

یک مقایسه از تئوری پاسخ آیتم و روش های آنالیتیک عامل تاییدی برای اندازه

گیری هم ارزی / تغییرناپذیری

اخیراً علاقه ای فزاینده به آزمون های اندازه گیری هم ارزی /عدم تغییر وجود دارد. این مطالعه از داده های شبیه سازی شده با ویژگی های دانسته برای دسترسی به تناسبات، تشابهات و تفاوت ها بین تحلیل عامل تاییدی و روش های تئوری پاسخ آیتم برای ارزیابی ME/I استفاده می کند. نتایج نشان می دهند که هرچند این روش بدون اشکال نیست، تئوری پاسخ آیتم مبتنی بر روش به نظر می رسد برای برخی از انواع تحلیل های ME/I بیان شود.

کلمات کلیدی: تغییرناپذیری اندازه گیری، هم ارزی اندازه گیری، تئوری پاسخ آیتم، تحلیل عامل تاییدی، مونت

کارلو

ترجمه فا

هم ترازی / تغییرناپذیری اندازه گیری (MI.:/I; Vandenberg 2002 J) می تواند به عنوان عملیات هایی برای نشان دادن اندازه ه های همان یک ویژگی و تحت شرایط مختلف تصور شود (McArdlc. 1992 & I lorn). این شرایط متفاوت شامل پایداری اندازه گیری در طول زمان (Golembiewski. Billingsley, & Yeager. 1976), می شود. همراه با جمعیت های مختلف (مثلاً فرهنگ ها -Eactcau Riordan & Vandenberg, 1994; - & Craig, 2001) یا در راسنه های مختلف از اجرای اندازه گیری (مثلاً اجرای بازبینی تحت وب به ازای اندازه های مداد-کاغذ (Taris. Hok. 6.: Meijer. 109). تحت این سه شرط آزمون های ME/I نوعاً توسط روش های تحلیل عامل تاییدی (CFA) انجام می شوند. این روش های متعاقباً در مدت 20 سال گذشته ارتقا یافته و به طور گسترده در وضعیت های مختلف استفاده شده اند (Vandenberg & I allCC, 200). هرچند، تئوری پاسخ آیتم (IRT) برای هر مقصد مشابه استفاده می شود و در برخی موارد می تواند اطلاعات مفیدتر و بالقوه را برای ایجاد تغییرناپذیری اندازه گیری تامین نماید (Maurer, Raju. & Collins. 1991):

(McDonald, 199':1; Raju, LalTille. & Byrne, 2(02

اخيراً، موضوع ME/I از توجه را در میان محققان و اقدام کنندگان جلب کرده است (Vandenberg, 2002). یک دلیل برای این توجه فزاینده درک بهتر در میان جوامع تحقیقاتی سازمانی و روانشناسی اهمیت ایجاد ME/I است. به علت برخی از افزایش در تعداد مقالات و مقالات کنفرانس و یک کتاب کلی (Schricshcirn & Neider. 200) (ا) متعهد به این موضوع. همراه با افزایشی در تعداد مقالات در پرینت موضوع، درک بهتر از مکانیک تحلیل های مورد نیاز برای ایجاد ME/I منجر به افزایشی در علاقه کلی به موضوع شده است.

به عنوان نتیجه، مطالعات اخیر سازمانی تحقیقات تحلیل های ME/I را برای ایجاد ME/I برای موجودی های آنلاین در مقابل مداد-گاغذ و آزمون ها هدایت نموده است. (Donovan, Meade, & Probst. :WO(): Drusgow. & Probst. 2003. Craig & raiser) (Lautenschlager, Michels. & Gentry. 2{()3 بازخورد چندمنبعی (Craig & raiser) 2003. Iromcn & .) (Mauivi2001 Craig و Faeten) (1998) مقایسه های نگرش های از راه دور و کار در محل (& Iromcn & .) (Raju, 2(JU3 پاسخ های کارمندان مرد و زن (Collins, Raju, & Edwards, 2000), مقایسه های عرضی (Collins et al., 2000), مقایسه ها در عرض فرهنگ (Ghorpacle. I laurup, & Lackritz. 1999; & Rogg, 2002; & Riordan & Vandenberg, 1994; Ployhart. Wiechmann, Schmitt. SdCU Riordan. Richardson. Schaffer,) (Steenkarnp, & Baumgartner, 1998) و چندین ارزیابی همراه (Riordan. Richardson. Schaffer,) (Steenkarnp, & Baumgartner, 1998) (& Vandenberg, 2001; Taris et al., 19':18

(Vandenberg (2002) فراخوانی برای تحقیقاتی فزاینده را در تحلیل های ME/I و منطق ورای آن صورت داده است، با این بیان که جنبه منفی این روش هرچند توسط سوال نکردن از باور به قسمت برخی از آن مشخص شود تحت تمام کیفیتها متعبر یا درست است. (صفحه 140). به طور مشابه، (Riordan ct al. (2001) برای افزایش در مطالعات مونت کارلو برای تعیین بازده روش های موجود برای بررسی ME/I درخواست شده است. بعلاوه، هرچند (and Raju ct al. (2002 (1993) Reise. Widarnan. and I'ugh) بررسی کلی از مشابهت ها و تفاوت ها در CFA و IRT برای ME/I فراهم می کند و این روش ها با استفاده از نمونه داده ها نشان داده می شوند. آنها به

طور واضح برای مطالعه مرتبط با داده های شبیه سازی درخواست می شوند به گونه ای که دو روش می تواند به طور مستقیم هدایت شود. مطالعه کنونی از داده های شبیه سازی شده با فقدان دانسته از ME/I در تلاشی برای تعیین شرایط استفاده می کند که تحت آن CFA و irt منتج به نتایج مختلف با توجه هم تراز می اندازد می شود.

آزمون های CFA برای ME/I

تحلیل های CFA می تواند به طور ریاضی توسط فرمول زیر توصیف شود

$$x_i = \tau_i + \lambda_i \xi + \delta_{i\epsilon} \quad (1)$$

که در آن x_i پاسخ مشاهده شده برای قلم i است، t قطع برای قلم i ، و λ_i و τ_i بارگذاری برای آیت i و ساخت جانبی است. و مورد باقیمانده/خطا برای قلم تحت این روش است. واضح است که پاسخ مشاهده شده ترکیب خطی از متغیر پنهان، قطع آیت، عامل بارگذاری و برخی امتیازات خطا/باقیمانده برای آیت است.

(WOO Vandenberg and Lance) یک بازنگری کلی نوشته های ME/I را هدایت نمود و تعدادی از مدل تو درتوی پیشنهاد شده را برای تشخیص ME/I ترسیم نمود. محققان به طور عادی اولین اجرا از آزمون عمومی برای معادل ماتریس کوواریانس انجام دادند. (Vandenberg & Lance, 2000)، اگر آزمون عمومی نشان دهد که هیچ تفاوتی در ماتریس های کوواریانس بین مجموعه داده ها وجود ندارد، آنگاه محققان می توانند نتیجه بگیرند که شرایط ME/I برای داده ها برقرار است. هرچند، برخی نویسندگان از مفید بودن این آزمون خاص سوال نموده اند (Wcrtx. & Flau;her. 197R Jorcskog. 1971; Muthcn, as cited in Raju et al., 2002; Ruck) در زمینه هایی که این آزمون می تواند نشان دهد که ME/I به طور معقول مقاوم است هنگامی که آزمون های خاص تر از Me/I در غیراینصورت یافت شوند. بدون توجه به اینکه آیا آزمون کلی تعادل ماتریس کوواریانس نشاندهنده سری Me/I از مدل تودرتو باشد آزمون های تفاوت مربع چپ می تواند برای تعیین منابع ممکن اختلاف تعیین شود.

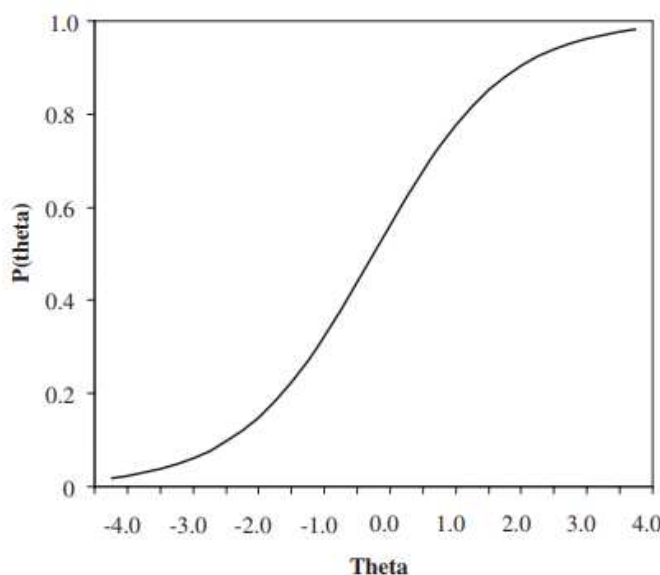
برای انجام آزمون های مدل تودرتو، داده ها (نشاندهنده گروه ها، دوره های زمانی و غیره) به طور همزمان آزمایش می شوند، با حفظ تنها الگوی تغییرناپذیری بارگذاری عامل. به بیانی دیگر؛ همان آیت ها برای بارگذاری در همان

عوامل ناچاراً استفاده می شوند، اما برآورده های پارامتر خودشان برای تغییر بین گروه ها مجاز هستند. مدل خط اصلی الگوهای عامل معادل یک مقدار مربع-چی را فراهم می کند که منعکس کننده تناسب برای پارامترهای آیتم برآورده شده به طور جداگانه برای هر وضعیت است و نشاندهنده یک آزمون برای تغییرناپذیری پیکربندی شده است (horn & Mc Ardle 1992). سپس، یک آزمون برای تغییرناپذیری بارگذاری عامل همراه با وضعیت ها توسط آزمون های یک مدل خط انجام می شود به جز بارگذاری عامل (یعنی در ماتریس AX) محدود به معادل بودن در آن وضعیت ها می شوند. تفاوت در خط اصلی و مدل محدودتر به عنوان آمار مربع-چی با درجات آزادی معادل با تعداد پارامترهای آزاد شده بیان می شود. هرچند آزمون های بعدی از بارگذاری های عامل آیتم منفرد می تواند هدایت شود، این به ندرت در عمل انجام می شود تقریباً هرگز انجام نمی شود مگر اینکه آزمون کلی ماتریس های AX متفاوت مهم باشد (Vandenberg & Lance, 2000)، بعد از آزمون های بارگذاری عامل، (Vandenberg & Lance, 2000)، آزمون هایی برای قطع آیتم پیشنهاد دادند که علاوه بر بارگذاری های عامل محدود شده در مرحله اول محدود می شدند. سپس، محقق انتخاب آزمون های اضافی برای بهترین تناسب نیازهای خود را کنار می گذارد. آزمون های ممکن می تواند شامل آزمون های میانگین های پنهان، آزمون ها موارد یکنواختی آیتم معادل و آزمونهایی برای واریانس های عامل و کوواریانس ها باشد، که هر کدام از آنها علاوه بر پارامترهای محدود شده در مدل های قبلی محدود می شوند زمانی که نیازهای محقق برآورده می شود.

چارچوب IRT

چارچوب IRT خطی لگاریتمی است، نه خطی، مدلی برای توصیف رابطه بین پاسخ های آیتم مشاهده شده و سطح رفتار پنهان مورد نظر. ماهیت دقیق این مدل توسط مجموعه ای از پارامترهای آیتم تعیین می شود که بالقوه برای هر آیتم واحد هستند. تخمین برای شخصی معین مبتنی بر پاسخ های آیتم مشاهده شده با توجه به پارامترهای آیتم است. دو نوع از پارامترهای آیتم اغلب برای هر قلم تخمین زده می شوند. تمایز یا پارامتر a ارائه دهنده شیب نمودار آیتم (منحنی مشخصه آیتم [ICC] یا تابع پاسخ آیتم نامیده می شود) که رابطه بین رفتار پنهان و امتیاز مشاهده شده را تعیین می کند (شکل 1 را ببینید). نوع دوم از پارامتر آیتم موقعیت آیتم یا پارامتر b است که

موقعیت افقی ICC را مبتنی بر خمیدگی ICC معین تعیین می کند. برای مدل های IRT دوبخشی، پارامتر b نوعاً به عنوان پارامتر آیتم مشکل ارجاع می شود (Lord, 1980. or Embretson & Reise, 2000). را برای مقدمه ای به روش های IRT ببینید).



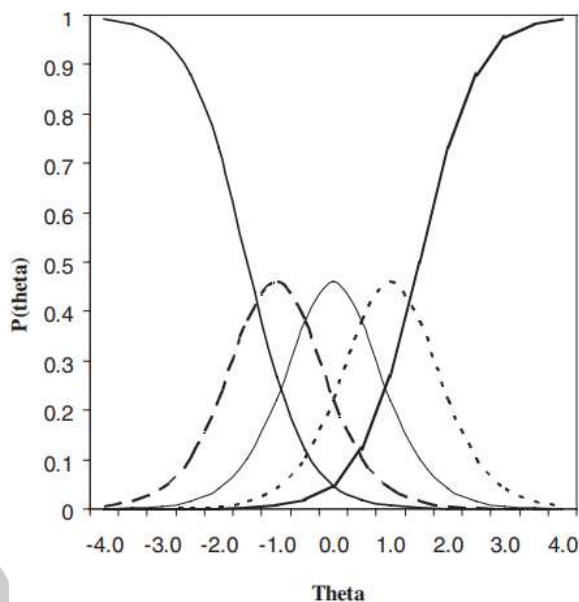
شکل منحنی مشخصه آیتم برای آیتم دوبخشی

در بسیاری موارد، محققان از مقیاس های چند نقطه ای پاسخ برای آیتم هایی غیر از امتیازبندی دوبخشی ساده استفاده می کنند. هرچند چندین مدل IRT برای هر آیتم چندبخشی وجود دارد، یکی از رایج ترین مدل های استفاده شده مدل پاسخ تدریجی است («RM: Samcjuua. 1965a»)، با GRM، رابطه بین احتمال یک شخص با رفتار پنهان تایید هر گزینه پاسخ آیتم خاص می تواند به صورت گرافیکی توسط تابع پاسخ رده آیتم نشان داده شود (CRF). فرمول چنین نموداری می تواند به طور کلی با معادله زیر نشان داده شود

$$P_{ik}(\theta_s) = \frac{e^{a_I(\theta_s - b_{ik+1})} - e^{a_I(\theta_s - b_{ik})}}{(1 + e^{a_I(\theta_s - b_{ik})})(1 + e^{a_I(\theta_s - b_{ik-1})})} \quad (2)$$

که در آن $P(0)$ نشاندهنده احتمال این مورد است که یک آزمایش با سطح معین رفتار به ایتام با رده k پاسخ می دهد. اصطلاحات نمایی جایگزین شده با 1 و 0 برای کمترین و بالاترین گزینه های پاسخ به ترتیب هستند. GRM می تواند یک مورد کلی تر برای مدل های ایتام های دوبخشی ذکر شده در بالا باشد.

نمونه ای از گراف یک آیتم با CRF را می توان در شکل 2 دید، که در آن هر خط متناظر با احتمال پاسخ برای هر گزینه های پنج گانه پاسخ رده برای آیتم است. در گراف، خط سمت چپ دورتر متناظر با احتمال پاسخدهی به کمترین گزینه پاسخ است (یعنی گزینه 1)



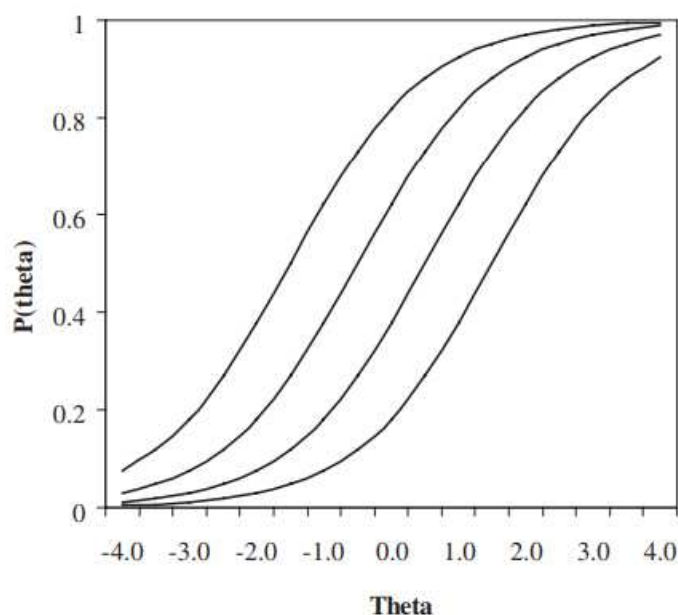
شکل 2. توابع پاسخ رده برای آیتم نوع لیکرت با پنج گزینه پاسخ

زمانی که سطح رفتار پنهان به منفی بینهایت می رسد، احتمال پاسخدهی با گزینه کمترین پاسخ به امی رسد. متناظراً مخالف این مورد برای بالاترین گزینه پاسخ درست است. گزینه های 2 تا 4 در جایی در میانه قرار می گیرند با مقادیر احتمال اول بالا برنده آنها که با افزایش سطوح رفتار پنهان مورد نظر افت می کند.

در GRM، رابطه بین سطح شرکت کننده برای یک رفتار پنهان و یک احتمال شرکت کننده از انتخاب پاسخ مشاهده شده افزایشی نوعاً توسط مجموعه ای از توابع پاسخ مرزی نشان داده می شود (BRFها). نمونه ای از BRFها برای یک آیتم با پنج رده پاسخ در شکل 3 نشان داده شده است. BRFها برای هر آیتم با استفاده از تابع زیر محاسبه می شود

$$P_{ik}^*(\theta_s) = \frac{e^{a_i(\theta_s - b_{ik})}}{1 + e^{a_i(\theta_s - b_{ik})}}, \quad (3)$$

که در آن نشاندهنده احتمال اینست که یک نمونه با سطح احتمال a به آیت i در رده بالا باشد. پارامترهای کمتری برای هر BRF وجود دارند نسبت به رده های پاسخ آیت i و از قبل، پارامترهای a نشاندهنده پارامترهای تمایز آیت است. BRFها شبیه به CC نشان داده شده در شکل 1 هستند. به جز m_i ، BRFها برای هر آیت معین نیاز می شوند. تنها یک پارامترهای a برای هر آیت نیاز می شود زیرا این پارامتر برای معادل بودن با BRFها در یک آیت معین محدود می شود، هر چند پارامترهای a می توانند آیت ها را تحت GRM تغییر دهند.



شکل 3. توابع پاسخ منرزی برای آیت چندبخشی (نوع لیکرت)

پارامترهای به طور مفهومی مشابه هستند. و به طور ریاضی مرتبط با بارگذاری های فاکتور در روش CFA هستند (McDonald, 1999). پارامترهای آیت b هیچ معادلی در CFA ندارد، هر چند آیت CFA که تعریف کننده مقدار مشاهده شده برای آیت است زمانی که ساخت پنهان برابر صفر می شود، اغلب به طور مفهومی مشابه با پارامترهای IRT b هستند. هر چند در حالیکه تنها یک قطع آیت می تواند برای هر آیت در CFA تخمین زده شود، تخمین های GRM مبتنی بر IRT چندین پارامتر b را برای هر آیت b تخمین می زنند.

آزمون های IRT از Me/I وابسته به نوع مشخص روش استفاده شده تغییر می کنند. در این مطالعه ما نسبت احتمال آزمون R را در نظر می گیریم. آزمون های R در سطح آیت رخ می دهد و مانند روش های CFA برآورد

احتمال ماکزیمم برای تخمین پارامترهای آیتم استفاده می شود که منتج به مقدار تناسب مدل شناخته شده به عنوان تابع تناسب می شود. در IRT، این مقدار تابع تناسب شاخصی از این مورد است چگونه مدل معین (در این مورد مدل لجستیک) داده‌ها را به عنوان نتیجه روند تخمین احتمال ماکزیمم استفاده شده برای تخمین پارامترهای آیتم متناسب می کند (Cunulh & Shepard, 1994).

آزمون LR با مقایسه تناسب دو مدل ارتباط دارد. یک مدل فشرده خط اصلی و مدل مقایسه. اول، مدل خط اصلی ارزیابی می شود که در آن تمام پارامترهای آیتم برای تمام ایتم های آزمون با محدودیتی تخمین زده می شوند که پارامترهای آیتم برای چنین آیتم هایی (مثلاً گروه 1 گروه 2) معادل با وضعیت ها هستند (این مدل مقدار خط اصلی پارامتر آیتم برای مدل تناسب می یابد. بعد، هر آیتم تست شده، یکی برای هر بار، برای تابعیت دیفرانسیل آیتم (DIF). برای هر آیتم برای DIF، داده های جداگانه برای هر آیتم اجرا می شوند که در آن تمام آیتم ها مانند تخمین های پارامتری محدود به معادل بودن برای وضعیت ها می شوند (مثلاً دوره های زمانی، گروه های اشخاص). به استثنای پارامترهای آیتم آزمایش شده برای DIE. این مدل فراهم کننده مقدار احتمال مرتبط با تخمین پارامترهای آیتم برای آیتم i به طو رجداگانه برای هر گروه است. این مقدار احتمال می تواند با مقدار احتمال مدل فشرده مقایسه شود که در آن پارامترها برای تمام آیتم های مشابه به یکسان بودن در گروه ها محدود می شوند. فرمول آزمون LR به صورت زیر است:

$$LR_i = \frac{L_C}{L_A}, \quad (4)$$

که در آن LC تابع احتمال مدل فشرده است (که پارامترهای کمتری را محدود می کند) و LZ تابع احتمال مدل ارتقا یافته است که در آن پارامترهای آیتم i مجاز به تغییر در وضعیت های مختلف هستند.

$$\chi^2(M) = -2\ln(LR) = -2\ln L_C + 2\ln L_A, \quad (5)$$

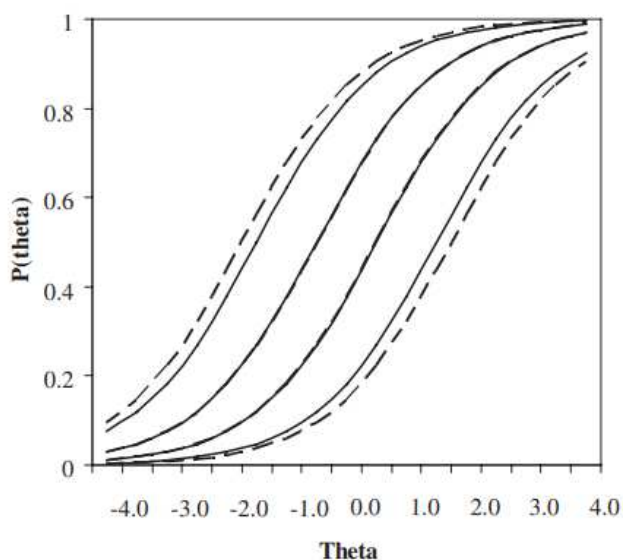
تبدیل لگاریتم طبیعی این تابع می تواند صورت گیرد و نتایج در یک آزمون آماری توزیع شده به صورت تحت فرضیه پوچ برابرست با M معادل با تفاوت رد تعداد پارامترهای آیتم تخمین زده شده در مدل فشرده به ازای مدل

ارتقا یافته (یعنی درجه هایی از آزادی بین مدل ها). این شاخص بدی تناسب است که در آن نتیجه ای چشمگیر تناسب های مدل فشرده را به طور چشمگیری ضعیف تر از مدل ارتقا یافته پیشنهاد می دهد. برای استفاده از آزمون LR, یک مقدار برای هر آیتام در آزمون محاسب می شود. این آیتام ها با مقادیر چشمگیر نشاندهنده Dif هستند (یعنی با استفاده از برآوردهای پارامتری آیتام مختلف تناسب مدل کلی بهبود می یابد).

هرچند روش LR به طور سنتی توسط اجزاهای چندگانه از برنامه MUTILOG انجام می شود (Thissen, 1991), ما در نظر می گیریم که (Thissen, 1991), پیاده سازی روش LR را با کمتر نمودن تخمیل از طریق [RTLDRIF] که یک برنامه است صورت داده است. بنابراین, پیاده سازی این روش ارائه شده در اینجا باید برای محققان سازمانی بیشتر در دسترس باشد که به طور سنتی این مورد وجود دارد. علاوه بر این Cohen. Kim. and Wollaek (1996) قابلیت آزمون LR را برای تشخیص DIF تحت تغییرات وضعیت های مختلف با استفاده از داده های شبیه سازی شده دید. هرچند, داده های شبیه سازی شده دوبخشی بودند نه چندبخشی در ماهیت خود, آنها دریافتند که نرخ های خطای نوع I برای بیشتر مدل های بررسی شده در سطوح مورد انتظار قرار می گیرد. در کل, آنها نتیجه گرفتند که این شاخص به طور معقولی رفتار می کند.

مرور کلی مطالعه

همانطور که قبلاً بیان شد, CFA و IRT روش هایی برای آزمون ME/I به طور مفهومی مشابه هستند اما در عمل متمایز هستند (Raju et al., 2002; Reise et al., 1993). هر کدام اطلاعات واحدی را با توجه به معادل بودن دو نمونه فراهم می کنند (Robic & see Zickar). یکی از تفاوت های اولیه بین این روش ها اینست که GRM در IRT استفاده از تعدادی از پارامترهای آیتام b را برای ایجاد تغییرناپذیری بین اندازه ها ایجاد می کند. هرچند آیتام CFA به طور مشخص, بیشتر مشابه این پارامترهای GRM b هست, آنها هنوز به طور قابل توجهی در تابعیت تنها به عنوان یک قطع در هر آیتام تخمین زده می شوند.



شکل 4. توابع پاسخ مرزی برای آیتم نشاندهنده تابعیت آیتم دیفرانسیل پارامتری **b**

توجه: توابع پاسخ مرزی گروه 2 به صورت خط های تکه تکه هستند.

از آن مهم تر، این تفاوت در پارامترهای آیتم **b** بین گروه ها از نظر تئوری نباید به طور مستقیم توسط روش های CFA استفاده شده به طور رایج مشخص شود. تنها از طریق IRT باشد تفاوت ها در تنها پارامترهای **b** (شکل 4 را ببینید) به طور مستقیم تعیین شود، در حالیکه تفاوت ها در پارامترهای **a** (شکل 5 را ببینید) می تواند از نظر تئوری توسط روش IRT یا GRM مشخص شود به این علت که تفاوت های پارامتر **b** باید توسط آزمون های CFA برای بارگذاری های فاکتور مشخص شود.

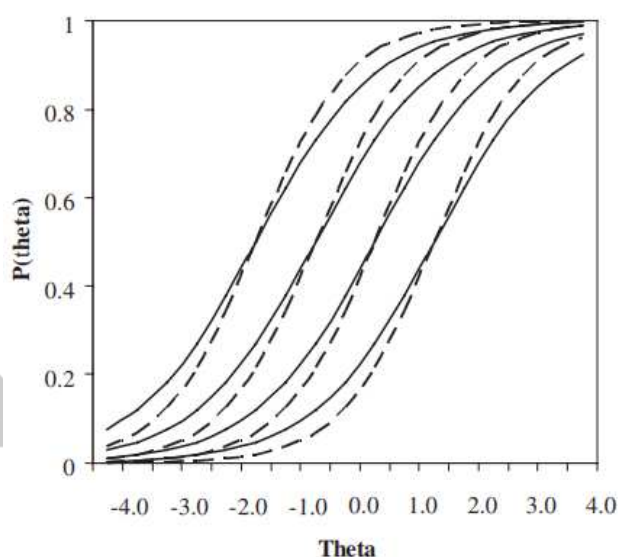
داده ها در این مطالعه در تلاشی برای تعیین این مورد شبیه سازی شدند که آیا اطلاعات اضافی از پارامترهای **b** در دسترس تنها توسط ارزیابی های مبتنی بر IRT برای **ME/I** تشخیص دقیق تر تفاوت های **ME/I** را با روش های CFA فراهم می سازد. دو فرضیه در زیر بیان شده است.

فرضیه 1: داده های شبیه سازی شده دارای تفاوت هایی در تنها پارامترهای **b** به طور مناسب توسط فقدان **ME/I** توسط روش های IRT آشکار می شوند اما **ME/I** با آزمون های مبتنی بر CFA نشان داده می شود.

فرضیه 2: داده ها با پارامترهای متفاوت **a** به طور مناسب توسط فقدان **ME/I** توسط آزمون های IRT و CFA آشکار می شوند.

ویژگی های داده ها

یک بازبینی کوتاه از شش آیتم برای ارائه یک مقیاس تک برای اندازه گیری یک ساختار تک شبیه سازی شد. هر چند تنها یک فاکتور در این مطالعه برای تعیین بازده IRT و CFA تحت ساده ترین شرایط تعیین شدند، این نتایج باشد برای تحلیل های تک بعدی مقیاس های منفرد در بازبینی های چندبعدی عمومی شوند.



شکل 5. توابع پاسخ مرزی برای هر آیتم نشاندهنده تابعیت آیتم دیفرانسیل پارامتر a

توجه: توابع پاسخ مرزی گروه 2 با خط های تکه تکه

سه وضعیت برای تعداد پاسخ دهندگان در مطالعه وجود داشت: 1000 و 150,500. علاوه بر این داده ها برای منعکس نمودن پاسخ های آیتم های از نوع لیکرت با پنج گزینه پاسخ شبیه سازی شد. هر چند نمونه های 150 در حدود نیمی از اندازه نمونه نوعی برای تحلیل های LR بودند (Cohen, 1997 & Kim) و تا حدودی کمتر از مورد پیشنهاد شده برای تحلیل های CFA ME/I (Meade & Lautenschlager, 2004) چندین مطالعه ME/I وجود دارد که از اندازه های نمونه مشابه هنگام انجام تحلیل های CFA ME/I استفاده نموده اند (e.g. Boles, Wang, 2000; Luczak, Raine, & Venables, 2001; Martin & Dean, Ricks, Short & Friedman, 2000; Schaubroeck & Green, 1989; Schmitt, 1982; Vandenberg & Self, 1988).

e.g. (1993; Yoo, 2002). علاوه بر این، اندازه های نمونه این دامنه در تحقیقات سازمانی غیرمعمول نیستند (e.g. Vandenberg & Self, 1993). برای کنترل خطای نمونه برداری، 100 نمونه برای هر وضعیت در مطالعه شبیه سازی شدند.

در یک مقیاس معین، این امکان وجود دارد که هر تعداد از آیتم های منفرد بتوانند فقدان ME/I را نشان دهند. در این مطالعه، دو یا چهار آیتم برای نشان دادن فقدان ME/I در وضعیت ها شبیه سازی شدند (اشاره به آیتم های DIF دارد). پارامترهای آیتم a و چهار پارامتر آیتم b برای شبیه سازی DIF دستکاری شدند. برای دستکاری میزان حضور DIF، پارامترهای آیتم گروه 1 شبیه سازی شده و سپس به روش های مختلف برای شبیه سازی DIF در داده های گروه 2 تغییر یافتند. یک توزیع نرمال تصادفی برای تولید مقادیر پارامتر b برای کمترین BRF برای داده های گروه 1 نمونه برداری شدند. ثوابت 1.2، 2.4 و 3.6 سپس به کمترین آستانه برای تولید پارامترهای آستانه سه BRF دیگر برای تولید داده های نوع لیکرت با پنج گزینه پاسخ رده اضافه شدند. این ثوابت برای فراهم نمودن جداسازی مناسب بین مقادیر آستانه BRF و برای نتیجه دادن گستره محتمل از 2.0 تا 3.3 برای هر آستانه برای یک آیتم معین به طور تقریبی انتخاب شدند.

پارامتر a برای هر آیتم گروه 1 نیز از توزیع تصادفی نرمال نمونه برداری شد. $N[\mu = 1.25, \sigma = 0.07]$ این توزیع برای خلق پارامترهای متمایز سازی آیتم انتخاب شد که دارای گستره محتمل 1 تا 1.5 بودند. تمام داده ها با استفاده از تولید کننده پاسخ آیتم GENIRV تولید شدند (BAKER 1994)

شبیه سازی تفاوت ها در داده ها

DIF توسط کم نمودن 0.25 از پارامتر a آیتم های گروه 1 برای هر آیتم DIF برای خلق پارامترهای a آیتم های DIF گروه 2 شبیه سازی شدند. هرچند پارامترهای b آیتم های DIF به چندین روش تغییر یافتند (در زیر بررسی شده است)، دامنه کلی تغییرات برای هر وضعیت یکسان بود. به طور خاص، برای هر آیتم DIF که در آن پارامترهای b تغییر یافته بودند، یک مقدار 0.40 به پارامترهای b گروه 1 برای خلق پارامترهای b گروه 2 اضافه یا کم شد. این مقدار برای سبب شدن تغییری قابل توجه در برآوردهای نمونه نتیجه شده از پارامترهای آیتم به اندازه کافی

بزرگ است اما نه به بزرگی سبب شدن بالقوه همپوشانی با دیگر پارامترهای آیتم (که 1.2 واحد دورتر هستند). آیتم های 3 و 4 آیتم های DIF برای شرایطی بودند که در آن تنها دو آیتم DIF شبیه سازی شدند در حالیکه آیتم های 3, 4 و 5 و 6 آیتم های DIF برای شرایط چهار آیتم DIF بودند.

سه شرط شبیه سازی شدند که در آن پارامترهای b آیتم تغییر یافتند. در اولی شرط، تنها یک پارامتر b برای هر آیتم DIF متغیر بود. در کل، این توسط اضافه نمودن 0.40 به آیتم های بزرگترین مقدار b انجام شد. این وضعیت نشاندهنده موردی است که در آن اغلب گزینه های بی کران (مانند نرخ بندی نوع لیکرت 5) به طور محتمل کمتر توسط گروه 2 استفاده می شوند تا گروه 1. این نوع از تفاوت می تواند برای مثال در نرخ بندی های عملکرد دیده شود اگر فرهنگ یک آپارتمان در یک سازمان (گروه 2) به ندرت بالاترین عملکرد ممکن را منصوب نماید در حالیکه دیگر آپارتمان (گروه 1) احتمالاً از بالاترین نرخ بندی استفاده می کند. توجه کنید که هرچند چنین تغییری در پارامتر b تک احتمال پاسخ برای یک گزینه پاسخ آیتم را در برخی از مقادیر تغییر می دهد، این وضعیت با هدف شبیه سازی نسبی اختلافات مختلف در داده ها است.

در وضعیت دوم، هر پارامتر b دو بزرگترین برای تفاوت بین گروه ها تنظیم شد. دوباره، این کار با اضافه نمودن 0.40 به بزرگترین دو پارامتر b برای آیتم های DIF گروه 2 انجام شد. این تفاوت شبیه سازی شده دوباره می تواند منعکس کننده نرخ بندی های عملکرد با تغییرات بیشتر برای یک گروه (گروه 1) نسبت به دیگری (گروه 2) باشد.

در وضعیت سوم، هر آیتم DIF برای دو پارامتر کرانی b برای تفاوت بین گروه ها شبیه سازی شد. این کار با اضافه نمودن 0.40 به بزرگترین دو پارامتر b انجام شد در حالیکه همزمان همین مقدار از کمترین پارامتر b آیتم کم می شد. این وضعیت نشاندهنده موردی است که در آن اشخاص در گروه 2 احتمالاً از گزینه های پاسخ بی کران تر (1 و 5) استفاده می کنند نسبت به اشخاص گروه 1. چنین الگوهای پاسخی می تواند زمان مقایسه با نتایج بازدید سازمان برای یک سازمان چند ملیتی همراه با فرهنگ ها دیده شود در حالیکه برخی شواهد وجود دارد که چند تفاوت در تمایلات پاسخ کرانی توسط فرهنگ وجود دارد (Clarke, 2000; watkins & Illui & Triandis, 1991).

(Cheung, 1995).<Downl;.;

References

- Baker, F. (1994). *GENIRV: Computer program for generating item response theory data*. Madison: University of Wisconsin, Laboratory of Experimental Design.
- Bentler, P. M., & Yuan, K.-H. (1999). Structural equation models with small samples: Test statistics. *Multivariate Behavioral Research, 34*, 181-197.
- Boles, J. S., Dean, D. H., Ricks, J. M., Short, J. C., & Wang, G. (2000). The dimensionality of the Maslach Burnout Inventory across small business owners and educators. *Journal of Vocational Behavior, 56*, 12-34.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research, 35*, 169-199.
- Clarke, L. III. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality, 15*, 137-152.
- 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, 20*, 15-26.
- Collins, W. C., Raju, N. S., & Edwards, J. E. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology, 85*, 451-461.
- Craig, S. B., & Kaiser, R. B. (2003, April). *Using item response theory to assess measurement equivalence of 360 performance ratings across organizational levels*. Paper presented at the 18th Annual Meeting of the Society for Industrial/Organizational Psychology, Orlando, FL.
- Donovan, M. A., Drasgow, F., & Probst, T. M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of Applied Psychology, 85*, 305-313.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86*, 215-227.
- Flowers, C. P., Raju, N. S., & Oshima, T. C. (2002, April). *A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Fromen, A., & Raju, N. (2003, April). *Telecommuters and non-telecommuters: Assessing measurement equivalence on an employee survey*. Paper presented at the 18th Annual Meeting of the Society for Industrial/Organizational Psychology, Orlando, FL.
- Ghorpade, J., Hattrup, K., & Lackritz, J. R. (1999). The use of personality measures in cross-cultural research: A test of three personality scales across two countries. *Journal of Applied Psychology, 84*, 670-679.
- Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science, 12*, 133-157.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*, 296-309.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: Users reference guide*. Chicago: Scientific Software International.
- Kim, S., & Cohen, A. S. (1997, March). *An investigation of the likelihood ratio test for detection of differential item functioning under the graded response model*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Luczak, S. E., Raine, A., & Venables, P. H. (2001). Invariance in the MAST across religious groups. *Journal of Studies on Alcohol, 62*, 834-837.
- Martin, L. R., & Friedman, H. S. (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality, 68*, 85-110.
- Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*, 693-702.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- McLaughlin, M. E., & Drasgow, F. (1987). Lord's chi-square test of item bias with estimated and known person parameters. *Applied Psychological Measurement, 21*, 161-173.
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11*, 60-72.
- Meade, A. W., Lautenschlager, G. J., Michels, L. C., & Gentry, W. (2003, April). *The equivalence of online and paper and pencil assessments*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods, 7*, 27-65.



leFa.Com

- Ployhart, R. E., Wiechmann, D., Schmitt, N., Sacco, J. M., & Rogg, K. (2002). The cross-cultural equivalence of job performance ratings. *Human Performance, 16*, 49-79.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517-529.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.
- Riordan, C. M., Richardson, H. A., Schaffer, B. S., & Vandenberg, R. J. (2001). Alpha, beta, and gamma change: A review of past research with recommendations for new directions. In

C. A. Schriesheim & L. L. Neider (Eds.), *Equivalence of measurement*. Greenwich, CT: Information Age Publishing.

- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643-671.
- Rock, D. A., Werts, C. E., & Flaugh, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research, 13*, 403-418.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*, 100-114.
- Schaubroeck, J., & Green, S. G. (1989). Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology, 74*, 892-900.
- Schmitt, M. J., Kihm, J. A., & Robie, C. (2000). Development of a global measure of personality. *Personnel Psychology, 53*(1), 153-193.
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research, 17*, 343-358.
- Schriesheim, C. A., & Neider, L. L. (Eds.). (2001). *Equivalence of measurement*. Greenwich, CT: Information Age Publishing.
- Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.
- Taris, T. W., Bok, I. A., & Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *Journal of Psychology, 132*, 301-316.
- Thissen, D. (1991). MULTIFIT users guide: Multiple categorical item analysis and test scoring using item response theory [Computer software]. Chicago: Scientific Software International.
- Thissen, D. (2001). IRTLRDIF v.2.02b: Software for the computation of the statistics involved in item response theory likelihood-ratio test for differential item functioning [Computer software]. Chapel Hill, NC: I.L. Thurstone Psychometric Laboratory.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum.
- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*, 139-158.
- 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*, 4-69.
- Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomer's changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology, 78*, 557-568.
- Watkins, D., & Cheung, S. (1995). Culture, gender, and response bias: An analysis of responses to the Self-Description Questionnaire. *Journal of Cross-Cultural Psychology, 26*, 490-504.
- Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology and Marketing, 19*, 357-368.
- Zackar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551-563.

