

ПРИМЕНА НА МОДЕРНАТА ПСИХОМЕТРИСКА ТЕОРИЈА ПРИ КОНСТРУКЦИЈА НА ПСИХОЛОШКИ ТЕСТОВИ

Ана Стојанов¹

Лабораторија за социјална когниција, Офис за психологија
Универзитетот во Отаго, Нов Зеланд

Кратка содржина

Во првиот дел од оваа статија се разгледуваат недостатоците на класичната психометриска теорија, како насоченост кон тестот, нестабилност на метриските карактеристики и претпоставката дека тестот ги дискриминира подеднакво добро сите испитаници, без оглед на нивото на латентниот психолошки конструкт. Во вториот дел е прикажана алтернатива на класичната психометрија – модерната психометрија. Во овој дел се претставени основните поими во модерната психометриска теорија (крива карактеристична за категоријата на одговор, функција на информацијата што ја дава ајтемот, функција на информацијата што ја дава тестот, како и основните претпоставки во модерната психометрија: претпоставка за едnodимензионалност и претпоставка за локална независност на ајтемите) со особен акцент на политомните модели. Во последниот дел е даден пример на селекција на ајтеми од Ликертов тип при конструкција на прашалник, но истите принципи можат да се употребат и за конструирање кратка верзија од веќе постоечки прашалник.

Клучни зборови: класична психометрија, ајтеми од Ликертов тип, модерна психометрија, полиномни модели

¹ ana.stojanov@postgrad.otago.ac.nz

Модерната психометрија (item response theory)² била развиена за да ги надмине недостатоците на класичната психометрија и претставува нејзино надополнување и надградување. Во оваа статија најпрво ќе бидат разгледани недостатоците на класичната психометриска теорија, а потоа ќе бидат претставени основните поими во модерната психометриска теорија. На крај ќе биде илустриран пример за анализа на ајтеми во рамки на модерната психометриска теорија со помош на програмата IRTPRO (Cai, Thissen & du Toit, 2017).

Ограничувања на класичната психометрија. Во класичната психометрија, особините на испитаниците не можат да се разделат од особините на тестот (Henard, 2000). Со други зборови, доколку тестот е тежок, изгледа како испитаниците да имаат ниски способности, а доколку е лесен, како да имаат високи способности. Но и обратното е точно - карактеристиките на тестот се одредени од пропорцијата на испитаници кои одговориле на одреден начин, и зависат од психолошкиот конструкт што се мери. За пример, да го разгледаме следниот ајтем „На нашиа земја ѝ треба силен лидер за да можат да се уништат радикалните и неморални струци што доминираат во нашето општество“, земен од скалата за мерење авторитарност (Zakrisson, 2005). За да се согласи некој сосема со овој ајтем, тој или таа треба биде авторитарен во поголема мера од некој друг испитаник кој воопшто нема да се согласи. Доколку овој ајтем се зададе на примерок кој е авторитарен, одговорите што ќе се добијат ќе бидат концентрирани околу опцијата „сосема се согласувам“. Со други зборови овој ајтем ќе се чини „лесен“, затоа што е веројатно сите „сосема да се согласат“ со него. Од друга страна, доколку ајтемот се зададе на примерок кој не е авторитарен, одговорите би гравитирале кон „воопшто не се согласувам“ и овој ајтем би изгледал многу „тежок“.³

Слично, местоположбата на испитаникот на психолошкиот конструкт зависи од ајтемите што се користат за да се мери конструктот, како и од другите испитаници што се тестираат. Во зависност од примерокот, даден испитаник ќе има различна позиција на латентната црта. На пример, доколку претходниот ајтем беше зададен на неавторитарна личност и на некоја умелно авторитарна личност, втората личност ќе изгледа како да е авторитарна

² Модерната психометрија најчесто се изедначува со теоријата на одговор на ајтеми (item response theory) која опфаќа семејство на модели (на пример, Kline, 2005; McDonald, 2013), затоа во оваа статија *item response theory* ќе биде преведено како *модерна психометриска теорија*.

³ Поимот за тежина на ајтем најдобро се разбира во контекст кога се разгледува некоја способност, но исто така може да се примени и на други конструкти како ставови, црти на личност или верувања. Кога се користи терминот „тежина“ во контекст на мерење на личноста тој означува дека испитаникот треба да го поседува психолошкиот конструкт во голема мера за да се согласи со дадениот ајтем.

во голема мера. Но, доколку тоа исто лице се тестира заедно со примерок кој е авторитарен, ќе изгледа како да не е воопшто авторитарен. Затоа, мерењето не е неваријантно (invariant), а метриските карактеристики што го карактеризираат ајтемот како тежина и дискриминативност, како и карактеристиките на испитаникот, можат да варираат од примерок до примерок. Во модерната психометрија, карактеристиките на ајтемите и на испитаниците се непроменливи и не зависат од примерокот. Заради тоа што тежината на ајтемите е земена предвид при пресметката на психолошкиот конструкт, процената на конструктот кај испитаникот е независна од ајтемите што се користат. Слично, кога претпоставките на моделот се задоволени, метриските карактеристики на ајтемот не варираат од примерок до примерок.

Друго ограничување на класичната психометрија е тоа што е ориентирана кон тестот (Henard, 2000). Со други зборови, за дадено ниво на конструктот, класичната психометрија не предвидува како испитаникот ќе одговори на даден ајтем. Да се предвиди како испитаникот ќе одговори на даден ајтем е важно во случаи каде целта е да се конструира скала со одредени карактеристики (на пример, тест наменет добро да дискриминира меѓу оние со многу висока способност, или тест наменет да дискриминира еднакво добро низ широк опсег на конструктот). Доколку намената на скалата е да се користи како прецизна мерка на латентниот конструкт, како што е случај при многу мерења на личноста, препорачливо е да се бираат ајтеми што покриваат широк опсег на конструктот. Ова овозможува фер интерпретација на скоровите, затоа што скоровите може да бидат „надуени“ доколку има висока концентрација на лесни ајтеми во даден регион на латентниот конструкт. За илустрација, може да се разгледа следново хипотетичко сценарио. Истражувачката сака да избере два ајтема за нејзиниот прашалник од вкупно три ајтеми. Да претпоставиме дека ајтемите имаат тежина како што е дадено во табела 1 (овие метриски карактеристики не ѝ се познати на истражувачката затоа што таа не ги анализира ајтемите во рамката на модерната психометрија).

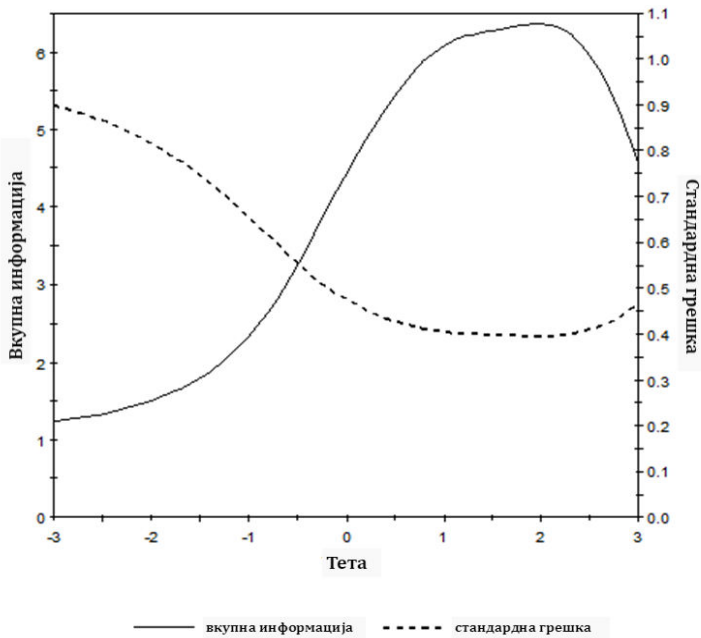
Табела 1. Метриската карактеристика тежина за фиктивни ајтеми

	b1	b2	b3
Ајтем 1	1.23	2.05	2.34
Ајтем 2	-2.03	0.12	1.89
Ајтем 3	0.78	1.34	2.09

Забелешка: b1, b2, b3 претставуваат тежина на ајтем, односно ги одразуваат нивоата на θ (изразени во стандардни отстапувања) на кои испитаниците имаат 50 % веројатност да преминат од еден одговор на следниот. Види го делот *Тежина и дискриминајтивност*.

Доколку истражувачката одлучи да ги задржи ајтемите број еден и три, тестот нема да биде дискриминативен меѓу оние со ниско ниво на способноста, односно тестот ќе биде „лесен“ и сите ќе имаат високи скорови. Но, овие метриски карактеристики не можат да бидат пресметани со помош на класичната психометрија. Затоа, потпирајќи се на класичната психометрија истражувачката може да не ги избере најдобрите ајтеми. Овие метриски карактеристики можат да бидат пресметани со помош на модерната психометрија и на тој начин можат да се изберат најсоодветните ајтеми за целта на тестот/скалата.

Друг недостаток на класичната психометрија е тоа што претпоставува дека грешката при мерењето и релијабилноста се еднакви за сите нивоа на конструктот (Emberson & Reise, 2000). Со други зборови, претпоставува дека тестот ги дискриминира подеднакво добро сите испитаници, без разлика на тоа во која мера го поседуваат конструктот. Модерната психометрија, пак, го признава фактот дека скоровите на тестот или скалата не се подеднакво прецизна мерка на конструктот. Затоа, во модерната психометрија стандардната грешка зависи од нивото на латентниот конструкт. За илустрација, на слика 1 е прикажана функција на информацијата што ја дава тестот (претставена со непрекинатата линија), како и стандардната грешка при мерењето (претставена со испрекинатата линија) за различни нивоа на θ (ознака за психолошкиот конструкт). Како што може да се види од слика 1, тестот дискриминира најмногу за оние нивоа на θ што се околу 1,5 стандардни отстапувања од аритметичката средина. За овие нивоа, стандардната грешка на мерењето е 0,42. Но, со приближување кон ниските нивоа на θ , дискриминативната моќ на тестот се намалува, а стандардната грешка при мерењето се зголемува. Затоа, во овој пример, испитаниците на +1.5SD се мерат со многу поголема прецизност отколку испитаниците на -1.5SD.



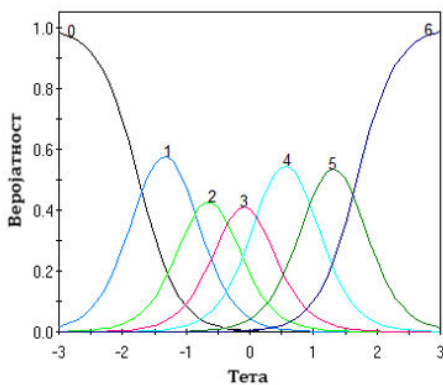
Слика 1. Илустрација на функција на информацијата што ја дава тестот и стандардната грешка на мерењето. Психолошкиот конструкт, θ , е даден во единици на стандардно отстапување. Сликата е добиена од необјавени сопствени истражувања.

Од досега изложеното може да се заклучи дека класичната психометрија има неколку ограничувања што се надминати со модерната психометрија.

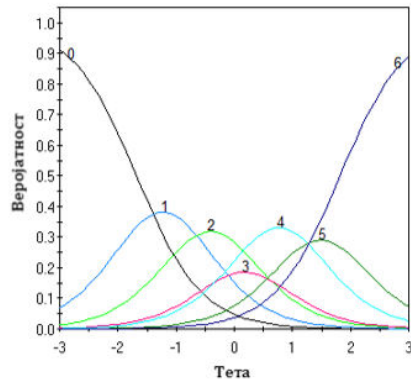
Основни поими во модерната психометрија. Модерната психометрија е теорија базирана на модел. Постојат неколку модели, во зависност од видот на ајтеми (на пример, дихотомни ајтеми – ајтеми со точен/погрешен одговор, или политомни ајтеми – ајтеми каде што има повеќе од два скорa, како на пример ајтеми од Ликертов тип) и од видот на метриски карактеристики што се пресметуваат. Најсоодветен модел за анализа на политомни ајтеми од Ликертов тип е моделот на подредени категории (graded response model) (Samejima, 1969, 1997). Во следниот дел ќе бидат прикажани поважните поими поврзани со модерната психометрија, со акцент на политомните ајтеми.

Крива карактеристична за категоријата на одговор. Кривата карактеристична за категоријата на одговор ја претставува веројатноста за бирање даден одговор како функција на θ . Рамни криви (на пример, слика 2 б) укажуваат дека ајтемот не дискриминира добро, бидејќи промените во θ се

следени со мали промени во веројатноста да се избере одреден одговор. Од друга страна, криви со голем наклон (на пример, слика 2 а) укажуваат на тоа дека ајтемот дискриминира добро, бидејќи мали промени во θ се следени со големи промени во веројатноста да се избере даден одговор. На пример, на слика 2 а и 2 б подолу, кривата обележана со 0 ја претставува веројатноста за избирање на опцијата „воопшто не се согласувам“ за два различни ајтеми соодветно. Кривата обележана со 1 ја претставува веројатноста да се избере опцијата „не се согласувам“, додека кривата обележана со 3 ја претставува можноста да се избере опцијата „делумно не се согласувам“ итн.



Слика 2 а.⁴ Криви карактеристични за категоријата на одговор за ајтем со голема дискриминативна моќ

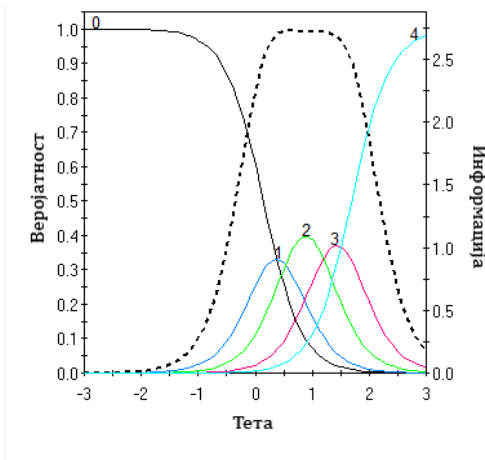


Слика 2 б. Криви карактеристични за категоријата на одговор за ајтем со ниска дискриминативна моќ

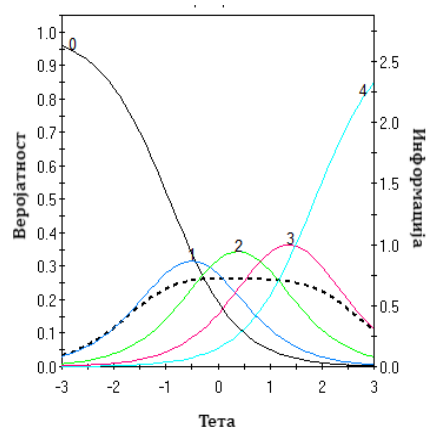
Крива на информација што ја дава ајтемот. Во модерната психометриска теорија, не се претпоставува дека ајтемите дискриминираат подеднакво добро помеѓу испитаници со различни вредности на θ . На пример, даден ајтем може да прави фина дистинкција меѓу испитаници со високи нивоа на конструктот, но не и меѓу испитаници со ниски нивоа на конструктот (види слика 3 а). Дискриминативната моќ може да се види од кривата на информација што ја дава ајтемот. Оваа крива всушност ја претставува дискриминативната моќ на ајтемот за разни нивоа на θ . Колку повеќе информација дава некој ајтем, толку повеќе дискриминира. На пример, на слика 3 а и 3 б се претставени ајтеми со високи и ниски функции

⁴ Сликите 2 а, 2 б, 3 а и 3 б се добиени од необјавени сопствени податоци. Сликите 4 а и 4 б се добиени од фиктивни податоци за илустративни цели.

на информација што ја дава ајтемот, што значи дека ајтемот од слика 3 а е подискриминативен од ајтемот даден на слика 3 б.



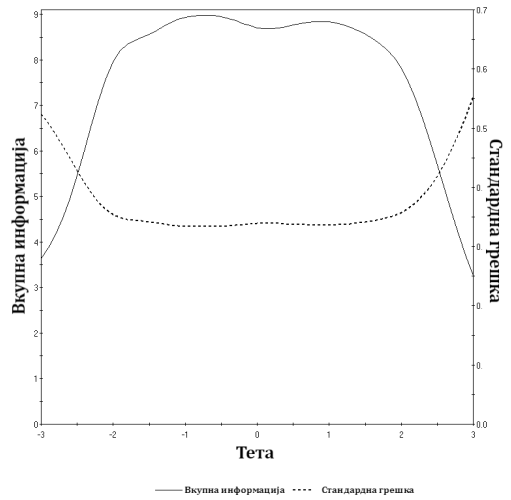
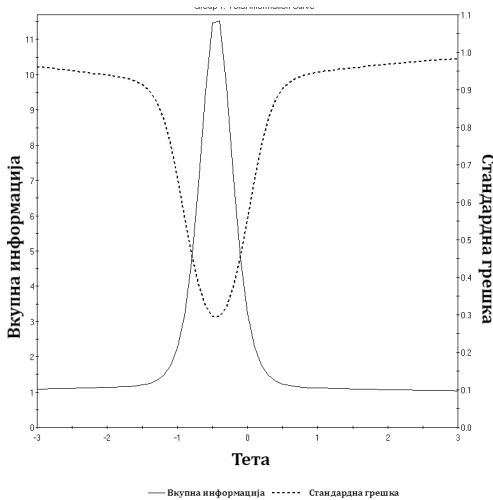
Слика 3 а. Крива на информација што ја дава ајтемот за ајтем што има висок коефициент на дискриминација, но дискриминира само во ограничен опсег



Слика 3 б. Крива на информација што ја дава ајтемот за ајтем што има понизок коефициент на дискриминација, но дискриминира долж поширок опсег

Крива на информација што ја дава ајтемот. Кривата на информација што ја дава ајтемот е адитивна, што значи дека може да се добие со сумирање на индивидуалните криви на информација што ја даваат ајтемите. Слично како и кривата на информација што ја дава ајтемот, кривата на информација што ја дава ајтемот ја преиспитува дискриминативната моќ на ајтемот како целина. За време на конструирање на прашалник, целта може да биде да се најрави прашалник што ќе дискриминира прецизно меѓу индивидуи со шесен ојсеј на θ , во кој случај ќе се тежнее кон тоа да се има крива на информација што ја дава тестот со врв околу вредностите на θ каде што тестот треба да биде дискриминативен (на пример слика 4 а). Доколку, пак, целта е да се има скала што ќе дискриминира меѓу широк опсег на θ , потребно е да се конструира скала што има крива на информација што ја дава тестот во форма на рамнина (на пример слика 4 б). Бидејќи информацијата што ја дава тестот на разни нивоа на θ не е иста, прецизноста на мерењето, исто така, варира долж континуумот. На пример, информација од 20 соодветствува на

алфа-коэффициент од 0,95, информација од 10 соодветствува на алфа од 0,9, а информација од 5 соодветствува на алфа од 0,8 (Reeve & Fayers, 2005).



Слика 4 а. Пример за крива на информација што ја дава тестот, кога тестот дискриминира добро за ограничен опсег на θ

Слика 4 б. Пример за крива на информација што ја дава тестот, кога тестот дискриминира добро за широк опсег на θ

Тежина и дискриминаџивноста. Моделот на поодредени категории обезбедува две метрички карактеристики - индекс на дискриминација, кој е концептуално сличен со корелацијата ајтем-вкупен скор, и тежина, уште познат и како локација. Индексот на дискриминација, a , го одразува степенот до кој ајтемот ги диференцира луѓето со различен степен на конструктив, па така во извесна смисла го претставува степенот до кој ајтемот го одразува психолошките конструкти (Furr, 2011). Вообичаено има вредност меѓу 1 и 3 (Reeve & Fayers, 2005), при што повисоки броеви значат подобра дискриминација. Тежината на ајтемот, b , во политомните модели ги одразува нивоата на θ на кои испитаниците имаат 50 % веројатност да преминат од еден одговор на следниот. Така, бројот на метрички карактеристики b (тежина) што се пресметува е еднаков на бројот на опции за одговор минус еден. На пример, за ајтем кој има четири можни одговори (воопшто не се согласувам, не се согласувам, се согласувам, сосема се согласувам) и следните b метрички карактеристики $b_1 = -1,23$, $b_2 = 0,78$ and $b_3 = 1,98$, метриската карактеристика b_1 укажува на тоа дека испитаник кој се наоѓа на $-1,23SD$ на латентниот конструкт, има 50 % веројатност да ја избере опцијата „не се согласувам“, а испитаник на $0,78SD$ има 50 % веројатност да ја избере опцијата „се согласувам“. Така, за да се сложат

испитаниците со тврдењето во повисок степен потребни се растечки вредности на психолошкиот конструкт.

Претпоставки во модерната психометриска теорија. Постојат две основни претпоставки во модерната психометриска теорија: локална независност и еднодимензионалност. Еднодимензионалност значи дека има само еден конструкт што се мери со скалата (Hambleton, Swaminathan, & Rogers, 1991). Во пракса, претпоставката за еднодимензионалност се задоволува со спроведување факторска анализа и добивање еден доминантен фактор што е одговорен за најмалку 20 % од варијансата (Nguyen, Han, Kim, & Chan, 2014), или пак добивање повеќе фактори, но соодносот на првата и втората ајген-вредност да е во пропорции 4:1 (Furr, 2011). Доколку конструктот е повеќедимензионален постојат специјално развиени модели за анализа на ајтеми што сочинуваат повеќедимензионална скала. Алтернативно, секоја супскала може да се анализира посебно со модели што подразбираат еднодимензионалност.

Локалната независност подразбира дека за дадено ниво на конструктот, одговорите на ајтемите се независни. Во суштина, тоа значи дека кога метриските карактеристики на моделот се земени предвид, не постои понатамошна врска помеѓу ајтемите (Emberson & Reise, 2000). Ова е слично на претпоставката за некорелирани грешки во класичната психометриска теорија и е поврзано со претпоставката за еднодимензионалност. Односно, доколку одговорите на дадени ајтеми не се независни, тоа значи дека некој друг фактор освен психолошкиот конструкт од интерес влијае на одговорите на ајтемите. На пример, доколку се тестира некоја способност, локалната независност може да не биде запазена доколку ајтемите кои мерат разбирање на текст се однесуваат на истиот параграф, или при мерење на личност доколку ајтемите имаат слична содржина (Chen, W.-H., & Thissen, 1997).

Во следниот дел ќе биде илустрирана употребата на моделот за подредени одговори при конструкција на прашалник од Ликертов тип.

Пример за анализа на ајтеми со моделот на подредени одговори. За илустрација на моделот на подредени одговори при анализа на ајтеми ќе се користат податоци добиени при конструкција на скалата за мерење на менталитет на заговор, поточно супскалата за мерење верувања во теориите на заговор (Stojanov & Halberstadt, во печат). Оваа супскала се состои од седум ајтеми, дадени во табела 3, кои имаат најдобри метриски карактеристики од вкупно 64 ајтеми користени за развој на прашалникот. Во овој илустративен пример

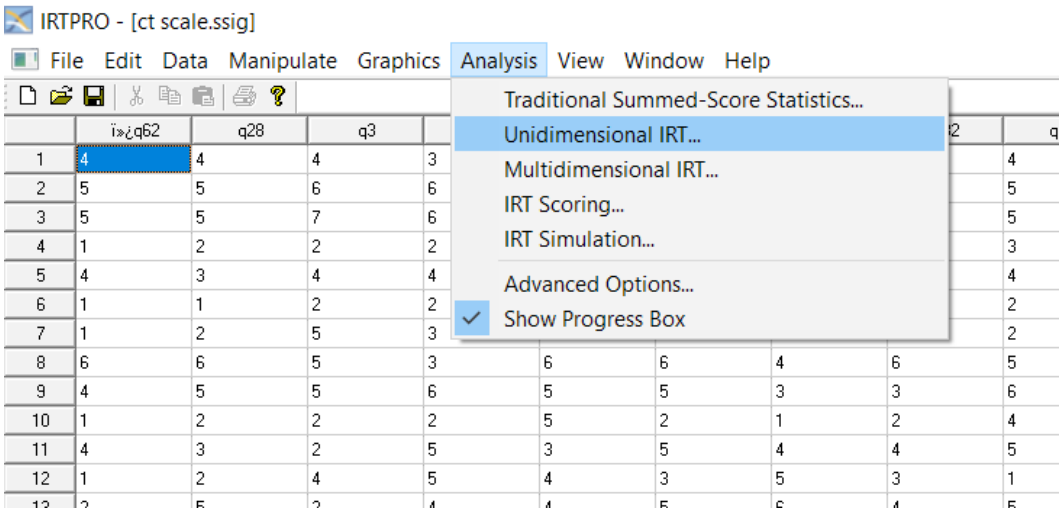
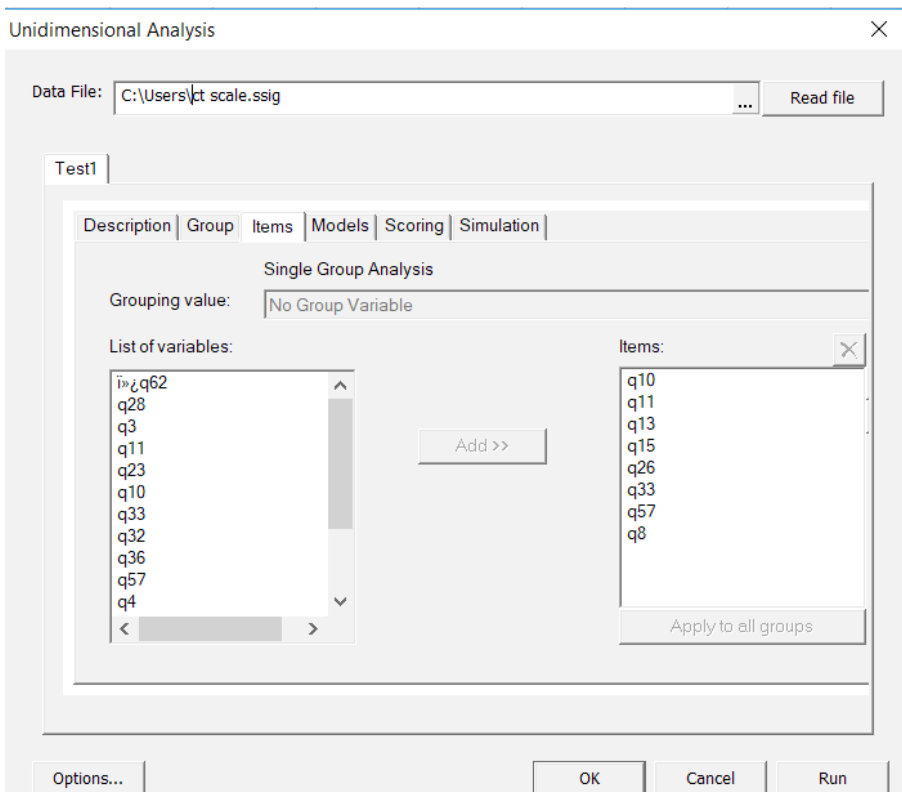
ќе се користат само 8 ајтеми, седумте ајтеми кои го сочинуваат прашалникот и уште еден ајтем кој е отфрлен од финалната верзија поради несоодветни метриски карактеристики, означен со * во табела 2.

Табела 2. Ајтеми од супскалата за мерење верувања во теоријата на заговор и ајтем што не е вклучен во финалната супскала

-
1. (q10) Алтернативните објаснувања за важни општествени настани се поблиску до вистината од официјалните информации.
 2. (q11) Владата или тајните организации се одговорни за необични или необјаснети настани.
 3. (q13) Многу ситуации или настани претставуваат нелегални или штетни дела од страна на владата или други моќни луѓе.
 4. (q15) Некои работи кои секој ги прифаќа како точни, всушност се измами креирани од моќни луѓе.
 5. (q26) Настани соопштени во вестите може никогаш да не се случиле.
 6. (q33) Многу таканаречени „случајности“ се всушност индиции за тоа како нештата навистина се случиле.
 7. (q57) Настаните низ историјата се внимателно планирани и организирани од лица за нивна лична корист.
 8. (q8) Зад сите важни настани има некоја причина. *
-

Одговорите се даваат на седумстепенa скала со опсег од 1 = воопшто не се согласувам до 7 = сосема се согласувам. За анализа ќе се користи програмата IRTPRO (Cai, Thissen, & du Toit, 2017).

Откако ќе се вчита датотеката, за да се пристапи кон анализа треба да се кликне на *Analysis->Unidimensional IRT...* како што е прикажано на слика 5. Потоа се појавува прозорец како на слика 6. Со селектирање на ајтемите и кликување на копчето *add* варијаблите што се предмет на анализа се префрлаат од белиот простор на левата страна во празниот простор на десната страна. Потоа се кликува на *run*.

Слика 5. Мени *Анализа*

Слика 6. Селекција на ајтеми за анализа

Исписот се состои од неколку дела. Прво се дадени пресметаните метриски карактеристики и нивните стандардни грешки (слика 7). Од слика 7 очигледно е дека ајтемот број осум не дискриминира добро ($a = 0,93$).

Graded Model Item Parameter Estimates for Group 1, logit: $a(\theta - b)$ (Back to TOC)

Item	Label	a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.	b_5	s.e.	b_6	s.e.	
1	q10	7	2.52	0.19	-1.83	0.13	-1.04	0.09	-0.42	0.07	0.32	0.07	1.09	0.09	1.94	0.13
2	q11	14	3.29	0.26	-1.76	0.12	-0.93	0.08	-0.36	0.06	0.19	0.06	0.94	0.07	1.66	0.11
3	q13	21	2.94	0.22	-1.90	0.13	-1.03	0.08	-0.52	0.07	0.10	0.06	0.85	0.07	1.64	0.11
4	q15	28	2.41	0.18	-1.65	0.12	-0.93	0.08	-0.44	0.07	0.09	0.06	0.83	0.08	1.70	0.12
5	q26	35	1.70	0.14	-1.73	0.15	-0.80	0.09	-0.04	0.08	0.40	0.08	1.19	0.11	1.87	0.15
6	q33	42	2.34	0.18	-1.86	0.14	-0.95	0.09	-0.42	0.07	0.13	0.07	0.98	0.08	1.89	0.13
7	q57	49	2.09	0.16	-1.95	0.15	-0.95	0.09	-0.28	0.07	0.22	0.07	1.12	0.09	1.90	0.14
8	q8	56	0.93	0.11	-4.56	0.58	-3.00	0.35	-1.90	0.23	-0.94	0.15	0.54	0.12	2.26	0.26

Слика 7. Дискриминативност и тежина на ајтемите

Следната информација во исписот покажува колку добро ајтемот се вклопува во моделот (слика 8). Значајни вредности укажуваат дека ајтемот значително отстапува од тоа што е предвидено со моделот. Бидејќи се прават повеќе тестирања на значајни разлики на истите податоци, треба да се контролира грешка од типот еден. Прилагодените веројатности со процедурата Бенџамини-Хохберг (Benjamini & Hochberg, 1995) не се значајни на ниво 0,05 што укажува на тоа дека ајтемите добро се вклопуваат во моделот.

S- χ^2 Item Level Diagnostic Statistics

Item	Label	χ^2	d.f.	Probability
1	q10	122.16	102	0.0846
2	q11	88.88	89	0.4844
3	q13	99.49	98	0.4398
4	q15	137.43	109	0.0341
5	q26	130.59	128	0.4194
6	q33	99.34	104	0.6115
7	q57	126.01	114	0.2080
8	q8	179.30	136	0.0076

Слика 8. Дијагностика на ајтемите

Следна е информацијата за локална независност на ајтемите, дадена на слика 9. Доколку вредностите се поголеми од 10 се смета дека ајтемите ја нарушуваат претпоставката за независност. Во овој случај нема такви ајтеми.

Marginal fit (χ^2) and Standardized LD χ^2 Statistics for Group 1 (Back to TOC)

Item	Label	Marginal								
		χ^2	1	2	3	4	5	6	7	
1	q10	2.0								
2	q11	2.2	2.8							
3	q13	1.0	4.2	5.1						
4	q15	0.6	3.6	2.3	2.6					
5	q26	0.7	-0.5	0.8	0.6	4.9				
6	q33	1.1	4.1	2.1	3.3	2.9	2.0			
7	q57	1.5	1.0	3.9	3.2	1.5	2.3	3.2		
8	q8	0.6	5.3	4.9	4.9	4.3	4.7	5.0	5.2	

Слика 9. Локална независност на ајтемите

На крај е дадена табелата која содржи информации за кривата на информација добиена од ајтемите. Оваа табела е прикажана на слика 10.

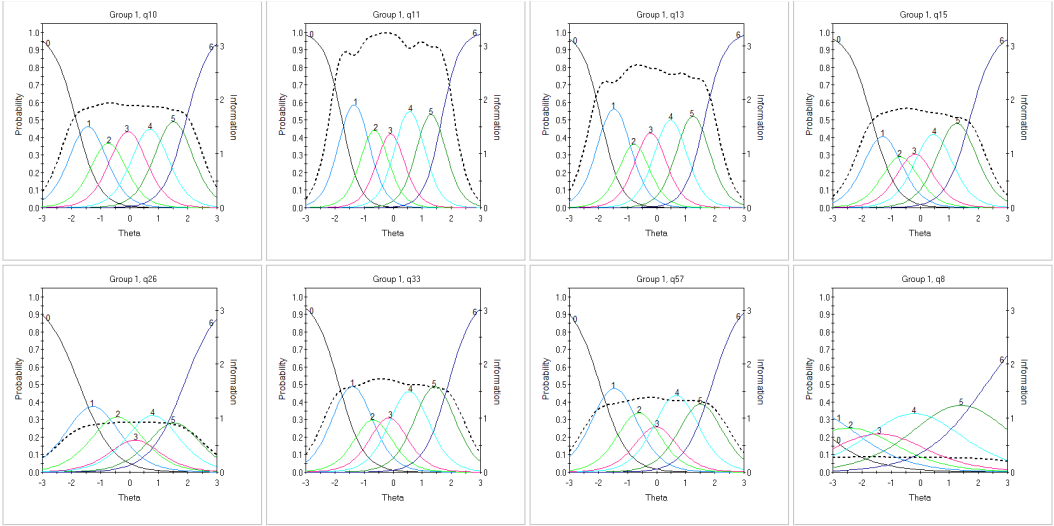
Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 (Back to TOC)

Item	Label	θ :														
		-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	q10	0.47	1.00	1.58	1.82	1.88	1.93	1.92	1.89	1.89	1.87	1.85	1.80	1.67	1.16	0.58
2	q11	0.33	1.04	2.33	2.89	2.89	3.17	3.24	3.21	3.02	3.03	3.00	2.95	2.04	0.81	0.25
3	q13	0.54	1.32	2.19	2.32	2.44	2.63	2.61	2.57	2.47	2.49	2.41	2.35	1.68	0.76	0.27
4	q15	0.32	0.71	1.25	1.64	1.76	1.82	1.84	1.81	1.76	1.73	1.68	1.62	1.31	0.77	0.36
5	q26	0.35	0.54	0.72	0.84	0.88	0.90	0.92	0.92	0.92	0.92	0.91	0.88	0.78	0.61	0.41
6	q33	0.50	0.95	1.40	1.56	1.61	1.70	1.72	1.70	1.63	1.61	1.59	1.56	1.42	0.99	0.52
7	q57	0.54	0.90	1.17	1.26	1.30	1.35	1.38	1.38	1.34	1.32	1.33	1.30	1.17	0.86	0.50
8	q8	0.28	0.28	0.28	0.28	0.28	0.28	0.27	0.27	0.27	0.27	0.26	0.26	0.25	0.24	0.21
Test Information:		4.32	7.73	11.93	13.60	14.04	14.77	14.90	14.75	14.30	14.24	14.03	13.71	11.31	7.19	4.11
Expected s.e.:		0.48	0.36	0.29	0.27	0.27	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.30	0.37	0.49

Marginal Reliability for Response Pattern Scores: 0.93

Слика 10. Вредности од кривата на информација што ја даваат ајтемите за вредности на θ од -2.8SD до 2.8SD

За да се видат кривите карактеристични за категоријата на одговор, како и кривата на информација, потребно е да се кликне на Analysis-> Graphs, при што ќе се добие следната слика (слика 11).



Слика 11. Криви карактеристични за категоријата на одговор и крива на информација што ја дава ајтемот за осумте ајтеми

Од графиконите на слика 11, слично како и од табелата на слика 7 и слика 10 се гледа дека осмиот ајтем (q8) не дискриминира меѓу испитаниците. Бидејќи овој ајтем не нуди голема информација за нивото на психолошкиот конструкт, најдобро е да се отфрли од скалата.

Литература

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289-300.
- Cai, L., Thissen, D., & du Toit, S. H. C. (2017). *IRTPRO 4.1 for Windows [Computer software]*. Skokie, IL: Scientific Software International, Inc.
- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265-289.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. London: Sage Publications.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Henard, D.H. (2000). Item response theory. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 67-97). Washington, DC: American Psychological Association.
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA, US: Sage Publications, Inc.
- McDonald, R.P. (2013). Modern test theory. In T.D. (Ed.), *The Oxford Handbook of quantitative methods in psychology, Vol. 1*. Oxford University Press, Retrieved 20 March 2019 from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199934874.001.0001/oxfordhb-9780199934874-e-007>.
- Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient*, 7, 23-35. doi:10.1007/s40271-013-0041-0
- Reeve B. B., & Fayers P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers, & R.D. Hays (Eds.), *Assessing quality of life in clinical trials* (2nd ed., pp. 55-73). New York, NY: Oxford University Press.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4, Pt. 2), 100. New York, NY.
- Samejima, F. (1997). Graded response model. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp.85-100). New York: Springer.
- Stojanov, A., & Halberstadt, J. (in press). The conspiracy mentality scale: Distinguishing between irrational and rational suspicion. *Social Psychology*.
- Zakrisson, I. (2005). Construction of a short version of the right-wing authoritarianism (RWA) scale. *Personality and Individual Differences*, 39(5), 863-872.

APPLICATION OF ITEM RESPONSE THEORY TO PSYCHOLOGICAL TEST CONSTRUCTION

Ana Stojanov

Abstract

In the first part of this paper the shortcomings of classical test theory, such as test-centeredness, dependence of parameter estimates on the sample used and the assumption that the test discriminates equally among all participants, are presented. In the second part, item response theory is offered as an alternative to classical test theory. Here, the basic concepts in item response theory are discussed (category characteristic curve, item information function, test information function, as well as the basic assumptions of item response theory: the assumption of unidimensionality and local independence of items) with particular focus on polytomous models. In the last part the use of item response theory is demonstrated for selection of Likert type items during test construction, but the same principles can be used in constructing a shorter version of an existing questionnaire.

Key words: *classical test theory, Likert type items, psychometrics, item response theory, polytomous models*