verbose = FALSE, itemtype = 'graded', SE = TRUE))

#Check the model fit

M2(mod1, type = "C2", calcNULL = FALSE)

```

#Check the item fit:

```

itemfit(mod1)

```

```

# Fit the partial credit model (PCM) and check model & item fit:
mod2 <- (mirt(scale, 1, itemtype="Rasch", SE=TRUE, verbose=FALSE))
coef.pcm <- coef(mod2, IRTpars=TRUE, simplify=TRUE)
items.pcm <- as.data.frame(coef.pcm$items)
print(items.pcm)
M2(mod2, type = "C2", calcNULL = FALSE)
itemfit(mod2)

# Fit the generalized partial credit model and check model & item fit:
mod3 <- (mirt(scale, 1, itemtype="gpcm", SE=TRUE, verbose=FALSE))
coef.gpcm <- coef(mod3, IRTpars=TRUE, simplify=TRUE)
items.gpcm <- as.data.frame(coef.gpcm$items)
print(items.gpcm)
M2(mod3, type = "C2", calcNULL = FALSE)
itemfit(mod3)

# Fit the rating scale model and check model & item fit:
mod4 <- (mirt(scale, 1, itemtype="rsm", SE=TRUE, verbose=FALSE))
coef.rsm <- coef(mod4, IRTpars=TRUE, simplify=TRUE)
items.rsm <- as.data.frame(coef.rsm$items)
print(items.rsm)
M2(mod4, type = "C2", calcNULL = FALSE)
itemfit(mod4)

## MGRM
# Fit the multidimensional graded response model:
mgrm.mod <- 'F1 = 1-11,
            F2 = 12-18,
            F3 = 19-27,
            COV = F1*F2*F3'
mgrm_fit <- mirt(scale, model = mgrm.mod,
                itemtype = "graded", SE = TRUE)
# Check model & item fit:
M2(mgrm_fit)
itemfit(mgrm_fit)

### Item parameters

```

```

##Uni-GRM
#Compute item parameters for GRM:
IRT_parms <- coef(mod1, IRTpars = TRUE, simplify = TRUE)
IRT_parms$items
#Compute factor loadings:
summary(mod1)
#Plot category characteristic curves:
plot(mod1, type='trace', which.item = c(1:27), facet_items=T,
      as.table = TRUE, auto.key=list(points=F, lines=T, columns=4, space = 'top',
      cex = .8),
      theta_lim = c(-3, 3),
      main = "")
#Plot item information curves:
plot(mod1, type='infotrace', which.item = c(1:27), facet_items=T,
      as.table = TRUE, auto.key=list(points=F, lines=T, columns=1, space = 'right',
      cex = .8),
      theta_lim = c(-3, 3),
      main="")
#Plot scale information and conditional standard errors:
plot(mod1, type = 'infoSE', theta_lim = c(-3, 3),
      main="")
#Plot conditional reliability:
plot(mod1, type = 'rxx', theta_lim = c(-3, 3),
      main="" )
#Compute marginal reliability:
marginal_rxx(mod1)
#Plot scale characteristic curve:
plot(mod1, type = 'score', theta_lim = c(-3, 3), main = "")
#Compute factor scores:
fscores(mod1, method = 'EAPsum', full.scores = FALSE)

##Multi-GRM
# Compute item parameters for MGRM:
mgrm_params <- coef(mgrm_fit, simplify = TRUE)
mgrm_params$items
# Compute multidimensional IRT item parameters:
mgrm_items <- cbind(MDISC(mgrm_fit),
                    MDIFF(mgrm_fit))

```

```

show(mgrm_items)
# Compute estimated variance-covariance (correlation) matrix:
mgrm_params$cov
# Plot item trace plots:
itemplot(mgrm_fit, type = "trace", item = 1)
itemplot(mgrm_fit, type = "trace", item = 13)
# Plot test information and SEM plots:
plot(mgrm_fit, type = "info")
plot(mgrm_fit, type = "SE")
plot(mgrm_fit, type = "score")

### Differential item functioning
## Testing multidimensional DIF with IRT-LR on MIData
model.mgrm <- mirt.model('F1 = 1-11
    F2 = 12-18
    F3 = 19-27
    COV = F1*F2, F2*F3, F1*F3')
model <- multipleGroup(MIData[,1:27], model.mgrm, itemtype = c("graded"),
    group=as.factor(MIData$gender), SE = TRUE)
DIF(model, c('a1', 'a2', 'a3', "d1", "d2", "d3", "d4"))
itemnames <- colnames(MIData)
model2 <- multipleGroup(MIData[,1:27], model.mgrm, itemtype = c("graded"),
    group=as.factor(MIData$gender), SE = TRUE,
    invariance = c(itemnames[c(9, 16, 26)], 'free_means',
    'free_var'))
DIF(model2, c('a1', 'a2', 'a3', "d1", "d2", "d3", "d4"), items2test=c(1:8, 10:15,
17:25, 27))
DIF(model2, c('a1', 'a2', 'a3'), items2test=c(1:8, 10:15, 17:25, 27))
DIF(model2, c("d1", "d2", "d3", "d4"), items2test=c(1:8, 10:15, 17:25, 27))

## Logistic regression:
library(MASS)
library(mirt)
library(dplyr)
library(magrittr)

ESCQ_27 <- read.csv(file.choose(), fileEncoding="UTF-8-BOM", header=T)
options(max.print=999999)
ESCQ_27 <- na.omit(ESCQ_27)
ESCQ_27$gender <- as.factor(ESCQ_27$gender)

```

```

ESCQ <- ESCQ_27

MIData <- ESCQ[!(ESCQ$country=="1" | ESCQ$country=="2" | ESCQ$country=="3" |
ESCQ$country=="4" | ESCQ$country=="7" | ESCQ$country=="9" | ESCQ$country=="11" |
ESCQ$country=="12" | ESCQ$country=="13"), ]

ESCQ_27 <- MIData

mgrm.mod <- 'F1 = 1, 3-6,8-11
            F2 = 13-17
            F3 = 19,21-27
            COV = F1*F2, F2*F3, F1*F3'

mod <- mirt(ESCQ_27[,1:27], model = mgrm.mod,
           itemtype = "graded", SE = TRUE)

theta <- fscores(mod, full.scores = TRUE, method = "EAP")

theta <- as.data.frame(theta)

raw <- data.frame(raw1=rowSums(ESCQ_27[,c(1, 3:6, 8:11)]), raw2=rowSums(ESCQ_27[,
c(13:17)]), raw3=rowSums(ESCQ_27[,c(19, 21:27)]))

data <- cbind(ESCQ_27[,1:27],theta,raw, ESCQ_27$gender)

colnames(data)[34] = "gender"

cols <- c(1:27)

data %<>% mutate_at(cols, factor, levels = c("1", "2", "3", "4", "5"), ordered =
TRUE)

data$gender = factor(data$gender, levels = c("1", "2"), ordered = TRUE)

#escq14

raw1model1= polr(data$escq14 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq14 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq14 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)

raw.DIF.escq14 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq14 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq14 <- anova(raw1model2, raw1model3, test = "Chisq")

raw.DIF.escq14
raw.uniformDIF.escq14
raw.nonuniformDIF.escq14

theta1model1= polr(data$escq14 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq14 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq14 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq14 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq14 <- anova(theta1model1, theta1model2, test = "Chisq")

```

```

theta.nonuniformDIF.escq14 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq14
theta.uniformDIF.escq14
theta.nonuniformDIF.escq14

#escq15
raw1model1= polr(data$escq15 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq15 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq15 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq15 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq15 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq15 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq15
raw.uniformDIF.escq15
raw.nonuniformDIF.escq15

theta1model1= polr(data$escq15 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq15 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq15 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq15 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq15 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq15 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq15
theta.uniformDIF.escq15
theta.nonuniformDIF.escq15

#escq25
raw1model1= polr(data$escq25 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq25 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq25 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq25 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq25 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq25 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq25
raw.uniformDIF.escq25

```

```
raw.nonuniformDIF.escq25
```

```
theta1model1= polr(data$escq25 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq25 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq25 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq25 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq25 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq25 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq25  
theta.uniformDIF.escq25  
theta.nonuniformDIF.escq25
```

```
#escq26
```

```
raw1model1= polr(data$escq26 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)  
raw1model2= polr(data$escq26 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =  
TRUE)  
raw1model3= polr(data$escq26 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =  
TRUE)  
raw.DIF.escq26 <- anova(raw1model1, raw1model3, test = "Chisq")  
raw.uniformDIF.escq26 <- anova(raw1model1, raw1model2, test = "Chisq")  
raw.nonuniformDIF.escq26 <- anova(raw1model2, raw1model3, test = "Chisq")  
raw.DIF.escq26  
raw.uniformDIF.escq26  
raw.nonuniformDIF.escq26
```

```
theta1model1= polr(data$escq26 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq26 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq26 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq26 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq26 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq26 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq26  
theta.uniformDIF.escq26  
theta.nonuniformDIF.escq26
```

```
#escq34
```

```
raw1model1= polr(data$escq34 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
```

```

raw1model2= polr(data$escq34 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq34 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq34 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq34 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq34 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq34
raw.uniformDIF.escq34
raw.nonuniformDIF.escq34

theta1model1= polr(data$escq34 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq34 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq34 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq34 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq34 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq34 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq34
theta.uniformDIF.escq34
theta.nonuniformDIF.escq34

#escq35
raw1model1= polr(data$escq35 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq35 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq35 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq35 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq35 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq35 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq35
raw.uniformDIF.escq35
raw.nonuniformDIF.escq35

theta1model1= polr(data$escq35 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq35 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq35 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq35 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq35 <- anova(theta1model1, theta1model2, test = "Chisq")

```



```

theta.nonuniformDIF.escq35 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq35
theta.uniformDIF.escq35
theta.nonuniformDIF.escq35

#escq36
raw1model1= polr(data$escq36 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq36 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq36 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq36 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq36 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq36 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq36
raw.uniformDIF.escq36
raw.nonuniformDIF.escq36

theta1model1= polr(data$escq36 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq36 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq36 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq36 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq36 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq36 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq36
theta.uniformDIF.escq36
theta.nonuniformDIF.escq36

#escq37
raw1model1= polr(data$escq37 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq37 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq37 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq37 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq37 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq37 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq37
raw.uniformDIF.escq37

```

```
raw.nonuniformDIF.escq37
```

```
theta1model1= polr(data$escq37 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq37 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq37 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq37 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq37 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq37 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq37  
theta.uniformDIF.escq37  
theta.nonuniformDIF.escq37
```

```
#escq38
```

```
raw1model1= polr(data$escq38 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)  
raw1model2= polr(data$escq38 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =  
TRUE)  
raw1model3= polr(data$escq38 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =  
TRUE)  
raw.DIF.escq38 <- anova(raw1model1, raw1model3, test = "Chisq")  
raw.uniformDIF.escq38 <- anova(raw1model1, raw1model2, test = "Chisq")  
raw.nonuniformDIF.escq38 <- anova(raw1model2, raw1model3, test = "Chisq")  
raw.DIF.escq38  
raw.uniformDIF.escq38  
raw.nonuniformDIF.escq38
```

```
theta1model1= polr(data$escq38 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq38 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq38 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq38 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq38 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq38 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq38  
theta.uniformDIF.escq38  
theta.nonuniformDIF.escq38
```

```
#escq39
```

```
raw1model1= polr(data$escq39 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
```

```

raw1model2= polr(data$escq39 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq39 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq39 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq39 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq39 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq39
raw.uniformDIF.escq39
raw.nonuniformDIF.escq39

theta1model1= polr(data$escq39 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq39 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq39 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq39 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq39 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq39 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq39
theta.uniformDIF.escq39
theta.nonuniformDIF.escq39

#escq42
raw1model1= polr(data$escq42 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq42 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq42 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq42 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq42 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq42 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq42
raw.uniformDIF.escq42
raw.nonuniformDIF.escq42

theta1model1= polr(data$escq42 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq42 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq42 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq42 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq42 <- anova(theta1model1, theta1model2, test = "Chisq")

```

```

theta.nonuniformDIF.escq42 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq42
theta.uniformDIF.escq42
theta.nonuniformDIF.escq42

#escq2
raw1model1= polr(data$escq2 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq2 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq2 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq2 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq2 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq2 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq2
raw.uniformDIF.escq2
raw.nonuniformDIF.escq2

theta1model1= polr(data$escq2 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq2 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq2 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq2 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq2 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq2 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq2
theta.uniformDIF.escq2
theta.nonuniformDIF.escq2

#escq17
raw1model1= polr(data$escq17 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq17 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq17 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq17 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq17 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq17 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq17
raw.uniformDIF.escq17

```

```
raw.nonuniformDIF.escq17
```

```
theta1model1= polr(data$escq17 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq17 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq17 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq17 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq17 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq17 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq17  
theta.uniformDIF.escq17  
theta.nonuniformDIF.escq17
```

```
#escq21
```

```
raw1model1= polr(data$escq21 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)  
raw1model2= polr(data$escq21 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =  
TRUE)  
raw1model3= polr(data$escq21 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =  
TRUE)  
raw.DIF.escq21 <- anova(raw1model1, raw1model3, test = "Chisq")  
raw.uniformDIF.escq21 <- anova(raw1model1, raw1model2, test = "Chisq")  
raw.nonuniformDIF.escq21 <- anova(raw1model2, raw1model3, test = "Chisq")  
raw.DIF.escq21  
raw.uniformDIF.escq21  
raw.nonuniformDIF.escq21
```

```
theta1model1= polr(data$escq21 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq21 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq21 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq21 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq21 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq21 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq21  
theta.uniformDIF.escq21  
theta.nonuniformDIF.escq21
```

```
#escq22
```

```
raw1model1= polr(data$escq22 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
```

```

raw1model2= polr(data$escq22 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq22 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq22 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq22 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq22 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq22
raw.uniformDIF.escq22
raw.nonuniformDIF.escq22

theta1model1= polr(data$escq22 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq22 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq22 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq22 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq22 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq22 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq22
theta.uniformDIF.escq22
theta.nonuniformDIF.escq22

#escq23
raw1model1= polr(data$escq23 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq23 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq23 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq23 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq23 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq23 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq23
raw.uniformDIF.escq23
raw.nonuniformDIF.escq23

theta1model1= polr(data$escq23 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq23 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq23 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq23 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq23 <- anova(theta1model1, theta1model2, test = "Chisq")

```

```

theta.nonuniformDIF.escq23 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq23
theta.uniformDIF.escq23
theta.nonuniformDIF.escq23

#escq43
raw1model1= polr(data$escq43 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq43 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq43 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq43 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq43 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq43 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq43
raw.uniformDIF.escq43
raw.nonuniformDIF.escq43

theta1model1= polr(data$escq43 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq43~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq43 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq43 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq43 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq43 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq43
theta.uniformDIF.escq43
theta.nonuniformDIF.escq43

#escq44
raw1model1= polr(data$escq44 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq44 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq44 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq44 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq44 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq44 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq44
raw.uniformDIF.escq44

```

```
raw.nonuniformDIF.escq44
```

```
theta1model1= polr(data$escq44 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq44 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq44 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq44 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq44 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq44 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq44  
theta.uniformDIF.escq44  
theta.nonuniformDIF.escq44
```

```
#escq4
```

```
raw1model1= polr(data$escq4 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)  
raw1model2= polr(data$escq4 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =  
TRUE)  
raw1model3= polr(data$escq4 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =  
TRUE)  
raw.DIF.escq4 <- anova(raw1model1, raw1model3, test = "Chisq")  
raw.uniformDIF.escq4 <- anova(raw1model1, raw1model2, test = "Chisq")  
raw.nonuniformDIF.escq4 <- anova(raw1model2, raw1model3, test = "Chisq")  
raw.DIF.escq4  
raw.uniformDIF.escq4  
raw.nonuniformDIF.escq4
```

```
theta1model1= polr(data$escq4 ~ F1 + F2 + F3, data = data, Hess = TRUE)  
theta1model2= polr(data$escq4 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)  
theta1model3= polr(data$escq4 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)  
theta.DIF.escq4 <- anova(theta1model1, theta1model3, test = "Chisq")  
theta.uniformDIF.escq4 <- anova(theta1model1, theta1model2, test = "Chisq")  
theta.nonuniformDIF.escq4 <- anova(theta1model2, theta1model3, test = "Chisq")  
theta.DIF.escq4  
theta.uniformDIF.escq4  
theta.nonuniformDIF.escq4
```

```
#escq5
```

```
raw1model1= polr(data$escq5 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
```



```

raw1model2= polr(data$escq5 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq5 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq5 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq5 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq5 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq5
raw.uniformDIF.escq5
raw.nonuniformDIF.escq5

theta1model1= polr(data$escq5 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq5 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq5 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq5 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq5 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq5 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq5
theta.uniformDIF.escq5
theta.nonuniformDIF.escq5

#escq8
raw1model1= polr(data$escq8 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq8 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq8 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq8 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq8 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq8 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq8
raw.uniformDIF.escq8
raw.nonuniformDIF.escq8

theta1model1= polr(data$escq8 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq8 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq8 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq8 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq8 <- anova(theta1model1, theta1model2, test = "Chisq")

```

```

theta.nonuniformDIF.escq8 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq8
theta.uniformDIF.escq8
theta.nonuniformDIF.escq8

#escq9
raw1model1= polr(data$escq9 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq9 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq9 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq9 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq9 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq9 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq9
raw.uniformDIF.escq9
raw.nonuniformDIF.escq9

theta1model1= polr(data$escq9 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq9 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq9 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq9 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq9 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq9 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq9
theta.uniformDIF.escq9
theta.nonuniformDIF.escq9

#escq11
raw1model1= polr(data$escq11 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq11 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq11 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq11 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq11 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq11 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq11
raw.uniformDIF.escq11

```

```

raw.nonuniformDIF.escq11

theta1model1= polr(data$escq11 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq11 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq11 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq11 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq11 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq11 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq11
theta.uniformDIF.escq11
theta.nonuniformDIF.escq11

#escq29
raw1model1= polr(data$escq29 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq29 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq29 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq29 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq29 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq29 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq29
raw.uniformDIF.escq29
raw.nonuniformDIF.escq29

theta1model1= polr(data$escq29 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq29 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq29 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq29 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq29 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq29 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq29
theta.uniformDIF.escq29
theta.nonuniformDIF.escq29

#escq30
raw1model1= polr(data$escq30 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)

```

```

raw1model2= polr(data$escq30 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq30 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq30 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq30 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq30 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq30
raw.uniformDIF.escq30
raw.nonuniformDIF.escq30

theta1model1= polr(data$escq30 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq30 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq30 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq30 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq30 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq30 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq30
theta.uniformDIF.escq30
theta.nonuniformDIF.escq30

#escq33
raw1model1= polr(data$escq33 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq33 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq33 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq33 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq33 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq33 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq33
raw.uniformDIF.escq33
raw.nonuniformDIF.escq33

theta1model1= polr(data$escq33 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq33 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq33 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq33 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq33 <- anova(theta1model1, theta1model2, test = "Chisq")

```

```

theta.nonuniformDIF.escq33 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq33
theta.uniformDIF.escq33
theta.nonuniformDIF.escq33

#escq40
raw1model1= polr(data$escq40 ~ raw1 + raw2 + raw3, data = data, Hess = TRUE)
raw1model2= polr(data$escq40 ~ (raw1 + raw2 + raw3) + gender, data = data, Hess =
TRUE)
raw1model3= polr(data$escq40 ~ (raw1 + raw2 + raw3) * gender, data = data, Hess =
TRUE)
raw.DIF.escq40 <- anova(raw1model1, raw1model3, test = "Chisq")
raw.uniformDIF.escq40 <- anova(raw1model1, raw1model2, test = "Chisq")
raw.nonuniformDIF.escq40 <- anova(raw1model2, raw1model3, test = "Chisq")
raw.DIF.escq40
raw.uniformDIF.escq40
raw.nonuniformDIF.escq40

theta1model1= polr(data$escq40 ~ F1 + F2 + F3, data = data, Hess = TRUE)
theta1model2= polr(data$escq40 ~ (F1 + F2 + F3) + gender, data = data, Hess = TRUE)
theta1model3= polr(data$escq40 ~ (F1 + F2 + F3) * gender, data = data, Hess = TRUE)
theta.DIF.escq40 <- anova(theta1model1, theta1model3, test = "Chisq")
theta.uniformDIF.escq40 <- anova(theta1model1, theta1model2, test = "Chisq")
theta.nonuniformDIF.escq40 <- anova(theta1model2, theta1model3, test = "Chisq")
theta.DIF.escq40
theta.uniformDIF.escq40
theta.nonuniformDIF.escq40

## MIMIC-interaction model
library(MplusAutomation)
ESCQ_27 <- read.csv(file.choose(),fileEncoding="UTF-8-BOM", header=T)
options(max.print=999999)
ESCQ_27 <- na.omit(ESCQ_27)
ESCQ_27$gender <- as.factor(ESCQ_27$gender)
ESCQ <- ESCQ_27
MIData <- ESCQ[!(ESCQ$country=="1" | ESCQ$country=="2" | ESCQ$country=="3" |
ESCQ$country=="4" | ESCQ$country=="7" | ESCQ$country=="9" | ESCQ$country=="11" |
ESCQ$country=="12" | ESCQ$country=="13"), ]
escq_mimic <- subset(MIData, select = -c(country) )

```

```

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  VARIABLE=
    "usevar = escq14-gender;
    CATEGORICAL ARE escq14-escq40;",
  MODEL = "
    !LATENT TRAIT 1;
    [factor_1@0];
    [escq14$4-escq42$4];

    factor_1 BY escq14* escq15-escq42;
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    [factor_2@0];
    [escq2$4-escq44$4];

    factor_2 BY escq2-escq44;
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    [factor_3@0];
    [escq4$4-escq40$4];

    factor_3 BY escq4-escq40;
    factor_3@1;
    factor_3 ON gender;

    ! UNIFORM DIF;
    escq14 ON gender;

    ! NONUNIFORM DIF;
    zxfact1 | gender XWITH factor_1;
    escq14 ON zxfact1;

```

```

",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus14.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus_fit
mimic_omnibus_fit[["results"]][["parameters"]][["unstandardized"]][31,6]
mimic_omnibus_fit[["results"]][["parameters"]][["unstandardized"]][32,6]
mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);
    factor_3@1;
    factor_3 ON gender;

    ! UNIFORM DIF;
    escq15 ON gender;

```

```

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq15 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus15.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;

```



```

escq25 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq25 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus25.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);
    factor_3@1;
    factor_3 ON gender;
  "
)

```

```

! UNIFORM DIF;
escq26 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq26 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus26.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

```

```

! UNIFORM DIF;
escq34 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq34 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus34.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;

```

```

factor_3 ON gender;

! UNIFORM DIF;
escq35 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq35 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus35.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);

```

```

factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq36 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq36 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus36.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;

```

```

escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq37 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq37 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus37.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;

```

```

factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq38 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq38 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus38.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;
  "
)

```

```

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq39 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq39 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus39.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

```



```

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq42 ON gender;

! NONUNIFORM DIF;
zxfact1 | gender XWITH factor_1;
escq42 ON zxfact1;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus42.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;

```

```

factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq2 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq2 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus2.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);

```

```

factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq17 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq17 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus17.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;

```

```

escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq21 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq21 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus21.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2

```

```

factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq22 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq22 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus22.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

```

```

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq23 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq23 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus23.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

```

```

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq43 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq43 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus43.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;

```

```

factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq44 ON gender;

! NONUNIFORM DIF;
zxfact2 | gender XWITH factor_2;
escq44 ON zxfact2;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus44.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);

```



```

factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq4 ON gender;

! NONUNIFORM DIF;
zxfact3 | gender XWITH factor_3;
escq4 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus4.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;

```

```

escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq5 ON gender;

! NONUNIFORM DIF;
zxfact3 | gender XWITH factor_3;
escq5 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus5.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;

```

```

factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq8 ON gender;

! NONUNIFORM DIF;
zxfact3 | gender XWITH factor_3;
escq8 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus8.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "

```

```

! LATENT TRAIT 1;
factor_1 BY escq14-escq42*;
escq25(1);
factor_1@1;
factor_1 ON gender;

! LATENT TRAIT 2
factor_2 BY escq2-escq44*;
escq23(1);
factor_2@1;
factor_2 ON gender;

! LATENT TRAIT 3;
factor_3 BY escq4-escq40;
escq33(1);
factor_3@1;
factor_3 ON gender;

! UNIFORM DIF;
escq9 ON gender;

! NONUNIFORM DIF;
zxfact3 | gender XWITH factor_3;
escq9 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus9.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",

```

```

MODEL = "
  ! LATENT TRAIT 1;
  factor_1 BY escq14-escq42*;
  escq25(1);
  factor_1@1;
  factor_1 ON gender;

  ! LATENT TRAIT 2
  factor_2 BY escq2-escq44*;
  escq23(1);
  factor_2@1;
  factor_2 ON gender;

  ! LATENT TRAIT 3;
  factor_3 BY escq4-escq40;
  escq33(1);
  factor_3@1;
  factor_3 ON gender;

  ! UNIFORM DIF;
  escq11 ON gender;

  ! NONUNIFORM DIF;
  zxfact3 | gender XWITH factor_3;
  escq11 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus11.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

mimic_omnibus <- mplusObject(

```

```

TITLE = "MIMIC - omnibus test;",
MODEL = "
  ! LATENT TRAIT 1;
  factor_1 BY escq14-escq42*;
  escq25(1);
  factor_1@1;
  factor_1 ON gender;

  ! LATENT TRAIT 2
  factor_2 BY escq2-escq44*;
  escq23(1);
  factor_2@1;
  factor_2 ON gender;

  ! LATENT TRAIT 3;
  factor_3 BY escq4-escq40;
  escq33(1);
  factor_3@1;
  factor_3 ON gender;

  ! UNIFORM DIF;
  escq29 ON gender;

  ! NONUNIFORM DIF;
  zxfact3 | gender XWITH factor_3;
  escq29 ON zxfact3;
",
ANALYSIS = "estimator = mlr; type = random;",
usevariables = colnames(escq_mimic),
rdata = escq_mimic,
hashfilename = FALSE,
writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
                                modelout="mimic_omnibus29.inp",
                                check=TRUE, run = TRUE, hashfilename = FALSE)

```

```

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);
    factor_3@1;
    factor_3 ON gender;

    ! UNIFORM DIF;
    escq30 ON gender;

    ! NONUNIFORM DIF;
    zxfact3 | gender XWITH factor_3;
    escq30 ON zxfact3;
  ",
  ANALYSIS = "estimator = mlr; type = random;",
  usevariables = colnames(escq_mimic),
  rdata = escq_mimic,
  hashfilename = FALSE,
  writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
  modelout="mimic_omnibus30.inp",
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

```

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);
    factor_3@1;
    factor_3 ON gender;

    ! UNIFORM DIF;
    escq33 ON gender;

    ! NONUNIFORM DIF;
    zxfact3 | gender XWITH factor_3;
    escq33 ON zxfact3;
  ",
  ANALYSIS = "estimator = mlr; type = random;",
  usevariables = colnames(escq_mimic),
  rdata = escq_mimic,
  hashfilename = FALSE,
  writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
  modelout="mimic_omnibus33.inp",
  check=TRUE, run = TRUE, hashfilename = FALSE)

```



```

mimic_omnibus <- mplusObject(
  TITLE = "MIMIC - omnibus test;",
  MODEL = "
    ! LATENT TRAIT 1;
    factor_1 BY escq14-escq42*;
    escq25(1);
    factor_1@1;
    factor_1 ON gender;

    ! LATENT TRAIT 2
    factor_2 BY escq2-escq44*;
    escq23(1);
    factor_2@1;
    factor_2 ON gender;

    ! LATENT TRAIT 3;
    factor_3 BY escq4-escq40;
    escq33(1);
    factor_3@1;
    factor_3 ON gender;

    ! UNIFORM DIF;
    escq40 ON gender;

    ! NONUNIFORM DIF;
    zxfact3 | gender XWITH factor_3;
    escq40 ON zxfact3;
",
  ANALYSIS = "estimator = mlr; type = random;",
  usevariables = colnames(escq_mimic),
  rdata = escq_mimic,
  hashfilename = FALSE,
  writeData = "always"
)
mimic_omnibus_fit <- mplusModeler(mimic_omnibus,
  modelout="mimic_omnibus40.inp",
  check=TRUE, run = TRUE, hashfilename = FALSE)

```

Privitak 2: Upitnik emocionalne kompetentnosti UEK-45

UPITNIK EMOCIONALNE KOMPETENTNOSTI

UEK-45

Autor: doc. dr. Vladimir Takšić, Odsjek za psihologiju, Filozofski fakultet u Rijeci ®

Ovo nije test kojim ispitujemo Vaše znanje i zato ne možete dati pogrešan odgovor. Zanima nas kako se obično osjećate i kako razmišljate. Na postavljena pitanja odgovarajte po prvom dojmu i nemojte previše razmišljati o njima.

Odgovarajte koliko se ponuđene tvrdnje **odnose na Vas** i to zaokruživanjem jednoga od brojeva koji imaju sljedeće značenje:

- 1 - uopće NE
- 2 - uglavnom NE
- 3 - kako kada
- 4 - uglavnom DA
- 5 - u potpunosti DA

Tvrdnja	procjena				
1. Dobro raspoloženje mogu zadržati i ako mi se nešto loše dogodi.	1	2	3	4	5
2. Gotovo uvijek mogu riječima opisati svoje osjećaje i emocije.	1	2	3	4	5
3. I kada je moja okolina loše raspoložena, mogu zadržati dobro raspoloženje.	1	2	3	4	5
4. Iz neugodnih iskustava učim kako se ubuduće ne treba ponašati.	1	2	3	4	5
5. Kada me netko pohvali (nagradi), radim s više elana.	1	2	3	4	5
6. Kada mi nešto ne odgovara, to odmah i pokažem.	1	2	3	4	5
7. Kada mi se neka osoba sviđa, učinit ću sve da joj se i ja svidim.	1	2	3	4	5
8. Kada sam dobro raspoložen, teško me je oneraspoložiti.	1	2	3	4	5
9. Kada sam dobrog raspoloženja svaki mi se problem čini lako rješiv.	1	2	3	4	5
10. Kada sam s osobom koja me cijeni, pazim kako se ponašam.	1	2	3	4	5
11. Kada se osjećam sretno i raspoloženo, tada najbolje učim i pamtim.	1	2	3	4	5
12. Kada se zainatim, riješit ću i naizgled nerješiv problem.	1	2	3	4	5
13. Kada sretnem poznanika, odmah shvatim kako je raspoložen.	1	2	3	4	5
14. Kada vidim kako se netko osjeća, obično znam što mu se dogodilo.	1	2	3	4	5
15. Kod prijatelja mogu razlikovati kada je tužan, a kada razočaran.	1	2	3	4	5
16. Lako ću smisliti način da pridem osobi koja mi se sviđa.	1	2	3	4	5
17. Lako mogu nabrojiti emocije koje trenutno doživim.	1	2	3	4	5
18. Lako primijetim promjenu raspoloženja svoga prijatelja.	1	2	3	4	5
19. Lako se mogu domisliti kako obradovati prijatelja kojem idem na rođendan.	1	2	3	4	5
20. Lako uvjerim prijatelja da nema razloga za zabrinutost.	1	2	3	4	5

Okrenite stranicu i nastavite raditi →

1 - uopće NE; 2 - uglavnom NE; 3 - kako kada; 4 - uglavnom DA; 5 - u potpunosti DA

Tvrdnja	procjena				
21. Mogu dobro izraziti svoje emocije.	1	2	3	4	5
22. Mogu opisati kako se osjećam.	1	2	3	4	5
23. Mogu opisati svoje sadašnje stanje.	1	2	3	4	5
24. Mogu reći da znam puno o svom emocionalnom stanju.	1	2	3	4	5
25. Mogu točno odrediti emocije neke osobe ako je promatram u društvu.	1	2	3	4	5
26. Mogu zapaziti kada se netko osjeća bespomoćno.	1	2	3	4	5
27. Moje ponašanje odražava moje unutarnje osjećaje.	1	2	3	4	5
28. Na meni se uvijek vidi kakvog sam raspoloženja.	1	2	3	4	5
29. Nastojim ublažiti neugodne emocije, a pojačati pozitivne.	1	2	3	4	5
30. Nema ništa loše u tome kako se obično osjećam.	1	2	3	4	5
31. Obaveze ili zadatke radije odmah izvršim, nego da o njima mislim.	1	2	3	4	5
32. Obično razumijem zašto se loše osjećam.	1	2	3	4	5
33. Pokušavam održati dobro raspoloženje.	1	2	3	4	5
34. Prema izrazu lica mogu prepoznati nečija osjećanja.	1	2	3	4	5
35. Prepoznajem prikrivenu ljubomoru kod svojih prijatelja.	1	2	3	4	5
36. Primijetim kada netko nastoji prikriti loše raspoloženje.	1	2	3	4	5
37. Primijetim kada netko osjeća krivnju.	1	2	3	4	5
38. Primijetim kada netko prikriva svoje prave osjećaje.	1	2	3	4	5
39. Primijetim kada se netko osjeća potišteno.	1	2	3	4	5
40. Što se mene tiče, potpuno je u redu ovako se osjećati.	1	2	3	4	5
41. Uglavnom mi je bilo lako izraziti simpatije osobi suprotnog spola.	1	2	3	4	5
42. Uočim kada se netko ponaša različito od onoga kako je raspoložen.	1	2	3	4	5
43. Većinu svojih osjećaja lako mogu imenovati.	1	2	3	4	5
44. Većinu svojih osjećaja mogu prepoznati.	1	2	3	4	5
45. Znam kako mogu ugodno iznenaditi svakoga svoga prijatelja.	1	2	3	4	5

Ime i prezime _____

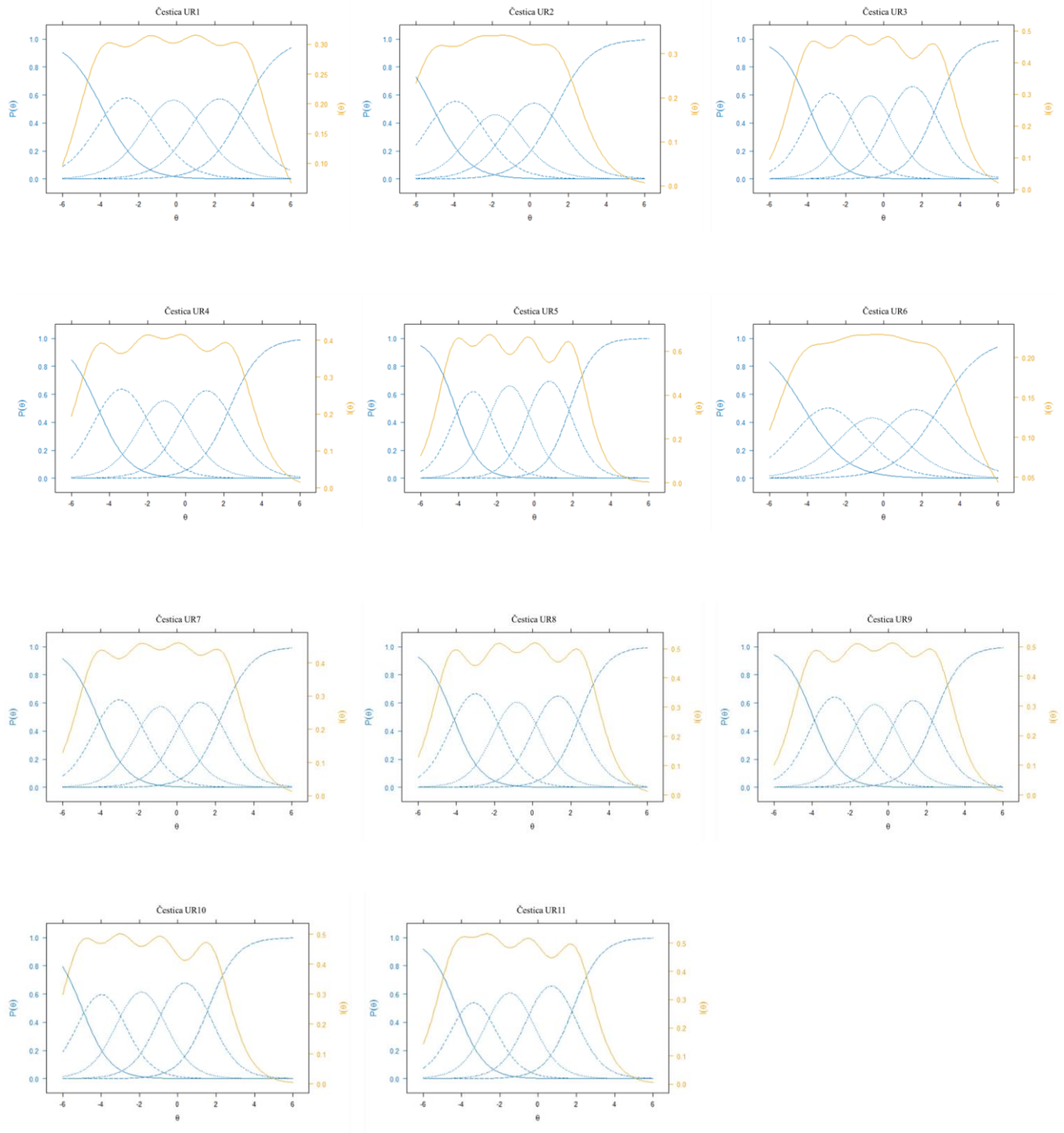
Spol Ž M

Dob _____

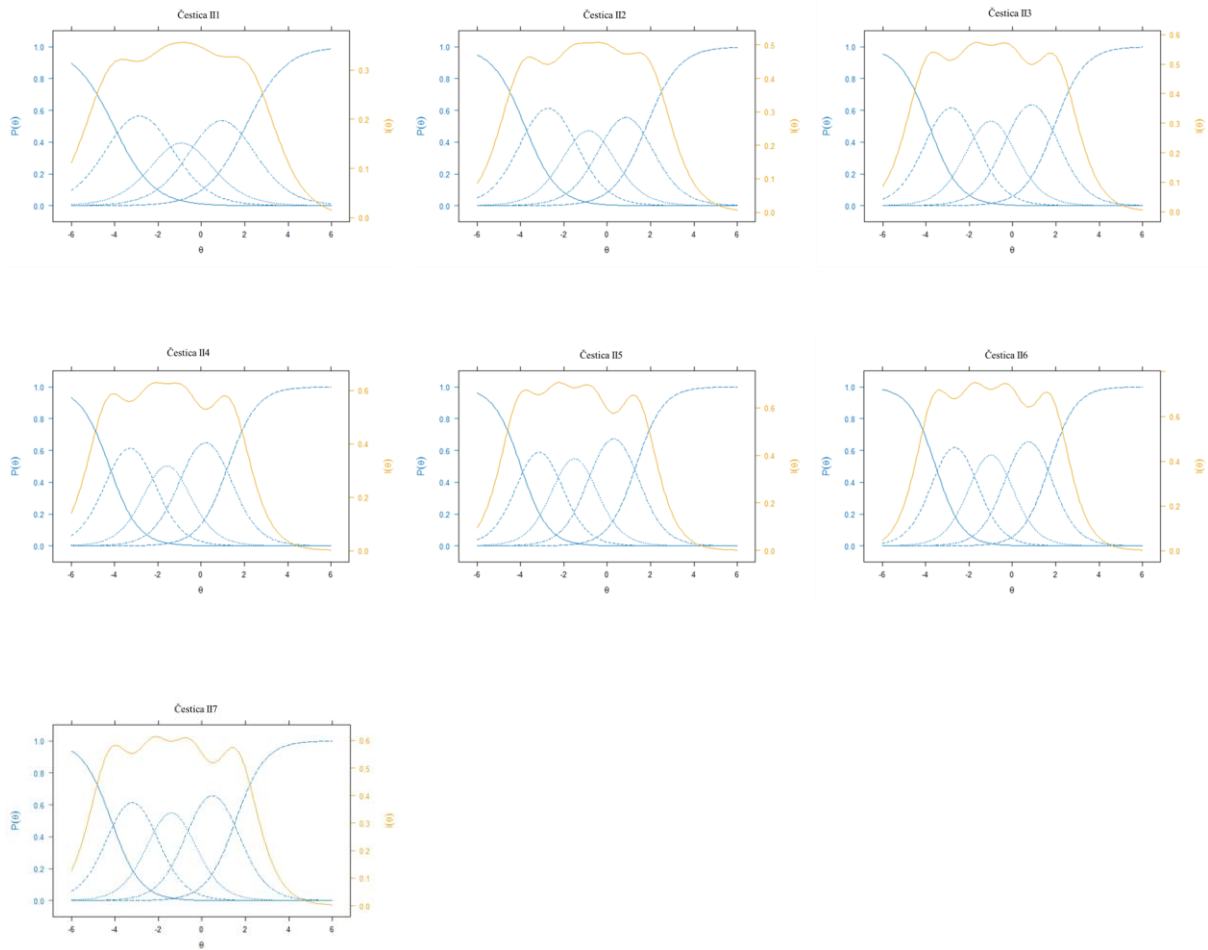
Stručna sprema _____

Zahvaljujemo na iskrenosti i suradnji !

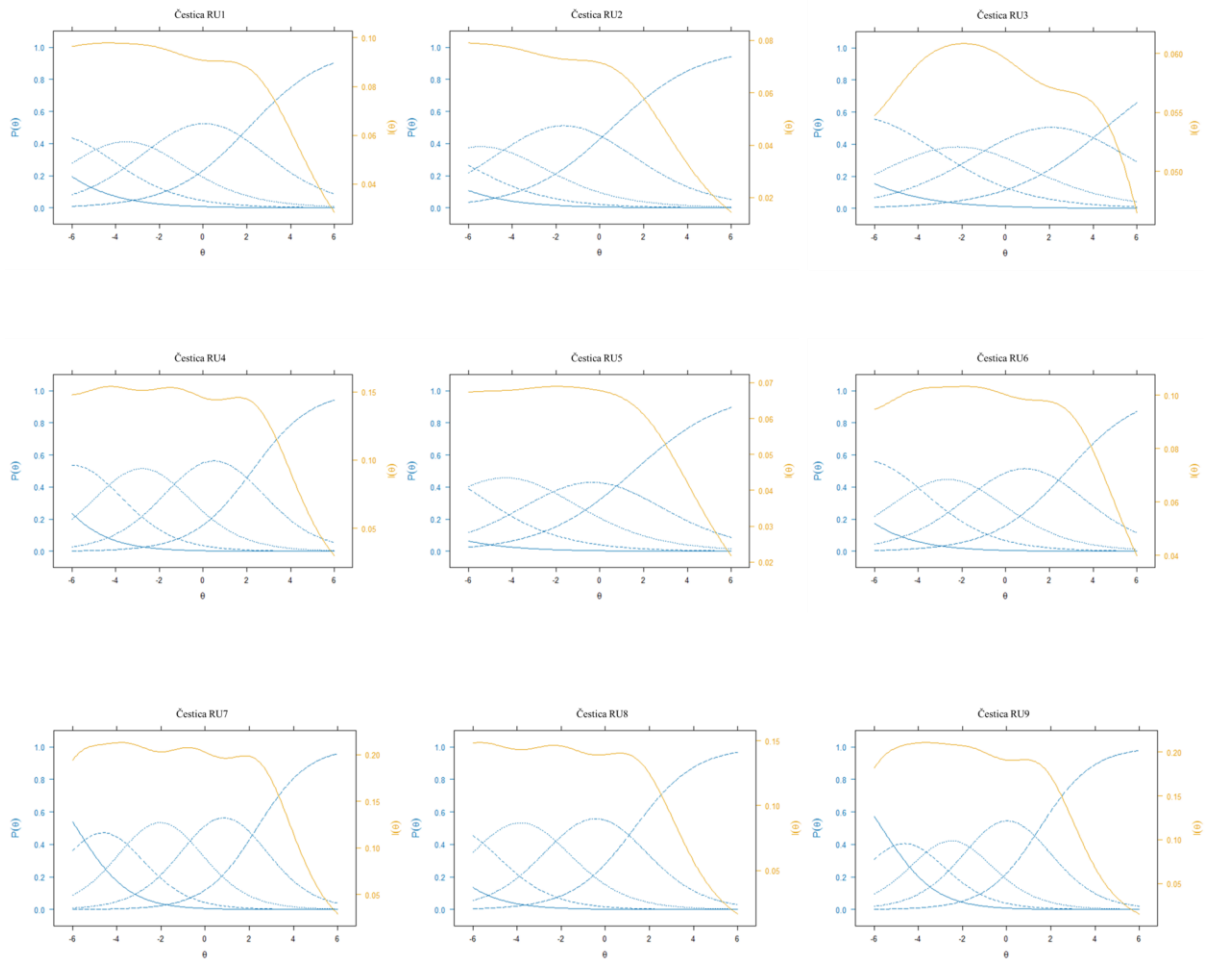
Privitak 3: Grafički prikazi čestica UEK-27



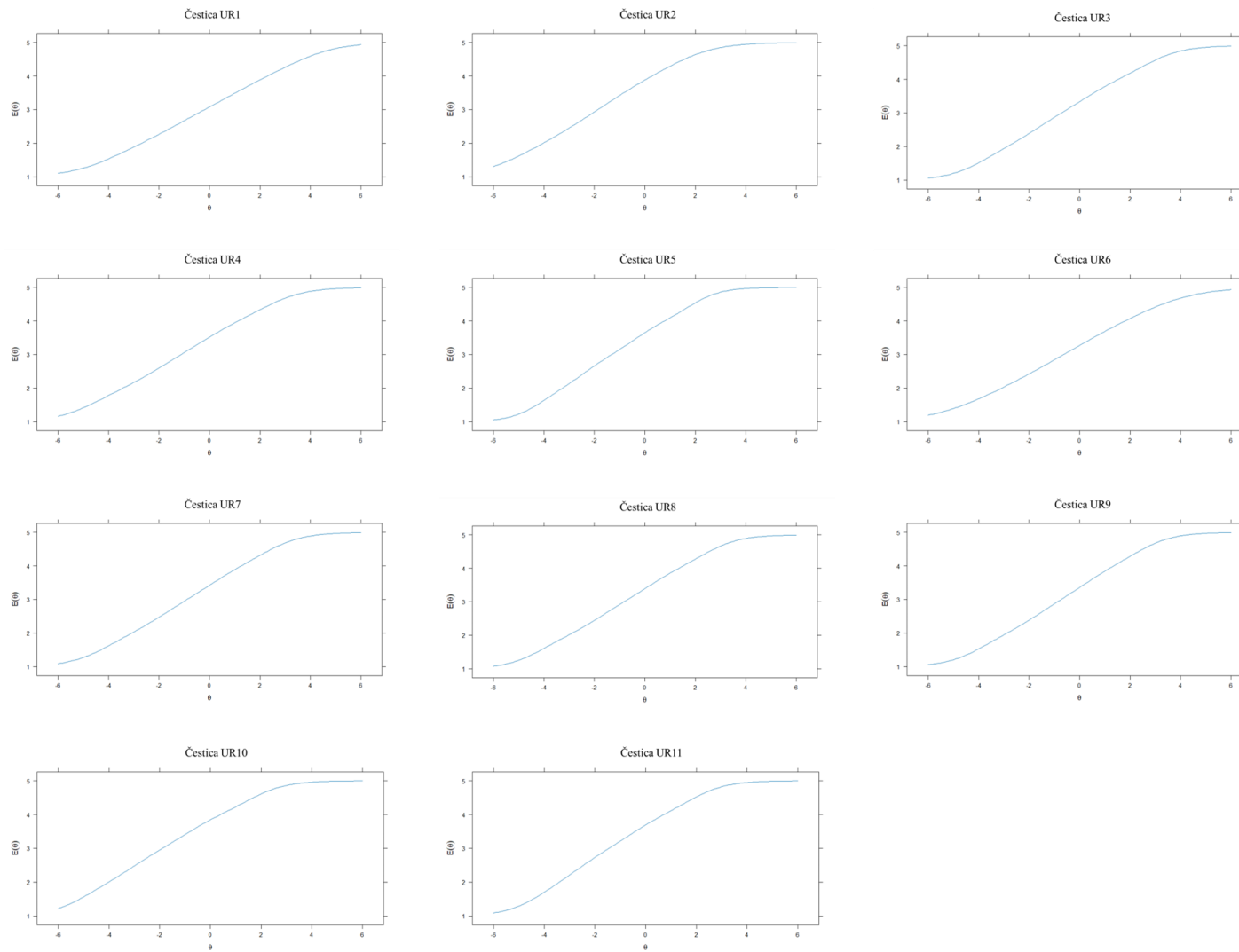
Slika 1. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti uočavanja i razumijevanja emocija



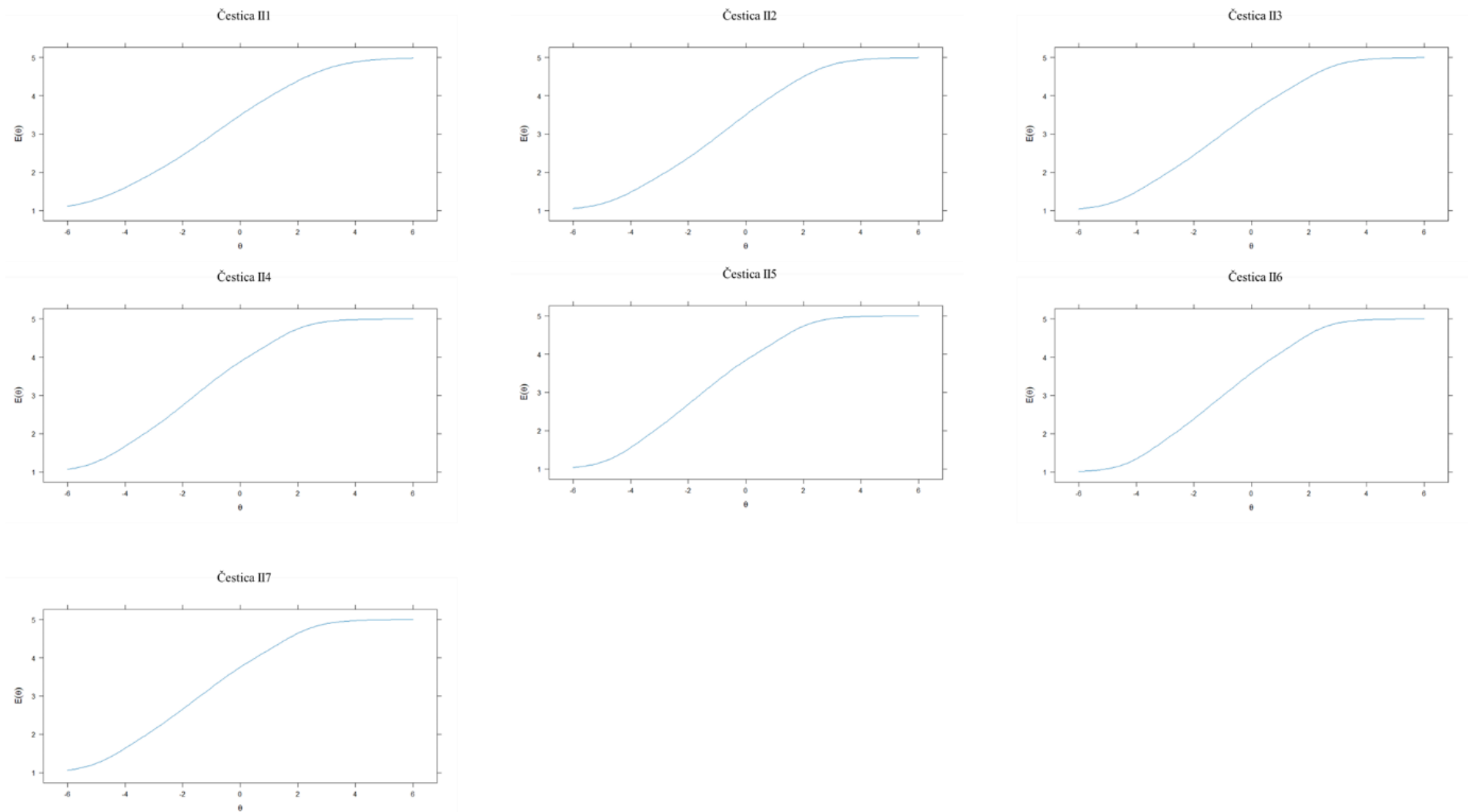
Slika 2. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti izražavanja i imenovanja emocija



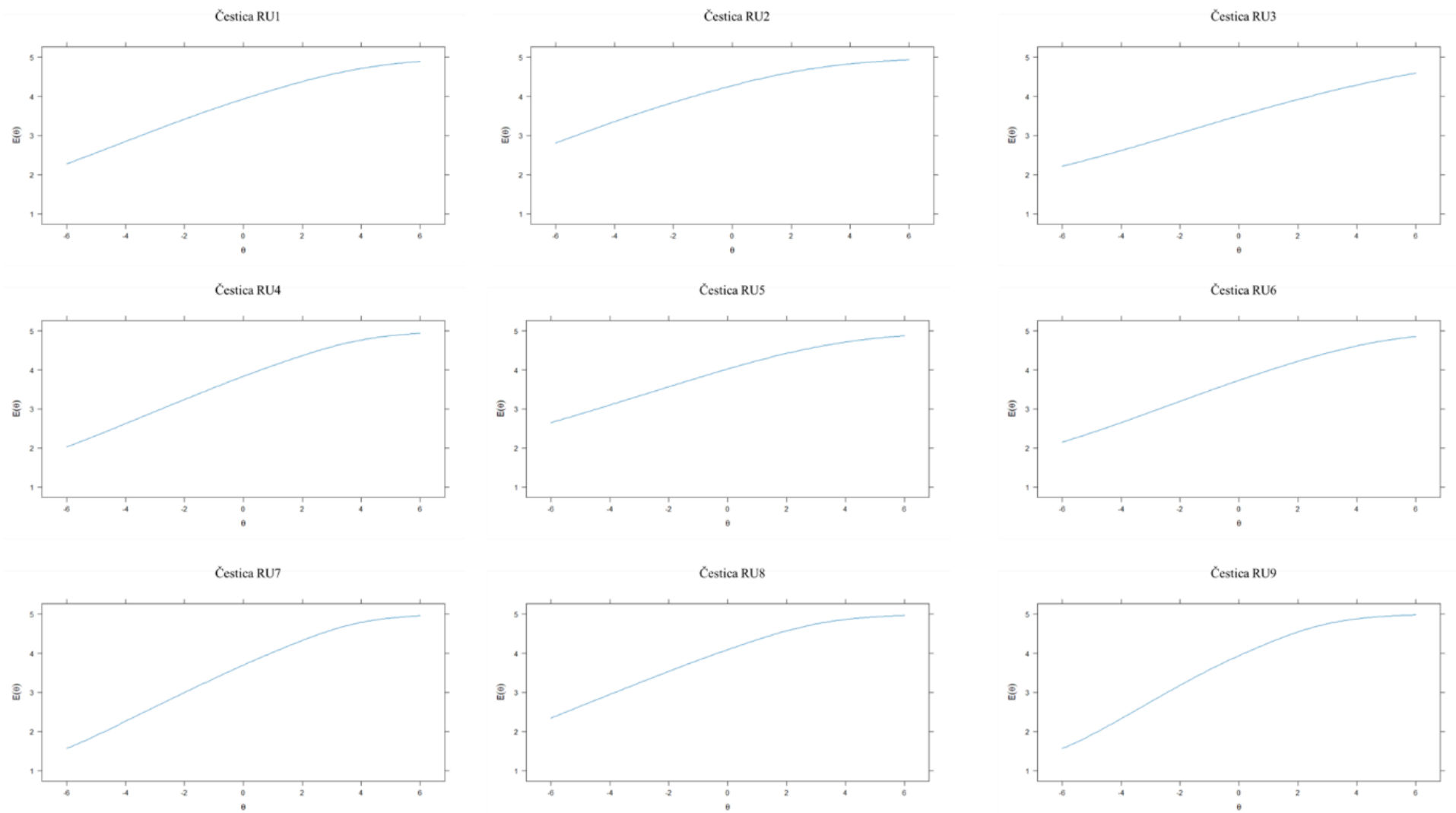
Slika 3. Grafički prikaz karakterističnih krivulja i informacijskih krivulja čestica sa Skale sposobnosti regulacije i upravljanja emocijama



Slika 4. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti uočavanja i razumijevanja emocija



Slika 5. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti izražavanja i imenovanja emocija

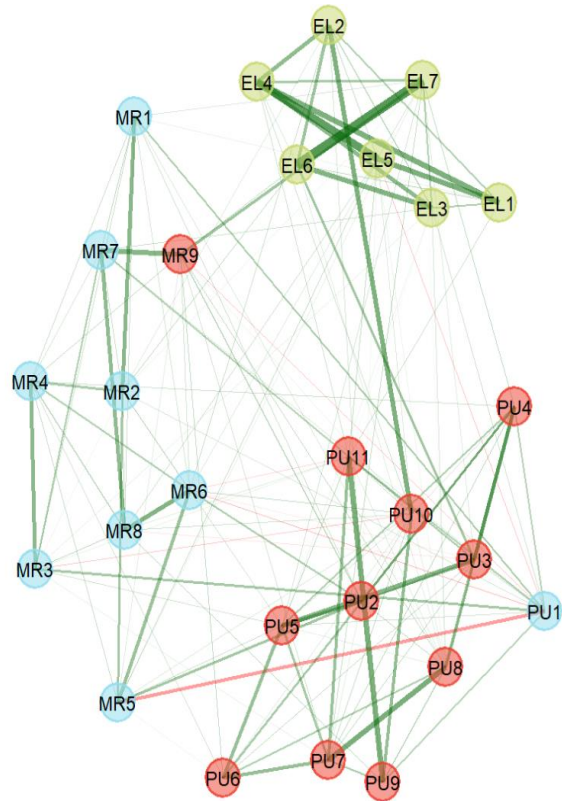


Slika 6. Grafički prikaz kumulativnih krivulja čestica sa Skale sposobnosti regulacije i upravljanja emocijama

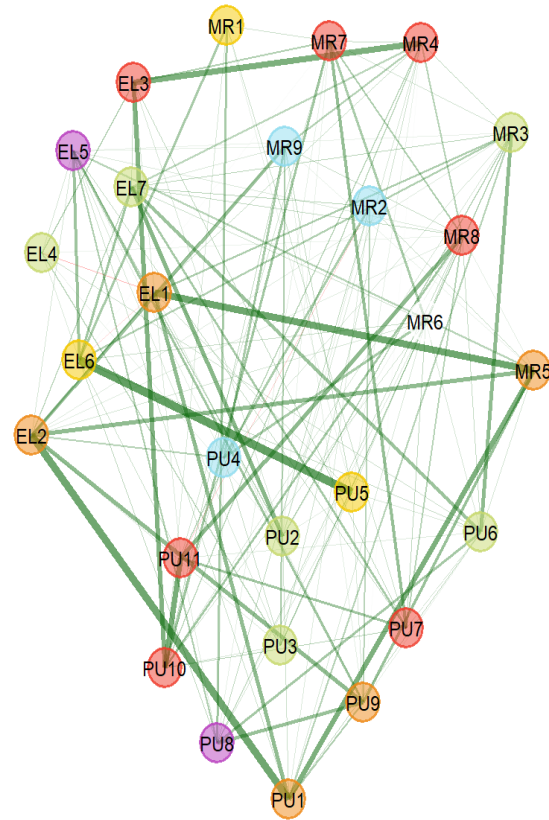
Privitak 4: Usporedni prikaz EGA rješenja svih 13 zemalja

314

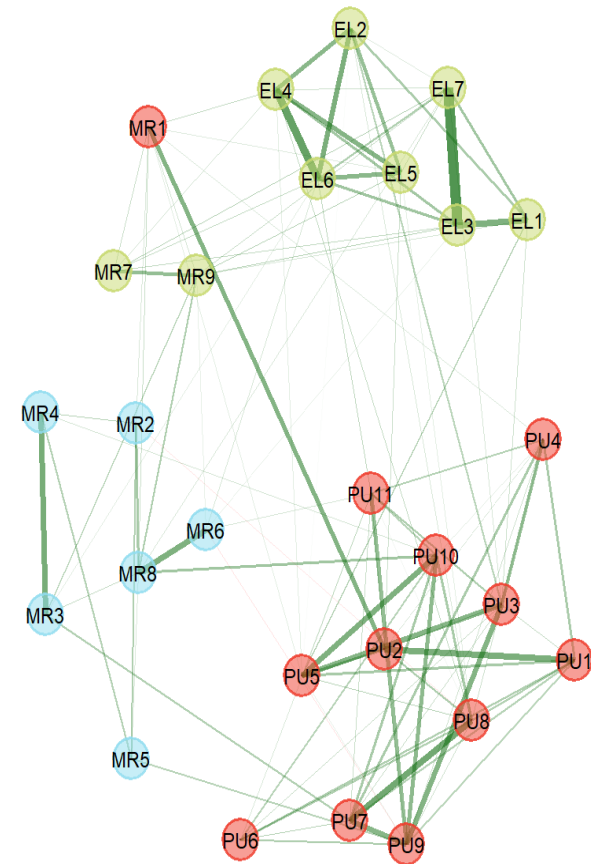
Hrvatska



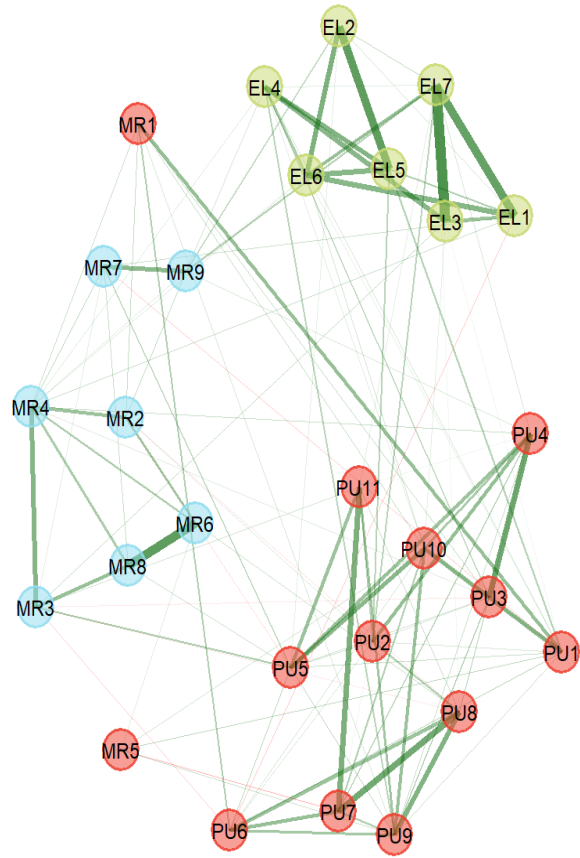
Portugal



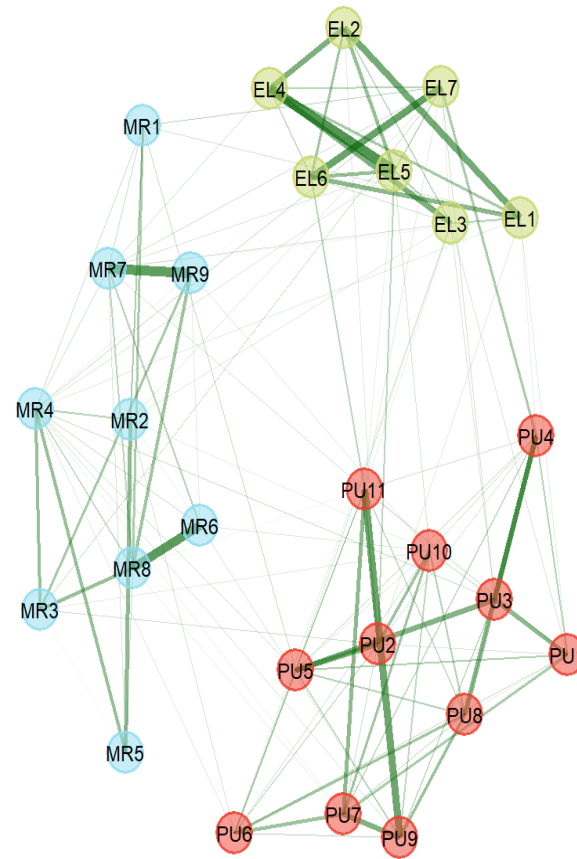
Finska



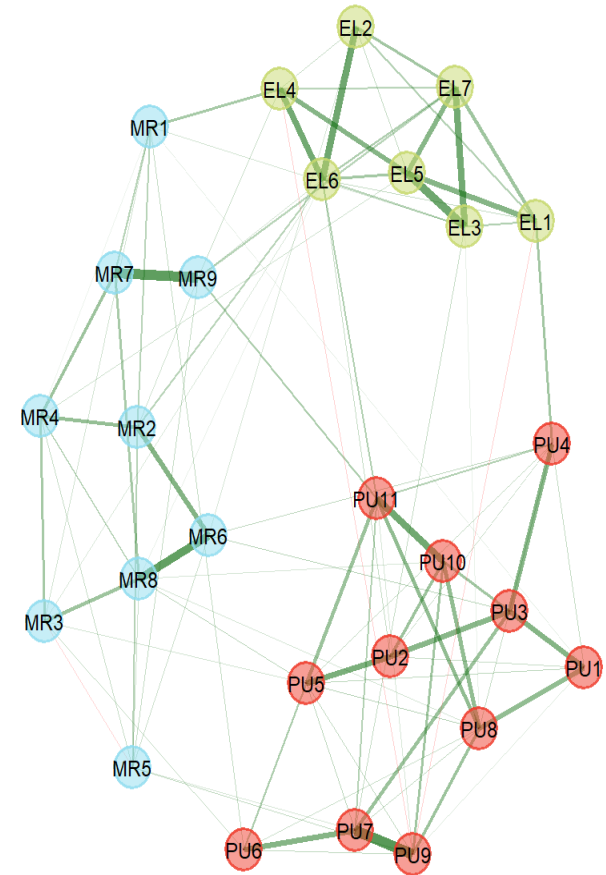
Švedska



Slovenija

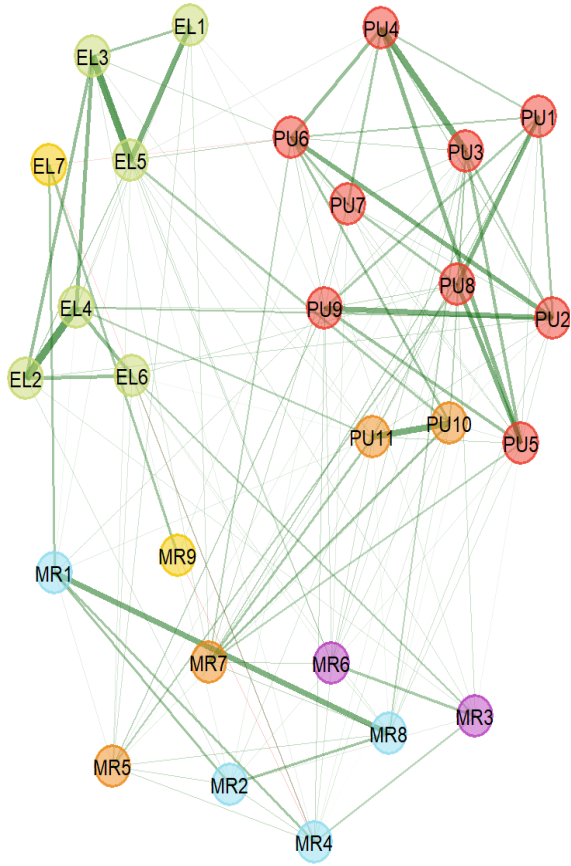


Španjolska

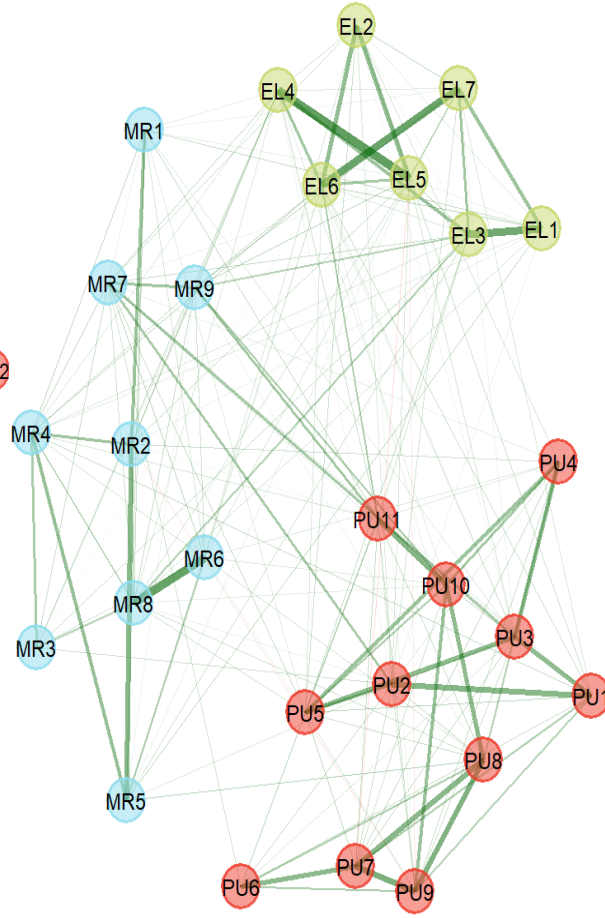


315

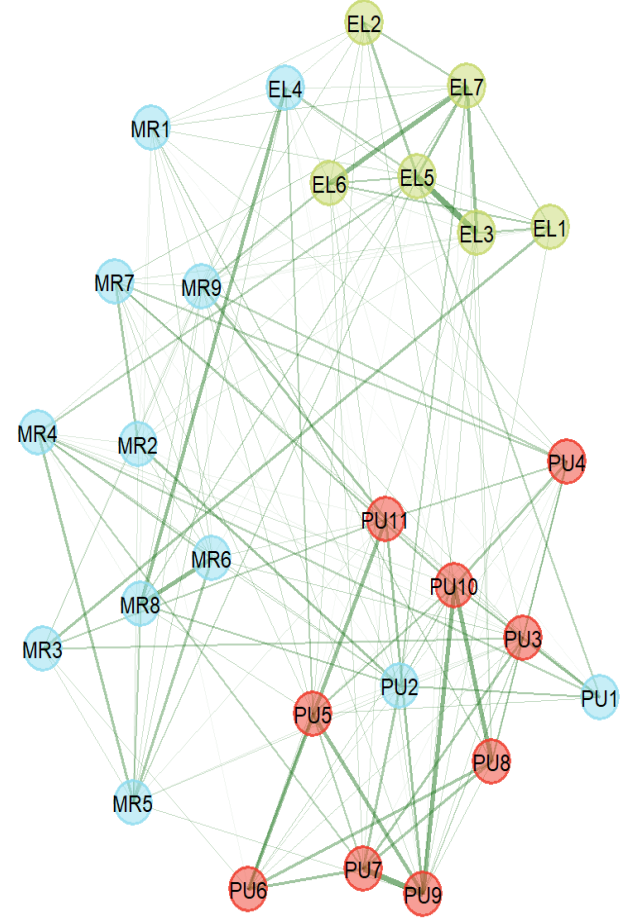
Japan



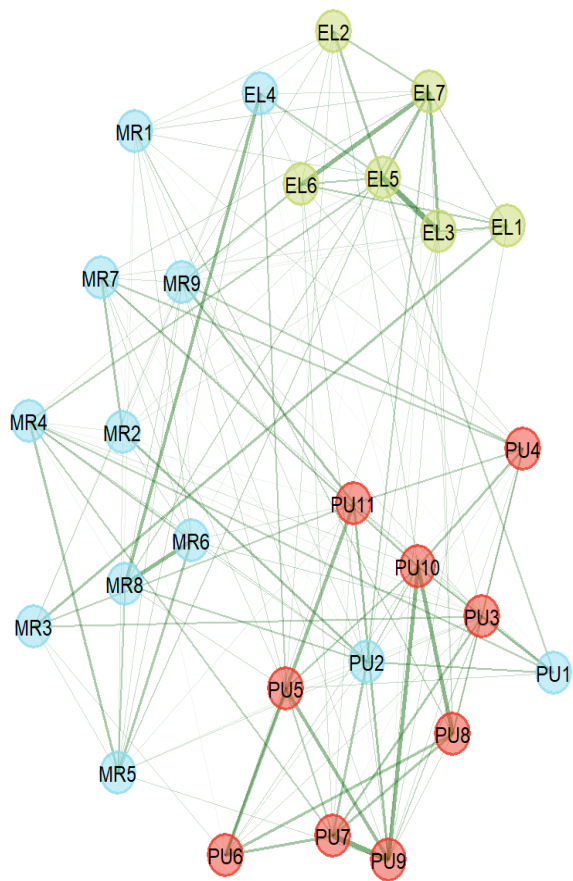
Kina



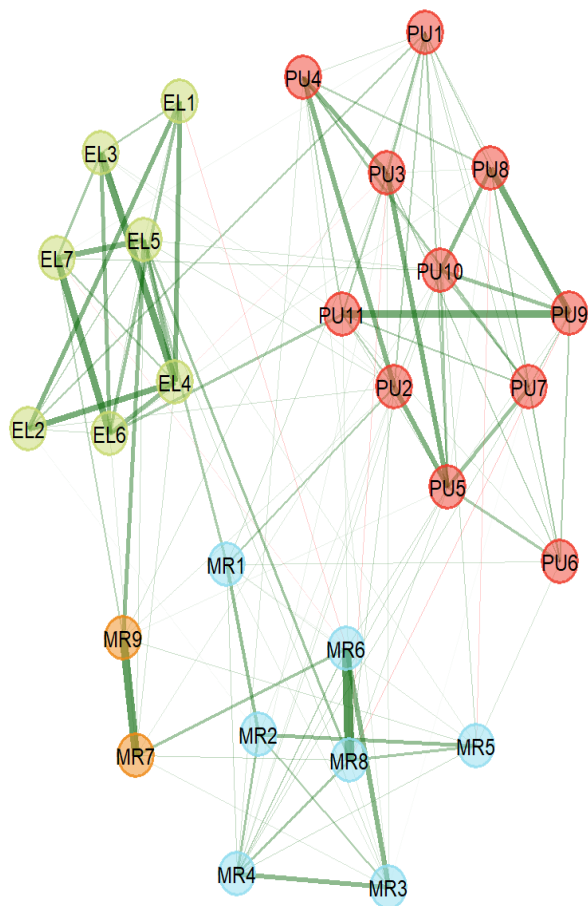
Indija



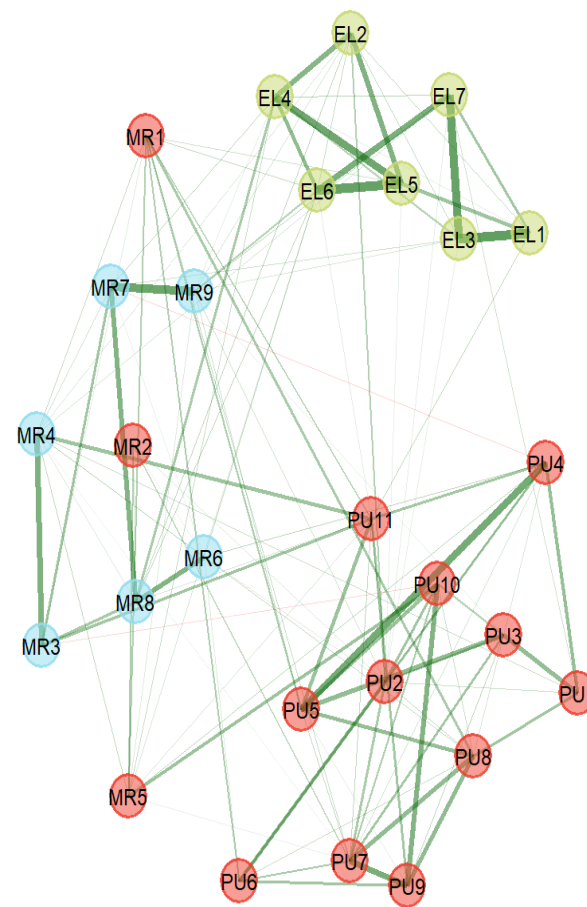
Argentina



Srbija

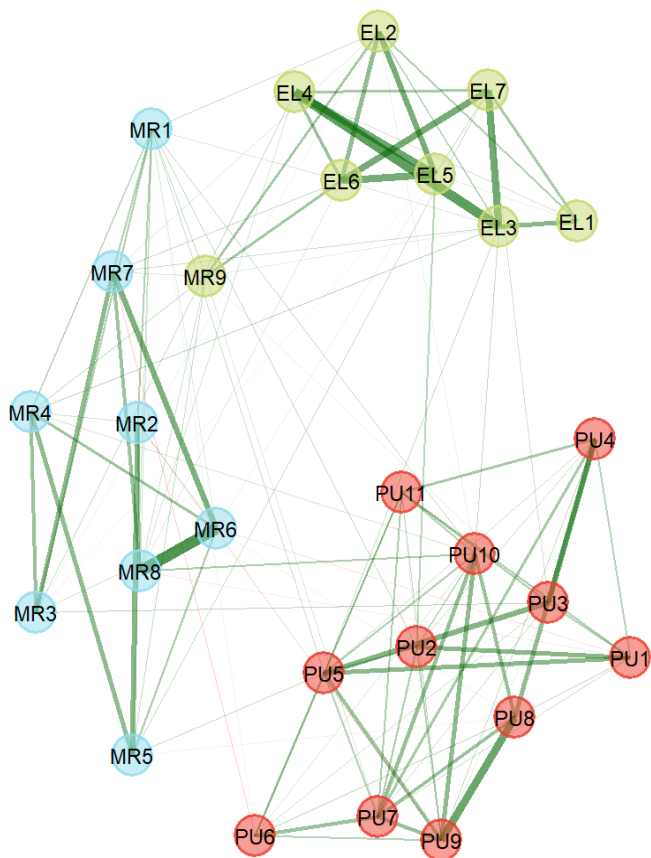


SAD






Turska

318



Napomena:

-  Skala sposobnosti uočavanja i razumijevanja emocija (UR)
-  Skala sposobnosti regulacija i upravljanja emocijama (RU)
-  Skala sposobnosti izražavanja i imenovanja emocija (II)