

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

مسعود گرامی پور<sup>۱</sup>، مریم مقدسین<sup>۲</sup>، ریحانه رضازاده<sup>۳</sup>

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 مجموعه‌ای 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 سؤالات 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 تبدیل شده 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<sup>۱</sup> آزمون بسیار مهم جلوه می‌کند و سنجش اعتبار به‌عنوان مشخصه‌ای از کیفیت در قلب نظریات ارزیابی آموزشی و روانی قرار می‌گیرد (کورتز<sup>۲</sup>، ۲۰۰۸: ۲۱۵). بحث ابعاد آزمون با اعتبار ارتباط تنگاتنگی دارد بر اساس چارچوبی که مسیک<sup>۳</sup> معرفی کرده است، ابعاد آزمون به‌عنوان قوی‌ترین نوع شاهد در حمایت از اعتبار سازه<sup>۴</sup> نام برده می‌شود. اهمیت متغیر (های) پنهان<sup>۵</sup> (مکنون)، به‌استفاده از مدل‌های چندبعدی پاسخ سؤال<sup>۶</sup> در حوزه‌های مختلف منجر می‌گردد که تاکنون روش‌های تحلیل و آزمون‌سازی با پیش فرض‌های نظریه کلاسیک که مفهوم بعد به‌صورت مستقیم مورد بحث نبوده و یا با به‌کاربردن 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 به‌اندازه‌گیری بیش از یک توانایی از خود نشان دادند (ریکیس، ۲۰۰۹: ۵۳). در نظریه پاسخ سؤال (گُرد<sup>۷</sup>، ۱۹۸۰) فرض بر گسسته بودن مشاهدات و در حالت چند وجهی<sup>۸</sup> فرض بر ترتیب است و بیشتر مدل‌های تعریف شده این نظریه بر توزیع نرمال توانایی پنهان تأکید

- 
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<sup>۱</sup>، 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<sup>۲</sup> 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<sup>۳</sup>، 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<sup>۴</sup> و ابعاد آزمون<sup>۵</sup> قوی‌ترین نوع شواهد در حمایت از اعتبار سازه بودند. تعریف ابعاد آزمون، بر اساس الگوی مدل‌سازی صورت می‌گیرد. (به‌عنوان مثال تحلیل عاملی یا نظریه پاسخ سؤال). در گذشته، 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 قابل بحث بوده است (آکرمن<sup>۶</sup>، ۱۹۹۶). یک روش بسیار معمول استفاده از شکل اسکری مقادیر ویژه به دست آمده از تحلیل مؤلفه‌های اصلی است (ریکیس، ۱۹۷۹). در این روش تصمیم‌گیری درباره این که آیا مقادیر ویژه دوم (و احتمالاً سوم) به اندازه کافی بزرگ هستند که به‌عنوان ابعاد دیگر مورد توجه قرار گیرند یا نه همواره مشکل اساسی بوده است. هورن<sup>۷</sup> (۱۹۶۵) و دارسگو و لیساک<sup>۸</sup> (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<sup>۱</sup> و همکاران، ۱۹۹۱). همچنین استوت<sup>۲</sup> 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<sup>۳</sup> (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<sup>۴</sup>، ۱۹۸۸) متناسب با مدل اجیو نرمال<sup>۵</sup> با روش حداقل مربعات که قادر به برآورد پارامتر تمیز و دشواری سؤال‌ها و TESTFACT (ویلسون<sup>۶</sup> 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<sup>۷</sup>، ۱۹۴۵؛ هالین<sup>۸</sup> و همکاران، ۱۹۸۳). زوویک<sup>۹</sup>، ۱۹۸۷ به 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

واقعي و شبیه‌سازی شده (همبلتون و رووینلی<sup>۱</sup>، ۱۹۸۶)، بررسی و مقایسه ابعاد غالب برای مجموعه‌ای از سؤالات (نانداکومار<sup>۲</sup>، 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<sup>۳</sup>، ۱۹۹۸).

در پی تحلیل عاملی مدل‌های رایج و پرکاربرد نظریه پاسخ سؤال نیز که علاوه بر مفروضه‌های اصلی استقلال موضعی و تکنوایی بر تک‌بعدي بودن که به‌منظور برآورد استقلال موضعی تعریف می‌شود (لرد و نوبک<sup>۴</sup>، 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 تک‌بعدي 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 تک‌بعدي برای دستاوردهای آموزشی نامناسب جلوه می‌کند (تراب و لام<sup>۵</sup>، ۱۹۸۵: ۲۲).

نظریه پاسخ سؤال چندبعدي یک آرمان‌گری از واقعیت است که به‌عنوان حالت خاصی از نظریه پاسخ سؤال تک‌بعدي مدلی ریاضی است که بردارها به‌عنوان 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. به‌کارگیری این نظریه تعیین جنبه ساختاری (عملکرد سؤالات آزمون) و ضمنی (بردار توصیف فرد) سازه را امکان‌پذیر می‌سازد (همبلتون و سوامینگتون<sup>۵</sup>، ۱۹۸۵).

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

از آنجا که بسیاری از آزمون‌های آموزشی و روانی ذاتاً چندبعدی هستند و دو یا چندسازه یا بعد را اندازه‌گیری می‌کند، علاوه بر این پژوهش‌های علوم شناختی به عنوان مثال (فردریکسون<sup>۱</sup> 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 به عنوان 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 (اکرم، ۱۹۸۹؛ بردلو<sup>۲</sup> و همکاران، ۱۹۹۹؛ دوچاپلین، ۱۹۹۶؛ شلی و استوت<sup>۳</sup>، ۱۹۹۳؛ یین<sup>۴</sup>، ۱۹۹۳؛ چن و تیسسن<sup>۵</sup>، ۱۹۹۷؛ سیرسی و همکاران، ۱۹۹۱؛ ون در لیدن<sup>۶</sup>، ۱۹۹۶؛ وی<sup>۷</sup> و همکاران، ۱۹۸۸؛ واکر و برتواس<sup>۸</sup>، ۲۰۰۳).

نظریه 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.

جدول ۱. طبقه بندی چهارگانه متغیر پنهان بر اساس مقیاس اندازه گیری

متغیرهای پنهان (مکنون)		
طبقه ای	پیوسته	متغیرهای آشکار (مشاهده شده)
تحلیل مشخصات نهفته <sup>۱</sup>	تحلیل عاملی	پیوسته
تحلیل طبقه (کلاس) پنهان <sup>۱</sup>	تحلیل ویژگی نهفته <sup>۱</sup>	طبقه ای <sup>۱</sup>

به عنوان مثال در جدول فوق طبقه بندی چهارگانه (بارتولمئو 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) جدول ۲. طبقه بندی از محبوب ترین مدل های 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	شبکه های بیزی مدل های سنجش شناختی	تحلیل طبقه (کلاس) پنهان	نظریه پاسخ سؤال چند بعدی	نظریه پاسخ سؤال
چند و جهی بدون ترتیب	شبکه های بیزی مدل های سنجش شناختی	تحلیل طبقه (کلاس) پنهان	نظریه پاسخ سؤال	نظریه پاسخ سؤال
چند وجهی با ترتیب	شبکه های بیزی مدل های 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 مدل های روان سنجی مورد توجه متخصصین قرار گرفت. نظریه پاسخ سؤال (گرد، ۱۹۸۰) فرض بر گسسته بودن مشاهدات و در حالت چند وجهی<sup>۱</sup> فرض بر ترتیب است، به علاوه بیشتر مدل های تعریف شده این نظریه بر توزیع نرمال توانایی پنهان تاکید دارند.

مدل عمومی نظریه 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 می آید:

(۱) نوع تابع ربط<sup>۱</sup> (مشخصات تابع ربط (.)  $g$  به ماهیت سؤالها بستگی دارد).  
با توجه به پژوهش که بر داده های دو ارزشی (سؤالهایی که به صورت صحیح یا غلط  
نمره گذاری می شوند) تمرکز دارد. از تابع رابط دودویی<sup>۲</sup> 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 پارامتر ضریب تمیز<sup>۳</sup>  $\gamma_j$ .

(۳) فرمول بندی پارامتر سختی سؤال<sup>۴</sup>  $\beta_{jy}$ .

نشانه گذاری

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

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

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

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

$$\pi_u = P(U_i = \xi_u) : \pi_u$$

$\delta_{jd}$ : متغیر ساختگی<sup>۵</sup> برابر ۱ در صورتی که سؤال  $j$  ام ویژگی پنهان از نوع  $d$  را  
اندازه بگیرد و در غیر این صورت برابر صفر. ( $d=1, \dots, s$ )

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



جدول ۳. 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	$\# \pi_u$	$\# \xi_{ud}$	$\# \beta_{jv}$	$\# \gamma_j$
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	$(k - 1)$	$sk$	$[\sum_{j=1}^J (r_j - 1) - s]$	
محدود	محدود	$(k - 1)$	$sk$	$[(J - s) + (r - 2)]$	

کریستنسن<sup>۱</sup> 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 دلیو<sup>۲</sup>، 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<sup>۳</sup>، ۱۹۹۰). پژوهش‌های متعددی توسط وایتلی<sup>۴</sup> (۱۹۸۰) مسترز<sup>۵</sup> (۱۹۸۵)، راست (۱۹۸۸)، هاینن<sup>۶</sup> (۱۹۹۶) و لنگه‌این و

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

راست<sup>۱</sup> (۲۰۱۳) در بکارگیری و مقایسه رویکرد (مدل‌های مختلف طبقه پنهان صورت گرفته است. همچنین در حوزه فضای چندبعدي ژانگ<sup>۲</sup> به مقایسه مدل‌های تک بعدی و چند بعدی پاسخ سؤال در برآورد پارامتر پرداخته است.

بارتولوچی<sup>۳</sup> 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<sup>۴</sup>، ۱۹۷۴). پژوهشی توسط هابرم<sup>۵</sup> 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<sup>۶</sup> را از طریق الگوریتم بیشینه کردن انتظار<sup>۷</sup> آسان می‌کند.

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<sup>۸</sup>، 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 و آزمودنی 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<sup>۱</sup> (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 INVALSI منعکس کننده ی اهداف ارزیابی تعریف شده در سطح ملی 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 روان‌سنجی 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 جنبه کاربردی و هم جنبه ی بنیادی دارند. جامعه آماری این پژوهش، به دلیل زیاد بودن آزمون‌ها در گروه‌های 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 پژوهش 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 (ارلاندا، ۲۰۰۴). تتاکاوا و جانسن<sup>۲</sup> (۱۹۹۰) و هوانگ<sup>۳</sup> (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 (ین<sup>۴</sup>، ۱۹۸۷؛ کیم<sup>۵</sup>، ۲۰۰۶).

با توجه به مدل "چندبعدي" 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 چندبعدي کافی است، ولی با توجه به حساسیت مدل‌های چندبعدي 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 چندبعدي پاسخ سؤال که در ادامه به تفصیل آورده شده است برازش داده شد. در بخش تجزیه و تحلیل کمی پژوهش 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.

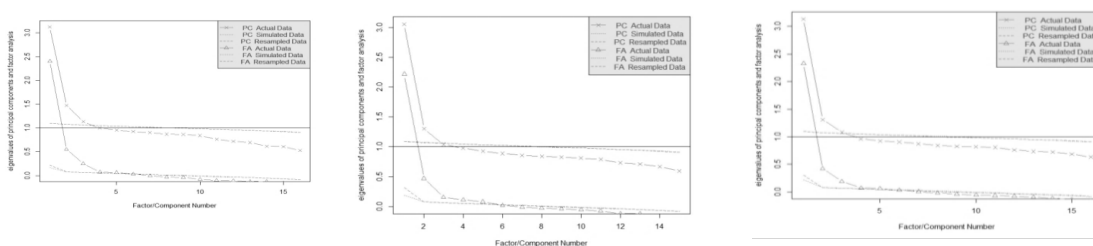
### نتایج

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 سال‌های ۸۷، 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.

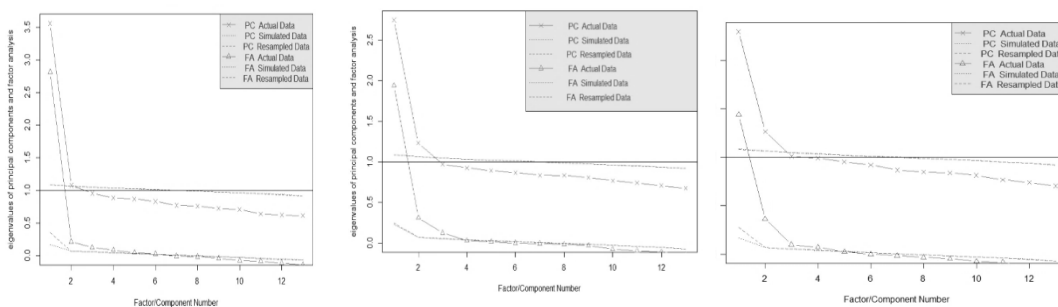
با توجه به نتایج حاصل از پژوهش برای هر یک از سال‌ها مشخص شد که شاخص‌های BIC در تمامی سال‌ها، مدل‌هایی دارای ابعاد بیشتر از یک را ترجیح می‌دهند. این در حالی است که برای هر یک از خرده آزمون‌های ریاضی در این سه سال قبل از شروع به تحلیل توسط مدل چند بعدی، رد تک بعدی بودن با استفاده از تحلیل موازی وحدت یافته (هیتون، الن و اسکارپلو، 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

۹۴، ۶ و ۳ و برای خرده آزمون جبر نیز به ترتیب برای هر یک از سال‌ها ۳ و ۲ و ۴ و ۶ و ۲ بود.



شکل ۱. تحلیل موازی وحدت یافته خرده آزمون هندسه درس ریاضی کنکور در سال‌های ۸۷، ۹۰ و ۹۴

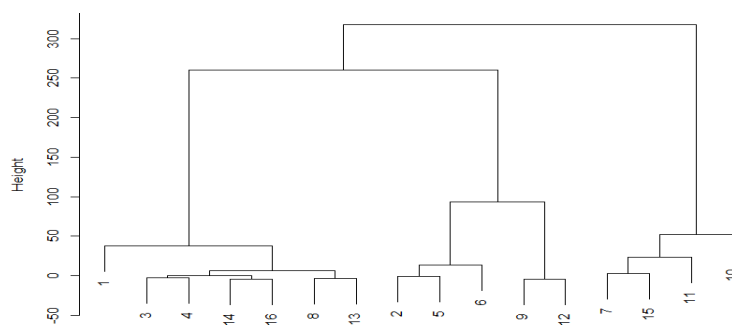


شکل ۲. تحلیل موازی وحدت یافته خرده آزمون جبر درس ریاضی کنکور در سال‌های ۸۷، ۹۰ و ۹۴

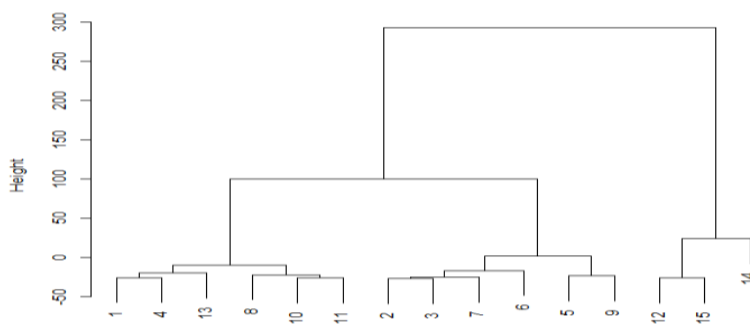
در ادامه، تحلیل ابعاد با استفاده از مدل طبقه پنهان چند بعدی پاسخ سؤال برای خرد آزمون هندسه به ترتیب در سال‌های ۸۷، ۹۰ و ۹۴ به صورتی بود که واکنش مدل به افزایش تعداد ابعاد به صورت معناداری تا ۴ بعد و دو پارامتر برای سال‌های ۸۷ و ۹۰ و ۵ بعد و دو پارامتر برای سال ۹۴ به صورت اکتشافی صورت گرفت که در درجه اول در راستای رد فرضیه تک بعدی بودن برای این تمامی این 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.

جدول ۴. مقایسه میزان برازش مدل طبقه پنهان چند بعدی پاسخ سوال در تعیین ابعاد خردآزمون‌های  
هندسه در سال‌های ۸۷، ۹۰ و ۹۴

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

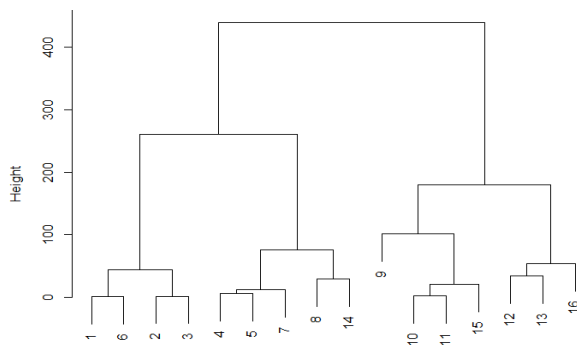


شکل ۳. خوشه بندی خرده آزمون درس هندسه سال ۸۷



شکل ۴. خوشه بندی خرده آزمون درس هندسه سال ۹۰





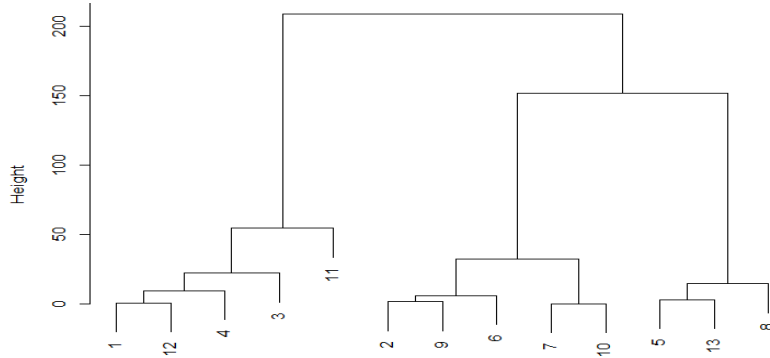
شکل ۵. خوشه بندی خرده آزمون درس هندسه سال ۹۴

و نیز، یافته‌های تحلیل ابعاد با استفاده از مدل طبقه پنهان چند بعدی پاسخ سؤال برای خرده آزمون‌های درس جبر نیز در سال‌های ۸۷، ۹۰ و ۹۴ با افزایش تعداد ابعاد به صورت معناداری 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 نتیجه‌ای 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.

جدول ۵. مقایسه میزان برازش مدل طبقه پنهان چند بعدی پاسخ سوال در تعیین ابعاد خرده‌آزمون‌های

جبر

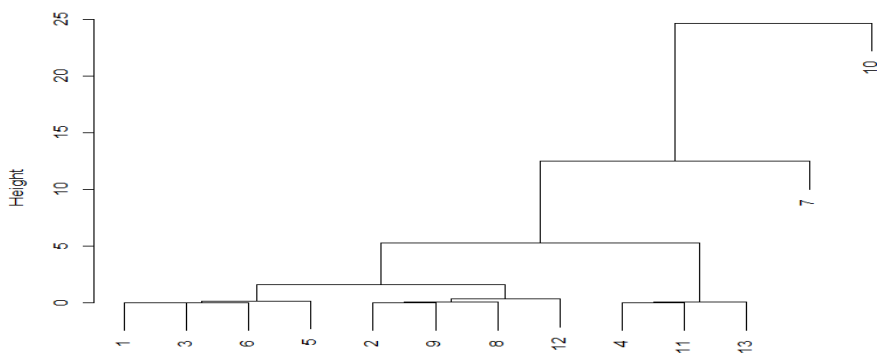
مدل	خرده آزمون	سال	طبقه	log-lik.	np	BIC
طبقه پنهان چند بعدی پاسخ سوال جبر	۸۷	۴	۴	-۲۵۰۷۵.۴۳	۳۵	۵۰۴۴۸.۹۵
	۹۰	۴	۴	-۱۵۷۷۵.۹۴	۳۵	۳۱۸۴۹.۹۸
	۹۴	۳	۳	-۱۵۰۹۴.۱۷	۳۳	۳۰۴۶۹.۴۰



شکل ۶. خوشه بندی خرده آزمون درس جبر سال ۸۷



شکل ۷. خوشه بندی خرده آزمون درس جبر سال ۹۰



شکل ۸. خوشه بندی 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.

جدول ۶. مقایسه ابعاد شناختی و خوشه‌های خرده آزمون‌ها هندسه درس ریاضی کنکور در طی

سال‌های ۸۷، ۹۰ و ۹۴

مدل	log-lik.	np	BIC	زمان اجرا (ساعت)
مدل ۱- ۸۷	-۲۵۸۹۰.۱۱	۳۷	۵۲۰۹۵.۳۵	۱۰.۸۴۴۸۲
مدل ۲- ۹۰	-۱۵۹۰۷.۶۵	۳۵	۳۲۱۱۳.۴۱	۶.۹۰۲۵۰۷
مدل ۳- ۹۴	-۱۸۶۹۲.۷۷	۳۹	۳۷۷۱۷.۷۲	۳۵.۶۸۴۶۴

جدول ۷. مقایسه ابعاد شناختی و خوشه‌های خرده آزمون‌ها جبر درس ریاضی کنکور در طی سال‌های

۸۷، ۹۰ و ۹۴

مدل	log-lik.	np	BIC	زمان اجرا (ساعت)
مدل ۱- ۸۷	-۲۲۸۰۷.۸۶	۳۱	۴۵۸۷۹.۷۴	۴.۲۰۹۵۴
مدل ۲- ۹۰	-۱۳۹۹۸.۹۸	۳۱	۲۸۲۶۱.۹۸	۱۲.۷۸۷۱۹
مدل ۳- 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 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 (براک و همکاران، ۲۰۱۳)؛ و تصور فضایی از اشکال هندسی قابل پاسخ‌گویی بودند، سؤالات ۱۲۶، ۱۲۹، ۱۳۰ و ۱۳۶ در یک بعد قرار گرفته اند و بنا بر تحقیقات بارتولوچی (۲۰۰۷)، سؤالات که از لحاظ شناختی یک قابلیت را ارزیابی می‌کنند در یک خوشه قرار می‌گیرند، که در بین آن‌ها سؤالات ۱۲۹ و ۱۳۳ دارای کمترین درصد پاسخ‌گویی در این خرده آزمون بودند که در هر دو این سؤال‌ها گزینه‌های به گونه‌ای 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 پاسخ‌گویی 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 و مهارت‌های 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.

صحيح پيروي مي كند سؤالات ۱۲۷، ۱۴۰ و ۱۳۷ به ترتيب مقاطع مخروطي متفاوت، دستگاه معادلات خطي، دترمینان، مثلث و كانون‌های بيضي طرح شده اند و در بُعد ديگر سؤالاتي قرار گرفته اند كه پاسخ گويي به آنها درك مطلب بر روی تك محتواها حائز اهميت بوده به عنوان مثال سؤال ۱۳۱ با درك مفاهيم ابتدائي خط و صفحه به راحتی قابل پاسخ گويي خواهد بود، سؤال ۱۳۴ تنها به دانستن رابطه مساحت متوازي الاضلاع با استفاده از بردار اکتفا مي كند و سؤال ۱۳۵ مفهومي از خط و صفحه را مورد سؤال قرار داده است و آن چنان كه بايد نياز به استدلال در سطح ابعاد گروه اول سؤالات بحث شده ندارد و نيز از نظر محتوایی نيز بيشتر مفاهيم در حوزه خط و صفحه، بردار و ماتريس را دربر مي گيرد كه نياز مند سطوح فرايند شناختي به ياد آوردن و فهميدن هستند (آندرسون، كراتول، و همكاران، ۲۰۰۱).

جدول ۸. ابعاد شناختي و خوشه‌های خرده آزمون هندسه درس رياضي كنكور سال ۸۷

سؤالات	ابعاد
۱۲۶، ۱۲۹، ۱۳۰، ۱۳۳ و ۱۳۶	استدلال و تصور فضايي از اشكال هندسي
۱۲۷، ۱۴۰ و ۱۴۷	ساختار شناختي تحليلي و حل مسائله در حوزه‌های محتوایی مختلف
۱۳۴، ۱۳۱ و ۱۳۵	درك مطلب بر روی تك محتواها

در سال ۹۰ نيز در خرده آزمون هندسه سؤالات ۱۲۵، ۱۳۸، ۱۳۶ بيشتر بر روی مهارت حل مساله در محتواهای مختلف تاكيد داشته و نسبت به ساير سؤال‌ها در دسته دشواري متوسط تقسيم بندي مي شوند كه سؤال ۱۳۲ نيز در سطح شناختي دانشي با بالاترين درصد پاسخ گويي تنها با اطلاع از مفاهيم هندسه فضايي قابل به پاسخ گويي است و زير خوشه دوم با در بر گيري سؤالات ۱۲۶، ۱۲۷ و ۱۳۴ سطح بالاتري از استدلال و تحليل را در فضاي چند بعدي طلب مي كند كه اين بعد داراي سؤالاتي با سطح دشواري بيشتر بوده و حل سؤالات و انتخاب گزينه صحيح از ميان گزينه‌های موجود تبحر بالاتري را طلب مي كند، به عنوان مثال سؤال ۱۳۰ با در بر گيرنده چندين مفهوم هندسي از جمله مماس مشترك داخلي، خارجي و تسلط بر روی مفهوم ميانه وارد بر وتر و ویژگی‌های مثلث قائمه داراي كمتري در صد پاسخ در بين سؤالات اين خرده آزمون است. با مقايسه اين دو زير خوشه اين طور استنتاج مي شود كه هر دو بعد از نظر شناختي روی استدلال تمرکز داشته ولي زير خوشه اول بيشتر بر مهارت حل مساله تاكيد دارد و در نهايت خوشه دوم كه 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 می شوند در بر گیرنده 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. بنا بر 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 کمی دشوار تر از سال های قبل طراحی شده است در بخش محتوایی نیز سهمیه ماتریس ها 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 می شود در یک خوشه اصلی قرار گرفته اند که خود شامل دو زیر خوشه است.

جدول ۹. ابعاد شناختی و خوشه های خرده آزمون هندسه درس ریاضی کنکور سال ۹۰.

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

و در نهایت در سال ۹۴، همان طور که در دو سال گذشته نیز بحث شد سؤالات از جنبه های شناختی در ابعاد تقسیم شده اند که در این آزمون با توجه به در بر داشتن سؤالات از هندسه پایه و هندسه تحلیلی می توان 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 می رسند. در مرحله توصیف ابعاد دو خوشه اصلی و چهار زیر خوشه وجود دارد که با مقایسه آنها استنتاج می شود که خوشه اول بیشتر روی مهارت های تحلیل و استدلال تمرکز داشته این در حالی است که خوشه دوم مهارت حل مساله و درک مفاهیم را مورد آزمون قرار می دهد. در ادامه تحلیل سؤالات هر یک از خوشه های اصلی می توان گفت که سؤالات 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 می کنند و در مجموع سؤالات در دو زیر خوشه اولین خوشه در مقایسه با دیگر سؤالات این خرده آزمون در سطح دشوار دسته بندی می شود 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 و صفحه، ماتریس که نسبت به سطوح سؤالات در سطح متوسط و گاهاً ساده سؤال ۱۴۰ مورد ارزشیابی قرار می گیرند حضور دارند.

جدول ۱۰. ابعاد شناختی و خوشه های خرده آزمون هندسه درس ریاضی کنکور سال ۹۴

سؤالات	ابعاد
۱۲۵، ۱۳۰ و ۱۲۶ ۱۳۱، ۱۳۲	مهارت های تحلیل و استدلال
۱۳۳، ۱۳۹	مهارت حل مساله و درک
۱۴۰	مفاهیم

در ادامه بحث برای خرده آزمون جبر در سال ۸۷ نیز به صورتی است که می توان گفت که خوشه اول با مجموع ۵ سؤال محتواهای آموزشی را در این خرده آزمون مورد سؤال قرار می دهد 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 سؤال های ۱۴۵، ۱۴۶ و ۱۵۳ است که به ترتیب از مباحث هم ارزی، مجموعه ها 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، ۱۵۲ و ۱۴۴ که در آن حل مسائل

نیز به چشم می خورد 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 می گیرند.

جدول ۱۱. ابعاد شناختی و خوشه های خرده آزمون جبر درس ریاضی کنکور سال ۸۷

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

بنا بر نتایج، در توصیف ابعاد جبر خرده آزمون جبر سال ۹۰ نیز می توان 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 می شود 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 می شود 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 می کنند 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.



جدول ۱۲. ابعاد شناختي و خوشه‌هاي خرده آزمون جبر درس رياضي کنکور سال ۹۰

سؤالات	ابعاد
۱۵۰ و ۱۴۸، ۱۴۷، ۱۵۲	دقت و سرعت (سخت) استدلال و تحليل روي محتواهاي مختلف
۱۵۴ و ۱۴۶، ۱۵۵	تاکيد بر روي سختي سؤال‌ها + حل مساله
۱۴۴، ۱۵۱	ساده تر، براي ارزيابي از مفاهيم خاص (فاکتوريل و تقسيم) سطح دانش

و در نهايت ابعاد خرده آزمون جبر سال ۹۴ به گونه ای متفاوت از ساير سال‌ها و خرده آزمون‌ها بود، به صورتی که در کل ۱۳ سؤال سؤالات ۱۵۲ به عنوان سخت ترين سؤال از محتوای نظريه اعداد و با در بر گيري سطح بالایی از بعد شناختی که نیاز به تحليل و به کار بستن مفاهيم داشته و نیز آفریدن (خلاقیت)؛ تعريف واحدی برای آفرينندگی به دست نیامده است و به صورت توانایی انديشيدن درباره ی امور به راه‌های تازه و غير معمول و رسيدن به راه حل‌های منحصر به فرد برای مسائل تعريف شده است (سانتروک، ۲۰۰۴)؛ در کنار سرعت به حل اين سؤال کمک بسزایی می کند و در مقابل سؤال ۱۴۹ به عنوان ساده ترين سؤال مفهوم گراف که تنها با دانش مفاهيم گراف قادر به حل بوده است، طرح شده بود. ما بقي سؤالات را که در يک خوشه ديگر 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 در پاسخ گویی به سؤالات این گروه بود. به عنوان مثال سؤالات ۱۴۳، ۱۴۵ و ۱۴۷ که هر یک به گونه ای نیاز به مهارت استدلال مفاهيم مختلف در پاسخ گویی آن‌ها به چشم می خورد و در زیر خوشه دوم نسبت به زیر خوشه نخست سختی بیشتری داشته و بر روي محتواهاي نظريه اعداد طراحی شده اند که این مباحث در حل نیاز مهارت حل مساله و استدلال در مفاهيم 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.

جدول ۱۳. ابعاد شناختی و خوشه‌های خرده آزمون جبر درس ریاضی کنکور سال ۹۴

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

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 (شبیه‌سازی 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 کارآمد 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 multidimensional 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. New York: 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
- Dragow, 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*, 45 (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.