روش‌های سنجش تک بعدی بودن سؤال‌ها در مدل‌های دو ارزشی IRT

اصغر مینائی[1]
دکتر محمدرضا فلسفی نژاد[2]

تاریخ وصول: ۸۹/۷/۲۱ تاریخ پذیرش: ۸۹/۱۲/۱۵

چکیده

مدل‌های رایج و پرکاربرد IRT بر ۳ مفروضه اصلی تک بعدی بودن، تکنوا بودن و استقلال موضعی یا استقلال شرطی استوار هستند. تکنوا بودن به این معنا است که احتمال پاسخ درست به سؤال با بالا رفتن توانایی افزایش یابد؛ در استقلال موضعی فرض بر این است که پاسخ‌های سؤال‌های مختلف یک آزمون، برای یک سطح خاص از توانایی (θ)، مستقل از یکدیگرند؛ و تک بُعدی بودن به این معنا است که تمام سؤال‌های آزمون، تنها یک‌صفت یا ویژگی مکنون را اندازه می‌گیرند.

نقض مفروضه تک بعُدی بودن، می‌تواند برآورد پارامترهای سؤال و توانایی را به‌طور جدی دچار تورش نماید. علاوه بر این، تک بعدی بودن سؤال‌ها برای تفسیر روشن و بدون ابهام نمره‌ها و انجام مقایسه‌های بین‌فردی و همچنین برای سایر فعالیت‌های روان‌سنجی مانند همتراز سازی مهم است.

علی رغم اهمیت مفروضه ابعاد برای مدل‌های IRT، در خصوص تعریف تک بعدی بودن و روش‌های سنجش آن بین متخصصان روان‌سنجی، توافق وجود ندارد و تعاریفی که ارائه می‌شود نوعاً انتزاعی و غیرعملیاتی هستند.

باتوجه به نقش و اهمیت مفروضه استقلال‌موضعی در تعریف و شکل‌گیری روش‌های جدید سنجش ابعاد آزمون، در این مقاله، ضمن مرور ادبیات مربوط به مفروضه ابعاد در

۱- هیئت علمی پژوهشگاه آموزش و پرورش و دانشجوی دکتری سنجش و اندازه گیری، دانشگاه علامه طباطبائی
۲- عضو هیئت علمی دانشگاه علامه طباطبائی

مدل‌های دو ارزشی IRT و معرفی برخی از روش‌های سنجش تک بعدی بودن، دو روش NOHARM و DIMTEST در مورد یک مجموعه از داده‌های تستی به‌کار رفت و هر دو روش نتیجه مشابهی به‌دست دادند.

براساس مرور ادبیات موجود و نتایج حاصل از NOHARM و DIMTEST می‌توان گفت زمان آن فرا رسیده است که پژوهشگران از کاربرد شاخص‌های اکتشافی مانند «درصد واریانس تبیین شده توسط عامل اول» یا «نسبت ارزش ویژه عامل اول به عامل دوم» به‌عنوان شاخصی برای تک‌بعدی بودن اجتناب کنند و به استفاده از روش‌های جدید مانند NOHARM، TESTFACT، DIMTEST، DETECT، H&R و MSP رو بیاورند.

واژگان کلیدی: IRT، نظریه سوال- پاسخ، دوارزشی، تک بعدی.

مقدمه

پیدایش نظریهٔ سؤال– پاسخ (IRT) در مدل‌پردازی داده‌های تستی، یک پیشرفت جدید و مهم محسوب می‌شود. به کمک این نظریه، می‌توان بر پایه پاسخ‌های فرد به سؤال‌های آزمون، سطح توانایی وی را استنباط کرد. مدل‌های رایج و پرکاربرد IRT بر **۳** مفروضه اصلی تک بعدی بودن[1]، تکنوا بودن[2] و استقلال موضعی[3] یا استقلال شرطی استوار هستند. تکنوا بودن به این معنا است که احتمال پاسخ درست به سؤال با بالا رفتن توانایی افزایش یابد؛ در استقلال موضعی فرض بر این است که پاسخ‌های سؤال‌های مختلف یک آزمون، برای یک سطح خاص از توانایی (θ)، مستقل از یکدیگرند؛ و تک بُعدی بودن به این معنا است که تمام سؤال‌های آزمون، تنها یک صفت یا ویژگی مکنون را اندازه می‌گیرند.

اگرچه می‌توان استدلال کرد که عملکرد در آزمون، غالباً تحت تأثیر چندین صفت مکنون قرار می‌گیرد، لکن به‌نظر می‌آید اکثر محققان بر این نکته توافق دارند که یک آزمون یا پرسشنامه باید ترجیحاً فقط یک صفت مکنون را اندازه‌گیری کند. وجود تعداد

1. Unidimensionality
2. Monotonicity
3. Local Independence

زیادی مدل تک بُعدی IRT در کنار تعداد معدودی مدل چندبعدی (به‌عنوان مثال، کلدرمن و ریجکس[1]، ۱۹۹۴؛ رکاس[2]، ۱۹۹۷) شاهدی بر این مدعاست.

نقض مفروضه تک بعُدی بودن، می‌تواند برآورد پارامترهای سؤال و توانایی را به‌طور جدی دچار تورش[3] نماید (آکرمن[4]، ۱۹۸۹؛ کریسی و سو[5]، ۱۹۹۵). علاوه بر این، تک بعدی بودن سؤال‌ها برای تفسیر روشن و بدون ابهام نمره‌ها و انجام مقایسه‌های بین فردی (مک نمار[6]، ۱۹۴۶؛ هتی[7]، ۱۹۸۴) و همچنین‌برای سایر فعالیت‌های روان‌سنجی مانند همتراز همتراز سازی (گسارولی و دی چمپلین[8]، ۲۰۰۵) مهم است.

علی‌رغم اهمیت مفروضه ابعاد[9] برای مدل‌های IRT، در خصوص تعریف تک بعدی بعدی بودن و روش‌های سنجش آن بین متخصصان روان‌سنجی، توافق وجود ندارد و تعاریفی که ارائه می‌شود نوعاً انتزاعی و غیرعملیاتی هستند (همبلتون[10] و راونیلی، ۱۹۸۶). در گذشته، مفروضه ابعاد براساس همگنی سؤال‌ها و اعتبار تعریف می‌شد (به‌عنوان مثال، گیج و دارمین[11]، ۱۹۵۰؛ فریمن[12]، ۱۹۶۲؛ هارست[13]، ۱۹۵۳؛ گیلفورد[14]، ۱۹۶۵). تعاریف جدید بر برخی از اَشکال اصل استقلال موضعی، متکی هستند (به‌عنوان مثال، مک دونالد[15] ۱۹۸۱؛ مک دونالد و فریزر، ۱۹۸۵؛ استوت[16]، ۱۹۸۷، ۱۹۹۰).

باتوجه به نقش و اهمیت مفروضه استقلال‌موضعی در تعریف و شکل‌گیری روش‌های جدید سنجش ابعاد آزمون، در ادامه تعاریف مختلفی را که از این مفروضه ارائه شده است

1. Kelderman & Rijkes
2. Reckase
3. Bias
4. Ackerman
5. Kirisci & Hsu
6. McNemar
7. Hattie
8. Gessaroli & De Champlain
9. Dimensionality
10. Hambleton & Rovinelli
11. Gage & Darmin
12. Freeman
13. Horst
14. Guilford
15. McDonald
16. Stout

بیان می‌کنیم و سپس به بررسی روش‌هایی که براساس این تعاریف برای سنجش ابعاد آزمون شکل گرفته‌اند، می‌پردازیم.

تعاریف استقلال موضعی

مفهوم استقلال موضعی یا شرطی نمره‌های سؤال، به صورت‌های گوناگونی در نظریه کلاسیک نمره واقعی (CTT)، تحلیل عاملی، تحلیل طبقه مکنون[1]، و IRT (لرد و ناویک[2]، ناویک[2]، ۱۹۶۸) به‌کار برده می‌شود. برای مثال، در CTT فرض بر این است که با توجه به به نمره واقعی یک آزمودنی، خطاهای اندازه‌گیری ناهمبسته هستند (ین[3] و فیتزپاتریک، ۲۰۰۶). در IRT استقلال موضعی به این معنا است که احتمال پاسخ درست آزمودنی به یک سؤال تحت‌تأثیر پاسخ‌های وی به سؤال‌های دیگر آزمون نیست. شق دیگر استقلال موضعی این است که سؤال‌ها برای افرادی که در یک سطح از توانایی قرار دارند ناهمبسته هستند. بنابراین آنچه که این مفروضه مطرح می‌کند این است که عملکرد در آزمون تحت تأثیر ویژگی‌های آزمودنی و سؤال است. بیان فنی این مفروضه را از زبان لرد و ناویک (۱۹۶۸، ص ۳۶۱) مطرح می‌کنیم. آنها می‌گویند استقلال موضعی به این معنا است که در درون هر گروهی از آزمودنی‌ها که همگی از توانایی یکسان $\theta_1, \theta_2, \ldots \theta_K$ برخوردارند، توزیع‌های شرطی نمره‌های همه سؤال‌ها مستقل از یکدیگر باشند. بیان این مفروضه از لحاظ ریاضی به صورت زیر است:

$$
\begin{aligned}
&P(Y_1 = y_1, Y_2 = y_2, \cdots, Y_k = y_k \mid \Theta = \theta) \\
&= P(Y_1 = y_1 \mid \Theta = \theta) P(Y_2 = y_2), \ldots\ldots\ldots, P(Y_k = y_k \mid \Theta = \theta) \\
&= (P_1^{y_1} Q_1^{1-y_1})(P_2^{y_2} Q_2^{1-y_2}), \ldots\ldots\ldots, (P_k^{y_k} q^{1-y_k}) \\
&= \prod_{i=1}^{k} P_i^{y_i} Q_i^{1-y_i} \qquad (1)
\end{aligned}
$$

در معادلهٔ فوق، y_i یا صفر و یا یک است.

1. Latent Class Analysis
2. Lord & Novick
3. Yen

تعریف فوق از استقلال موضعی به استقلال موضعی قوی[1] (SLI) معروف است، زیرا تمام 2^K تعامل مرتبه بالاتر میان سؤال‌ها را در بر می‌گیرد(گسارولی و دی چمپلین، ۲۰۰۵).

مک دونالد (۱۹۷۹) یک تعریف کمتر سختگیرانه‌ای از استقلال موضعی ارائه می‌دهد که به استقلال موضعی ضعیف[2] (WLI)، یا استقلال موضعی جفتی[3] (ژانگ[4] و استوت، ۱۹۹۹) معروف است، زیرا مستلزم این است که صرفاً کوواریانس‌های شرطی‌بین تمام جفت سؤال‌های آزمون، برای تمام مقادیر یا ارزش‌های ثابت صفت مکنون (توانایی) صفر باشد. برای یک جفت از سؤال‌ها، این مفهوم به شکل زیر تعریف می‌شود:

$$P(Y_i = y_i, Y_j = y_j|\Theta = \theta) = P(Y_i = y_i|\Theta = \theta) \times P(Y_j = y_j|\Theta = \theta) \qquad (۲)$$

$i \neq j;\ j,i=1,\ldots k$

در عمل، فرمول فوق به شکل زیر مورد ارزیابی قرار می‌گیرد:

$$\sum_{j=1}^{k} \sum_{i=1}^{k} Cov(Y_i, Y_j|\Theta = \theta) = 0 \qquad (۳)$$

$i \neq j$

برای درک بهتر مفروضه استقلال موضعی، توجه به نحوه استخراج عامل‌ها در تحلیل عاملی، می‌تواند سودمند باشد. در تحلیل عاملی ابتدا عامل اول، استخراج وتأثیر آن از داده‌ها برداشته می‌شود. این عمل، مشابه مشروط کردن روی θ است. اگر کلیه ضرایب همبستگی‌های پس ماندهٔ میان سؤال‌ها صفر باشد، در این‌صورت عامل دیگری استخراج نمی‌شود. با این حال، اگر ضرایب همبستگی‌های پس ماندهٔ مجموعه‌ای از سؤال‌ها، صفر یا ناچیز نباشد، عامل دوم استخراج می‌شود.

بین مفروضه ابعاد و استقلال موضعی یک ارتباط نزدیک و تنگاتنگ وجود دارد. مک دونالد (۱۹۸۱، ۱۹۸۲) مفهوم ابعاد آزمون را به مفهوم استقلال موضعی سؤال مرتبط

1. Strong Local Independence
2. Weak Local Independence
3. Pairwise Local Independence
4. Zhang

ساخت. او ابعاد آزمون را از طریق تعداد صفات مکنونی که بایستی مد نظر قرار گیرند تا به استقلال موضعی ضعیف بین سؤال‌ها دست یافت، تعریف نمود. تعریف مک دونالد، جالب و جذاب است زیرا ابعاد را براساس مفهوم استقلال موضعی سؤال‌ها که برای IRT و روش آن اساسی است تعریف می‌کند. همچنین تعریف او از ابعاد، یک تعریف عملیاتی است، زیرا براساس‌این تعریف، ابعاد آزمون را می‌توان از طریق‌بررسی کوواریانس‌های‌شرطی بین سؤال‌ها تعیین کرد.

تعریف صفات مکنون و ابعاد آزمون با استفاده از SLI یا WLI یک تعریف ریاضی است و بر الزامات نظری استقلال آماری مبتنی هستند. به عبارت دقیق‌تر، انعکاس اصل کلی‌تر استقلال موضعی دقیق، هستند. برخی از افراد مانند گلداشتاین[1] (۱۹۸۰، ۱۹۸۹) استدلال‌می‌کنند که این تعاریف، گاهی اوقات تمام وابستگی‌های میان پاسخ‌های‌سؤال‌ها را در بر نمی‌گیرند. علاوه بر این، این امکان وجود دارد که تعریف ریاضی مفروضه ابعاد برقرار باشد ولی تمام متغیرهای روان‌شناختی که پاسخ‌های سؤال‌ها را تحت‌تأثیر قرار می‌دهند به‌طور کامل تبیین نشوند. برای مثال، فرض کنید در یک آزمون درک مطلب که به تئاتر یونان باستان مربوط می‌شود، از آزمودنی‌ها خواسته شود که پس از خواندن متن به یک‌سری سؤال‌های‌مرتبط با متن پاسخ بدهند. این امکان کاملاً وجود دارد که وابستگی‌های ریاضی میان سؤال‌ها فقط توسط یک صفت تبیین شود. با این حال، اگر چه به لحاظ ریاضی، فقط یک صفت وجود دارد و استقلال موضعی در معنای ریاضی آن برقرار است، لکن این امکان وجود دارد که پاسخ‌های افراد به سؤال‌ها تحت‌تأثیر دو صفت روان‌شناختی باشند. پاسخ یک آزمودنی به یک سؤال ممکن است تحت‌تأثیر توانایی کلی درک خواندن و همچنین دانش خاص وی در زمینه تئاتر یونان باستان باشد.

فردی به نام ویلیام استوت (۱۹۸۷، ۱۹۹۰) با کمک گرفتن از این عقاید، نظریه استقلال اساسی سؤال[2] را پیشنهاد کرد و بر این اساس، تعریفی از ابعاد اساسی[3] ارائه کرد.

1. Goldstein
2. Essential Item Independence
3. Essential dimensionality

استوت و برخی دیگر از محققان مانند ناندا‌کومار[1] (۱۹۹۱) اظهار می‌دارند که تعریف ابعاد آزمون براساس استقلال موضعی قوی، خیلی سختگیرانه است، زیرا هم ابعاد اصلی و هم ابعاد فرعی را دربر می‌گیرد. آنها استدلال می‌کنند که ابعاد فرعی، اگرچه به لحاظ ریاضی وجود دارند، اما تأثیر مهمی روی پاسخ‌های افراد ندارند و احتمالاً در توصیف ابعاد زیربنایی پاسخ‌های سؤال، بی اهمیت هستند. این ایده، پایه و اساس تعریف ابعاد اساسی (استوت، ۱۹۸۹، ۱۹۹۰؛ ناندا‌کومار، ۱۹۹۱) را تشکیل می‌دهد.

تعریف استوت از استقلال اساسی[2] (EI) در مقایسه با SLI یا WLI، الزامات نسبتاً ضعیف‌تری را طلب می‌کند. از تعریف WLI چنین برمی‌آید زمانی که استقلال موضعی برقرار باشد، کوواریانس‌های شرطی بین تمام جفت سؤال‌ها برای تمام θها برابر با صفر خواهد بود؛ در حالی‌که استقلال اساسی صرفاً مستلزم این است که متوسط قدر مطلق این کوواریانس‌ها، با افزایش طول آزمون، به صفر میل نماید (ین و فیتزپاتریک، ۲۰۰۶). علاوه بر این، تعریف WLI تلویحاً به این معنا است که صفات مورد استفاده توسط مدل مورد نظر، کوواریانس‌های بین تمام جفت سؤال‌ها را به‌طور کامل تبیین می‌کنند، در حالیکه EI به‌طور ضمنی اشاره به این دارد که صفات مورد استفاده توسط مدل، بایستی صرفاً صفات غالب باشند (ناندا‌کومار، ۱۹۹۱). بیان ریاضی EI به شکل زیر است (استوت، ۱۹۹۰):

$$\frac{۲}{P(P-۱)}\sum_{1\le j<k\le p}\left|C\hat{o}v(Y_j,Y_k)\middle|\Theta=\theta\right| \qquad (۴)$$

$$\to 0 \text{ اگر } P\to\infty$$

استقلال اساسی از این لحاظ که صرفاً به وابستگی‌های جفتی میان سؤال‌ها توجه دارد، مشابه WLI است. با این حال، برخلاف WLI در EI لازم نیست که کوواریانس‌های شرطی سؤال‌ها مساوی صفر باشد، بلکه همان‌طور که قبلاً نیز گفته شد، EI مستلزم این است که میانگین قدرمطلق کوواریانس‌های تمام سؤال‌ها با افزایش طول آزمون‌به صفر نزدیک شود. در نتیجه، EI صرفاً ابعاد غالب را مورد توجه قرار می‌دهد در حالی‌که در WLI برای اینکه

1. Nandakumar
2. Essential Independence

معادله‌های ۲ یا ۳ برآورده شوند، به لحاظ نظری لازم است تا تمام ابعاد، حتی ابعاد فرعی، مورد توجه قرارگیرند. بنابراین، بُعدی بودن اساسی[1] (d_{EI}) را می‌توان به صورت کمترین تعداد ابعاد مورد نیاز برای برقراری EI تعریف نمود (استوت، ۱۹۹۰؛ گسارولی و دی چمپلین، ۲۰۰۵). از تعاریف فوق چنین برمی‌آید که $d_{EI} \leq d_{LI}$ است. در مواردی که $d_{EI} = ۱$ است، ماتریس پاسخ سؤال‌ها، اساساً تک بعدی[2] تلقی می‌شود.

روش‌های سنجش ابعاد

زمانی که برای تحلیل داده‌ها تستی از مدل‌های تک بعدی IRT استفاده‌می‌کنیم، روایی این مفروضه که فضای مکنون کامل، تک بعدی است یک مسأله کاملاً اساسی است. برای بررسی‌مفروضه تک‌بعدی بودن سؤال‌های آزمون، رویکردها و روش‌های‌مختلفی ارائه شده است. هتی(۱۹۸۴، ۱۹۸۵) بیش از ۸۰ شاخص را که برای سنجش تک بعدی بودن سؤال‌ها به‌کار می‌روند شناسایی کرد. او بیان می‌کند که اکثر شاخص‌های مورد استفاده (مانند شاخص‌های مبتنی بر اعتبار، همگنی، مؤلفه‌های اصلی) ماهیتاً پیش تجربی[3] بوده و بر هیچ تعریف رسمی و روشنی از ابعاد، استوار نیستند. او همچنین اظهار می‌دارد که در خصوص مفاهیم همگنی، همسانی درونی و ابعاد یک نوع سردرگمی و بی نظمی وجود دارد (برای مطالعه در خصوص تفاوت‌هایی‌که بین این مفاهیم وجود دارد به کورتینا[4]، ۱۹۹۳؛ اسمیت[5]، ۱۹۹۵، مراجعه نمایید). برای مثال، اگرچه تصور بر این است که بین میزان اعتبار آزمون و ابعاد آن ارتباط وجود دارد، (یعنی بالا بودن اعتبار، به مثابه تک بعدی بودن آن تلقی می‌شود)، گرین، لیستیز و مولیک[6] (۱۹۷۷) نشان داده‌اند که این امکان وجود دارد که حتی ضریب آلفای یک آزمونِ ۵ بعدی نیز بالا باشد.

1. Essential Dimensionality
2. Essentially Uidimensional
3. ad hoc
4. Cortina
5. Schmitt
6. Green, Lisstz & Mulaik

روش‌هایی که امروزه برای سنجش ابعاد به‌کار می‌روند از مبنای نظری دقیق‌تر و محکم‌تری برخوردارند، زیرا یا بر استقلال موضعی و یا بر استقلال اساسی متکی هستند. روسوس، استوت و ماردن[1] (۱۹۹۸) اظهار می‌دارند که روش‌های سنجش ابعاد آزمون را می‌توان به طور کلی براساس یک طرح طبقه‌بندی ۲ × ۲ تقسیم‌بندی کرد. اولاً روش‌ها را می‌توان به صورت پارامتریک یا غیرپارامتریک طبقه‌بندی کرد. در روش‌های پارامتریک فرض بر این است که تابع سؤال- پاسخ (IRF) از یک مدل پارامتریک خاص، تبعیت می‌کند، در حالی‌که در روش‌های‌غیرپارامتریک صرفاً فرض‌بر این است که IRF به‌صورت تکنوا است. ثانیاً این روش‌ها یا تلاش می‌کنند که ابعاد آزمون را بطور کامل برآورد کنند، (یعنی تعداد ابعاد و سؤال‌های مربوط به هر بعد را مشخص نمایند)، و یا تلاش می‌کنند که صرفاً فقدان تک بعدی بودن، (یعنی اینکه آیا آزمون تک بعدی است یا نه)، را برآورد یا کشف نمایند.

این روش‌ها به یک نحوی به اصول استقلال اساسی مربوط‌می‌شوند، زیرا شاخص‌هایی فراهم می‌کنند که میزان فقدان استقلال‌شرطی داده‌ها را نشان‌می‌دهند. برخی از این روش‌ها شاخص‌های کلی بدست می‌دهند و برخی دیگر نیز تک بعدی بودن را از طریق سنجش میزان وابستگی موجود بین جفت سؤال‌ها، بیان می‌کنند (گسارولی و دی چمپلین، ۲۰۰۵). در ادامه، روش‌های مورد استفاده در سنجش ابعاد آزمون‌ها ارائه می‌شود.

الف- روش‌های مبتنی بر استقلال موضعی: از جمله این روش‌ها می‌توان به تحلیل عامل خطی[2] (همبلتون و تراب، ۱۹۷۳؛ رکاس، ۱۹۷۹)، تحلیل عاملی غیرخطی[3] (گسارولی و دی چمپلین، ۱۹۹۶)، و تحلیل عاملی با اطلاعات کامل[4] (مک لاود، سوئی گرت، و تیسن[5]، ۲۰۰۱) اشاره کرد.

1. Roussos, Stout & Marden
2. Linear Factor Analysis
3. Non-Linear
4. Full Information Factor Analysis
5. McLeod, Swygert & Thissen

در تحلیل عاملی یک متغیر مشاهده شده، روی یک یا چند متغیر مکنون برگشت داده می‌شود. در سؤال‌های دو ارزشی، Y یک متغیر دو ارزشی است که مقدار آن برای پاسخ‌های نادرست، صفر و برای پاسخ‌های درست ۱ است. متغیرهای مکنون، صفات مشاهده ناپذیر هستند که برای ارائه پاسخ درست به سؤال‌ها ضروری هستند.

برای تحلیل عاملی خطی سؤال‌های دو ارزشی، از ماتریس همبستگی فی یا تتراکوریک استفاده می‌شود (ناندا‌کومار و همکاران، ۱۹۹۸). پژوهش‌ها نشان داده‌اند که استفاده از ماتریس ضرایب فی، تعداد ابعاد زیر بنایی آزمون را بیش برآورد می‌کند (همبلتون و راونیلی، ۱۹۸۶). ضرایب‌همبستگی فی، تحت‌تأثیر فراوانی‌های‌حاشیه‌ای سؤال‌ها قرار می‌گیرد (پانتر، سوئی گرت و دال استروم[1]، ۱۹۹۷)، از این رو تحلیل عاملی ماتریس همبستگی‌های ضرایب فی، منجر به عامل یا عامل‌های تصنعی می‌شود که ناشی از غیرخطی بودن رابطه بین سؤال‌ها و پیوستارتوانایی است (سوامیناتان[2]، همبلتون و راجرز، ۲۰۰۷). این عامل‌های‌تصنعی یا غیرواقعی را گاهی‌اوقات، عامل‌دشواری[3] نیزمی‌نامند (فرگسون[4]، ۱۹۴۱؛ مک دونالد، ۱۹۶۷).

همبستگی تتراکوریک بر این فرض مبتنی است که متغیرهای دو ارزشی مشاهده شده، معرف متغیرها یا صفات مکنونی هستند که توزیع آنها نرمال است. ضعف استفاده از ماتریس همبستگی‌های تتراکوریک این است که وجود یک عامل مشترک، شرط کافی برای تک بعدی بودن یک مجموع از سؤال‌ها است، اما شرط لازم نیست (لرد و ناویک، ۱۹۶۸، ص ۳۸۲). علاوه بر این، ماتریس‌های همبستگی تتراکوریک، الزاماً معین مثبت[5] نیستند (دی آیالا و هرتزوگ[6]، ۱۹۸۹؛ میسلوی[7]، ۱۹۸۶؛ هتی، ۱۹۸۴؛ همبلتون و سوامیناتان، سوامیناتان، ۱۹۸۵). بنابراین، می‌توان‌گفت که روش‌های مرسوم تحلیل عاملی، برای سنجش ابعاد داده‌های تستی مناسب نیستند.

1. Panter, Swygert & Dahlstrom
2. Swaminathan
3. Difficulty Factor
4. Ferguson
5. Positive Definite
6. De Ayala & Hertzoy
7. Mislevy

مشکل اصلی در تحلیل عاملی خطی، خطی بودن آن است. با عنایت به این واقعیت و مشکلات ناشی از استفاده از ماتریس‌های همبستگی فی و تتراکوریک، همبلتون و راونیلی (۱۹۸۶) پیشنهاد می‌کنند که برای بررسی تک بعدی بودن داده‌های تستی از روش تحلیل عاملی غیرخطی که توسط مک دونالد (۱۹۶۷، ۱۹۸۲) ارائه شده است استفاده شود. این رویکرد که از طریق برنامه کامپیوتری NOHARM (فریزر[1]، ۱۹۸۸) قابل اجراست، یک روش خوش‌آتیه برای سنجش تک بعدی بودن داده‌های تستی است (نانداکومار، یو، لی، واستوت، ۱۹۹۸).

از آنجایی‌که برای سنجش ابعاد آزمون، در برنامه NOHARM شیوه‌های آماری ارائه نمی‌شود، لذا گسارولی و دی چمپلین (۱۹۹۴) آمارهٔ $\chi^2_{G/D}$ را که براساس همبستگی‌های پس مانده بین سؤال‌ها محاسبه می‌شود، پیشنهاد می‌کنند. پس مانده‌ها را می‌توان با استفاده از روش Z فیشر تبدیل کرد و آمارهٔ مورد نظر را به شکل زیر محاسبه نمود:

$$\chi^2_{G/D} = (N-3)\sum_{i=1}^{n}\sum_{j=1}^{i-1} Z^2_{ij}$$

در فرمول فوق، N حجم نمونه است. فرض بر این است که آماره فوق دارای توزیع χ^2 با $\frac{1}{2}[n(n-1)-n(1+d)]$ درجه آزادی است که n طول آزمون و d تعداد ابعاد آزمون است. برای آزمون مفروضه تک بعدی بودن، d برابر با ۱ و درجه آزادی برابر با $n\frac{1}{2}[(n-5)]$ خواهد بود.

روش دیگر برای تحلیل عاملی سؤال‌های دو ارزشی و تبیین ابعاد آنها، تحلیل عاملی با اطلاعات کامل (FIFA) است که توسط باک، گیبونز و موراکی[2] (۱۹۸۸)، پرورش یافت و از طریق برنامه کامپیوتری TESTFACT (ویلسون، وود، و گیبونز[3]، ۱۹۹۱) قابل اجرا است. FIFA یک تکنیک تحلیل عاملی براساس IRT است. این رویکرد از عامل‌های

1. Fraser
2. Bock, Gibbons & Muraki
3. Wilson, Wood & Gibbons

تصنعی و تمام مشکلات‌مرتبط با تحلیل عاملی ماتریس‌های همبستگی‌اجتناب می‌کند (مک لاود، سوئی گرت، و تیسن، ۲۰۰۱). در این روش از همبستگی‌های درونی سؤال‌ها استفاده نمی‌شود، بلکه از اطلاعات موجود در تمام الگوها یا بردارهای پاسخ استفاده می‌شود. به همین دلیل این رویکرد را تحلیل عاملی با اطلاعات کامل می‌نامند (بارتلومیو[1]، ۱۹۸۰). به عبارت دقیق‌تر، این رویکرد به‌طور مستقیم از داده‌های سطح سؤال، استفاده می‌کند. البته، در این روش از ماتریس همبستگی‌های تتراکوریک استفاده می‌شود، اما صرفاً برای محاسبه مقادیر آغازین برای روش برآورد به شیوه ازسرگیری.

در واقع، در FIFA برازش مدل چند عاملی ترستون با داده‌ها با استفاده از روش برآورد بیشینه درستنمایی کناری (MML) مورد آزمون قرار می‌گیرد (مک لاود، سوئی گرت، و تیسن، ۲۰۰۱).

تحلیل عاملی خطی، تحیل عامل غیرخطی و تحلیل عاملی با اطلاعات کامل، جزء روش‌های پارامتریک هستند که تلاش می‌کنند ابعاد آزمون را به طور کامل برآورد نمایند. همان‌طور که قبلاً نیز ذکر شد، پژوهش‌ها نشان داده‌اند که تحلیل عاملی خطی برای سنجش ابعاد داده‌های تستی مناسب نیست. یافته‌های برخی از پژوهش‌ها (به‌عنوان مثال، دی چمپلین و گساورلی، ۱۹۹۸؛ فینچ وهابینگ[2]، ۲۰۰۵؛ نول و برگر[3]، ۱۹۹۱) نشان داده است که در مطالعات بازیابی[4] ابعاد، NOHARM در مقایسه با TESTFACT بهتر عمل می‌کند. از این رو برنامه NOHARM را با استفاده از یک مثال، بطور مفصل‌تر توضیح می‌دهیم.

برای این منظور از داده‌های مربوط به درس ریاضی کنکور سال ۱۳۸۵ برای گروه ریاضی و فیزیک استفاده گردید. از بین کل داوطلبان این گروه، تعداد ۲۰۰۰ نفر به‌طور کاملاً تصادفی انتخاب و پاسخ‌های آنها به سؤالات درس ریاضی استخراج گردید. پاسخنامه‌ها با استفاده از کلید مربوطه، نمره‌گذاری شدند. پس از نمره‌گذاری و تبدیل پاسخ‌ها به داده‌های دو ارزشی (۰ یا ۱) ابتدا ماتریس گشتاور حاصلضربی نمره‌های خام، $\underline{P}$،

1. Bartholomew
2. Finch & Habing
3. Knol & Berger
4. Recovery

با استفاده از برنامه PRODMOM (فریزر، ۱۹۸۸) برآورد گردید. این ماتریس از طریق $(1/N)\,\underline{X}'\underline{X}$ بدست می‌آید که N برابر با حجم نمونه و $\underline{X}'$ ماتریس داده‌های خام است. اگر X پاسخ‌های دو ارزشی باشد، در این صورت قطر اصلی ماتریس $\underline{P}$ را میانگین یا همان درجه دشواری سؤال‌ها تشکیل می‌دهد و عناصر غیرقطری نیز میانگین مجموع مضروبات است. به‌عنوان مثال، برای یک آزمون ۳ سؤال ماتریس $\underline{P}$ به صورت زیر خواهد بود.

$$\underline{P} = \begin{bmatrix} P_1 & \sum X_1X_2/N & \sum X_1X_3/N \\ \sum X_2X_1/N & P_2 & \sum X_2X_3/N \\ \sum X_3X_1/N & \sum X_3X_2/N & P_3 \end{bmatrix}$$

فایل درونداد برای برنامه NOHARM برای انجام یک تحلیل تک بعدی در جدول ۱ ارائه شده است. سطر اول این فایل، عنوان تحلیل را نشان می‌دهد. سطر ۲ و ۳ دستورهای لازم برای انجام تحلیل را مشخص می‌کنند. برای نمونه، سطر۲ نشان می‌دهد که تعداد سؤال‌ها ۳۰ و تعداد آزمودنی‌ها ۲۰۰۰ است، و فایل درون داد یک ماتریس گشتاور حاصل‌ضربی است، و تحلیل‌مورد نظر یک تحلیل‌اکتشافی است، و ماتریس‌های همبستگی، کوواریانس و پس مانده‌ها باید تولید و نمایش داده شوند. سطر ۳ به کاربر اجازه می‌دهد تا مقدار پارامتر مجانب پایین را برای هر سؤال مشخص نماید. در این تحلیل، مقدار این پارمتر برای تمام سؤال‌ها برابر با صفر قرار داده شده است، زیرا فرض بر این بود که به دلیل اِعمال نمره منفی، داوطلبان دست به حدس نمی‌زنند. در انتهای جدول، بخش پایینی ماتریس $\underline{P}$ ارائه شده است. از آنجایی که ماتریس $\underline{P}$ یک ماتریس متقارن است لذا از بخش پایینی ماتریس، به‌عنوان درونداد استفاده می‌شود.

جدول ۲ خلاصه‌ای از خروجی تحلیل را نشان می‌دهد. در ابتدای جدول یا فایل خروجی، همان اطلاعاتی که در فایل درونداد توسط کاربر مشخص گردیده است نشان داده شده است. فایده این اطلاعات در این است که با استفاده از آنها می‌توان بررسی کرد که آیا برنامه، فایل درونداد را به‌طور صحیح، بازخوانی کرده است یا نه. اطلاعات مربوط به برازش مدل در پایین ماتریس پس‌مانده، Residual Matrix، ارائه شده است. این

ماتریس، از طریق تفاضل بین کوواریانس‌های مشاهده‌شده و کوواریانس‌های پیش‌بینی‌شده، توسط مدل، بدست می‌آید. بررسی این ماتریس نشان‌می‌دهد که کوواریانس‌های پس مانده در مقایسه با کوواریانس‌های مشاهده شده بین سؤال‌ها، نسبتاً کوچک هستند.

جدول ۱. فایل درونداد برای انجام تحلیل تک بُعدی بودن در NOHARM

```
Dimensionality of Mathematics Test
30 1 2000 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
.30
.05 .18
.09 .06 .25
.13 .08 .12 .38
.10 .05 .08 .13 .24
.09 .08 .08 .13 .09 .28
.09 .06 .07 .12 .07 .08 .25
.06 .05 .06 .10 .06 .07 .07 .18
.06 .04 .06 .09 .06 .06 .07 .05 .18
.09 .07 .09 .13 .08 .08 .09 .07 .07 .25
.08 .05 .06 .10 .07 .08 .07 .05 .05 .07 .17
.13 .05 .06 .15 .11 .09 .09 .07 .07 .09 .07 .28
.11 .08 .11 .18 .10 .13 .11 .09 .08 .12 .09 .11 .34
.06 .03 .06 .08 .05 .06 .06 .05 .05 .07 .05 .06 .08 .16
.05 .03 .05 .08 .05 .05 .05 .04 .04 .05 .04 .06 .07 .04 .14
.10 .07 .09 .16 .09 .13 .09 .08 .07 .10 .09 .11 .15 .07 .06 .20
.11 .07 .10 .14 .08 .10 .10 .08 .08 .09 .07 .11 .13 .06 .05 .11 .30
.07 .04 .07 .11 .07 .07 .08 .06 .06 .08 .06 .08 .10 .05 .04 .09 .09 .20
.07 .04 .06 .09 .07 .06 .06 .05 .05 .07 .06 .07 .08 .04 .04 .07 .08 .06 .17
.06 .04 .05 .09 .06 .05 .06 .05 .05 .06 .05 .06 .09 .05 .04 .07 .07 .05 .05 .18
.09 .06 .08 .14 .08 .08 .09 .07 .06 .09 .07 .09 .12 .06 .06 .10 .10 .08 .06 .07 .23
.07 .05 .06 .11 .07 .07 .07 .06 .06 .08 .06 .09 .10 .05 .05 .09 .08 .07 .06 .06 .08 .20
.07 .04 .07 .11 .06 .07 .07 .06 .06 .08 .06 .08 .09 .05 .04 .08 .08 .07 .05 .06 .08 .06  .18
.10 .06 .09 .16 .09 .10 .09 .08 .08 .11 .09 .11 .14 .07 .06 .12 .11 .09 .09 .07 .11 .10 .09 .27
.07 .05 .06 .10 .07 .07 .06 .06 .05 .06 .06 .07 .08 .04 .04 .08 .07 .06 .06 .05 .07 .06 .05 .08 .17
.12 .05 .08 .14 .09 .09 .08 .07 .06 .08 .07 .13 .12 .06 .06 .10 .10 .07 .07 .06 .09 .08 .07 .11 .07 .28
.09 .05 .08 .13 .08 .08 .08 .07 .07 .09 .07 .09 .11 .06 .06 .10 .09 .08 .07 .06 .09 .08 .07 .11 .07 .09 .22
.06 .04 .05 .09 .06 .05 .06 .05 .04 .06 .04 .07 .07 .04 .04 .07 .06 .05 .05 .05 .07 .06 .05 .07 .05 .07 .06 .18
.07 .06 .08 .11 .07 .08 .08 .06 .06 .08 .06 .08 .09 .05 .05 .09 .09 .07 .06 .06 .07 .07 .07 .10 .06 .08 .08 .06 .21
.09 .07 .08 .11 .07 .10 .08 .06 .06 .09 .07 .08 .12 .05 .05 .10 .11 .07 .06 .06 .08 .07 .06 .10 .07 .09 .08 .06 .07 0.27
```

جدول ۲: خلاصه‌ای از خروجی برنامه NOHARM **مربوط به راه حل تک بعدی**

```
N O H A R M

Fitting a (multidimensional) Normal Ogive
by Harmonic Analysis - Robust Method

Input File : D:\ Math.dat
Title : Dimensionality of Mathematics Test

Number of items       = 30
Number of dimensions  = 1
Number of subjects    = 2000

An exploratory solution has been requested.

Sample Product-Moment Matrix
```

```
.30
.05 .18
.09 .06 .25
.13 .08 .12 .38
.10 .05 .08 .13 .24
.09 .08 .08 .13 .09 .28
.09 .06 .07 .12 .07 .08 .25
.06 .05 .06 .10 .06 .07 .07 .18
.06 .04 .06 .09 .06 .06 .07 .05 .18
.09 .07 .09 .13 .08 .08 .09 .07 .07 .25
.08 .05 .06 .10 .07 .08 .07 .05 .05 .07 .17
.13 .05 .06 .15 .11 .09 .09 .07 .07 .09 .07 .28
.11 .08 .11 .18 .10 .13 .11 .09 .08 .12 .09 .11 .34
.06 .03 .06 .08 .05 .06 .06 .05 .05 .07 .05 .06 .08 .16
.05 .03 .05 .08 .05 .05 .05 .04 .04 .05 .04 .06 .07 .04 .14
.10 .07 .09 .16 .09 .13 .09 .08 .07 .10 .09 .11 .15 .07 .06 .20
.11 .07 .10 .14 .08 .10 .10 .08 .08 .09 .07 .11 .13 .06 .05 .11 .30
.07 .04 .07 .11 .07 .07 .08 .06 .06 .08 .06 .08 .10 .05 .04 .09 .09 .20
.07 .04 .06 .09 .07 .06 .06 .05 .05 .07 .06 .07 .08 .04 .04 .07 .08 .06 .17
.06 .04 .05 .09 .06 .05 .06 .05 .05 .06 .05 .06 .09 .05 .04 .07 .07 .05 .05 .18
.09 .06 .08 .14 .08 .08 .09 .07 .06 .09 .07 .09 .12 .06 .06 .10 .10 .08 .06 .07 .23
.07 .05 .06 .11 .07 .07 .07 .06 .06 .08 .06 .09 .10 .05 .05 .09 .08 .07 .06 .06 .08 .20
.07 .04 .07 .11 .06 .07 .07 .06 .06 .08 .06 .08 .09 .05 .04 .08 .08 .07 .05 .06 .08 .06 .18
.10 .06 .09 .16 .09 .10 .09 .08 .08 .11 .09 .11 .14 .07 .06 .12 .11 .09 .09 .07 .11 .10 .09 .27
.07 .05 .06 .10 .07 .07 .06 .06 .05 .06 .06 .07 .08 .04 .04 .08 .07 .06 .06 .05 .07 .06 .05 .08 .17
.12 .05 .08 .14 .09 .09 .08 .07 .06 .08 .07 .13 .12 .06 .06 .10 .10 .07 .07 .06 .09 .08 .07 .11 .07 .28
.09 .05 .08 .13 .08 .08 .08 .07 .07 .09 .07 .09 .11 .06 .06 .10 .09 .08 .07 .06 .09 .08 .07 .11 .07 .09 .22
.06 .04 .05 .09 .06 .05 .06 .05 .04 .06 .04 .07 .07 .04 .04 .07 .06 .05 .05 .05 .07 .06 .05 .07 .05 .07 .06 .18
.07 .06 .08 .11 .07 .08 .08 .06 .06 .08 .06 .08 .09 .05 .05 .09 .09 .07 .06 .06 .07 .07 .07 .10 .06 .08 .08 .06 .21
.09 .07 .08 .11 .07 .10 .08 .06 .06 .09 .07 .08 .12 .05 .05 .10 .11 .07 .06 .06 .08 .07 .06 .10 .07 .09 .08 .06 .07 0.27

=======
Results
=======

Success. The job converged to the specified criterion.
.
.
.
Residual Matrix (lower off-diagonals)
.
.
.
Sum of squares of residuals (lower off-diagonals)      =   0.0130941
Root mean square of residuals (lower off-diagonals)  =   0.0054865      RMSR
Tanaka index of goodness of fit                                =   0.9818874      GFI
```

برنامه NOHARM ریشه دوم میانگین مجذورات پس مانده‌ها[1] (RMSR) را محاسبه و به‌عنوان شاخصی برای برازش مدل ارائه می‌دهد. در واقع، RMSR برابر است با ریشه دوم میانگین مجذورات تفاوت بین کوواریانس‌های مشاهده شده و کوواریانس‌های پیش‌بینی شده. بنابراین، مقادیر کوچک RMSR حاکی از برازش مدل با داده‌ها است. یک ملاک

1. Essential Dimensionality

برای تفسیر RMSR این است که آن را با چهار برابر معکوس ریشه دوم حجم نمونه، (یعنی خطاهای‌استاندارد پس مانده‌ها) مقایسه کرد (مک دونالد، ۱۹۹۷). مقدار این ملاک برای تحلیل حاضر ۰/۰۲۲ است. شاخص دیگر برای بررسی برازش مدل، شاخص خوبی برازندگی تاناکا[1] (۱۹۹۳) است. مک دونالد (۱۹۹۷) پیشنهاد می‌کند مقدار ۰/۹۰ برای این شاخص حاکی از برازش قابل قبول و مقدار ۰/۹۵ بیانگر «برازش خوب» مدل با داده‌ها است. اگر GFI =۱ باشد بیانگر برازش کامل است. بنابراین، با توجه به این شاخص‌ها و با درنظرگرفتن ماتریس پس‌مانده‌ها، به‌نظر می‌آید که شواهد کافی برای رد فرضیه تک بعدی بودن آزمون ریاضی وجود ندارد.

به منظور بررسی دقیق‌تر مسئله و اطمینان از نتیجه‌گیری فوق، با اصلاح سطر دوم فایل درونداد برنامه NOHARM داده‌های مورد نظر براساس یک راه حل دو بعدی، یک بار دیگر مورد تحلیل قرار گرفت. پس‌مانده‌های راه حل دو بعدی نیز همانند پس مانده‌های راه حل تک بعدی بسیار کوچک بودند مقدار RMSR مربوط به راه حل دو بعدی ۰/۰۰۴۲ به دست آمد که در مقایسه با RMSR مربوط به راه حل تک بعدی، کاهش ناچیزی را نشان می‌دهد. علاوه بر این شاخص تاناکا که مقدار آن برابر با ۰/۹۸۹ است، در مقایسه با مقدار این شاخص در راه‌حل تک بعدی افزایش بسیار اندکی را نشان می‌دهد. بنابراین با اطمینان می‌توان نتیجه گرفت که آزمون مورد بحث تک بعدی است و می‌توان از مدل‌های تک بعدی IRT برای مقیاس‌پردازی آن استفاده کرد.

روش‌های دیگری نیز، به‌ویژه برای سنجش وابستگی شرطی دو به دوی سؤال‌ها تحت فرض تک بعدی بودن، پرورش یافته است. از جمله این روش‌ها می‌توان به آماره Q_3 (ین، ۱۹۸۴) و شاخص‌های وابستگی موضعی که چن[2] و تیسن(۱۹۹۷) بیان می‌کنند اشاره کرد. این آماره‌ها از طریق مقایسه فراوانی‌های مشاهده شده و فراوانیهای مورد انتظار براساس یک مدل IRT، محاسبه می‌شوند. برای آگاهی از نتایج حاصل از بررسی عملکرد این آماره‌ها می‌توانید به چن و تیسن(۱۹۹۷)، ین (۱۹۸۴) و زویک[3] (۱۹۸۷) مراجعه کنید.

1. Tanaka
2. Chen
3. Zwick

روش‌ها و آماره‌هایی که تا اینجا مورد بحث قرار گرفت جزء روش‌های پارامتریک هستند که بر اصل WLI مبتنی می‌باشند. سودمندی این روش‌ها در سنجش ابعاد آزمون به مدل IRT مورد استفاده و مفروضه‌های شکل توزیع صفت (صفات) مکنون وابسته است. برقرار نبودن مفروضه‌ها یا فرمولبندی[1] نادرست مدل، می‌تواند به نتیجه‌گیری نادرست راجع به ابعاد آزمون منجر گردد.

ب- **روشهای مبتنی بر استقلال اساسی**: در طول ۲۰ سال گذشته، استوت و همکارانش فعالیتهای زیادی را برای پرورش روش‌های غیرپارامتریک براساس اصل استقلال‌اساسی که قبلاً توضیح داده شد، انجام داده‌اند. تلاش‌های استوت و همکارانش در پرورش روش‌هایی برای شناسایی ماهیت و میزان چند بعدی بودن یک آزمون که هنوز نیز ادامه دارد، منجربه ارائه ۳ روش شده است. این روش‌ها عبارت است از :

۱- DIMTEST (استوت، ۱۹۸۷، ۱۹۹۰)؛ ۲- DETECT (کیم[2]، ۱۹۹۴؛ ژانگ و استوت، ۱۹۹۹)؛ ۳- HCA/CCPROX (روسوس، استوت و ماردن، ۱۹۹۸).

DIMTEST که مشهورترین و پرکاربردترین این روش‌ها است یک روش‌آماری برای سنجش تک بعدی بودن یک مجموعه از داده‌های تستی دو ارزشی است (ناندا کومار و یو، ۱۹۹۶). این روش برای اولین بار توسط استوت (۱۹۸۷) پرورش یافت و بعدها توسط نانداکومار و استوت (۱۹۹۳) اصلاح گردید. DIMTEST یک آزمون یا روش آماری ناپارامتریک است، زیرا آماره T استوت بر هیچ فرضی راجع به شکل توزیع توانایی زیربنایی یا نوع تابع سؤال- پاسخ متکی نیست.

در DIMTEST آزمون فرضیه صفر تک‌بعدی بودن اساسی داده‌ها در دو مرحله صورت می‌گیرد (نانداکومار و آکرمن، ۲۰۰۴). در گام اول، n سؤال آزمون به دو خرده آزمون به نام آزمون ارزشیابی[3] (AT) و آزمون گروه‌بندی[4] (PT) تقسیم می‌شوند. خرده آزمون AT دارای m سؤال (نصف طول آزمون > m ≤۴) و خرده آزمون PT دارای n-m

1. Specification
2. Kim
3. Assessment Test
4. Partitioning Test

سؤال است. خرده آزمون AT متشکل از سؤال‌هایی است که براساس قضاوت متخصص یا تحلیل عاملی ماتریس همبستگی‌های تتراکوریک (فرالیچ، ۲۰۰۰؛ ناندا‌کومار و استوت، ۱۹۹۳) تک بعدی هستند و خرده آزمون PT نیز از مابقی سؤال‌ها تشکیل می‌شود.

به‌خاطر روش انتخاب سؤال‌ها برای خرده آزمون AT، حتی زمانی که آزمون اصلی چند بعدی است، سؤال‌هایی که در خرده آزمون AT قرار می‌گیرند یک سازه تک‌بعدی را اندازه خواهند گرفت، در حالی‌که سؤال‌های خرده آزمون PT ماهیتاً چندبعدی خواهند بود. از طرف دیگر اگر آزمون اصلی اساساً تک بعدی باشد، در این صورت سؤال‌های هر دو خرده آزمون AT و PT یک سازه تک بعدی را اندازه‌گیری خواهند کرد.

در گام دوم، آماره T به شکل زیر محاسبه می‌شود. آزمودنی‌ها براساس نمره‌ای که در خرده آزمون PT بدست آورده‌اند به چند زیرگروه تقسیم می‌شوند. K امین زیر گروه متشکل از آزمودنی‌هایی است که نمره کل آنها در خرده آزمون PT که با X_{PT} نشان داده می‌شود، برابر با K است. برای هر زیر گروه با استفاده از نمره‌های خرده آزمون AT، دو مؤلفه واریانس، $\hat{\sigma}_k^2$ و $\hat{\sigma}_{u,k}^2$، محاسبه می‌شود.

$$\hat{\sigma}_k^2 = \frac{1}{J_k}\sum_{j=1}^{J_k}(Y_j^{(K)} - \bar{Y}^{(K)})^2$$

و

$$\hat{\sigma}_{u,k}^2 = \sum_{i=1}^{m}\hat{P}_i^{(K)}(1-\hat{P}_i^{(K)})$$

که

$$Y_j^{(k)} = \sum_{i=1}^{m} U_{ij}^{(K)}$$

$$\bar{Y}^{(k)} = \frac{1}{J_k}\sum_{j=1}^{J_k} Y_j^{(K)}$$

$$\hat{P}_i^{(k)} = \frac{1}{J_k}\sum_{j=1}^{J_k} U_{ij}^{(K)}$$

درمعادله‌های فوق، U_{ij}^{k} بیانگر پاسخ آزمودنی J از زیر گروه K به i امین سؤال خرده آزمون AT و J_k بیانگر تعداد آزمودنی‌های زیر گروه K است. البته زیر گروه‌هایی را که تعداد افراد آنها بسیار کم است باید حذف کرد و آماره T را براساس بقیه زیرگروه‌ها محاسبه نمود. به هر حال، K بیانگر تعداد زیرگروه‌هایی است که در محاسبه آمارهٔ T مورد استفاده قرار می‌گیرند.

برای هر زیر گروه K، آماره زیر محاسبه می‌شود:

$$\hat{\sigma}_{k}^{۲} - \hat{\sigma}_{u,k}^{۲} = ۲ \sum_{i<l \in AT}^{m} \hat{Cov}(U_i, U_l \mid X_{PT} = K)\, T_{L,K=}$$

که $\hat{Cov}(U_i, U_l \mid X_{PT} = K)$ برآوردی است از کوواریانس بین سؤال‌های u_i و u_l برای آزمودنی‌هایی که نمره آنها در خرده‌آزمون PT برابر با K است. آماره T_L به صورت زیر محاسبه می‌شود (L به معنای طولانی بودن آزمون است):

$$T_L = \frac{\sum_{k=۱}^{k} T_{l,k}}{\sqrt{\sum_{k=۱}^{k} S_k^{۲}}}$$

$S_k^{۲}$ واریانس مجانبی آماره T_L است که به شکل زیر محاسبه می‌شود:

$$J_k S_k^{۲} = \left[(\hat{\mu}_{۴,k} - \hat{\sigma}_{۴}^{k}) + \hat{\delta}_{۴,k} / M^{۴} + ۲\sqrt{(\hat{\mu}_{۴,k} - \hat{\sigma}_{۴}^{k})\hat{\delta}_{۴,k} / M^{۴}} \right]$$

$$\hat{\mu}_{۴,k} = \sum_{j=1}^{J_k} (Y_j^k - \bar{Y}^{(k)})^{۴}\ J_k \Big/$$

$$\hat{\sigma}_{۴,k} = \sum_{i=۱}^{m} \hat{P}_i^{(k)} (۱ - \hat{P}_i^{(k)})(۱ - ۲\hat{P}_i^{k})^{۲}$$

زمانی که حجم نمونه و تعداد سؤالها زیاد باشد آماره T_L از توزیع نرمال استاندارد تبعیت می‌کند. بنابراین اگر مقدار T_L بدست آمده از مقدار Z متناظر با صدک $(۱-\alpha)۱۰۰$ توزیع نرمال، بزرگتر باشد فرضیه صفر مبتنی بر تک بعدی بودن آزمون در سطح α رد می‌شود.

استوت (۱۹۸۷) نشان‌داد که آماره T_L برای آزمون‌های کوتاه (آزمون‌هایی‌با طول ۳۰ سؤال و کمتر) دارای تورش‌مثبت است؛ یعنی سطح واقعی‌خطای نوع‌اول این آماره از سطح اسمی آن بیشتر است. برای رفع این مشکل، استوت (۱۹۸۷) و فرولیچ (۲۰۰۰) راه حل‌هایی ارائه داده‌اند.

در راه حلی که استوت (۱۹۸۷) مطرح ساخت، علاوه بر خرده آزمون AT که در گام اول شکل گرفت و اینک AT_1 نامیده می‌شود، خرده آزمون دیگری با نام AT_2 تشکیل می‌شود (برای آگاهی از نحوه تشکیل AT_2 به استوت، ۱۹۸۷، مراجعه کنید). سپس آماره T_B، (B بیانگر اصلاح برای تورش است)، به همان صورت T_L محاسبه و آماره T استوت به شکل زیر به دست می‌آید:

$$T = \frac{T_L - T_B}{\sqrt{2}}$$

در اینجا نیز اگر مقدار T مشاهده شده از مقدار Z متناظر با صدک $(1-\alpha)100$ توزیع نرمال، بزرگتر باشد فرضیه صفر مبتنی بر تک بعدی بودن آزمون در سطح α رد می‌شود.

در روشی که فرالیچ (۲۰۰۰) مطرح ساخت، تابع سؤال– پاسخ (IRF) سؤال‌های آزمون، تحت فرض تک بعدی بودن با استفاده از روش ناپارامتریک هموارسازی هسته[1]، که توسط نادارایا[2] (۱۹۶۴) و واتسون[3] (۱۹۶۴) ارائه گردید و رامزی[4] (۱۹۹۱) برای اولین بار از آن در برآورد IRF استفاده کرد، برآورد می‌شود. در گام بعد، با استفاده از IRFهای برآورد شده، پاسخ‌های آزمودنی‌ها به سؤال‌های آزمون تولید می‌شود. آنگاه با استفاده از داده‌های تولید شده و خرده آزمونهای AT و PT اولیه، آماره دیگری با نام T_G محاسبه می‌شود، (برای جزئیات بیشتر به فرولیچ، ۲۰۰۰، مراجعه کنید). این فرایند تولید تصادفی داده‌های تک بعدی و محاسبه T_G، N بار تکرار می‌شود، (N معادل با تعداد افراد نمونه است)، و مقدار متوسط T_G، که به صورت $\overline{T_G}$ نشان داده می‌شود، بدست می‌آید. در پایان، مقدار آماره T به شکل زیر محاسبه می‌شود:

1. Kernel-Smoothing
2. Nadaraya
3. Watson
4. Ramsay

$$T = \frac{T_L - \overline{T}_G}{\sqrt{1 + 1/N}}$$

اگر مقدار T بدست آمده از مقدار Z متناظر با صدک $(\alpha - 1)100$ توزیع نرمال، بزرگتر باشد فرضیه صفر رد می‌شود و نتیجه می‌گیریم که سؤال‌های آزمون، تک بعدی نیست. مقادیر کوچک آماره T حاکی از تأیید فرضیه‌صفر و تک‌بعدی بودن سؤال‌ها است.

نتایج برخی از مطالعات نشان می‌دهد که DIMTEST یک روش معتبر برای سنجش تک بعدی بودن است. این روش‌همچنین در شناسایی و کشف چند بعدی بودن سؤال‌ها در مقایسه با سایر روش‌ها از توان آماری فوق العاده‌ای برخوردار است. (هتی و همکاران، ۱۹۹۶؛ نانداکومار، ۱۹۹۳، ۱۹۹۴؛ نانداکومار و استوت، ۱۹۹۳). هتی و همکاران (۱۹۹۶) بررسی جامعی در خصوص عملکرد DIMTEST انجام داده‌اند.

با توجه به اینکه DIMTEST پرکاربردترین روش برای سنجش تک بعدی بودن داده‌ها است (ین و فیتزپاتریک، ۲۰۰۶) و به منظور آشنایی خوانندگان با چگونگی کاربرد این برنامه و تفسیر خروجی حاصل از آن، داده‌های مثال قبل با استفاده از DIMTEST مورد تحلیل قرار گرفت. برای این منظور، ابتدا ماتریس همبستگی‌های تتراکوریک سؤال‌ها با استفاده از برنامه FACTOR (لورنزو- سوا، و فراندو[1]، ۲۰۰۵) مورد تحلیل عاملی اکتشافی قرار گرفت. برای استخراج عامل‌ها از روش کمترین مجذورات بدون وزن[2] (ULS) و برای تعیین تعداد عامل‌های قابل استخراج از روش تحلیل موازی (PA؛[3] تامپسون[4]، ۲۰۰۴) استفاده گردید. ماتریس ساختار عاملی با استفاده از روش Direct Oblimin (کلاین[5]، ۱۹۹۴) چرخش داده شد. براساس ماتریس عاملی چرخش یافته، تعداد ۱۱ سؤال که روی عامل اول، بار بالا و روی عامل‌های دیگر بار ناچیز داشتند به‌عنوان خرده آزمون AT مشخص شدند. پس از مشخص شدن سؤال‌های خرده‌آزمون AT، داده‌ها با استفاده از برنامه DIMTEST تحلیل گردید. خروجی حاصل از این برنامه در جدول ۳

1. Lorenzo-Seva & Ferrando
2. Unweighted Least Squares
3. Parallel Analysis
4. Thompson
5. Kline

ارائه شده است. مقدار مشاهده شدهٔ آماره T برابر با ۱/۹۰ است که از مقدار بحرانی Z در سطح ۰/۰۵ = α، (۱/۶۴)، بالاتر بوده و در سطح ۰/۰۲۸= α به لحاظ آماری معنادار است.

جدول ۳. خروجی برنامه DIMTEST

```
---------------------------------------------------
DIMTEST SUMMARY OUTPUT
---------------------------------------------------
Original Data Set:  D:\ Math..dat
Number of Items Used:       30
Number of Examinees Used to
Calculate DIMTEST Statistic:      2000
Minimum Cell Size for
Calculating DIMTEST Statistic:       2
Number of Examinees After
Deleting Sparse Cells:     1936
Proportion of Examinees Used to
Calculate DIMTEST Statistic:    0.9680
Number of Simulations Used
to Calculate TGbar:      100
Randomization Seed:     99991
Estimate of Examinee
Guessing on Test:    0.0000
---------------------------------------------------
AT List     PT List
---------------------------------------------------
2        1   4   5   6   7  12
3       14  15  16  17  18  19
8       20  22  25  26  27  28
9       29
10
11
13
21
23
24
30
---------------------------------------------------
---------------------------------------------------
TL=sum(TL,k)/sqrt(sum(S2,k)) {using original data}
TG=sum(TL,k)/sqrt(sum(S2,k)) {using simulated data}
TGbar = mean of ** TGs
T=(TL-TGbar)/sqrt(1+1/**)
---------------------------------------------------
DIMTEST STATISTIC
---------------------------------------------------
TL       TGbar       T       p-value
---------------------------------------------------
6.9285     5.0143     1.9047     0.0284
```

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

اگرچه این یافته با نتیجه حاصل از برنامه NOHARM همخوانی ندارد لکن، اولاً مقدار مشاهده شده T تفاوت زیادی با مقدار بحرانی ندارد و ثانیاً تحلیلی که با استفاده از برنامه DETECT (کیم، ۱۹۹۴؛ استوت و همکاران ۱۹۹۶) صورت گرفت نشان داد که میزان چندبعدی بودن داده‌ها ضعیف است. مقدار مشاهده شده آماره DETECT برابر با ۰/۲۱ است که براساس نظر روسوس و اُزبک (۲۰۰۶) بیانگر چندبعدی بودن ضعیف تا متوسط است. بنابراین با عنایت به این مطلب و با در نظر گرفتن اصل ایجاز، می‌توان ادعا کرد که آزمون مورد بحث اساساً تک بعدی است و کاربرد مدلهای تک بعدی IRT برای مقیاس پردازی این سؤال‌ها مجاز است.

در مواردی که مفروضه تک بعدی بودن اساسی، برقرار نباشد بایستی یا از نظریه بسته بندی سؤال[1] و یا از مدل‌های چندبعدی جبرانی استفاده کرد و یا آزمون را به چند خرده آزمون که اساساً تک بعدی هستند تقسیم کرد(نانداکومار، ۱۹۹۱ ، ص ۱۱۲).

روش‌های ناپارامتریک دیگری نیز برای بررسی مفروضه تک بعدی بودن ارائه شده است. از جمله این روش‌ها می‌توان به روش‌هالند و روزنبام(۱۹۸۶)، که به روش H&R معروف است،اشاره کرد. هالند و روزنبام[2] (۱۹۸۶) برای سنجش استقلال شرطی میان جفت سؤال‌ها رویکردی را ارائه می‌دهند که بر مدلهای IRT مبتنی نیست. روش آنها بر کارهالند (۱۹۸۸) و روزنبام (۱۹۸۴) استوار است. رابطه شرطی هر جفت سؤال با استفاده از آماره منتل- هانسزل(۱۹۵۹) مورد آزمون قرارمی‌گیرد. در خصوص کارآیی این روش، تحقیقات بسیار کمی صورت‌گرفته است. برای آشنایی با این روش به ایپ[3] (۲۰۰۱) مراجعه کنید.

تحلیل مقیاس ماکن[4] (MSA؛ ماکن، ۱۹۷۱؛ ۱۹۹۷؛ مولنار، و سی جتسما[5]، ۲۰۰۰؛ همچنین به همکر[6]، سی جتسما و مولنار، ۱۹۹۵؛ سی جتسما، ۱۹۹۸، نگاه کنید) که از طریق

1. Theory of Testlet
2. Holland & Rosebaum
3. IP
4. Mokken Scale Analysis
5. Molenaar & Sijtsma
6. Hemker

طریق برنامه MSP (مولنار، و سی جتسما، ۲۰۰۰) قابل‌اجرا است، روش ناپارامتریک دیگری است که برای بررسی مفروضه ابعاد مورد استفاده قرار می‌گیرد.

بحث و نتیجه‌گیری

به‌خاطر گسترش روز افزون کاربرد مدلهای IRT در سنجش‌های روانی- تربیتی، استفاده نادرست از این مدل‌ها می‌تواند به خطاهای آماری منجر گردد و در نتیجه در سنجش تفاوت‌های فردی و دستیابی به اهداف آموزشی متنوع، خطاهای‌جدی به‌وجود آورد (ناندا‌کومار، ۱۹۹۴). یکی از مفروضه‌های اصلی مدل‌های رایج و پرکاربرد IRT تک بعدی بودن داده‌ها است. به‌عبارت دیگر در این مدل‌ها فرض بر این است که برای تبیین واریانس مشترک میان سؤال‌ها صرفاً یک صفت مکنون کافی است (امبرتسون و ریس، ۲۰۰۰).

دیگر مفروضه مهم این مدل‌ها که در راستای مفروضه ابعاد مطرح می‌شود، مفروضه استقلال موضعی است. استقلال موضعی به این معنا است که با کنترل سطح توانایی آزمودنی‌ها، سؤال‌ها مستقل از یکدیگر بوده و همبستگی ندارند. به بیان دیگر، این مفروضه به آن معنا است که احتمال پاسخ درست به یک سؤال، صرفاً به سطح توانایی آزمودنی بستگی دارد و نه عوامل دیگر (همبلتون، سوامیناتان و راجرز، ۱۹۸۵؛ همبلتون، ۱۹۸۹).

برای سنجش‌ابعاد داده‌های تستی روش‌های‌آماری زیادی ارائه شده است که اکثر آنها یا ارتباطی به بحث ابعاد ندارند و یا ناکارآمد هستند (امبرتسون و ریس، ۲۰۰۰؛ هتی، ۱۹۸۴، ۱۹۸۵). هتی(۱۹۸۴) بیان می‌دارد که بسیاری از روش‌های مورد استفاده در ارزیابی ابعاد داده‌های تستی، دارای مشکلات زیر هستند.

۱- توزیع نمونه‌گیری آنها نامعلوم است؛ ۲- عملکرد آنها در شرایط‌مختلف نامشخص است؛ ۳- فاقد منطق و زیربنای نظری هستند؛ ۴- نمی‌توانند به روشنی بین داده‌های تک بعدی و دو بعدی تمایز قائل شوند.

یافته‌های حاصل از پژوهش‌ها نشان می‌دهد که NOHARM، TESTFACT، DIMTEST، DETECT، H&R و MSP روش‌های مناسب و خوش آتیه‌ای برای سنجش ابعاد آزمون‌ها هستند (هتی، ۱۹۸۴، ۱۹۸۵؛ استوت، ۱۹۸۷؛ ناندا‌کومار، ۱۹۹۶، ۱۹۹۴؛ وان

آبسود[1] و همکاران، ۲۰۰۴ ؛ وان شور[2]، ۲۰۰۳؛ ویس میجر[3] و همکاران، ۲۰۰۸). البته تحقیق در زمینه این روش‌ها هنوز ادامه دارد.

در این مقاله، ضمن مرور ادبیات مربوط به مفروضه ابعاد در مدل‌های دو ارزشی IRT و معرفی برخی از روش‌های سنجش تک بعدی بودن، دو روش NOHARM و DIMTEST در مورد یک مجموعه از داده‌های تستی بکار رفت و هر دو روش نتیجه مشابهی بدست دادند.

به‌عنوان نکته پایانی باید گفت که ما نیز هماهنگ و همسو با امبرتسون و ریس(۲۰۰۰) معتقدیم زمان آن فرا رسیده است که پژوهشگران از کاربرد شاخص‌های اکتشافی مانند «درصد واریانس تبیین شده توسط عامل اول» یا «نسبت ارزش ویژه عامل اول به عامل دوم» به‌عنوان شاخصی برای تک بعدی بودن اجتناب کنند و به استفاده از روش‌های جدید مانند NOHARM، TESTFACT، DIMTEST، DETECT، H&R و MSP رو بیاورند.

منابع

Ackerman, T, A.(1989).Unidimentional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement,12(2),* 112-127.

Bartholomew, D. J.(1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B, 42,* 293-321.

Bock, R. D., Gibbons, R.,& Muraki, E.(1988). Full- information factor analysis. *Applied Psychological Measurement, 26(4),* 261-280.

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.

Cortina, J. M.(1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104.

De Ayala, R. J., & Hertzog, M. A.(1989). A comparison of methods for assessing dimensionakity for use in item response theory. Paper presented at the annual meeting of the National Council on the Measurement in Education, San Francisco.

1. Van Abswoude
2. Van Schuur
3. Wismeijer

De Champlain, A. F., & Gessaroli, M. F.(1998). Assessing the dimensionality of item response matrices with small sample size and short test lenghts. *Applied Measurement in Education, 11,* 231-253.

Embretson, S. E., & Reise, S. P.(2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Ferguson, G. A.(1941). The factorial interpretation of test difficulty. *Psychometrika, 6,* 323-329.

Finch, H., & Habing, B.(2005). Comparison of NOHARM and DETECT in item cluster recovery: Counting dimensions and allocating items. *Journal of Educational Measurement, 42(2),* 149-170.

Fraser, C.(1988). NOHARM: An IBM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory [computer software]. Armidale, Australia: The University of New England.

Freeman, F. S.(1962). *Theory and Ppractice of Psychological Testing (3rd ed.)*. New York: Henry Holt.

Froelich, A. G.(2000). *Assessing Unidimensionality of Test Items and Some Asymptotics of Parametric Item Response Theory*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.

Gage, N. L., & Darmin, D. E.(1950). Reliability, homogeneity, and number of choices. *Journal of Educational Psychology, 41,* 385-404.

Gessaroli, M. E., & De Champlain, A. F.(1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement, 33(2),* 157-179.

Gessaroli, M. E., & De Champlain, A. F.(2005). Assessment of test dimensionality. In

Goldstein, H.(1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology, 33,* 234-246.

Goldstein, H., & Wood, R.(1989). Five decades of item response theory. *British Journal of Mathematical and Statistical Psychology, 42,* 139-167.

Green, S. B., Lissitz, R. W. & Mulaik, S. A.(1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-838.

Guilford, J. P.(1965). *Fundamental Statistics in Psychology and Education(4th ed.)*. New York: McGraw-Hill

Hambleton, R. K.(1989). Principles and selected applications of item response theory. In R. L. Linn(Ed.), *Educational Measurement* (3rd ed., pp. 147-200). New York: Macmillan.

Hambleton, R. K. & Rovinelli, R. J.(1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10(3),* 287-302.

Hambleton, R. K. & Swaminathan, H.(1985). *Item Response Theory: Principles and Applications.* Boston: Kluwer-Nijhoff Publishing.

Hambleton, R. K. & bTraub, R. E.(1973). Analysis of emprical data using two logistic latent trait models. *British Journal of Mathematical and Statistical Psychology,24,* 273-281.

Hattie, J.(1984). An emprical study of various indices for determining unidimensionality. *Multivariate Behavioral research, 19,* 49-78.

Hattie, J.(1985). Methodology review: Assessing unidimensionality if tests and items. *Applied Psychological Measurement, 9*, 139-164.

Hattir, J., Krakowski, K., Rogers, H. J. & Swaminathan, H.(1996). An assessment of Stout's index of essential dimensionality. *Applied Psychological Measurement, 20*, 1-14.

Hemker, B. T., Sijtsma, K. & Molenaar, I. W.(1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19*, 337-352.

Holland, P.W.(1981). When are item response models consistent with observed data? *Psychometrika, 46*, 79-92.

Holland, P. W. & Rosenbaum, P. R.(1986).Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14*, 1523-1543.

Horst, J. L.(1953). Correcting the Kuder-Richardson reliability for dispersion of item difficulties. *Psychological Bulletin, 50*, 371-374.

Ip, E. H.(2001). Testing for local dependency in dichotomous and polytomous item response models. *Psychometrika, 66*, 109-132.

Kelderman, H. & Rijkes, C. P.(1994). Loglinear multidimensional IRT models for polytomous scored items. *Psychometrika, 59(2)*, 149-176.

Kim, H. R.(1994). New techniques for the dimensionality assessment of standardized test data. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Kline, P.(1994). *An Eeasy Guide to Factor Analysis.*

Knol, D. L. & Berger, P. F.(1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

Lord, F. M. & Novick, M. R.(1968). *Statistical Theories of Mmental Test Scores.* Reading MA: Addison-Wesley.

Lorenzo-Seva, U. & Ferrando, P.J.(2005). FACTOR: A computer program to fit the exploratory factor analysis models. *Behavior Research Methods, 38*, 88-91.

Mantel, N. & Haenszel, W.(1959). Statistical aspects of the retrospective study of disease. *Journal of the National Cancer Institute,22*,719-748.

McDonald, R. P.(1967). Nonlinear factor analysis. *Psychometric Monographs* (No. 15).

McDonald, R. P.(1979). The structural analysis of multivariate data: a sketch of general theory. *Multivariate Behavioral Research, 14*, 21-38.

McDonald, R. P.(1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.

McDonald, R. P.(1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6*, 379-396.

McDonald, R. P.(1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleto(Ed.), *Handbook of Modern Item Response Theory*(pp. 258-269). New York: Springer Verlag.

McLeod, L. D., Swygert, K. A. & Thissen, D.(2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer.(Ed.), *Test Scoring.* Mahwah, NJ: Lawrence Erlbaum.

Mislevy, R. J.(1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics,11*, 3-31.

McNemar, Q.(1946). Opinion-attitude methodology. *Psychological Bulletin, 43*, 289-374.

Mokken, R. J.(1971). A theory and procedure of scale analysis with applications in political research. Berlin, Germany: de Gruyter.

Mokken, R. J.(1997). Nonparametric models for dichotomous response. In W. J. van der Linden & R. K. Hambleton(Eds.), *Handbook of Modern Item Response Theory* (pp. 351-367). New York: Springer-Verlag.

Molenaar, I. W. & Sijtsma, K.(2000). *User's Manual MSP5 for Windows*. Griningen, The Netherland: iecProGAMMA.

Nadaraya, E. A.(1964). On estimating regression. *Theory of Probability and its Applications, 10*, 186-190.

Nandakumar, R.(1991). Traditional dimensionality versus essesntial dimensionality. *Journal of Educational Measurement, 28(2)*, 99-117.

Nandakumar, R.(1994). Assessing dimensionality of a set of item responses-Comparison of different approaches. *Journal of Educational Measurement,31(1)*, 41-68.

Nandakumar, R. & Ackerman, T.(2004). Test modeling. In D. Kaplan(ed.), *The Sage Handbook of Quantitative Methodology for Tthe Social Ssciences.*Thousand Oaks, CA: Sage.

Nandakumar, R. & Stout, W.(1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics, 18(1)*, 41-68.

Panter, A. T., Swygert, K. A. & Dahlstrom, W. G.(1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment,68(3)*, 561-589.

Ramsay, J. O.(1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation, *Psychometrika, 56*, 611-630.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tets: Results and implications. *Journal of Educational Statistics,4*, 207-230.

Reckase, M. D.(1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25-36.

Rosenbaum, P. R.(1984). Testing the conditional independence and monotonicity assumption of item response trheory. *Psychometrika, 49(3)*, 425-435.

Roussos, L. A., & Ozbeck, O. Y.(2006). Formulation of DETECT population parameter and evaluation of DETECT estimator bias. *Journal of Educational Measurement, 43(3)*, 215-243.

Roussos, L. A., Stout, W. F. & Marden, J. I.(1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35(1)*, 1-30.

Schmitt, N.(1995).Uses and abuses of coefficient alpha. *Psychological Assessment,8(4)*, 350-353.

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement,22(1)*, 3-31.

Stout, W.(1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52(4)*, 589-617.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55(2)*, 293-325.

Tanaks, J. S.(1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage

Thompson, B.(2004). *Exploratory And Confirmatory Factor Analysis: Understanding Concepts And Applications*. Washington, DC: American Psychological association.

Yen, W. M.(1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement,8(2),* 125-145.

Yen, W. M. & Fitzpatrick, A. P.(2006). Item reponse theory. In R. L. Brennan(Ed.), *Educational Measurement* (4th ed.). Westport, CT: Greenwood.

Van Abswoude, A. A. H., Van der Ark, L. A. & Sijtsma, K.(2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28(1),* 3-24.

Van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis, 11,* 139-163.

Wismeujer, A. A. J., Sijtsma, K., Van Assen, M. A. L. M. & Vingerhoets, A. J. J. M. (2008). A comparative study of the dimensionality of the self-concealment scale using principal component analysis and Mokken scale analysis. *Journal of Personality Assessment,90(4),* 323-334.

Wilson, D. T., Wood, R. & Gibbons, R. (1991). *TESTFACT computer manual*. Chicago Scientific Software.

Zhang, J. & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213-249.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement,24,* 293-308.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی