فصلنامه اندازه‌گیری تربیتی
دانشگاه علامه طباطبائی
سال نهم، شماره ۳۶، تابستان ۹۸، ص ۱۳۷ تا ۱۶۶

Quarterly of Educational Measurement
Allameh Tabataba'i University
Vol. 9, No. 36, Summer 2019

کاربرد مدل طبقه پنهان چندبعدی مبتنی بر نظریه پاسخ سوال (LCMIRT) درمطالعه ابعاد شناختی و خوشه بندی سؤال‌های آزمون ریاضی: مطالعه موردی درس ریاضی کنکور سراسری ریاضی فیزیک

مسعود گرامی‌پور[۱]، مریم مقدسین[۲]، ریحانه رضازاده[۳]

تاریخ دریافت: ۹۸/۰۱/۱۳
تاریخ پذیرش: ۹۸/۰۴/۰۱

چکیده

مهارت ریاضی شامل حوزه‌های محتوایی و فرایندهای شناختی مختلف است که بیانگر پیچیدگی توانایی ریاضی و ویژگی‌های پنهان آن است. تاکنون مطالعه این پیچیدگی‌ها با روش‌های سنتی تحلیل داده‌ها یا به صورت ذهنی انجام شده است. بنابراین، پژوهش حاضر با استفاده از مدل چندبعدی طبقه پنهان پاسخ سوال (LCMIRT) به مطالعه ابعاد شناختی و متغیرهای پنهان درس ریاضی آزمون سراسری ورود به دانشگاه پرداخت. در این راستا، داده‌های درس ریاضی آزمون سراسری سال‌های ۸۷، ۹۰ و ۹۴ مورد مطالعه قرار گرفت. یافته‌ها نشان داد که آزمون ریاضی، به عنوان یکی از دروس سرنوشت ساز در آزمون سراسری ورود به دانشگاه، شامل مجموعه ای از ویژگی‌های شناختی چندبعدی است. نتایج روش تحلیل موازی وحدت یافته نشان داد که آزمون‌ها با مدل تک بعدی برازش ندارند و اضافه کردن ابعاد بیشتر، برازش مدل را به صورت معنی داری بهبود می‌بخشد. بعلاوه، حوزه‌های شناختی درک، حل مسئله و استدلال سه سازه اساسی در تبیین توانایی ریاضی شناسایی شدند که در خوشه بندی و تحلیل سؤال‌های آزمون، اطلاعات دقیق تری از کیفیت سوال‌ها به دست می‌دهند. این ویژگی LCMIRT، در مقایسه با سایر رویکردها، تعداد زیر خوشه‌های شناختی را افزایش می‌دهد. در نهایت توصیه می‌شود که در ساخت و تحلیل آزمون‌های

---

۱. استادیار، تحقیقات آموزشی، دانشگاه خوارزمی، تهران، ایران (نویسنده مسئول) mgramipour@yahoo.com
۲. استادیار، تحقیقات آموزشی، دانشگاه خوارزمی، تهران، ایران.
۳. کارشناسی ارشد، تحقیقات آموزشی، دانشگاه خوارزمی، تهران، ایران.

روانی و تربیتی، مدل‌های LCMIRT مد نظر قرار گیرند تا آشکارسازی توانایی‌های شناختی پنهان آزمون‌ها از اعتبار بیشتری برخوردار باشد.

واژگان کلیدی: ابعاد شناختی، مدل چندبعدی طبقه پنهان پاسخ سوال، خوشه بندی سؤال‌های آزمون، ریاضیات، آزمون‌های سرنوشت ساز.

مقدمه

در عصر پیشرفت و تحول در زمینه‌های مختلف ارزیابی که آزمون‌ها به جزئی از فرایندهای هر نوع آموزشی تبدیل‌شده است و گاه به‌عنوان معیاری برای ورود به مشاغل و مقاطع تحصیلی، تأیید تسلط بر مهارتی خاص و ارزیابی دوره‌های مختلف آموزشی استفاده می‌شود، نقش اعتبار[1] آزمون بسیار مهم جلوه می‌کند و سنجش اعتبار به عنوان مشخصه ای از کیفیت در قلب نظریات ارزیابی آموزشی و روانی قرار می‌گیرد (کورتز[2]، ۲۰۰۸: ۲۱۵). بحث ابعاد آزمون با اعتبار ارتباط تنگاتنگی دارد بر اساس چارچوبی که مسیک[3] معرفی کرده است، ابعاد آزمون به عنوان قوی ترین نوع شاهد در حمایت از اعتبارِسازه[4] نام برده می‌شود. اهمیت متغیر (های) پنهان[5] (مکنون)، به استفاده از مدل‌های چندبعدیِ پاسخ سؤال[6] در حوزه‌های مختلف منجر می‌گردد که تاکنون روش‌های تحلیل و آزمون سازی با پیش فرض‌های نظریه کلاسیک که مفهوم بعد به صورت مستقیم مورد بحث نبوده و یا با به کاربردن تحلیل عاملی، انجام می‌گرفته با گذشت زمان و گسترش دانش در حوزه‌های مختلف و در حوزه آزمون با روی کارآمدن مدل‌ها و نظریه‌ها، آزمون‌ها نیز تمایل بیشتری به‌اندازه گیری بیش از یک توانایی از خود نشان دادند (ریکِیس، ۲۰۰۹: ۵۳) .در نظریه پاسخ سؤال (لُرد[7]، ۱۹۸۰) فرض بر گسسته بودن مشاهدات و در حالت چند وجهی[8] فرض بر ترتیب است و بیشتر مدل‌های تعریف شده این نظریه بر توزیع نرمال توانایی پنهان تاکید

1. Validity
2. Koretz
3. Messick
4. Construct validity
5. Latent variable
6. Multidimensional item response theory
7. Lord
8. polytomous

دارند. در مقابل تحلیل طبقه پنهان و مدل‌های مرتبط متغیرهای قابل مشاهده و پنهان گسسته در نظر گرفته می‌شوند

(لازارسفلد و هنری[1]، ۱۹۶۸). با توجه پژوهش حاضر که در حوزه مدل‌های چند متغیری است، لازم به ذکر است که این مدل شباهت بسیاری به مدل‌های شناختی تشخیصی[2] و مدل‌های طبقه پنهان چند بعدی پاسخ سؤال دارند. با این تفاوت که مدل‌های چند بعدی بیشتر به صورت اکتشافی مورد بحث قرار می‌گیرند، در حالی که مدل‌های شناختی تشخیصی حالتی تاییدی دارند. به علاوه مدل‌های شناختی تشخیصی و مدل‌های طبقه پنهان چندبعدی پاسخ سؤال، توانایی‌های زیربنایی را به صورت طبقه پنهان در نظر می‌گیرند که یکی از تفاوت‌های عمده و حائز اهمیت آن با مدل‌های چند بعدی پیشین است که در آن توانایی پنهان به صورت پیوسته فرض می‌شود (گو[3]، ۲۰۱۱).

مفهوم بعد و مطالعه آن: در زیرمجموعه اعتبار بر اساس شواهد، پنج نوع شاهد تجربی در حمایت از اعتبار سازه آورده شد که مهارت در آزمون[4] و ابعاد آزمون[5] قوی‌ترین نوع شواهد در حمایت از اعتبار سازه بودند. تعریف ابعاد آزمون، بر اساس الگوی مدل‌سازی صورت می‌گیرد. (به‌عنوان‌مثال تحلیل عاملی یا نظریه پاسخ سؤال). درگذشته، تعریف ابعاد آزمون به دنبال پاسخ به این پرسش بود که چند متغیر پنهان زیربنای سؤال‌های آزمون را تشکیل می‌دهند و مایل به تبیین واریانس و کوواریانس سؤال‌های یک آزمون بوده است. تعیین تعداد ابعاد آزمون همیشه یک موضوع مهم و قابل‌بحث بوده است (اَکرمَن[6]، ۱۹۹۶). یک روش بسیار معمول استفاده از شکل اِسکری مقادیر ویژه به‌دست‌آمده از تحلیل مؤلفه‌های اصلی است (ریکیس، ۱۹۷۹). در این روش تصمیم‌گیری درباره این‌که آیا مقادیر ویژه دوم (و احتمالاً سوم) به‌اندازه کافی بزرگ هستند که به‌عنوان ابعاد دیگر مورد توجه قرار گیرند یا نه همواره مشکل اساسی بوده است. هورن[7] (۱۹۶۵) و دارسگو و لیساک[8] (۱۹۸۳) پیشنهاد داده‌اند که تفسیر درباره مقادیر ویژه با مقایسه شکل اسکری

1. Lazarsfeld and Henry
2. Cognitive diagnosis models
3. Gu
4. Test proficiency
5. Test dimensionality
6. Ackerman
7. Horn
8. Drasgow & Lissak

داده‌های واقعی با شکل حاصل از تحلیل عاملی داده‌های تولید شده تصادفی با همان تعداد سؤال بهبود پیدا می‌کند. روش دیگر برای ارزیابی ابعاد، بررسی فرض استقلال موضعی با استفاده از ماتریس کوواریانس برای داوطلبان در فواصل مختلف پیوستار خصیصه مورد مطالعه است (روزنوفسکی[1] و همکاران، ۱۹۹۱). همچنین استوت[2] یک برنامه کامپیوتری به نام DIMTEST (استوت، ۱۹۸۷) ارائه داد که به کاربران آزمون امکان استفاده از نمونه‌های بزرگ و آزمون‌های آماری را برای تعیین و تمیز خوشه‌های سؤال را می‌دهد. در همین راستا کیم و استوت[3] (۱۹۹۴) روشی آماری بر اساس مقادیر کوواریانس شرطی برای شناسایی ابعاد داده‌های تولیدشده با ساختارهای مختلف همبستگی بین ابعاد صفت پدید آوردند.

زمانی که ابعاد غالب به لحاظ آماری مورد تأیید قرار گرفتند، کاربران آزمون می‌توانند از برنامه‌های مختلف نظریه پاسخ سؤال چندبعدی مانند: NOHARM (فریزر و مکدونالد[4]، ۱۹۸۸) متناسب با مدل اجیو نرمال[5] با روش حداقل مربعات که قادر به برآورد پارامتر تمیز و دشواری سؤال‌ها و TESTFACT (ویلسون[6] و همکاران، ۱۹۸۷) برای تخمین پارامترها اقدام کنند.

در گذشته استفاده از تحلیل عاملی خطی برای ارزیابی ابعاد فضای پنهان زیربنایی مجموعه‌ای از سؤال‌ها مشکلات فنی و روش‌شناختی به وجود می‌آورد. به عنوان مثال سطح دشواری سؤال‌ها و ضریب حدس سؤال در آزمون‌های چندگزینه‌ای که نقش اساسی در ساختار عاملی دارند در تحلیل عاملی خطی نادیده گرفته می‌شد (دوبراکاوا و روی، ۲۰۱۴؛ کارول[7]، ۱۹۴۵؛هالین[8] و همکاران، ۱۹۸۳).زوویک[9]، ۱۹۸۷به تبع این مشکلات، تلاش‌های بسیار زیادی در توسعه و مقایسه روش‌های ارزیابی ابعاد توسط پژوهشگران صورت گرفته است به عنوان مثال: بررسی روش‌های تعیین ابعاد مجموعه‌ای از سؤالات برای داده‌های

1. Roznowski
2. Stout
3. Kim & Stout
4. Fraser and McDonald
5. Normal ogive model
6. Wilson
7. Carroll
8. Hulin
9. Zwick

واقعی و شبیه‌سازی‌شده (همبلتون و روواینلی[1]، ۱۹۸۶)، بررسی و مقایسه ابعاد غالب برای مجموعه‌ای از سؤالات (نانداکومار[2]، ۱۹۹۴) و معرفی روش‌های تعیین بعد و بررسی ابعاد ماتریس پاسخ سؤال برای آزمون‌هایی با حجم نمونه کوچک و طول آزمون کوتاه انجام شد (دو چاپلین و جسرولی[3]، ۱۹۹۸).

در پی تحلیل عاملی مدل‌های رایج و پرکاربرد نظریه پاسخ سؤال نیز که علاوه بر مفروضه‌های اصلی استقلال موضعی و تکنوایی بر تک‌بعدی بودن که به‌منظور برآورد استقلال موضعی تعریف می‌شود (لرد و نویک ۱۹۶۸) برای تفسیر روشن و بدون ابهام نمره‌ها، انجام مقایسه‌های بین فردی و همچنین برای سایر فعالیت‌های روان‌سنجی مانند هم‌تراز سازی مهم است روی کارآمدند. علی‌رغم اهمیت مفروضه ابعاد برای مدل‌های پاسخ سؤال، در خصوص تعریف تک‌بعدی بودن و روش‌های سنجش آن بین متخصصان روان‌سنجی، توافق وجود ندارد.

آزمون‌های استاندارد، مانند آزمون کنکور که به‌عنوان مهم‌ترین شاخص ارزیابی دوره متوسطه و یکی از تعیین‌کننده‌ترین عوامل برای ورود به دانشگاه‌ها بوده است مجموعه‌ای از ویژگی‌های پنهان در ساختار، نهفته است و فرض ساختاری تک‌بعدی برای دستاوردهای آموزشی نامناسب جلوه می‌کند (تراب ولام[4]، ۱۹۸۵: ۲۲).

نظریه پاسخ سؤال چندبعدی یک آرمان گری از واقعیت است که به عنوان حالت خاصی از نظریه پاسخ سؤال تک بعدی مدلی ریاضی است که بردارها به عنوان پارامترهای آن فرض می‌شوند. این پارامترها، بردار ویژگی‌های مختلف فرد که توصیف توانایی‌ها و دانش فرد در آزمون است و بردار ویژگی‌های سؤال که به توصیف دشواری سؤال و ضریب تمییز می‌پردازد است. به کارگیری این نظریه تعیین جنبه ساختاری (عملکرد سؤالات آزمون) و ضمنی (بردار توصیف فرد) سازه را امکان پذیر می‌سازد (همبلتون و سوامینگتون[5]، ۱۹۸۵).

1. Hambleton and Rovinelli
2. Nandakumar
3. De Champlain and Gessaroli
4. Traub and Lam
5. Hambleton and Swamination

ازآنجاکه بسیاری از آزمون‌های آموزشی و روانی ذاتاً چندبعدی هستند و دو یا چندسازه یا بعد را اندازه‌گیری می‌کند، علاوه براین پژوهش‌های علوم شناختی به عنوان مثال (فردریکسُن[1] و همکاران، ۱۹۹۰) نشان داده شده است که برای پاسخ به سؤالات یک آزمون مهارت‌های مختلفی مورد نیاز است.

از این رو مفهوم ارزیابی بعد به‌عنوان بررسی فرض"تک‌بعدی بودن "با روی کار آمدن مدل‌های چندبعدی کمرنگ شد. همچنین پژوهش‌های متعدد نشان داد که استفاده از مدل‌های تک‌بعدی برای سازه‌های چندبعدی سلسله‌ای از مشکلات را در کاربرد و تفسیر نتایج به وجود خواهد آورد (اَکرمن، ۱۹۸۹؛ بردلو[2] و همکاران، ۱۹۹۹؛ دوچاپلین، ۱۹۹۶؛ شِلی و استوت[3]، ۱۹۹۳؛ یِن[4]، ۱۹۹۳؛ چن و تیسن[5]، ۱۹۹۷؛ سیرسی و همکاران، ۱۹۹۱؛ وَن دِر لیندن[6]،۱۹۹۶؛ وِی[7] و همکاران، ۱۹۸۸؛ واکر و بِرتواس[8]، ۲۰۰۳).

نظَریه طبقه پنهان چند بعدی و مدل پاسخ سوال

نامگذاری روش‌ها بر اساس پیشینه تاریخی و اهداف مرتبط با آن و مفروضه‌ها و چارچوب‌های برآورد صورت گرفته است، با گسترش چارچوب‌های پیشرفته مدل سازی سیر تاریخی در پرده ای از ابهام رفته است.

جدول۱. طبقه بندی چهارگانه متغیر پنهان بر اساس مقیاس اندازه گیری

متغیرهای پنهان (مکنون)		
طبقه ای	پیوسته	متغیرهای آشکار (مشاهده شده)
تحلیل مشخصات نهفته[1]	تحلیل عاملی	پیوسته
تحلیل طبقه (کلاس) پنهان[1]	تحلیل ویژگی نهفته[1]	طبقه ای[1]

به عنوان مثال در جدول فوق طبقه بندی چهارگانه (بارتولُمئو و نات، ۱۹۹۹) متغیرهای پنهان بر اساس مقیاس اندازه گیری متغیرهای مشاهده شده و مکنون ارائه

1. Frederikson
2. Bradlow
3. Shealy and Stout
4. Yen
5. Chen and Thissen
6. Van der Linden
7. Way
8. Walker and Beretvas

شده است. تحلیل عاملی، تحلیل ویژگی نهفته (LT)، تجزیه و تحلیل مشخصات نهفته (LP) و تجزیه و تحلیل کلاس (طبقه) پنهان (LC)

جدول۲. طبقه بندی از محبوب ترین مدل‌های روان سنجی (بروکس و همکاران، ۲۰۱۱: ۵۳۳)

	متغیر (های) پنهان			
متغیرهای قابل مشاهده	گسسته		پیوسته	
	چند متغیره	تک متغیره	چند متغیره	تک متغیره
دو وجهی	شبکه‌های بیزی مدل‌های سنجش شناختی	تحلیل طبقه (کلاس) پنهان	نظریه پاسخ سؤال چند بعدی	نظریه پاسخ سؤال
چند و جهی بدون ترتیب	شبکه‌های بیزی مدل‌های سنجش شناختی	تحلیل طبقه (کلاس) پنهان		نظریه پاسخ سؤال
چند وجهی با ترتیب	شبکه‌های بیزی مدل‌های سنجش شناختی	تحلیل طبقه (کلاس) پنهان	نظریه پاسخ سؤال چند بعدی	نظریه پاسخ سؤال
معمولی			تحلیل عاملی مدل یابی معادلات ساختاری	تحلیل عاملی مدل یابی معادلات ساختاری

با گذشت زمان و اهمیت متغیرهای گسسته، طبقه بندی از محبوب ترین مدل‌های روان سنجی مورد توجه متخصصین قرارگرفت. نظریه پاسخ سؤال (لُرد، ۱۹۸۰) فرض بر گسسته بودن مشاهدات و در حالت چند وجهی[1] فرض بر ترتیب است، به علاوه بیشتر مدل‌های تعریف شده این نظریه بر توزیع نرمال توانایی پنهان تاکید دارند.

مدل عمومی‌نظریه پاسخ سؤال برای احتمال شرطی پاسخ y با توجه به u کلاس پنهان:

$$g[P(Y_{ij} = y|U_i = \xi_u)] = \gamma_j \left(\sum_{d=1}^{S} \delta_{jd} \xi_{ud} - \beta_{jy} \right)$$

1. polytomous

فرمول بندی مختلفی با توجه به مشخصات زیر به دست می‌آید:

۱) نوع تابع ربط[1] (مشخصات تابع ربط (.) g به ماهیت سؤال‌ها بستگی دارد).

با توجه به پژوهش که بر داده‌های دو ارزشی (سؤال‌هایی که به صورت صحیح یا غلط نمره گذاری می‌شوند) تمرکز دارد. ازتابع رابط دودویی[2] استفاده شد:

$$g[P(Y_{ij}=y|U_i=\xi_u)]=\log\frac{P(Y_{ij}=1|U_i=\xi_u)}{P(Y_{ij}=0|U_i=\xi_u)}$$

۲) محدودیت بر پارامترضریب تمیز[3] γ_j.

۳) فرمول بندی پارامتر سختی سؤال[4] β_{jy}.

نشانه گذاری

S: تعداد صفات نهفته مختلف اندازه گیری شده توسط J سؤال.

U_i : $U_i = (U_{1i} \ldots . U_{si})'$ بردار متغیرهای پنهان مربوط به ویژگی‌های پنهان .

u_i : $u_i = (u_{1i} \ldots . u_{si})'$ یک تحقق احتمالی از U_i.

ξ_u : $\xi_u = \{\xi_{ud}\}$ بردار ارزش‌ها (نقاط حمایت) فرض شده توسط U_i، با $u=1, \ldots, k$ و $d=1, \ldots, s$

π_u : $\pi_u = P(U_i = \xi_u)$

δ_{jd} : متغیر ساختگی[5] برابر ۱ در صورتی که سؤال jام ویژگی پنهان از نوع d را اندازه بگیرد و در غیر این صورت برابر صفر. (d=1,…, s)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Link function
2. Binary
3. Discriminating parameters
4. Difficulty parameters
5. Dummy variable

جدول۳. فرمول بندی‌های مختلف با توجه به وجود و عدم وجود محدودیت بر شاخص تمیز و سطوح دشواری (باچی و همکاران، ۲۰۱۴)

شاخص تمیز	سطح دشواری	$\#\ \pi_u$	$\#\ \xi_{ud}$	$\#\ \beta_{jy}$	$\#\ \gamma_j$
آزاد	آزاد	$k-1)$	sk	$[\sum_{j=1}^{J}(r_j-1)-s]$	$J-s)$
آزاد	محدود	$k-1)$	sk	$[(J-s)+(r-2)]$	$J-s)$
محدود	آزاد	$k-1)$	sk	$[\sum_{j=1}^{r}(r_j-1)-s]$	
محدود	محدود	$k-1)$	sk	$[(J-s)+(r-2)]$	

کریستنسن[1] و همکاران (۲۰۰۲) مطالعه‌ای شبیه سازی شده به منظور ارائه مشکلات محاسباتی حین برآورد در مدل‌های چند بعدی که توزیع توانایی در آن نرمال فرض شده است، پرداخته اند. در تحلیل طبقه پنهان و مدل‌های مرتبط متغیرهای قابل مشاهده و پنهان گسسته در نظر گرفته می‌شوند (لازارسفلد و هنری، ۱۹۶۸).

همان طور که در قبل نیز اشاره شد، مدل‌های چند بعدی متغیر پنهان برای داده‌های گسسته به صورت توسعه ای از مدل‌های تک بعدی نظریه پاسخ سؤال یا به عنوان کاربرد ی از تحلیل عوامل مشترک بر روی داده‌های گسسته قابل تعریف هستند (تکانه و د لیو[2]، ۱۹۸۷).

به علاوه روش‌ها را می‌توان به صورت متنوع با هم ترکیب کرد، مانند ترکیب به تازگی توسعه یافته ای از مدل‌های نظریه پاسخ سؤال که در جریان آن مدل‌های پاسخ سؤال با طبقه پنهان ترکیب شده اند (راست[3]، ۱۹۹۰). پژوهش‌های متعددی توسط وایتلی[4] (۱۹۸۰) مسترز[5] (۱۹۸۵)، راست (۱۹۸۸)،هاینن[6] (۱۹۹۶) و لنگهاین و

1. Christensen
2. Takane and de Leeuw
3. Rost
4. Whitely
5. Masters
6. Heinen

راست[1] (۲۰۱۳) دربکارگیری و مقایسه رویکرد (مدل)های مختلف طبقه پنهان صورت گرفته است. همچنین در حوزه فضای چندبعدی ژانگ[2] به مقایسه مدل‌های تک بعدی و چند بعدی پاسخ سؤال در برآورد پارامتر پرداخته است.

بارتولوچی[3] در سال ۲۰۰۷ یک دسته از مدل‌های طبقه پنهان چند بعدی پاسخ سؤال برای سؤال‌های دو ارزشی پیشنهاد کرد که در آن علاوه بر این که آزمون و سؤال‌های آزمون را در فضای چند بعدی مورد توجه قرار می‌دهد (آدامز و همکاران، ۱۹۹۷؛ ژانگ، ۲۰۰۴)، توانایی پنهان به صورت گسسته در نظر گرفته می‌شود، به طوری که هر نقطه از این توزیع مربوط به طبقه ای از افراد است که از نظر توانایی همگن هستند که این گروه‌ها به صورت طبقه (کلاس)های پنهان نام گذاری شده اند (لازارسفلد و هنری، ۱۹۶۸؛ گودمن[4]، ۱۹۷۴). پژوهشی توسط‌هابرمن[5] و همکاران از مقایسه مدل‌های پاسخ سؤال چند بعدی بر اساس مقیاس صفات نهفته (گسسته و پیوسته) صورت گرفته است (هابرمن و همکاران، ۲۰۰۸). همچنین در نظر گرفتن توزیع گسسته برای متغیر پنهان محاسبات برآوردهای بیشینه[6] را از طریق الگوریتم بیشینه کردن انتظار[7] آسان می کند.

پژوهشی توسط آدامز و همکاران صورت گرفته که به صورت گسترده مدل‌های راش را مورد توجه قرار داده است (آدامز و همکاران،۱۹۹۷). طرح مشابهی توسط (وُندویه[8]، ۲۰۰۸) به عنوان مدل‌های تشخیصی معرفی شد که بر خلاف مدل یاد شده بر توانایی‌های پنهان ثابت تاکید دارند.

گنالدی نیز در سال ۲۰۱۶ پژوهشی به عنوان رویکرد چند بعدی تعیین ابعاد آزمون استاندارد ریاضی دانش آموزان انجام داده است که برای اولین بار در حوزه

1. Langheine and Rost
2. Zhang
3. Bartolucci
4. Goodman
5. Haberman
6. Maximum estimates
7. Expectation–maximization algorithm
8. Von Davier

آموزش از مدل‌های طبقه پنهان چند بعدی برای برآورد توانایی آزمون وآزمودنی به کار گرفته شده است و درس ریاضی به عنوان مهارتی انتزاعی شامل چندین حوزه محتوایی مختلف و دارای فرایندهای شناختی پیچیده مورد تحلیل قرار گرفته است. پژوهشگر معتقد بوده که آزمون استاندارد که مجموعه ای از سازه‌های پنهان را مورد آزمایش و ارزیابی قرار می‌دهد. همچنین گرامی‌پور و شهمیرزادی[1] (۲۰۱۸) نیز توانستند با کاربرد رویکرد LCMIRT، ابعاد شناختی پنهان در بخش درک مطلب و کلوز تست درس زبان انگلیسی کنکور سراسری ایران را آشکار نمایند. آنها نشان دادند که ابعاد شناختی شامل درک، حل مسأله و استدلال، سه بعد پنهان اساسی هستند که به صورت سلسله مراتبی سؤال‌های آزمون در دل آنها جای می‌گیرند و الگوی‌های احتمال پاسخ صحیح به این سؤال‌ها را تبیین می‌نمایند.

مدل به کار رفته دسته ای از مدل‌های چند بعدی طبقه پنهان سؤال پاسخ بوده اند که این امکان را فراهم آورده اند که تعیین ابعاد بر اساس رویکرد اکتشافی صورت گیرد که به طور همزمان امکان اعمال محدودیت‌هایی برای پارامتر تمیز و امکان صورت بندی برای متغیر پنهان گسسته را نیز ایجاد کرده است. نتایج این پژوهش اذعان دارد که توانایی نهفته در آزمون INVALSI۱ منعکس کننده ی اهداف ارزیابی تعریف شده در سطح ملیِ برنامه درسی ریاضیات است که حاکی از آن است که ابعاد این آزمون ۵۰ سؤالی از ریاضی دارای ابعاد حل مسئله و فهم است که توسط ۲۰ سؤال حمایت می‌شوند و ۳۰ سؤال مربوط به توانایی استدلال بوده که با در برگیری ۳۰ سؤال به عنوان بعد دیگر این آزمون مورد حمایت قرار می‌گیرد.

ویژگی‌های زیر بنایی آزمون سراسری ورود به دانشگاه که سالانه تعداد بسیاری از دانش آموزان را مورد ارزیابی قرار می‌دهد و در تعیین سرنوشت آنان نقشی اساسی دارد، حائز اهمیت است. همچنین در مطالعه این شاخصه مهم و سرنوشت ساز در ارزیابی دوره متوسطه و شناخت پیچیدگی‌های زیر بنایی موجود در داده‌های آن برخلاف گذشته (نظریه کلاسیک که در پی ساختار ساده بود که این پیش فرض

1. Geramipour & Shahmirzadi

سادگی تقریباً هیچ‌گاه در داده‌های محقق نیست) باید استفاده مدل‌های آماری چند بعدی و روان سنجی پیشرفته که بازنمایی بهتری از داده‌های حاصل از اجرای آزمون‌ها را فراهم می‌کنند مورد استفاده قرار گیرد. این پژوهش با بکارگیری مدل چندبعدی طبقه پنهان سؤال به مطالعه بعدیت (تعداد ابعاد شناختی آزمون) و خوشه بندی خرده آزمون‌های درس ریاضی آزمون سراسری ورود به دانشگاه پرداخته است.

روش

پژوهش توصیفی و از نوع روانسنجی بوده، تحقیقات توصیفی هم جنبهٔ کاربردی و هم جنبه ی بنیادی دارند. جامعه آماری این پژوهش، به دلیل زیاد بودن آزمون‌ها درگروه‌های آزمایشی یکی از گروه (گروه آزمایشی ریاضی و فنی) و در گروه ریاضی بنا به اهمیت و محدودیت اجرایی بسته‌های موجود، خرده آزمون‌های درس ریاضی مورد پژوش قرار گرفت که شامل ۵۰۰۰ پاسخ نامه شرکت کنندگان گروه ریاضی (درس ریاضی) در آزمون سال‌های ۸۷، ۹۰ و ۹۴ است که به صورت تصادفی ساده انتخاب شده بودند. با توجه به مدل به کار رفته در پژوهش حاضر هیچ راه حل قطعی برای تعیین حجم نمونه در دسترس نیست. در حالت کلی هر اندازه حجم نمونه بیشتر باشد، دقت برآوردهای بدست آمده بالاتر می‌رود. در تعیین میزان دقت لازم برای برآوردهای یک پژوهش متخصصین همواره هدف‌های تحقیق را مد نظر قرار می‌دهند. در تحقیقی همبلتون بر اساس نظریه رایت و استون مطرح می‌سازد که حداقل حجم نمونه برای مدل‌های تک پارامتری، دو پارامتری و سه پارامتری به ترتیب ۲۰۰، ۵۰۰ و ۱۰۰۰ است (همبلتون و کوک، ۱۹۹۷). با این حال به طور جامع می‌توان گفت، اکثر متخصصین در انتخاب حجم نمونه به نکات زیر توجه دارند: (۱) هر چه تعداد پارامترهای مدل بیشتر باشد حجم نمونه بیشتری لازم است. (۲) برآورد پارامترها با افزایش حجم نمونه با ثبات‌تر خواهند شد. (۳) توجه به هدفی که تحلیل برای رسیدن به آن انجام می‌شود نیز تاثیرگذار است. مثلاً تحلیل سؤال برای گنجاندن آن‌ها در بانک سؤال در مقایسه با تحلیل سؤال برای ارزیابی آزمون‌ها و

پرسشنامه‌ها به نمونه بیشتری نیاز دارد. (۴) توزیع نمونه‌گیری افراد از نظر میزان تجانس. مثلاً در گروه‌های نامتجانسی که کل پیوستار صفت پنهان را می‌پوشانند نمونه بزرگ‌تری مورد نیاز است. (۵) تعداد سؤال‌ها، گزینه‌ها و رابطه سؤال‌ها با سازه زیربنایی (اُرلاندُ[1]، ۲۰۰۴). تتاکاوا و جانسُن[2] (۱۹۹۰) و هُوانگ[3] (۲۰۰۲) برای برآورد دقیق پارامترها حداقل حجم نمونه را ۵۰۰ نفر ذکر کرده است. شواهد موجود در پیشینه مدل‌های نظریه پاسخ سؤال همگی بر نمونه‌های دست‌کم به حجم ۱۰۰۰ نفر (به‌خصوص برای مدل‌های حاوی سه پارامتر) توافق دارند (یِن[4]، ۱۹۸۷؛ کیم[5]، ۲۰۰۶).

با توجه به مدل "چندبعدی"استفاده شده در پژوهش حاضر که نسبت به مدل‌های متداول پارامترهای بیشتری دارند، حجم نمونه بیش از ۱۰۰۰ در نظر گرفته شد. اگر چه بر اساس نظر متخصصان نمونه‌های ۲۰۰۰ تا ۲۵۰۰ برای برآورد پارامترهای مدل‌های چندبعدی کافی است، ولی با توجه به حساسیت مدل‌های چندبعدی در برآورد پارامترها (به‌خصوص وقتی که تعداد ابعاد بیشتر از دو بعد باشد) که با استناد به پژوهش‌های پیشین مشخص شد. تصمیم گرفته شد که از نمونه‌های ۵۰۰۰ آزمودنی برای پژوهش حاضر استفاده شود. بر همین اساس برای تحلیل داده‌های آزمون درس ریاضی در هریک از سه سال مورد نظر (۸۷، ۹۰ و ۹۴)، نمونه ۵۰۰۰ نفری به شیوه تصادفی ساده انتخاب شد.

از آنجایی که تحلیل داده‌ها باید در راستای اهداف پژوهش باشد، داده‌های واقعی کنکور با مدل طبقه پنهان چندبعدی پاسخ سؤال که در ادامه به تفصیل آورده شده است برازش داده شد. در بخش تجزیه و تحلیل کمی‌پژوهش از نرم افزار رایگان R که زبان برنامه نویسی برای محاسبات آماری است، استفاده شد. جهت تحلیل داده‌ها توسط این نرم افزار برای بررسی تک بعدی بودن آزمون‌ها از بسته

1. Orlando
2. Tsutakawa & Johnson
3. Hwang
4. Yen
5. Kim

Paran و به منظور مطالعه مدل طبقه پنهان چندبعدی پاسخ سؤال بسته MultiLCIRT به کار گرفته شد.

نتایج

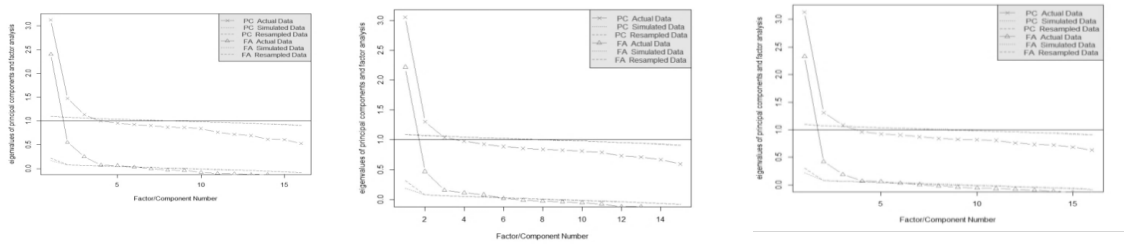
از آن جایی که هدف پژوهش حاضر استفاده از مدل طبقه پنهان پاسخ سؤال بر روی ابعاد شناختی و محتوایی آزمون کنکور سراسری بود، این پژوهش به کاربرد جدیدترین مدل پاسخ سؤال (مدل چندبعدی طبقه پنهان پاسخ سؤال) در مطالعه بعدیت آزمون و خوشه بندی خرده آزمون‌های (هندسه و جبر) و در نتیجه ابعاد شناختی درس ریاضی آزمون سراسری دانشگاه در سال‌های۸۷، ۹۰ و ۹۴ پرداخته است، همچنین کارایی مدل طبقه پنهان چندبعدی را در تعیین ابعاد آزمون و خوشه بندی سؤال‌ها مقایسه و برازش مدل‌ها را توسط داده‌های واقعی مورد بررسی قرار داد.

با توجه به نتایج حاصل از پژوهش برای هر یک از سال‌ها مشخص شد که شاخص‌های BIC در تمامی‌سال‌ها، مدل‌هایی دارای ابعاد بیشتر از یک را ترجیح می‌دهند. این در حالی است که برای هریک از خرده آزمون‌های ریاضی در این سه سال قبل از شروع به تحلیل توسط مدل چند بعدی، رد تک بعدی بودن با استفاده از تحلیل موازی وحدت یافته (هیتون، الن و اسکارپلو[1]، ۲۰۰۴) انجام شد که خود ماهیت پیچیده ی چند بعدی بودن را برای هر یک از سال‌ها تایید می‌کنند. این مدل با توجه به فرایندهای شناختی داده‌های آزمون ریاضی است و نیاز است بیشتر مورد توجه قرار گیرد.

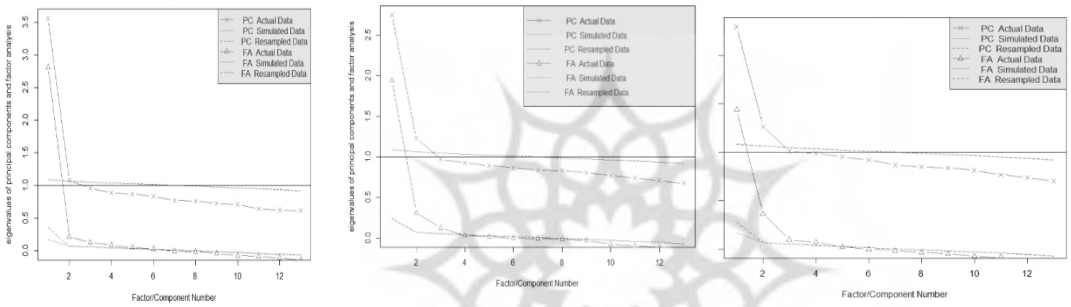
نتایج به دست آمده از تحلیل موازی وحدت یافته برای خرده آزمون‌های ریاضی در این سه سال نشان می‌دهد که تعداد عوامل و مولفه‌های اصلی نیز به ترتیب برای خرده آزمون هندسه به ترتیب در هر یک از سال‌های ۸۷، ۵ و ۳ ؛ ۹۰، ۵ و ۲؛ و سال

1. Hayton, Allen & Scarpello

۹۴، ۶ و ۳ و برای خرده آزمون جبر نیز به ترتیب برای هر یک از سال‌ها ۳ و۲، ۴و۲ و ۶ و۲ بود.



شکل ۱. تحلیل موازی وحدت یافته خرده آزمون هندسه درس ریاضی کنکور در سال‌های ۸۷، ۹۰ و۹۴

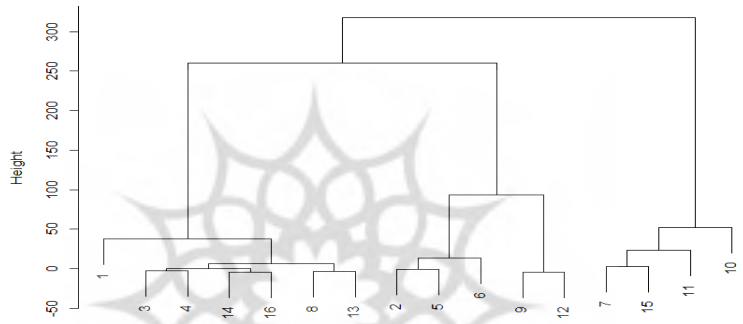


شکل ۲. تحلیل موازی وحدت یافته خرده آزمون جبر درس ریاضی کنکور در سال‌های ۸۷، ۹۰ و۹۴

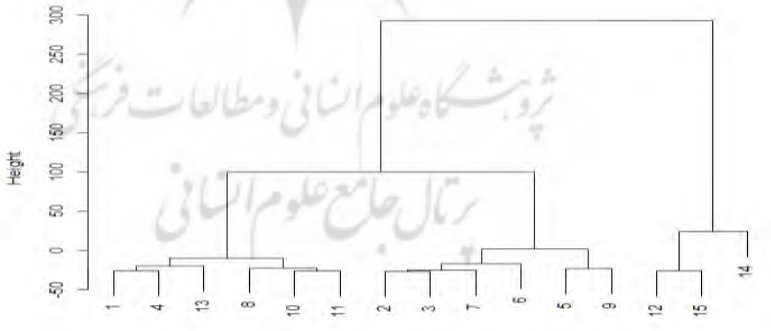
در ادامه، تحلیل ابعاد با استفاده از مدل طبقه پنهان چند بعدی پاسخ سؤال برای خرد آزمون هندسه به ترتیب در سال‌های۸۷، ۹۰ و ۹۴ به صورتی بود که واکنش مدل به افزایش تعداد ابعاد به صورت معنا داری تا ۴ بعد و دو پارامتربرای سال‌های۸۷ و۹۰ و ۵ بعد و دو پارامتر برای سال ۹۴ به صورت اکتشافی صورت گرفت که دردرجه ی اول در راستای رد فرضیه تک بعدی بودن برای این تمامی‌این خرده آزمون‌ها و همسو با نتایج تحلیل موازی است که با توجه به اهمیت این خرده آزمون‌ها در ارزیابی دوره متوسطه به عنوان یکی از عوامل تعیین کننده ورود به دانشگاه‌ها؛ حضور مجموعه ویژگی‌های پنهان در ساختار دور از ذهن نیست نتایج خوشه بندی سلسله مراتبی سؤال‌ها در هر یک از سال‌ها در شکل‌های زیر ارائه شده است.

جدول۴. مقایسه میزان برازش مدل طبقه پنهان چند بعدی پاسخ سوال در تعیین ابعاد خردآزمون‌های هندسه در سال‌های ۸۷، ۹۰ و۹۴

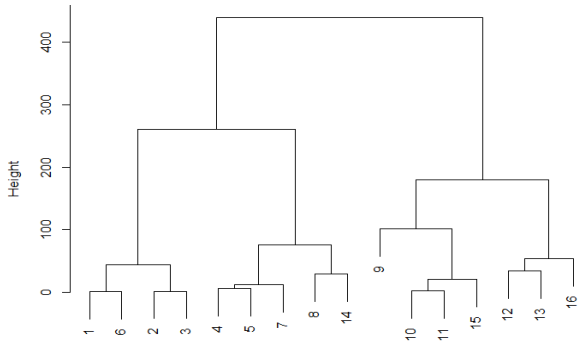
مدل	خرده آزمون	سال	طبقه	log-lik.	np	BIC
طبقه پنهان چند بعدی پاسخ سوال	هندسه	۸۷	۴	-۲۵۸۹۰ .۱۱	۳۷	۵۲۰۹۵ .۳۵
		۹۰	۴	-۱۵۹۰۷ .۶۵	۳۵	۳۲۱۱۳ .۴۱
		۹۴	۵	-۱۸۶۹۲ .۷۷	۳۹	۳۷۷۱۷ .۷۲



شکل۳. خوشه بندی خرده آزمون درس هندسه سال ۸۷



شکل۴. خوشه بندی خرده آزمون درس هندسه سال ۹۰

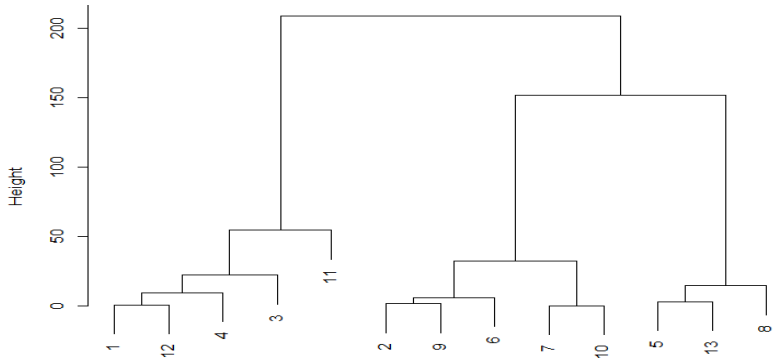


شکل۵. خوشه بندی خرده آزمون درس هندسه سال ۹۴

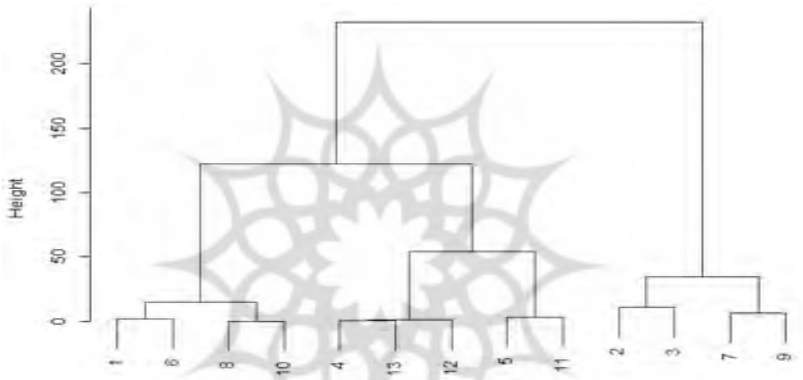
و نیز، یافته‌های تحلیل ابعاد با استفاده از مدل طبقه پنهان چند بعدی پاسخ سؤال برای خرد آزمون‌های درس جبر نیز در سال‌های۸۷، ۹۰ و ۹۴ با افزایش تعداد ابعاد به صورت معنا داری تا ۴ بعد و دو پارامتر برای سال‌های ۸۷ و ۹۰ و ۳ بعد و دو پارامتر برای سال ۹۴ به صورت اکتشافی صورت گرفت که با توجه به آن چه برای خرده آزمون درس هندسه نیز گفته شد در راستای رد فرضیه تک بعدی بودن برای این درس جبر در آزمون سرنوشت ساز کنکور است که همانند خرده آزمون هندسه از اهمیت به سزایی در ارزیابی دوره متوسطه و تعیین کننده برای ورود به دانشگاه‌ها به حساب می‌آید؛ که حضور مجموعه ویژگی‌های پنهان در ساختار آن نیز از دور از ذهن نیست و نتیجه ای همسو با روش تحلیل موازی در مطالعه ابعاد این خرده آزمون را نیز نشان می‌دهد که بر اساس این تحلیل چندبعدی نتایج خوشه بندی سلسله مراتبی سؤال‌ها جبر برای اطلاعات بیشتر در شکل زیر ارائه شده است.

جدول۵. مقایسه میزان برازش مدل طبقه پنهان چند بعدی پاسخ سوال در تعیین ابعاد خردآزمون‌های جبر

مدل	خرده آزمون	سال	طبقه	log-lik.	np	BIC
طبقه پنهان چند بعدی پاسخ سوال	جبر	۸۷	۴	-۲۵۰۷۵.۴۳	۳۵	۵۰۴۴۸.۹۵
		۹۰	۴	-۱۵۷۷۵.۹۴	۳۵	۳۱۸۴۹.۹۸
		۹۴	۳	-۱۵۰۹۴.۱۷	۳۳	۳۰۴۶۹.۴۰



شکل ۶. خوشه بندی خرده آزمون درس جبر سال ۸۷



شکل ۷. خوشه بندی خرده آزمون درس جبر سال ۹۰

شکل ۸. خوشه‌بندی خرده آزمون درس جبر سال ۹۴

سرانجام، تحلیل ابعاد و خوشه بندی آزمون هندسه و جبر درس ریاض در سال‌های ۸۷، ۹۰ و ۹۴ با استفاده از مدل طبقه پنهان و انتخاب مدل برای این خرده آزمون، با استفاده از شاخص BIC صورت گرفت. بر اساس معیار انتخاب، (کوچکترین BIC) برای درس هندسه سال ۹۰ و برای درس جبر سال ۹۴ به عنوان بهترین مدل طبقه پنهان برآورد شد. با توجه به نتایج بدست آمده می‌توان گفت واکنش مدل به افزایش ابعاد و تاثیر آن بر پارامترها متفاوت است و با افزودن بعد جدید شاخص‌های برازش بهتری را نشان می‌دهند. همچنین با استفاده از این مدل افزایش ابعاد باعث می‌شود برآورد با مشکلات مختلفی از جمله طولانی شدن اجرا و زمان بر شدن همگرایی همراه باشد.

جدول۶. مقایسه ابعاد شناختی و خوشه‌های خرده آزمون‌ها هندسه درس ریاضی کنکور در طی سال‌های ۸۷، ۹۰ و ۹۴

	log-lik.	np	BIC	زمان اجرا (ساعت)
مدل ۱ -۸۷	-۲۵۸۹۰.۱۱	۳۷	۵۲۰۹۵.۳۵	۱۰ . ۸۴۴۸۲
مدل۲ -۹۰	-۱۵۹۰۷.۶۵	۳۵	۳۲۱۱۳.۴۱	۶ . ۹۰۲۵۰۷
مدل۳ -۹۴	-۱۸۶۹۲.۷۷	۳۹	۳۷۷۱۷.۷۲	۳۵ . ۶۸۴۶۴

جدول۷. مقایسه ابعاد شناختی و خوشه‌های خرده آزمون‌ها جبر درس ریاضی کنکور در طی سال‌های ۸۷، ۹۰ و ۹۴

	log-lik.	np	BIC	زمان اجرا (ساعت)
مدل۱ -۸۷	-۲۲۸۰۷.۸۶	۳۱	۴۵۸۷۹.۷۴	۴ . ۲۰۹۵۴
مدل۲ -۹۰	-۱۳۹۹۸.۹۸	۳۱	۲۸۲۶۱.۹۸	۱۲ . ۷۸۷۱۹
مدل۳ -۹۴	-۱۳۳۰۲.۱۵	۲۹	۲۶۸۵۱.۲۹	۰ . ۶۹۶۹۱۱

بحث و نتیجه‌گیری

همانند آنچه در پژوهش گنالدی ۲۰۱۶ به عنوان رویکرد چند بعدی تعیین ابعاد آزمون استاندارد ریاضی دانش آموزان انجام شده است درس ریاضی به عنوان مهارتی انتزاعی شامل چندین حوزه محتوایی مختلف و دارای فرایندهای شناختی پیچیده مورد تحلیل قرار گرفته است. در پژوهش حاضر نیز مسئله ابعاد آزمون که همان صفات و زیربنای نهفته آن

است با استفاده از مدل‌های آماری در چهارچوب نظریه پاسخ سؤال مورد تحلیل قرار گرفت که نتایج حاکی از آن است که این آزمون دارای ابعاد حل مسئله و فهم است و سؤالاتی نیز در سطح توانایی استدلال بوده که به عنوان بعد دیگر این آزمون قرار می‌گیرد. همچنین یافته‌های پژوهش حاضر، از حیث مهارت‌های اساسی شناختی مبین پاسخ دهی به سؤال‌های آزمون، همسو با یافته‌های پژوهش گرامی‌پور و شهمیرزادی (۲۰۱۸) است.

در ادامه نتایج پژوهش حاضر می‌توان گفت در سال ۸۷ درس هندسه سؤالات بیشتر از لحاظ ابعاد شناختی مختلف تقسیم شده اند، مطالعه حوزه‌های شناختی را می‌توان برای مقاصد مختلفی نظیر تدوین برنامه درسی، تحلیل اهداف یادگیری مورد استفاده قرار داد (پین[1]۲۰۰۳). این موضوع که برخی از محتواهای آموزشی برای طرح سؤالات در تشخیص ابعاد شناختی مختلف مناسب هستند نیز باید مورد توجه قرار گیرد. در توصیف ابعاد، سؤالاتی که با استفاده از استدلال؛ به عنوان عمیق‌ترین بنیاد در درک ریاضیات (براگ[2]و همکاران، ۲۰۱۳) ؛ و تصور فضایی از اشکال هندسی قابل پاسخ گویی بودند، سؤالات ۱۲۶، ۱۲۹، ۱۳۰ و ۱۳۶ در یک بعد قرار گرفته اند و بنا بر تحقیقات بارتولوچی ۲۰۰۷، سوالات که از لحاظ شناختی یک قابلیت را ارزیابی می کنند در یک خوشه قرار می‌گیرند، که در بین آن‌ها سؤالات ۱۲۹ و ۱۳۳ دارای کمترین درصد پاسخ گویی در این خرده آزمون بودند که در هر دو این سؤال‌ها گزینه‌های به گونه ای طراحی شده بود که دانش آموزان در انتخاب گزینه صحیح دچار مشکل بوده به صورتی که درصد پاسخ‌گویی درست (گزینه ۳) به سؤال ۱۲۹، نسبت به گزینه‌های دیگر پایین تر بوده که با توجه به مشکل بودن سؤال و راه حل‌های مختلف که برای آن وجود دارد ممکن است در مسیر مختلف پاسخ گویی به سؤال رخ داده باشد.

بُعد دیگر که زیر مجموعه همین خوشه بود نیز باز هم شاهد ساختار شناختی تحلیلی و حل مسائله در حوزه‌های محتوایی مختلف بودیم حل مساله، کاربست دانش ومهارت‌های قبلا آموخته شده در موقعیت‌های تازه است (اسلاوین ۲۰۰۶).

عنوان مثال در سؤالاتی که چندین محتوای آموزشی در راستای یکدیگر مورد سؤال قرار گرفته اند به طبع از ساختار شناختی پیچیده تری برای حل مساله و رسیدن به پاسخ

1. payne
2. Bragg, et al.

صحیح پیروی می‌کند سؤالات ۱۲۷، ۱۴۰و۱۳۷ به ترتیب مقاطع مخروطی متفاوت، دستگاه معادلات خطی، دترمینان، مثلث و کانون‌های بیضی طرح شده اند و در بُعد دیگر سؤالاتی قرار گرفته اند که پاسخ گویی به آن‌ها درک مطلب بر روی تک محتواها حائز اهمیت بوده به عنوان مثال سؤال ۱۳۱ با درک مفاهیم ابتدایی خط وصفحه به راحتی قابل پاسخ گویی خواهد بود، سؤال ۱۳۴ تنها به دانستن رابطه مساحت متوازی الاضلاع با استفاده از بردار اکتفا می‌کند و سؤال ۱۳۵ مفهموی از خط و صفحه را مورد سؤال قرار داده است و آن چنان که باید نیاز به استدلال در سطح ابعاد گروه اول سؤالات بحث شده ندارد و نیز از نظر محتوایی نیز بیشتر مفاهیم در حوزه خط و صفحه، بردار و ماتریس را دربر می‌گیرد که نیازمند سطوح فرایند شناختی به یاد آوردن و فهمیدن هستند (آندرسون، کراتول، و همکاران، ۲۰۰۱).

جدول۸. ابعاد شناختی و خوشه‌های خرده آزمون هندسه درس ریاضی کنکور سال ۸۷

ابعاد	سؤالات
استدلال و تصور فضایی از اشکال هندسی	۱۲۶، ۱۲۹، ۱۳۰، ۱۳۳و ۱۳۶
ساختار شناختی تحلیلی و حل مسائله در حوزه‌های محتوایی مختلف	۱۲۷، ۱۴۰و۱۴۷
درک مطلب بر روی تک محتواها	۱۳۱،۱۳۴ و ۱۳۵

در سال ۹۰ نیز در خرده آزمون هندسه سؤالات ۱۲۵، ۱۳۸، ۱۳۶ بیشتر بر روی مهارت حل مساله در محتواهای مختلف تاکید داشته و نسبت به سایر سؤال‌ها در دسته دشواری متوسط تقسیم بندی می‌شوند که سؤال ۱۳۲ نیز در سطح شناختی دانشی با بالاترین درصد پاسخ گویی تنها با اطلاع از مفاهیم هندسه فضایی قابل به پاسخ گویی است و زیر خوشه دوم با در بر گیری سؤالات ۱۲۶، ۱۲۷ و ۱۳۴ سطح بالاتری از استدلال و تحلیل را در فضای چند بعدی طلب می‌کند که این بعد دارای سؤالاتی با سطح دشواری بیشتر بوده و حل سؤالات و انتخاب گزینه صحیح از میان گزینه‌های موجود تبحر بالاتری را طلب می‌کند، به عنوان مثال سؤال ۱۳۰ با در برگیرنده چندین مفهوم هندسی از جمله مماس مشترک داخلی، خارجی و تسلط بر روی مفهوم میانه وارد بر وتر وویژگی‌های مثلث قائمه دارای کمترین در صد پاسخ در بین سؤالات این خرده آزمون است. با مقایسه این دو زیر خوشه این طور استنتاج می‌شود که هر دو بعد از نظر شناختی روی استدلال تمرکز داشته ولی زیر خوشه اول بیشتر بر مهارت حل مساله تاکید دارد و در نهایت خوشه دوم که در بر

گیرنده بعد سوم این خرده آزمون است، سؤالات بیشتر برای درک مفاهیم و سنجش دانش از محتواهای مختلف طرح شده اند تا در پاسخ گویی به آن‌ها سطوح شناختی پایین ترین سنجش شود و در مقایسه با سؤالات دیگر در سطح پاسخ گویی آسان دسته بندی می‌شوند در برگیرنده سؤالات ۱۳۷، ۱۴۰ و ۱۳۹ است که این سؤالات به ترتیب در باره وضعیت دو دایره نسبت به هم، پیدا کردن یک درایه از یک ماتریس و نوشتن ماتریس به صورت مجموع دو ماتریس است و این بعد کمترین تعداد سؤال را در خود جای داده است.

بنابر نتایج همان طور که در سال ۸۷ نیز گفته شد سؤالات بیشتر از لحاظ شناختی در ابعاد متفاوت جای گرفته اند و نیز در این خرده آزمون سؤالات کمی‌دشوار تر از سال‌های قبل طراحی شده است در بخش محتوایی نیز سهمیه ماتریس‌ها که معمولا در سطح متوسط طراحی شده است کم تر شده است. در توصیف ابعاد، سؤالاتی که در حل آن‌ها به نحوی نیاز به استدلال و تحلیل در فضای دو یا چند بعدی دیده می‌شود در یک خوشه اصلی قرار گرفته اند که خود شامل دو زیر خوشه است.

جدول۹. ابعاد شناختی و خوشه‌های خرده آزمون هندسه درس ریاضی کنکور سال ۹۰.

ابعاد	سؤالات
مهارت حل مساله در محتواهای مختلف (دشواری متوسط)	۱۲۵، ۱۳۸، ۱۳۶
سطح بالایی از استدلال و تحلیل در فضای چند بعدی	۱۲۶، ۱۳۰،۱۲۷ و ۱۳۴
درک مفاهیم و سنجش دانش سطح پاسخ گویی آسان	۱۳۷، ۱۴۰ و ۱۳۹

و در نهایت در سال ۹۴،همان طور که در دو سال گذشته نیز بحث شد سؤالات از جنبه‌های شناختی در ابعاد تقسیم شده اند که در این آزمون با توجه به در بر داشتن سؤالات از هندسه پایه و هندسه تحلیلی می‌توان گفت سخت ترین سؤالات از هندسه پایه است به عنوان مثال سؤالات ۱۲۵ و ۱۲۶ نیاز به رسم شکل و تحلیل و استدلال در سطح گسترده داشته و با توجه به محدودیت زمانی آزمون بسیار دشوار به نظر می‌رسند. در مرحله توصیف ابعاد دو خوشه اصلی و چهار زیر خوشه وجود دارد که با مقایسه آن‌ها استنتاج می‌شود که خوشه اول بیشتر روی مهارت‌های تحلیل و استدلال تمرکز داشته این در حالی است که خوشه دوم مهارت حل مساله و درک مفاهیم را مورد آزمون قرار می‌دهد. در ادامه تحلیل سؤالات هر یک از خوشه‌های اصلی می‌توان گفت که سئوالات ۱۲۵، ۱۳۰ و ۱۳۲ در خوشه اول دارای کمترین درصدهای پاسخ گویی به سؤالات هستند و این سؤالات

سطح بالایی از استدلال و تحلیل را در فضای دو بعدی و صرف زمان بیشتری را برای رسیدن به جواب صحیح طلب می‌کنند ودر مجموع سؤالات در دو زیر خوشه اولین خوشه در مقایسه با دیگر سؤالات این خرده آزمون در سطح دشوار دسته بندی می‌شود که در بین آن‌ها سؤال ۱۳۱ غیر استاندارد بوده و از مطالب کتاب نظام قدیم در درس هندسه طراحی شده است. در خوشه دوم سؤالات ۱۳۳، ۱۳۹ به ترتیب بر روی حل مساله در محتواهای فضا و سنجش مفاهیم ماتریس دارد و در هر دو زیر خوشه آن حل مسائل در محتواهای مختلف از جمله خط وصفحه، ماتریس که نسبت به سطوح سؤالات در سطح متوسط وگاهاً ساده سؤال ۱۴۰ مورد ارزشیابی قرار می‌گیرند حضور دارند.

جدول۱۰. ابعاد شناختی و خوشه‌های خرده آزمون هندسه درس ریاضی کنکور سال ۹۴

ابعاد	سؤالات
مهارت‌های تحلیل و استدلال	۱۲۵، ۱۳۰ و۱۲۶ ۱۳۱،۱۳۲
مهارت حل مساله و درک مفاهیم	۱۳۳، ۱۳۹ ۱۴۰

در ادامه بحث برای خرده آزمون جبر در سال ۸۷ نیز به صورتی است که می‌توان گفت که خوشه اول با مجموع ۵ سؤال محتواهای آموزشی را در این خرده آزمون مورد سؤال قرار می‌دهد که از لحاظ شناختی بیشتر درک مطلب و مفاهیم جبر را مورد سؤال قرار داده است و درجه سختی پایین تری دارند این در حالی است که این بعد شامل سؤال‌های ۱۴۵، ۱۴۶ و ۱۵۳ است که به ترتیب از مباحث هم ارزی، مجموعه‌ها و تاس و احتمالات طراحی شده اند که دانش آموز با درک کلی از مفاهیم و روابط کلی گفته شده در این مباحث قادر به پاسخ گویی این سؤالات است. در خوشه دوم این خرده آزمون که خود شامل دو زیر خوشه است این طور به دست آمده است درجه سختی سؤالات نسبت به زیر خوشه اول بالا تر بوده و برای حل سؤالات نیاز به ابعاد شناختی پیچیده تری در سطوح بالاتری از شناخت مورد نیاز است، که زیر خوشه اول (بعد دوم) بیشتر استدلال در مفاهیم مختلف را در سطح پیشرفته مورد سؤال قرار داده است سؤالات ۱۵۱، ۱۵۲و ۱۴۴که در آن حل مسائل

نیز به چشم می‌خورد سؤالات ۱۴۹ و ۱۴۸. زیر خوشه دوم (بعد سوم) با در بر گرفتن سه سؤال که پایین ترین درصدهای پاسخ گویی را به خود اختصاص داده اند و در حوزه شناختی در طبقه به کار بستن مفاهیم و تحلیل قرار می‌گیرند.

جدول۱۱. ابعاد شناختی و خوشه‌های خرده آزمون جبر درس ریاضی کنکور سال ۸۷

ابعاد	سؤالات
درک مطلب و مفاهیم و روابط کلی جبر ودرجه سختی پایین تر	۱۴۵، ۱۴۶ و ۱۵۳
استدلال در مفاهیم مختلف را در سطح پیشرفته	۱۵۱، ۱۵۲و ۱۴۴
به کار بستن مفاهیم و تحلیل	۱۴۷،۱۵۵ و ۱۵۰

بنا بر نتایج، در توصیف ابعاد جبر خرده آزمون جبر سال ۹۰ نیز می‌توان گفت در خوشه اول، که خود شامل ۲ زیر خوشه است. سؤالات با درجه سختی بالاتری نسبت به خوشه دوم مطرح شده است. به عنوان مثال سؤالات ۱۵۲، ۱۴۷ و ۱۵۰ در خوشه اول سخت بوده و در حل آن‌ها نیاز به دقت و سرعت دیده می‌شود و در مقابل سؤالات ۱۵۱، ۱۴۴ در خوشه دوم ساده تر بوده و تنها برای ارزیابی از مفاهیم خاص (فاکتوریل و تقسیم) طراحی شده اند. علاوه بر این این طور به نظر استنتاج می‌شود که زیر خوشه اول (بعد اول) با توجه به سؤالات موجود در آن بیشتر بر روی استدلال و تحلیل روی محتواهای مختلف تاکید کرده است و نیز زیر خوشه دوم (بعد دوم) نیز با تاکید بر روی سختی سؤال‌ها حل مساله را نیز مورد آزمون قرار می_ دهد این دو زیر خوشه در مجموع ۹ سؤال از سؤالات جبر را به خود اختصاص داده اند که در میان آن‌ها سؤالاتی زمان بر ۱۴۶، ۱۴۸، ۱۵۲ از همه سؤالات بیشتر خود نمایی می‌کنند و نیز خوشه دوم (بعد سوم) بیشتر بر روی سطح دانش تاکید داشته و سؤالات در محتواهای مختلف طراحی شده اند که با دانشی نسبی نسبت به مفاهیم جبر قابل پاسخگویی هستند.

جدول۱۲. ابعاد شناختی و خوشه‌های خرده آزمون جبر درس ریاضی کنکور سال ۹۰

ابعاد	سؤالات
دقت و سرعت (سخت)استدلال و تحلیل روی محتواهای مختلف	۱۵۲، ۱۴۸،۱۴۷ و ۱۵۰
تاکید بر روی سختی سؤال‌ها+ حل مساله	۱۴۶،۱۵۵و۱۵۴
ساده تر،برای ارزیابی از مفاهیم خاص (فاکتوریل و تقسیم) سطح دانش	۱۵۱، ۱۴۴

و در نهایت ابعاد خرده آزمون جبر سال ۹۴ به گونه ای متفاوت از سایر سال‌ها و خرده آزمون‌ها بود، به صورتی که در کل ۱۳ سؤال سؤالات ۱۵۲ به عنوان سخت ترین سؤال از محتوای نظریه اعداد و با در برگیری سطح بالایی از بعد شناختی که نیاز به تحلیل و به کاربستن مفاهیم داشته ونیز آفریدن (خلاقیت)؛ تعریف واحدی برای آفرینندگی به دست نیامده است و به صورت توانایی اندیشیدن درباره ی امور به راه‌های تازه و غیر معمول و رسیدن به راه حل‌های منحصر به فرد برای مسائل تعریف شده است (سانتروک،۲۰۰۴)؛ در کنار سرعت به حل این سؤال کمک بسزایی می کند و در مقابل سؤال ۱۴۹ به عنوان ساده ترین سؤال مفهوم گراف که تنها با دانش مفاهیم گراف قادر به حل بوده است، طرح شده بود. ما بقی سؤالات را که در یک خوشه دیگر که خود دارای سه زیر خوشه است، گنجانده شده اند و از لحاظ درجه سختی سؤالات زیر خوشه اول در سطح متوسط قرار گرفتند و از لحاظ شناختی استدلال بیشترین گزینه محتمل درپاسخ گویی به سؤالات این گروه بود. به عنوان مثال سؤالات ۱۴۳، ۱۴۵ و ۱۴۷ که هر یک به گونه ای نیاز به مهارت استدلال مفاهیم مختلف در پاسخ گویی آن‌ها به چشم می‌خورد و در زیر خوشه دوم نسبت به زیر خوشه نخست سختی بیشتری داشته و بر روی محتواهای نظریه اعداد طراحی شده اند که این مباحث در حل نیاز مهارت حل مساله و استدلال در مفاهیم مختلف دارند؛ زیر خوشه سوم دارای سه سؤال ۱۴۶، ۱۵۳ و ۱۵۵ دارای سطح دشواری بیشتر نسبت به دو خوشه قبل بوده و حل آن‌ها علاوه بر استدلال نیاز به سرعت در حل و دقت در پاسخ گویی نیز دارد.

جدول۱۳. ابعاد شناختی و خوشه‌های خرده آزمون جبر درس ریاضی کنکور سال ۹۴

ابعاد	سؤالات
سخت ترین سؤال از محتوای نظریه اعداد تحلیل و به کاربستن مفاهیم آفریدن (خلاقیت)	۱۵۲
مهارت استدلال (سطح متوسط)	۱۴۳، ۱۴۵ و ۱۴۷
نیاز مهارت حل مساله و استدلال در مفاهیم مختلف (نسبت به زیر خوشه قبل دشوار تر)	۱۴۴،۱۵۱و ۱۵۴
استدلال+سرعت در حل و دقت در پاسخ گویی	۱۴۶، ۱۵۳ و ۱۵۵
ساده ترین سؤال دانش مفاهیم گراف	۱۴۹

سرانجام، پژوهش حاضر توانست به صورت موفقیت آمیزی ابعاد شناختی پنهان در آزمون ریاضی کنکور سراسری را با استفاده از رویکرد LCMIRT پدیدار نماید و به تبیین فرایند شناختی پاسخ به سؤال‌های آزمون به واسطه ارائه دندوگرام کمک کند. پژوهش‌های آتی به لحاظ روش شناختی می‌توانند مطالعات مونت کارلو (شبیه سازی داده‌ها) را به منظور بررسی کارایی و قدرت LCMIRT در مقایسه با سایر روش‌های تعیین ابعاد آزمون بکار گیرند. همچنین در نهایت، سازمانها و نهادهای ارزیابی آموزشی و روانی می‌توانند رویکرد پژوهش حاضر را به عنوان یک شیوه بسیار کارامد و منعطف جهت شناسایی ابعاد پیچیده شناختی آزمون‌های تربیتی و روانی بکار گیرند.

منابع

Ackerman, T. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13* (2), 113-127.

Ackerman, T. (1996). Graphical representation of multidmensional item response theory analyses. *Applied Psychological Measurement, 20*, 311–329.

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied psychological measurement, 21* (1), 1-23.

Anderson,L.W.,&Krathwohl,D.R. (2001).A Taxonomy for learning,teaching,and assessing: A Revision of Bloom's Taxonomy of educational objectives.NewYork:Longman.

Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*, 37-48.

Bacci, S., Pandolfi, S., & Pennoni, F. (2014). A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification, 8* (2), 125-145.

Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika, 72* (2), 141-157.

Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement, 27* (6), 395-414.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64* (2), 153-168.

Bragg, L. A., Vale, C., Herbert, S., Loong, E., Widjaja, W., Williams, G., & Mousley, J. (2013, January). Promoting awareness of reasoning in the primary mathematics classroom. In *MAV 2013: Mathematics of the planet earth: Proceedings of the MAV 50th Annual Conference 2013* (pp. 23-30). Mathematical Association of Victoria.

Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (Eds.). (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10* (1), 1-19.

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.

Christensen, K. B., Bjorner, J. B., Kreiner, S., & Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika, 67* (4), 563-574.

De Champlain, A. F. (1996). The effect of multidimensionality on IRT true-score equating for subgroups of examinees. *Journal of Educational Measurement*, 33, 181-201.

De Champlain, A., & Gessaroli, M. E. (1998). Assessing the dimensionality of item response matrices with small sample sizes and short test lengths. *Applied Measurement in Education, 11* (3), 231-253

Drasgow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied psychology, 68* (3), 363-373.

Dubravka Svetina & Roy Levy (2014) A Framework for Dimensionality Assessment for Multidimensional Item Response Models, *Educational Assessment*, 19:1, 35-57.

Frederiksen, N. (Ed.). (1990). *Diagnostic Monitoring of Skill and Knowledge Acquisition*. Psychology Press.

Fraser, C., & McDonald, R. P. (1988). NOHARM II: A FORTRAN program for fitting unidimensional and multidimensional normal ogive models

of latent trait theory. *The University of New England, Armidale, Australia*. : University of New England, Centre for Behavioral Studies.
Geramipour, M., & Shahmirzadi, N. (2018). Application and Comparison of Multidimensional Latent Class Item Response Theory on Clustering Items in Comprehension Tests. *The Journal of AsiaTEFL*, 15 (2), 479-490.
Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61* (2), 215-231.
Gu, Z. (2011). *Maximizing the potential of multiple-choice items for Cognitive Diagnostic Assessment* (Doctoral dissertation, University of Toronto).
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* (Vol. 7): Springer.
Hayton J. C., Allen D. G., and Scarpello V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*. 7 (2): 191–205
Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Sage Publications, Inc.
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30* (2), 179-185.
Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Irwin.
Kim, H. R., & Stout, W. F. (1994). A new index for assessing the amount of multidimensionality and/or simple structure present in test data. In *annual meeting of the American Educational Research Association, New Orleans*.
Koretz, D. (2008). Measuring Up: What Educational Testing Really Tells Us. Cambridge, MA: Harvard University Press. pp. 7-14.
Langeheine, R., & Rost, J. (Eds.). (2013). *Latent trait and latent class models*. Springer Science & Business Media.
Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin Co.
Lord, F. M. (1980). *Applications of item response theory to practical testing problems*:Mahwah, NJ: Erlbaum.
Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). Statistical theories of mental test scores. (pp.359-382). Reading, MA: Addison-Wesley.
Masters, G. N. (1985). A comparison of latent trait and latent class analyses of Likert-type data. *Psychometrika*, *50* (1), 69-82.
Nandakumar, R. (1994). Assessing essential dimensionality of set of items. *Journal of Educational Measurement*,*31*:1.
Payne, DA (2003). Applied Educational Assessment (2nd ed.). US: Wads Worth.
Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*, *4* (3), 207-230.

Reckase, M. (2009). *Multidimensional item response theory* (Vol. 150). New York: Springer.

Rost, J. (1988). Rating scale analysis with latent class models. *Psychometrika*, *53* (3), 327-348.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14* (3), 271-282.

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement*, *15* (2), 109-127.

Santrock, J. W. (2004). *Educational Psychology*. New York: McGraw-Hill.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, *58* (2), 159-194.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28* (3), 237-247.

Slavin, R. E., & Davis, N. (2006). Educational psychology: Theory and practice.

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

Sympson, J. B. (1978). *A model for testing with multidimensional items.* Paper presented at the Proceedings of the 1977 Computerized Adaptive Testing Conference, Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52* (3), 393-408.

Traub, R. E., & Lam, Y. R. (1985). Latent structure and item sampling models for testing. *Annual review of psychology*, *36* (1), 19-48.

Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, *40* (3), 255-275.

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, *12* (3), 239-252.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, ۴۵ (4), 479-494.

Wilson, D., Wood, R., & Gibbons, R. (1987). TESTFACT [Computer program]. Mooresville IN: Scientific Software.

Van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*: Springer.

von Davier, A. A., & Wilson, C. (2008). Investigating the population sensitivity assumption of item response theory true-score equating across two subgroups of examinees and two test formats. *Applied Psychological Measurement*, *32* (1), 11-26.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement*, *30* (3), 187-214.

Zhang, J. (2004). Comparison of unidimensional and multidimensional approaches to IRT parameter estimation. *ETS Research Report Series*, *2004* (2), i-40.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement*, *24* (4), 293-308.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی