تغییرناپذیری پارامترها در نظریه سؤال پاسخ و تحلیل عاملی تأییدی: توصیف و مقایسه دو رویکرد براساس داده‌های تجربی

مجتبی حبیبی[۱]
فاطمه مردای[۲]
بلال ایزانلو[۳]

تاریخ وصول: ۹۰/۷/۸ تاریخ پذیرش: ۹۰/۱۲/۲٤

چکیده

زمینه: بحث تغییرناپذیری سؤال‌ها و آزمون‌ها در سنجش از اهمیت فراوانی برخوردار است. **هدف:** پژوهش حاضر با هدف مقایسه تغییرناپذیری پارامترها براساس دو رویکرد سؤال پاسخ و تحلیل عاملی تأییدی انجام گرفت. **روش:** ابتدا مبانی هر دو رویکرد پیرامون تغییرناپذیری اندازه‌گیری مورد بحث قرار گرفت و سپس شیوه هر دو رویکرد در تغییرناپذیری پارامترهای سؤال براساس یک نمونه داده تجربه حاصل از آزمون بین‌المللی پیشرفت در خواندن (PIRLS) که پاسخ ۵۰۰۰ دانش‌آموز ایرانی (۲۵۰۰دختر و۲۵۰۰ پسر) در سال ۲۰۰۶ به ۶ سؤال مقیاس نگرش نسبت به خواندن بود مقایسه گردید. **یافته‌ها:** نتایج تحلیل داده‌ها نشان داد که سؤال ۶ براساس هر دو رویکرد دارای سوگیری است، هر چند که نتایج هر دو رویکرد در خصوص سؤال‌های ۱، ۳ و ۴ متفاوت بود. سؤال ۱ براساس رویکرد سؤال پاسخ و سؤال‌های ۳ و۴ براساس رویکرد تحلیل عاملی تأییدی دارای سوگیری تشخیص داده شدند. **بحث و نتیجه‌گیری:** یافته‌های پژوهش حاکی از آن بود که هنگام بررسی‌های مربوط به تغییرناپذیری پارامترها، از

۱- استادیار پژوهشکده خانواده دانشگاه شهید بهشتی (نویسنده مسئول) babakhabibius@gmail.com
۲- کارشناس ارشد روانشناسی عمومی دانشگاه آزاد اسلامی واحد روهن ntmoradi@gmail.com
۳- دانشجوی دکتری تخصصی سنجش آموزش، دانشگاه تهران (نویسنده مسئول) b.ezanloo@gmail.com

روش‌های موجود در هر دو رویکرد سؤال پاسخ و تحلیل عاملی تأییدی استفاده شود و صرف قضاوت براساس نتایج یک روش می‌تواند بسیار گمراه کننده باشد. به علاوه، در نظر گرفتن عرض از مبدأ در مدل اندازه‌گیری به هنگام بررسی تغییرناپذیری و تفاوت توزیع توانایی گروه‌ها و تأثیر آنها روی تغییرناپذیری نیز باید مورد توجه قرار گیرد.

واژگان کلیدی: تحلیل عاملی تأییدی، نظریه سؤال- پاسخ، کارکرد متفاوت سؤال، تغییرناپذیری پارامترها، آزمون بین المللی پیشرفت در خواندن (پرلز).

مقدمه

بحث سوگیری[1] سؤال یا آزمون که در پیشینه نظریه کلاسیک اندازه‌گیری ریشه دارد،در نظریه سؤال پاسخ به کارکرد متفاوت سؤال[2] (امبرستون و رایز[3]، ۲۰۰۰) و در مدل‌یابی معادلات ساختاری به تغییرناپذیری یا یکسان بودن اندازه‌گیری[4] معروف است (برن[5]، ۱۹۹۸). یک اصل اولیه در بررسی تغییرناپذیری اندازه‌گیری در سطح سؤال یا آزمون،قابل مقایسه بودن افرادی است که از گروه‌ها یا جوامع مختلف برای تشخیص سوگیری در سطح سؤال یا آزمون مورد استفاده قرار می‌گیرند. به طور معمول، رایج‌ترین ملاکی که برای تعیین قابل مقایسه بودن افراد مورد استفاده قرار می‌گیرد، نمرات آزمونی است که خصیصه مورد نظر را اندازه‌گیری می‌کند. در صورتی که این نکته مورد توجه قرار نگیرد، آنچه تحت عنوان کارکرد متفاوت سؤال در گروه‌ها به دست می‌آید حاکی از سوگیری نیست، بلکه اندازه‌ای از اثر سؤال[6] است. یعنی، گروه‌های مورد مقایسه در توانایی اندازه‌گیری شده به وسیله سؤال با هم تفاوت واقعی دارند

1. Bias
2. Differential item function(DIF)
3. Embereston & Reise
4. Measurement invariant/equivalence
5. Byrne
6. Item impact

(اکرمن[1]، ۱۹۹۲). به عنوان مثال، اگر نسبت پاسخ درست به یک سؤال را در بین دو گروه مقایسه‌ناپذیر، که از نظر صفت مورد نظر همسان نشده‌اند، مقایسه کنیم نتیجه این مقایسه نشانگر میزانی از اثر سؤال است نه کارکرد متفاوت سؤال. بنابراین، تحت این شرایط علاوه بر ویژگی‌های سؤال، توانایی ناهمخوان افراد گروه‌ها نیز در نتیجه تحلیل تاثیرگذاراست. پس نمی‌توان نتیجه آن را به عنوان کارکرد متفاوت سؤال در نظر گرفت (واینر و براون[2]، ۱۹۸۸).

در صورت عدم رعایت اصل فوق، اگر افرادی که در سازه زیربنایی (در سطح سؤال، نمره کل آزمون یا هر دو) دارای نمرات یکسان هستند، دارای نمره خام مورد انتظار (نمره واقعی) یکسانی نیز باشند، گفته می‌شود که آزمون یا سؤال دربین گروه‌ها یا جوامع دارای تغییرناپذیری اندازه‌گیری است (دراسگو و کانفر[3]، ۱۹۸۵). در صورتی که تغییرناپذیری اندازه‌گیری وجود نداشته باشد تفاوت بین میانگین نمرات مشاهده شده گروه‌ها را نمی‌توان به طور معنی‌داری تفسیر کرد. زیرا این مقدار تفاوت، نه تنها منعکس‌کننده تفاوت‌های واقعی بین گروه‌ها است، بلکه تفاوت در رابطه بین متغیر پنهان و نمره مشاهده شده را نیز که در بین‌گروه‌ها یکسان نیست،آشکار می‌سازد. روشن است که اگر تغییرناپذیری اندازه‌گیری وجود داشته باشد رابطه بین متغیر مکنون و متغیر مشاهده شده در گروه‌ها یا جوامع ثابت خواهد ماند. بنابراین، تحت این شرایط می‌توان تفاوت‌های مشاهده شده بین گروه‌ها را به عنوان تفاوت‌های واقعی در نظر گرفت. به عبارتی دیگر، اگر در بررسی رابطه متغیر مکنونو مشاهده شده عاملی غیر از توانایی زیربنایی موجب تفاوت در عملکرد مشاهده افراد در گروه‌های مختلف شود، می‌توان گفت که تغییرناپذیری وجود ندارد و در نتیجه سؤال یا آزمون دارای سوگیری خواهد بود (راجو، لافیت و برن[4]، ۲۰۰۲).

1. Ackerman
2. Wainer and Braun
3. Drasgow& Kanfer
4. Raju & Laffitte and Byrne

اگرچه از نظر تاریخی روش‌های تشخیص سوگیری سؤال با استفاده از مفاهیم نظریه کلاسیک اندازه‌گیری توسط انگاف[1] شکل گرفت، ولی به تدریج و با توجه به نقطه ضعف‌های نظریه کلاسیک و جایگزین شدن نظریه سؤال پاسخ[2] و مدل‌یابی معادلات ساختاری، روش‌های بررسی تغییر ناپذیری پارامترها نیز تغییر یافت. هدف این مقاله بررسی و مقایسه دو رویکرد نظریه سؤال پاسخ و مدل‌یابی معادلات ساختاری در خصوص بررسی تغییرناپذیری اندازه‌گیری است. در ادامه ابتدا روش‌شناسی هر یک از دو رویکرد در این خصوص توصیف شده و سپس با یک نمونه داده تجربی هر دو روش مقایسه می‌شوند. در چارچوب نظریه سؤال پاسخ تغییرناپذیری پارامترها تحت عنوان کارکرد متفاوت سؤال یا آزمون[3] مورد اشاره قرار می‌گیرد. در این نظریه فرض بر این است که در پاسخ‌گویی افراد به سؤال یک توانایی زیربنایی و مشاهده‌ناپذیرموثر است. افرادی که در این توانایی زیربنایی دارای موقعیت بالاتری باشند (یعنی مقدار بیشتری از توانایی یا صفت را دارا باشند) احتمال بیشتری وجود دارد که در بین گزینه‌های مرتب شده مختلف یک سؤال، گزینه مربوط به طبقه بالاتر را انتخاب کنند. یعنی انتخاب گزینه پاسخ از بین طیف گزینه‌های یک سؤال تحت تأثیر مستقیم موقعیت فرد در پیوستار خصیصه یا صفتی است که سؤال اندازه‌گیری می‌کند (بیکر و کیم[4]، ۲۰۰۴). مدل‌های بسیار زیادی در نظریه سؤال پاسخ شکل گرفته‌اند. در یک تقسیم‌بندی می‌توان این مدل‌ها را به دومقوله‌ای و چندمقوله‌ای[5] تقسیم کرد (ایزانلو و حبیبی، ۱۳۸۶). به طور معمول مدل‌های دومقوله‌ای نظریه سؤال پاسخ که برای تحلیل داده‌های دوارزشی مورد استفاده قرار می‌گیرند از سایر مدل‌های آن متداول‌تر هستند.این مدل‌ها عبارتند از: مدل یک، دو و سه پارامتری. در بین این سه مدل، درعمل، مدل دوپارامتری بیشتر از همه مورد استفاده قرار می‌گیرد (همبلتون،

1. Angof
2. Item response theory
3. Differential functioning of items and tests
4. Baker & Kim
5. Dichotomous and polytomous

سوامیناتان و راجرز[1]، ۱۹۹۱). با توجه به این که بیشتر داده‌هایی که پژوهشگران حوزه علوم اجتماعی و روان‌شناسی با آنها سرو کار دارند از نوع رتبه‌ای است (مقیاس لیکرت) و این موضوع که داده‌های دوارزشی مبتنی بر آزمون‌های روان‌شناختی- مثل آزمون‌های پیشرفت تحصیلی- نیز در واقع نوعی داده رتبه‌ای دو طبقه‌ای محسوب می‌شوند. بنابراین، مدل پاسخ مدرج سمجیما[2] برای توصیف، تحلیل و مقایسه دو رویکرد سؤال پاسخ و مدل‌یابی معادلات ساختاری انتخاب شده است. مدل‌های یک و دو پارامتری که در فوق توصیف شدند در واقع، موارد خاصی از مدل پاسخ مدرج هستند. این مدل قادر است سؤالات طیف لیکرت و دوارزشی را پوشش دهد. در عمل نیز یکی از مدل‌هایی که در مقایسه با سایر مدل‌های چندارزشی بیشتر مورد استفاده قرار می‌گیرد، مدل پاسخ مدرج سمجیما است (تویت[3]، ۲۰۰۳). در این مدل لازم نیست که همه سؤال‌های موجود در یک مقیاس n سؤالی دارای تعداد مقوله‌های پاسخ یکسان باشند، بلکه فقط کافی است مقوله‌های پاسخ هر یک از سؤال‌ها نسبت به هم مرتب باشند. به علاوه فرض می‌شود که همه سؤال‌های موجود در یک مقیاس فقط یک خصیصه را اندازه‌گیری می‌کنند (همبلتون و سوامیناتان، ۱۹۸۵).

با فرض اندازه‌گیری صفت یا توانایی θ در جامعه‌ای از افراد به عنوان هدف اندازه‌گیری، و این که صفت θ مشاهده‌ناپذیر و پیوسته است و مقادیر آن در دامنه $+\infty$ تا $-\infty$ قرار دارد،اگر کوچکترین واحد اندازه‌گیری صفت θ هر یک از n سؤال یک مقیاس باشند، پاسخ افراد به هر یک از n سؤال در دامنه $+\infty$ تا $-\infty$ قرار خواهد گرفت. در صورتی که پاسخ پیوسته افراد به سؤال $i\,(i=1,2,...,n)$ را به m مقوله $(m=m_1,m_2,...,m_j)$ تقسیم کنیم طبقات حاصل دارای نظم ترتیبی خواهند بود. زیرا اگر چه به لحاظ نظری پاسخ افراد پیوسته است ولی در عمل ما این پاسخ پیوسته در هر یک از n سؤال را در غالب m مقوله پاسخ جمع‌آوری می‌کنیم. بنابراین، پاسخ افرادبه هر یک از سؤالات فقط و فقط به یکی از m مقوله اختصاص

1. Hambleton; Swaminathan & Rogers
2. Samejima s graded-response model
3. Toit

خواهد یافت (بیکر و کیم، ۲۰۰۴). برای محاسبه احتمال پاسخ به یکی از m مقوله دو مرحله وجود دارد. در اولین مرحله باید منحنی‌های مشخصه عملگر[1] را محاسبه کرد. برای درک ساده‌تر مطلب در این مرحله، می‌توان گفت که آنچه در مدل پاسخ مدرج روی می‌دهد این است که با m گزینه هر سؤال، به صورت مجموعه‌ای از $m-1$ گزینه دو ارزشی برخورد می‌شود. سپس،بر اساس تابع زیر،برای هر سؤال m مقوله‌ای، $m-1$ پارامتر آستانه β_{ij} و یک ضریب شیب α_i محاسبه می‌شود (امبرستون و رایز، ۲۰۰۰).

$$p_{ij}^{*}(\theta)=\frac{1}{1+e^{-\alpha_i(\theta-\beta_{ij})}} \quad (1)$$

این تابع به شرط توانایی ثابت، احتمال آنکه نمره فرد در سؤال i بیشتر از jامین مقوله پاسخ باشد را محاسبه می‌کند. در نتیجه برای یک سؤال m مقوله‌ای، $m-1$ تابع خواهیم داشت، که هر یک از آنها شبیه تابع مدل دوپارامتری در سؤال‌های دو ارزشی است و همانطور که مدل دوپارامتری احتمال انتخاب پاسخ درست (گزینه بالاتر) به شرط θ را محاسبه می‌کنند، این توابع نیز احتمال آنکه نمره خام فرد در سؤال بالاتر از نمره خام jباشد را تعیین می‌کنند و عبارتند از:

$$\begin{aligned} p_{i1}^{*}(\theta)&=\frac{1}{1+e^{-\alpha_i(\theta-\beta_{i1})}} \\ &\vdots \\ p_{i(m-1)}^{*}(\theta)&=\frac{1}{1+e^{-\alpha_i(\theta-\beta_{i(m-1)})}} \end{aligned} \quad (2)$$

به دنبال محاسبه این توابع، در مرحله دوم می‌توان به کمک آنها توابع مقوله پاسخ[2] را برای m مقوله هر سؤال که احتمال به دست آوردن نمره مقوله j به شرط θ را برآورد می‌کنند،

1. Operating characteristic curves
2. Category response functions

محاسبه نمود. با توجه به اینکه احتمال پاسخ (انتخاب)به پایین‌ترین مقوله یا مقوله‌های بالاتر از آن برابر با یک و احتمال آنکه پاسخ فرد بالاتر از بیشترین مقوله قرار بگیرد برابر با صفر است، می‌توان توابع مقوله پاسخ را به صورت زیر بیان کرد (امبرستون و رایز، ۲۰۰۰؛ راجو، لافیت و برن، ۲۰۰۲):

$$
\begin{aligned}
&p_{i1}(\theta)=1-p^{*}_{i1}\\
&p_{i2}(\theta)=p^{*}_{i1}-p^{*}_{i2}\\
&\vdots \qquad (۳)\\
&p_{i(m_j-1)}(\theta)=p^{*}_{i(m_j-2)}-p^{*}_{i(m_j-1)}\\
&p_{i(m_j)}=p^{*}_{i(m_i-1)}-۰
\end{aligned}
$$

با توجه به آنچه ذکر شد،اکنون می‌توان نمره واقعی یا مورد انتظار $t_i(\theta)$ یک فرد با θ معین در سؤال i را به صورت زیر بیان کرد:

$$(۴)\quad t_i(\theta)=(۱)p_{i۱}+(۲)p_{i۲}+\ldots+(m_j-۱)p_{i(m_j-۱)}+(m_i)p_{i(m_j)}.$$

یعنی برای به دست آوردن نمره واقعی فردی با θ معین در یک سؤال، ابتدا با استفاده از توابع مقوله پاسخ، احتمال پاسخ به هر یک از گزینه‌ها را برای وی محاسبه کرده و سپس این احتمال‌ها را در وزن‌های مربوط به مقوله‌های پاسخ ضرب و در نهایت جمع کردتا نمره واقعی یا مورد انتظار فرد به دست آید. روشن است که نمره واقعی فرد در یک آزمون n سؤالی برابر است با:

$$(۵)\quad T(\theta)=\sum_{i=1}^{n}t_i(\theta)$$

با در نظر گرفتن معادله‌های ۴ و ۵، نمره مشاهده شده فرد درسؤال i برابر با $x_i = t_i(\theta) + e_i$ خواهد بود. در نتیجه نمره مشاهده شده فرد در کل آزمون برابر است با:

$$X = (x_1 + \cdots + x_n) = \sum_{i=1}^{n} t_i(\theta) + \sum_{i=1}^{n} e_i = T(\theta) + e \quad (۶)$$

با این فرض که امید ریاضی خطا برابر با صفر است (وندرلیندن و همبلتون، ۱۹۹۷)،ارزش مورد انتظار نمرات مشاهده شده و واریانس آنها درسؤال i به ترتیب برابر با $E(x_i) = t_i(\theta)$ و $\sigma_{xi}^2 = \sigma_{ti}^2 + \sigma_{ei}^2$ و در کل آزمون به ترتیب برابر با $E(X) = T(\theta)$ و $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ خواهد بود. براساس نظریه سؤال پاسخ در صورت تغییرناپذیری پارامترهای سؤال در بین دو یا چند جامعه، گفته می‌شود که سؤال دارای تغییری‌ناپذیری است(در اصطلاح این نظریه دارای کارکرد متفاوت نیست)، در نتیجه نمره واقعی در سؤال برای دو فرد از دو جامعه با نمره یکسان، برابر خواهد بود. در چارچوب نظریه سؤال پاسخ برای بررسی این موضوع از روش‌های مختلفی استفاده می‌شود که هر یک مزایا و معایب خاص خود را دارند. اکثر این روش‌ها تغییرناپذیری پارامترها را در دو جامعه مرجع[1] و هدف[2] بررسی می کنند. گروه مرجع گروهی است که مبنای مقایسه بوده و پارامترهای گروه یا گروه‌های هدف با آن مقایسه می‌شوند (سایرسی و آلالوف[3]، ۲۰۰۳). برای اینکه نتایج حاصل از تحلیل براساس نظریه سؤال پاسخ با نتایج تحلیل عاملی تأییدی قابل مقایسه باشد،از رویکرد راجو، وندرلیندن و فلیر[4] (۱۹۹۵) استفاده می‌کنیم. این رویکرد که کارکرد متفاوت سؤال‌ها و آزمون‌ها نام گرفته است،بر تغییرناپذیری نمرات واقعی افراد دارای θ یکسان در سطح سؤال و آزمون تأکید دارد. براساس معادله ۴ در صورتی که پارامترهای سؤال i برای s امین فرد در گروه مرجع و هدف در همه مقادیر θ یکسان باشند، در

1. Reference group
2. Focal group
3. Sireci and Allalouf
4. Fleer

این صورت داریم $d_{is}=t_{isF}-t_{isR}=0$ و بر اساس معادله ۵ در سطح آزمون خواهیم داشت $D_s=(T_{sF}-T_{sR})=d_{1s}+\cdots+d_{ns}$. اکنون براساس دو رابطه اخیر شاخص NCDIF[1] و DTF[2] به ترتیب در سطح سؤال و آزمون به صورت زیر تعریف می‌شوند(اوشیما، راجو و ناندا[3]، ۲۰۰۶).

$$(۷)\quad NCDIF=\mathrm{E}(d^2)=(t_{isF}-t_{isR})^2=\mu_{d^2}=\sigma_d^2+\mu_d^2$$

$$(۸)\quad DTF=\mathrm{E}(D^2)=\mu_{D^2}=\sigma_D^2+\mu_D^2$$

علاوه بر بررسی معنی‌داری این شاخص‌ها از طریق آزمون کای‌دو، برای آماره NCDIF نمره برش نیز وجود دارد که شبیه شاخص‌های برازش در معادلات ساختاری است و بر اساس تعداد گزینه‌ها محاسبه می‌گردد. در سؤال‌های دو ارزشی نقطه برش ۰/۰۰۶، در سؤال‌های چهار گزینه‌ای ۰/۰۵۴، در سؤال‌های ۵ گزینه‌ای ۰/۰۹۶ و در سؤال‌های ۶ گزینه‌ای ۰/۱۵ است. برای DTF نیز نقطه برش از طریق مجموع مقادیر برش NCDIF در همه سؤال‌های آزمون محاسبه می‌شود (مورالز و همکاران[4]، ۲۰۰۶).به عنوان مثال، در آزمون حاضر ۶ سؤال ۴ گزینه‌ای وجود دارند که بر اساس آن نقطه برش برای آماره DTF برابر با ۰/۳۲۹ است. هر یک از این روش‌ها دارای معایب و مزایای خاص خود هستند. اگرچه از هر دو روش برای بررسی معنی‌داری آماری استفاده می‌شود و سؤال‌ها یا آزمون‌هایی که مقادیر آماره آنها از نقطه برش بیشترند سودار تشخیص داده می‌شوند، ولی امروزه برای معنی‌داری آماری بیشتر از شیوه‌های مبتنی بر تکرار پارامترهای سؤال[5] به تعداد زیاد (به عنوان مثال ۱۰۰۰ بار) برای تعیین نقطه برش در

1. Noncompensatory differential item function
2.Differential test function
3.Oshima & Nanda
4. Morales et al
5. Item parameter replication

سطوح مختلف استفاده می‌شود که توسط نرم‌افزار صورت می‌گیرد (اوشیما، کوشوبار، اسکوت و راجو[1]، ۲۰۰۹). در پژوهش حاضر نیز از رویکرد جدید برای معناداری آماره‌های فوق استفاده شده است.

تحلیل عاملی تأییدی

مدل‌های رایج نظریه سؤال پاسخ تحت مفروضه تک بعدی بودن عمل می‌کنند و برای اینکه بتوان نتایج تغییرناپذیری پارامترها در دو رویکرد تحلیل عاملی تأییدی و سؤال پاسخ را با هم مقایسه کرد،باید مدل تک عاملی تحلیل عاملی تأییدی را در اینجا بررسی کرد. فرض کنید که آزمون n سؤالی که هر کدام دارای m گزینه است، برای اندازه‌گیری خصیصه ξ مورد استفاده قرار می‌گیرد. در چارچوب مدل‌یابی معادلات ساختاری رابطه بین نمره مشاهده شده و سازه ξدر سؤال i به صورت $x_i = \lambda_i \xi + \delta_i$ ودر کل آزمون به صورت $X = \lambda\xi + \delta$ تعریف می‌شود (راجو، لافیت و برن، ۲۰۰۲). در معادله اول، x_i نمره مشاهده شده در سؤال i، λ_iبار عاملی این سؤال روی سازه ξو δ_i باقی مانده یا خطای مربوط به این سؤال است.در معادله دوم $X = x_1 + \cdots + x_n$ ، $\lambda = \lambda_1 + \cdots + \lambda_n$ و $\delta = \delta_1 + \cdots + \delta_n$ است. بر اساس این روابط و با توجه به اینکه ارزش مورد انتظار خطا برابر با صفر بوده و باξهمبسته نیست و نیز رابطه بین نمره مشاهده شده با متغیر مکنون خطی است، نمره واقعی و واریانس نمرات مشاهده شده در سؤال i به ترتیب برابر با $E(x_i) = \lambda_i \xi$ و $\sigma^2_{xi} = \lambda_i^2 \sigma^2_\xi + \sigma^2_\delta$و در آزمون به ترتیب برابر است با $E(x) = \lambda\xi$ و $\sigma^2_x = \lambda^2 \sigma^2_\xi + \sigma^2_\delta$ خواهد بود.

بنابراین اگر نمره دو فرد در سازه زیربنایی یکسان باشد،در این حالت نمره واقعی آنها در سطح سؤال و به تبع آن در سطح آزمون نیز برابر خواهد بود. به عبارتی،نمره واقعی در سطح آزمون برابر است با مجموع نمرات واقعی در سطح سؤال‌ها. در صورتی که هدف مطالعه‌ای بررسی تغییرناپذیری در بین جوامع یا زیر گروه‌ها باشد به طور معمول به دنبال تعیین تعداد

1. Oshima, Kushubar, Scott& Raju

عامل‌ها و الگوی روابط بین آنها، چندین آزمون برای این منظور اجرا می‌شود که عبارتند از: آزمون یکسان بودن ماتریس Λ_x، آزمون یکسان بودن ماتریس Φ و آزمون یکسان بودن ماتریس Θ_δ. بین این آزمون‌ها یک رابطه سلسله مراتبی وجود دارد. به عنوان مثال آزمون یکسان بودن ماتریس Θ_δ تنها در صورتی اجرا می‌شود که یکسانی ماتریس Λ_x به دست آید و یا آزمون یکسانی ماتریس Φ زمانی اجرا می‌شود که ماتریس Λ_x در بین جوامع قابل مقایسه باشد. مطلبی که بسیاری از پژوهشگران در مورد آن توافق دارند آن است که برای بررسی تغییرناپذیری اندازه‌گیری تنها یکسانی ماتریس Λ_x لازم است (برن، ۱۹۹۸؛ رایز، ویدامن و پوگ[1]، ۱۹۹۳).در کل برای بررسی تغییرناپذیری یک مدل سلسله مراتبی ۸ مرحله‌ای که ۵ مرحله اول به تغییرناپذیری اندازه‌گیری و ۳ مرحله بعدی به تغییرناپذیری ساختاری مربوط است، وجود دارد که بسته به هدف پژوهش تنها چند مرحله از آن باید مورد بررسی قرار گیرند (واندبرگ و لانس[2]، ۲۰۰۰).

در چارچوب مدل‌یابی معادلات ساختاری بررسی تغییرناپذیری فقط بر تحلیل ساختار کوواریانس متمرکز شده و به همین دلیل نیز به هنگام بررسی تغییرناپذیری، باید از ماتریس کوواریانس به جای ماتریس همبستگی استفاده گردد. ولی برخی از پژوهشگران معتقدند عرض از مبدا یا همان میانگین مشاهده شده گروه‌ها در متغیرها نیز باید از نظر تغییرناپذیری مورد بررسی قرار گیرد (راجو، لافیت و برن، ۲۰۰۲)، زیرا یکسان بودن ضرایب عاملی به هیچ وجه برابری مقادیر عرض از مبدأ را تضمین نمی‌کند. بنابراین،اگر بارهای عاملی سؤال‌ها روی متغیر زیر بنایی در دو گروه یکسان باشد،امکان اینکه میانگین سؤال یا عامل در بین گروه‌ها متفاوت باشد، وجود دارد. با این حال، بررسی تفاوت بین میانگین گروه‌ها در متغیرهای مشاهده شده یا مکنون در پاسخ به سؤال‌هایی که به بررسی تغییرناپذیری مربوط نیستند مطرح می‌شود. تفاوت میانگین گروه‌ها در عامل مکنون در چارچوب کارکرد متفاوت سؤال تحت عنوان اثر مورد

1. Reise & Widaman and Pugh
2. Vandenberg & Lance

اشاره قرار می‌گیرد. در اثر، تفاوت عملکرد گروه‌ها در سؤال، حاکی از تفاوت در توزیع توانایی کلی است ولی کارکرد متفاوت سؤال یک تفاوت غیرمنتظره در بین افرادی است که از نظر ویژگی که سؤال یا آزمون اندازه‌گیری می‌کند با هم یکسان شده‌اند. بنابراین به نظر می‌رسد که در کارکرد متفاوت سؤال بیشتر رابطه بین میزان خصیصه و نمره واقعی متغیر موردنظر باشد تا تفاوت بین میانگین‌ها.علاوه بر این، برای اینکه نتایج حاصل از نظریه سؤال پاسخ و تحلیل عاملی تأییدی قابل مقایسه باشند و نیز برای جلوگیری از طولانی شدن بیش از حد این مقاله از بررسی مقادیر عرض از مبدا در تحلیل عاملی تأییدی و مقایسه آنها با مقادیر دشواری سؤال‌ها در نظریه سؤال پاسخ خودداری شده است.امروزه بر سر این موضوع که بدون تحقق یافتن تغییرناپذیری[1] انجام مقایسه‌های بین گروهی از هر نوع که باشد بی معنی و محل اشکال است، توافق وجود دارد (اشمیت و کولجانین[2]،۲۰۰۸).

روش

برای مقایسه دو رویکرد سؤال پاسخ و تحلیل عاملی تأییدی در تغییرناپذیری،از داده‌های مطالعه بین المللی پیشرفت در سواد خواندن[3] (PIRLS) سال ۲۰۰۶ استفاده گردید. این مطالعه از وسیع‌ترین مطالعات تطبیقی بین المللی درزمینه پیشرفت تحصیلی است که زیر نظر انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی[4] (IEA)اجرا گردید (مولیس و همکاران[5]، ۲۰۰۷). برای این منظور از داده‌های مربوط به دانش آموزان ایرانی در خرده مقیاس نگرش دانش آموز نسبت به خواندن که در پرسشنامه دانش آموز قرار دارد و دارای ۶ سؤال ۴ گزینه‌ای است استفاده شد. تعداد کل افراد شرکت کننده۵۴۱۱ نفر بودند که از این تعداد ۲۵۰۰ نفر دختر (گروه هدف) و ۲۹۱۱ نفر نیز پسر (گروه مرجع) بودند. با توجه به حجم کمتر دختران، از گروه پسران نیز به

1. invariance
2. Schmitt & Kuljanin,
3. Progress in International Reading Literacy Study
4. International Association for the Evaluation of Educational Achievement
5. Mullis and et al

تصادف ۲۵۰۰ انتخاب شدند. در این مطالعه تغییرناپذیری پارامترهای سؤال‌های این خرده مقیاس ۶ سؤالی در بین دو گروه پسر (گروه مرجع) و دختر (گروه هدف) براساس تحلیل عاملی تأییدی و نظریه سؤال پاسخ مورد بررسی قرار گرفته است.

یافته‌ها

تحلیل عاملی تأییدی

ابتدا مدل برای هر یک از دو گروه به صورت مجزا اجرا گردید که شاخص‌های برازش مربوطه برای هر دو گروه و بارهای عاملی هر یک از آنها در جدول ۱ گزارش شده‌اند. در همه موارد روش برآورد ML مورد استفاده قرار گرفت.

جدول ۱. بارهای عاملی و شاخص‌های برازش به تفکیک جنسیت

پسران			دختران		
سؤال	بار عاملی*	شاخص‌های برازش	سؤال	بار عاملی	شاخص‌های برازش
۱	۰/۲۸(۰/۰۲)	کای دو ۴۳/۶۳	۱	۰/۲۷(۰۲)	کای دو ۸۶/۳۶
۲	۱	درجه آزادی ۶	۲	۱	درجه آزادی ۶
۳	۲/۰۷(۰/۰۹)	معنی‌داری ۰/۰۰۰۰۱	۳	۲/۷۹(۰/۱)	معنی‌داری ۰/۰۰۰۰۱
۴	۰/۶(۰/۰۳)	CFI ۰/۹۹	۴	۰/۴۵(۰/۰۲)	CFI ۰/۹۹
۵	۲/۲۱(۰/۰۹)	GFI ۰/۹۹	۵	۲/۱۴(۰/۰۸)	GFI ۰/۹۹
۶	۲/۵۳(۰/۱)	RMSEA ۰/۰۵	۶	۳/۲۸(۰/۱۱)	RMSEA ۰/۰۷۳

*تمام بارهای عاملی، به جز بار عاملی سؤال ۲ که برای تعیین مقیاس روی مقدار عددی۱ تثبیت شد، در سطح ۰/۰۱ معنی‌دار هستند و مقادیر داخل پرانتز خطای معیار اندازه‌گیری هستند.

ابتدا هر دو گروه به طور همزمان با یک اجرا تحلیل شدند که نتایج آن در جدول ۲ در ردیف مدل مبنا آمده است. سپس همه بارهای عاملی گروه دوم (دختران)برابر با بارهای عاملی

گروه اول (پسران) در نظر گرفته شدند. نتایج حاصل در ردیف مدل محدود[1] جدول ۲ نشان می‌دهد که تفاوت بین کای دوی این مدل و مدل بدون محدودیت قبلی برابر است با $\Delta\chi^2 = ۱۵۵/۴۲, \Delta df = ۵$ که از کای دو بحرانی با همین درجه آزادی در آلفای ۰/۰۱ (۱۵/۰۹) بزرگتر است. بنابراین، می‌توان گفت که حداقل دریک سؤال بین بارهای عاملی گروه پسران و دختران تفاوت معنی‌داری وجود دارد، که حاکی از تغییرپذیری پارامترها در بین دو گروه است. سپس باید مشخص کرد که پارامترهای کدام سؤال‌ها در بین دو گروه تفاوت دارد. ابتدا بار عاملی سؤال اول در گروه دختران را برابربا بار عاملی همین سؤال در پسران قرار می‌دهیم و بار عاملی بقیه سؤال‌ها را آزاد می‌گذاریم.تفاوت کای‌دو ردیف سوم جدول ۲ از ردیف اول این جدول، $\Delta\chi^2 = ۰/۰۹, \Delta df = ۱$، از مقدار کای‌دو بحرانی با همین درجه آزادی در آلفای ۰/۰۱ (۶/۶۴) کوچکتر است، بنابراین بین کارکرد این سؤال در دو گروه تفاوت معنی‌داری وجود ندارد. روشن است که اگر نتایج مربوط به محدود کردن پارامترهای یک سؤال حاکی از تغییرناپذیری بار عاملی آن در دو گروه باشد پس آن محدودیت تا آخر حفظ می‌شود، ولی اگر تغییرناپذیری برای یک سؤال وجود نداشته باشد محدودیت اعمال شده بر آن حذف می‌شود تا پارامترهای آن به طور آزاد در دو گروه برآورد شوند. فرایند مشابهی برای بقیه سؤال‌ها نیز تکرار شد که نتایج آن در جدول ۲ قابل مشاهده است. معنی‌داری آماری دلتا کای‌دوی مربوط به سؤال‌های ۳، ۴ و ۶نشان می‌دهد که بار عاملی این سؤال‌ها در در دو گروه برابر نیست.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. restricted model

جدول ۲. نتایج آزمون کای دو در هر یک از مدل‌ها همراه با سایر شاخص‌های برازش

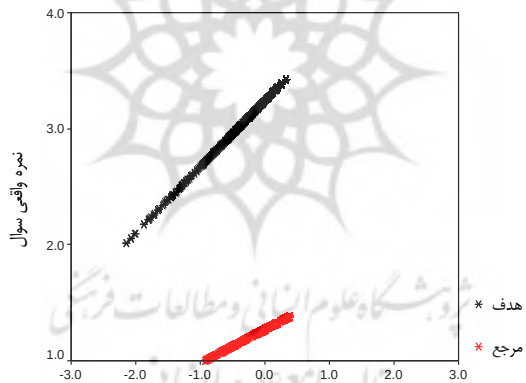
مدل	کای دو	درجه آزادی	معنی‌داری	$\Delta\chi^2$	Δdf	ECVI	CFI	GFI	RMSEA
مدل مبنا	۱۲۹/۹۹	۱۲	۰/۰۰۰۰۰۱	–	–	۰/۰۳۸	۰/۹۹	۰/۹۹	۰/۰۶۳
مدل محدود	۲۸۵/۴۱	۱۷	۰/۰۰۰۰۰۱	۱۵۵/۴۲*	۵	۰/۰۶۷	۰/۹۸	۰/۹۸	۰/۰۷۹
مدل محدود سؤال ۱	۱۳۰/۰۸	۱۳	۰/۰۰۰۰۰۱	۰/۰۹	۱	۰/۰۳۸	۰/۹۹	۰/۹۹	۰/۰۶
مدل محدود سؤال ۱ و ۲	۱۳۰/۰۸	۱۳	۰/۰۰۰۰۰۱	۰/۰۹	۱	۰/۰۳۸	۰/۹۹	۰/۹۹	۰/۰۶
مدل محدود سؤال ۱تا ۳	۱۶۱/۷۹	۱۴	۰/۰۰۰۰۰۱	۳۱/۰۸*	۲	۰/۰۴۴	۰/۹۹	۰/۹۹	۰/۰۶۵
مدل محدود سؤال ۱-۲ و۴	۱۵۰/۹۰	۱۴	۰/۰۰۰۰۰۱	۲۰/۹۱*	۲	۰/۰۴۱	۰/۹۹	۰/۹۹	۰/۰۶۳
مدل محدود سؤال۱-۲ و ۵	۱۳۰/۲۲	۱۴	۰/۰۰۰۰۰۱	۰/۲۳	۲	۰/۰۳۷	۰/۹۹	۰/۹۹	۰/۰۵۸
مدل محدود سؤال۱،۲، ۵، ۶	۱۸۳/۲۰	۱۵	۰/۰۰۰۰۰۱	۵۳/۲۱*	۳	۰/۰۴۷	۰/۹۹	۰/۹۹	۰/۰۶۷

* نشان دهنده معنی‌دار بودن کای دو

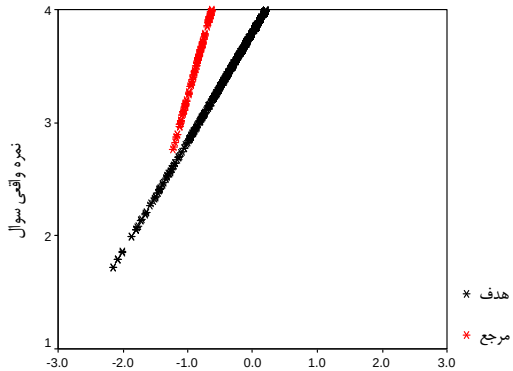
با توجه به نتایج به دست آمده از تحلیل،در سؤال‌های ۳ (اگر کسی کتابی به من هدیه دهد خوشحال خواهم شد[1])، ۴ (به نظر من مطالعه کردن کسل کننده است[2]) و ۶ (از خواندن لذت می‌برم[3]) تغییرپذیری وجود دارد. براساس نمودارهای ۲تا ۴ (که رابطه بین میزان خصیصه و نمره واقعی سؤال را نشان می‌دهند) می‌توان گفت که در سؤال ۳ پسران در کل پیوستار خصیصه در مقایسه با دختران در همان سطح خصیصه،گزینه‌های با وزن بالاتررا انتخاب می‌کنند. در سؤال ۴

1. I would be happy if someone gave me a book as a present
2. I think reading is boring
3. I enjoy reading

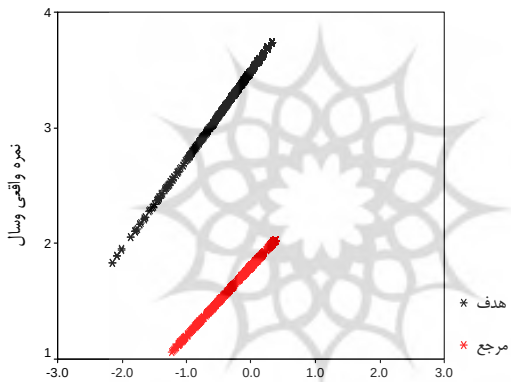
وضعیت بر عکس شده است، به طوری که دختران در مقایسه با پسران هم سطح خود، گزینه‌های با وزن بالاتر را انتخاب می‌کنند. البته بررسی نمودار سؤال ۴ نشان می‌دهد که این تفاوت بیشتر یک تفاوت در میانگین نمره واقعی است تا در رابطه نمره واقعی با میزان خصیصه، در تحلیل سؤال‌ها براساس نظریه سؤال پاسخ این سؤال جزو سؤال‌های سودار تشخیص داده نشد. شاید سؤال ۱ نیز (نمودار ۱) که بر اساس تحلیل عاملی تأییدی فاقد سوگیری ولی بر اساس نظریه سؤال پاسخ دارای سوگیری است چنین وضعیتی داشته باشد. در بخش نتیجه‌گیری بیشتر به این موضوع پرداخته می‌شود. در نهایت، سؤال ۶ که دارای سوگیری غیر یکنواخت است نشان می‌دهد که تا نمره ۱/۵- گروه هدف در مقایسه با گروه مرجع هم سطح خود تمایل بیشتری به انتخاب گزینه‌های بالاتر دارند در حالی که از نمره ۱/۵- بعد وضعیت بر عکس شده است.



نمودار ۱: رابطه میزان خصیصه و نمره واقعی در سوال ۱

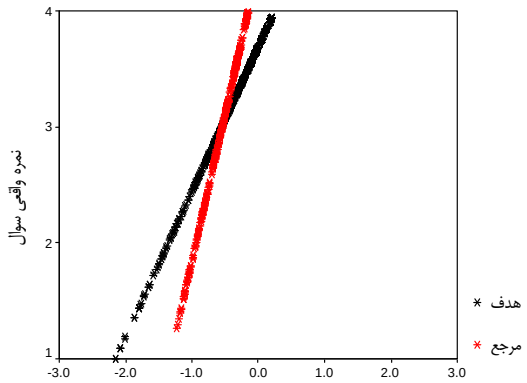


نمودار ۲: رابطه بین میزان خصیصه و نمره واقعی در سوال ۳



نمودار ۳: رابطه بین میزان خصیصه و نمره واقعی سوال ۴

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



نمودار ۴: رابطه بین میزان خصیصه و نمره واقعی سوال ۶

نظریه سؤال پاسخ

برای تحلیل داده‌ها براساس نظریه سؤال پاسخ از مدل پاسخ مدرج استفاد شده است. پارامترها به کمک نرم افزار مولتی‌لوگ[1] (تیسن[2]، ۱۹۹۱) برآورد شده‌اند. بعد از برآورد پارامترها و قبل از محاسبه آماره‌های سوگیری، پارامترها باید همتراز[3] شوند. برای این کار از نرم افزار ایکویت[4] (بیکر، ۱۹۹۵) استفاده شد و با استفاده از این نرم افزار پارامترهای گروه پسران (گروه مرجع) در مقیاس پارامترهای گروه دختران (گروه هدف) قرار گرفت. سپس برای محاسبه ماتریس واریانس-کوواریانس بین پارامترها از نرم‌افزار پلی‌کوو[5] (موریس[6]، ۲۰۰۹) استفاده شد. در نهایت آماره‌های NCDIF و DTF توسط نرم افزار دی‌فیت۸[7] (اوشیما، کوشوبار، اسکوت و راجو، ۲۰۰۹) محاسبه شدند که همراه با پارامترها در جدول ۳ ارائه شده است.

1. MULTILOG
2. Thissen
3. Equating
4. EQUATE
5. PolyCov
6. Morris
7. DFIT8

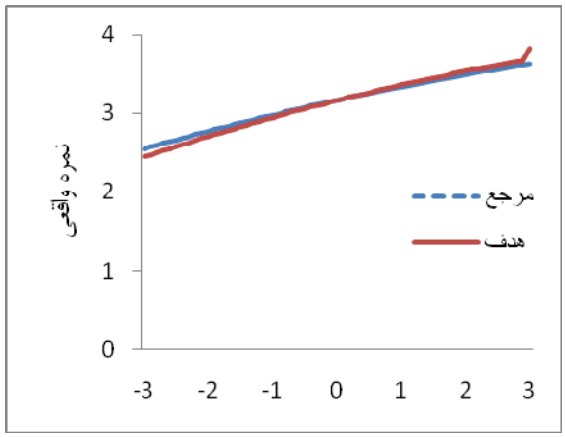
جدول ۳. پارامترهای دو گروه در هر یک از سؤالات

سؤال	گروه	شیب	آستانه۱	آستانه۲	آستانه۳	NCDIF		DTF
۱	پسر	۰/۳۶	-۵/۹۶	-۰/۳۵	۲/۲۲	۰/۰۱۹۵*	مقدار	۰/۰۰۲۹۳
	دختر	۰/۴۸	-۶/۲۴	-۰/۹۷	۱/۱۴		کای دو	۲۰۲۰/۹۴
۲	پسر	۰/۳۰	-۸/۲۵	-۴/۷۸	۰/۳۱	۰/۰۰۰۱	درجه آزادی	۱۷۴۵
	دختر	۰/۳۲	-۸/۵۵	-۵/۱۲	-۰/۰۷			
۳	پسر	۰/۵۱	-۶/۷۳	-۵/۲۷	-۳/۳۸	۰/۰۰۰۰۰۱	معنی‌داری	۰/۰۰۰۰۱
	دختر	۰/۵۵	-۶/۳۵	-۵/۳۱	-۳/۴۷			
۴	پسر	۰/۴۹	-۶/۴۷	-۲/۰۷	-۰/۰۹	۰/۰۰۰۸۹		
	دختر	۰/۵۹	-۶/۴۶	-۲/۲۶	-/۱۱			
۵	پسر	۰/۵۵	-۶/۶۸	-۵/۲۹	-۳/۳۱	۰/۰۰۰۰۹		
	دختر	۰/۴۹	-۷/۰۷	-۵/۸۲	-۳/۷۳			
۶	پسر	۰/۴۳	-۶/۱۷	-۴/۰۳	۱/۱	۰/۰۰۶۶*		
	دختر	۰/۵۱	-۵/۰۶	-۳/۷۵	۱/۱			

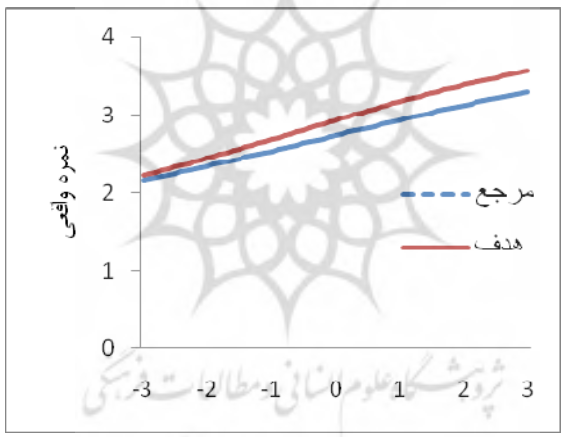
*نشان دهنده معنی‌دار بودن

مقدار آماره DTF برابر با ۰/۰۰۲۹به لحاظ آماری معنی‌دار است($P<۰/۰۱$). مقادیر آماره NCDIT فقط در سؤال ۱ (فقط زمانی که مجبور باشم مطالعه می‌کنم[1]) و سؤال ۶ از نظر آماری معنی‌دار هستند. بنابراین، می‌توان گفت که بین نمره واقعی دختران و پسران در این دو سؤال تفاوت معنی‌داری وجود دارد.نمودارهای ۱ و ۲ رابطه نمره افراد در خصیصه مورد نظر و نمره واقعی آنها در دو سؤال ۱ و ۶ را نشان می‌دهند. در سؤال ۱ دختران در مقایسه با پسران در کل پیوستار خصیصه گزینه‌های بالاتر سؤال را انتخاب کرده‌اند. در سؤال ۶ همانند تحلیل عاملی تأییدی حالت سوگیری غیر یکنواخت حفظ شده است به طوری که در سطوح پایین خصیصه، گروه مرجع در مقایسه با گروه هدف گزینه‌های بالاتر را انتخاب می‌کنند در حالی که در سطوح بالای خصیصه وضعیت برعکس شده، هر چند که این وضعیت بسیار ضعیف است.

1. I read only if I have to



نمودار ۵. رابطه میزان خصیصه و نمره واقعی در سؤال اول



نمودار ۶. رابطه میزان خصیصه و نمره واقعی در سؤال شش

بحث و نتیجه‌گیری

نتایج پژوهش حاضر نشان داد که براساس هر دو روش سؤال ۶ دارای سوگیری غیر یکنواخت است. سؤال ۱ فقط توسط نظریه سؤال پاسخ به عنوان سؤال دارای سوگیری تشخیص داده شد، این وضعیت در سایر مطالعات نیز مشاهده شده است (به عنوان مثال، راجو، لافیت و برن،

۲۰۰۲). به علاوه، بر اساس نتایج مطالعه فلورز، راجو و اوشیما[1] (۲۰۰۲) که در آن داده‌هایی براساس مدل پاسخ مدرج تولید شدند تا کارآمدی دو رویکرد تحلیل عاملی تأییدی و سؤال پاسخ در بررسی تغییرناپذیری مقایسه شود مشخص شد که روش معادلات ساختاری که برای بررسی تغییرناپذیری فقط بار عاملی سؤالات را مورد توجه قرار می‌دهد، قادر است سؤال‌هایی که بر اساس نظریه سؤال پاسخ دارای شیب متفاوتند را به خوبی مشخص کند، ولی قادر نیست سؤال‌های دارای پارامترهای b متفاوت را به درستی تفکیک نماید (این وضعیت در مورد سؤال ۱ صادق است و قبل از این نیز به آن اشاره کردیم). به علاوه مقادیر آستانه سؤال برای دو گروه پسر و دختر در جدول ۳ نشان می‌دهد که پارامترهای آستانه دختران و پسران تا حدی متفاوت است و به احتمال زیاد به همین دلیل این سؤال توسط روش تحلیل عاملی که در آن فقط شیب سؤال‌ها مقایسه شد به عنوان سؤال دارای تغییرپذیری مشخص نشد.

علاوه بر مطلب فوق نتایج مطالعه فلورز، راجو و اوشیما (۲۰۰۲) نشان داد که روش مبتنی بر شیب و عرض از مبدأ[2] در تحلیل عاملی تأییدی و روش NCDIF قادر است سؤال‌هایی با پارامترهای b متفاوت را نیز مشخص کنند.ولی هیچ یک از دو روش به سؤال‌هایی که فقط شیبب متفاوتی دارند زیاد حساس نیستند.

به علاوه، زمانی که گروه‌های مرجع و هدف توزیع توانایی متفاوتی (اثر) داشته باشند، روش مبتنی بر شیب و عرض از مبدادر تحلیل عاملی تأییدیدارای میزان خطای نوع دوم کمتری است. یعنی،در این روش نرخ قبول فرض صفر نادرست در شرایطی که تغییرپذیری وجود دارد، پایین یا قابل قبول است. در مقابل میزان خطای نوع اول آن بسیار بالا است. یعنی امکان این که تغییرناپذیری برقرار باشد ولی نتیجه این روش به رد تغییرناپذیری منجر شود بسیار بالا است. این در حالی است که روش NCDIF در مواردی که توزیع توانایی گروه‌ها متفاوت باشد یا نباشد دارای نرخ خطای نوع اول و دوم قابل قبولی است. شاید عدم وجود تغییرناپذیری در سؤال‌های

1. Flowers, Raju & Oshima
2. Intercept

۳ و ۴ را بر اساس رویکرد تحلیل عاملی تأییدی را بتوان براساس این مطلب توجیح کرد.با توجه به این مطلب تشخیص کارکرد متفاوت سؤال از یک سو با وضعیت پارامترهای b (همان میانگین سؤال‌ها) در گروه‌های مختلف وابسته است و از سوی دیگر به شدت به توزیع توانایی در گروه‌های مختلف بستگی دارد.بنابراین، در مطالعات بعدی باید این وضعیت و شرایط بیشتر مورد بررسی قرار گیرند تا درستی آنها تأیید گردد و وضعیت روشن‌تر شود.

در این مقاله تلاش شد تا مبانی دو رویکرد جدید اندازه‌گیری در بررسی تغییرناپذیری توصیف شود و چگونگی کار با آنها در عمل توصیف گردد. با اینکه از زمان شکل‌گیری نظریه‌های اندازه‌گیری روش‌های زیادی برای بررسی تغییرناپذیری پارامترهای سؤال‌ها در آزمون‌های مختلف شکل گرفته است، ولی اخیرا آنچه بیشتر از همه مورد استفاده قرار می‌گیرد روش‌های مبتنی بر مدل‌یابی معادلات ساختاری و نظریه سؤال پاسخ است. اگرچه بین این دو رویکرد شباهت‌های زیادی وجود دارد ولی تفاوت‌های عمده‌ای نیز باعث تمایز بین آنها شده است. به عنوان مثال روش‌های مبتنی بر مدل‌یابی معادلات ساختاری بر روی توزیع حاشیه‌ای خصیصه مکنون مورد نظر تمرکز دارند ولی روش‌های مبتنی بر نظریه سؤال پاسخ روی توزیع شرطی خصیصه مکنون عمل می‌کنند. به علاوه، در حالی که رویکرد اول روابط را خطی در نظر می‌گیرد، دومی سعی دارد در تبیین رابطه بین متغیرهای مشاهده شده و مکنون،به صورت غیرخطی عمل کند. در کل می‌توان گفت که تعیین سودار بودن یک سؤال باید براساس نتایج چندین روش صورت گیرد وصرف قضاوت براساس نتایج تحلیل عاملی تأییدی یا نظریه سؤال پاسخ می‌تواند بسیار گمراه کننده باشد. از طرفی دیگر، نمی‌توان گفت که یکی از این رویکردها بر دیگری برتری دارد و آنچه احتمال می‌رود درست باشد آن است که رویکردهای غیرخطی که هم در بطن مدل‌یابی معادلات ساختاری و هم نظریه سؤال پاسخ در حال توسعه هستند می‌توانند چشم انداز روشنی در خصوص بررسی تغییرناپذیری پارامترها فراهم کنند.

به عنوان آخرین نکته، با توجه به نتایج حاصل از این پژوهش و مطالعات موجود در پیشینه می‌توان گفت نتایج ضد و نقیض بین این دو رویکرد غالب تا حد زیادی از ناکارآمدی

مدل‌های موجود در آنها برای تبیین رابطه بین توانایی و پاسخ به سؤال نیز ناشی می‌شود. البته این مطلب تنها به بحث حاضر مربوط نیست. به نظر می‌رسد یکی از دلایل ناکارآمدی مدل‌های اندازه‌گیری این است که فقط به محرک و پاسخ آزمودنی به آن توجه دارند در حالی فرایند ما بین آنها نادیده گرفته می‌شود. این فرایند همان در نظر گرفتن ساختار شناختی افراد و چگونگی کارکرد آن در موقعیت‌های مختلف است که حداقل در حال حاضر در مدل‌های رایج موجود در دو نظریه جایگاهی ندارد.

منابع فارسی

ایزانلو، بلال؛ حبیبی، مجتبی. (۱۳۸٦). مقدمه‌ای برمبانی رویکردهای جدید اندازه‌گیری در حوزه روان‌شناسی وعلوم تربیتی. *فصلنامه روان‌شناسی و علوم تربیتی*. شماره ۸: ۱٦۵-۱۳۵.

منابع لاتین

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 .

Baker, F. (1995). EQUATE 2.1: Computer program for equating two metrics initem response theory [Computer program]. *Madison: University of Wisconsin, Laboratory of Experimental Design* .

Baker, F. B. (2004). *Item response theory: Parameter estimation techniques* (Vol. 176): CRC.

Braun, H. I. (1988). *Test validity*: Lawrence Erlbaum.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*: Lawrence Erlbaum.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations .*Journal of Applied Psychology, 70*(4), 662 .

Du Toit, M. (2003). *IRT from SSI: Bilog-MG, multilog, parscale, testfact*: Scientific Software.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists* (Vol. 4): Lawrence Erlbaum.

Flowers ,C. P., Raju, N. S., & Oshima, T. (2002). A Comparison of Measurement Equivalence Methods Based on Confirmatory Factor Analysis and Item Response Theory .

Hambleton, R. K. (1997). *Handbook of modern item response theory*: Springer Verlag.

Hambleton, R. K ,.Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2): Sage Publications, Inc.

Linden, W., & Hambleton, R. K. (1997). Handbook of modern item response theory. *New York* .

Morales, L. S., Flowers, C., Gutierrez, P., Kleinman, M., & Teresi, J. A. (2006). Item and scale differential functioning of the Mini-Mental State Exam assessed using the differential item and test functioning (DFIT) framework. *Medical care, 44*(11 Suppl 3), S143 .

Morris, B. S. (2009). Polycov [Computer program]. Chicago: Illinois Institute of Technology.

Mullis, I. V. S., Martin, M. O., Kennedy ,A. M., & Foy, P. (2007). PIRLS 2006 international report. *Boston: IEA (http://pirls. bc. edu/isc/publications. html# p06, 30.6. 2008).*

Oshima, T., Kushubar, S., Scott, J., & Raju, N. (2009). DFIT8 for Window User's Manual: Differential functioning of items and tests. *St. Paul MN: Assessment Systems Corporation* .

Oshima, T., Raju, N. S., & Nanda, A. O. (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of educational measurement, 43*(1), 1-17 .

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*(3), 517 .

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*(3), 552 .

Roju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*(4), 353-368 .

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222.

Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language testing, 20*(2), 148 .

Thissen, D. (1991). MULTILOG user's guide. *Chicago: Scientific Software* .

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 .