# تفکیک ابعاد متعامد از خوشه‌های سؤال بر اساس هشت روش تعیین بعد در داده‌های دو ارزشی: مورد آزمون ریاضی رشته ریاضی فیزیک کنکور ۹۲-۹۱*

بلال ایزانلو[۱]
عباس بازرگان[۲]
ولی ... فرزاد[۳]
ناهید صادقی[۴]
امیر کاوسی[۵]

تاریخ وصول: ۹۴/۸/۱۳ تاریخ پذیرش: ۹۳/۴//۵

## چکیده

زمینه: مشخص کردن ابعاد (عامل‌ها) موجود در آزمون‌ها اهمیت بسزایی برای سایر اهداف آزمون‌سازی دارد. این موضوع در داده‌های دو ارزشی اهمیت بیشتری داشته و با چالش‌های بیشتری نیز همراه است. هدف: هدف این مطالعه مشخص کردن ابعاد متعامد از خوشه‌های سوال بر اساس مقایسه نتایج هشت روش تعیین ابعاد داده‌های دو ارزشی است. روش: این روش‌ها عبارتند از: روش ناپارامتریک DIMTEST، روش ناپارامتریک DETECT، تحلیل محتوای سؤال‌ها براساس قضاوت متخصصان محتوایی و بررسی جواب‌های آن‌ها، تحلیل عاملی کل اطلاعات، تحلیل خوشه بر اساس زاویه بین بردار سؤال‌ها، تحلیل موازی، آزمون MAP و تحلیل

---

*برگرفته از رساله دکتری با عنوان"مقایسه کارایی مدل‌های جبرانی و غیرجبرانی چندبعدی نظریه سؤال پاسخ در برآورد پارامترهای سوال و فرد: مورد آزمون‌های سراسری ورود به دانشگاه" دانشکده روان‌شناسی، دانشگاه تهران

۱- استادیار گروه تحقیقات آموزشی دانشگاه خوارزمی (نویسنده مسئول) b.ezanloo@gmail.com
۲- استاد دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران
۳- دانشیار دانشکده روان‌شناسی و علوم تربیتی، دانشگاه خوارزمی
۴- استادیار دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران
۵- دانشیار دانشگاه علوم پزشکی شهید بهشتی

عاملی تأییدی. ابتدا روش‌ها به طور مختصر توصیف و سپس براساس داده‌های آزمون ۵۵ سوالی ریاضی رشته ریاضی فیزیک ۹۲-۹۱ مقایسه شده‌اند. یافته‌ها: نتایج تاییدی بر این مطلب است که باید بین تعداد خوشه و ابعاد متعامد تمایز قایل شد. بحث و نتیجه‌گیری: از آنجا که تعداد خوشه‌ها حد بالایی برای تعداد ابعاد است و اکثر روش‌های تعیین ابعاد منعکس کننده خوشه‌ها هستند تا ابعاد متعامد، پس بهتر است بسته به هدف مورد نظر، همراه با ملاحظات منطقی و محتوایی، از چندین روش برای این منظور استفاده کرد.

***واژگان کلیدی:*** خوشه، بعد، داده‌های دو ارزشی، آزمون ریاضی.

**مقدمه**

اندازه‌گیری هر شی یا هر چیز، فقط یک ویژگی شی اندازه‌گیری شده را توصیف می‌کند، این مشخصه کلی تمام اندازه‌گیری‌هاست (ترستون[1]، ۱۹۳۱ ص ۲۵۷). در واقع مقایسه کیفی و کمی افراد در آزمون‌ها تنها تحت شرایط تک‌بعدی بودن[2] معنا خواهد داشت (مک‌نمار، ۱۹۴۶). مسئله بعد[3] با مباحث روایی[4]، کارکرد متفاوت سؤال (تات[5]، ۲۰۰۳)، اعتبار (وکالتی، پنتانن و ترکنه[6]، ۲۰۰۹) و سنجش بدون سوگیری[7] (جیائو[8]، ۲۰۰۴) ارتباط دارد. این موضوع در نظریه کلاسیک نیز به طور غیرمستقیم در قالب اصل استقلال شرطی[9] (ریز[10]، ۱۹۹۹)، همبسته بودن خطاها (زیمرن و ویلیامز[11]، ۱۹۸۰) یا فرض همگنی[12] سوال‌ها مورد اشاره قرار گرفته (مک دونالد[13]، ۱۹۹۹). دلیل اهمیت بیشتر تک‌بعدی بودن در نظریه سؤال پاسخ آن است که همراه

1. Thurstone
2. Unidimensionality
3. Dimensionality
4. Validity
5. Tate
6. Vehkalahti, Puntanen & Tarkkonen
7. Unbiased testing
8. Jiao
9. Conditional (local) independence
10. Reese
11. Zimmerman and Williams
12. Homogeneity
13. McDonald

با دو پیش فرض افزایش احتمال پاسخ درست به سؤال‌ها همراه با افزایش توانایی و مستقل بودن سؤال‌ها پس از کنترل توانایی جزو پیش فرض‌های مدل‌های رایج این نظریه محسوب می‌شود (هابینگ و روسس[1]، ۲۰۰۳)، که نقض شدن آن منجر به سوگیری در برآورد پارامترها می‌گردد (ریکاس[2]، ۱۹۸۵؛ رایز، مور، و مای دو اولیوارس[3]، ۲۰۱۱). علت تاکید بر بحث بعد مشخص کردن ساختار پنهانی است که در طول آزمایش در پاسخ‌گویی به سؤال‌ها موثر است. پژوهش‌گر یا آزمون‌ساز با فرض آگاهی از ساختار پنهان زیربنایی که تفاوت در عملکرد افراد را تبیین می‌کند وارد عمل می‌شود. بر این اساس ساختار پنهان در صورتی به طور کامل مشخص خواهد شد که در هر گروهی از افراد که از نظر تمام صفات پنهان یکسانند متغیرهای مشاهده شده مستقل باشند. به این معنی که اگر متغیرهای تصادفی $X_{ga}, X_{ha}, \cdots$ نشان دهنده اندازه‌گیری فرد $a$ با سؤال‌های مختلف باشد. چون هدف اندازه‌گیری تفکیک جامعه به گروه‌های $G_1, G_2, \ldots$ است، پس با آگاهی از $a \in G_1$، متغیرهای تصادفی $X_{ga}, X_{ha}, \cdots$ از هم مستقل خواهند شد، به طوری که:

$$P(X_{ga} \leq x_{ga}, X_{ha} \leq x_{ha} | a \in G_1) = P(X_{ga} \leq x_{ga} | a \in G_1).P(X_{ha} \leq x_{ha} | a \in G_1)$$

در نتیجه با آگاهی از $a \in G_1$، شناخت ما از نتایج فرد $a$ در یک سوال اطلاعاتی در خصوص عملکردش در سایر سوال‌ها فراهم نمی‌کند. یعنی تمام استنباط‌ها و تفسیرهای روان‌شناسی در مورد فرد براساس آگاهی از عضویت گروهی وی بوده و زمانی که این عضویت مشخص شد اطلاعات بیشتر پیرامون متغیرهای مشاهده به لحاظ روان‌شناسی نامربوطند (نواکوفسکی[4]، ۱۹۸۳). بر این اساس فرض استقلال مشاهدات در گروهی از افراد که صفات پنهان آن‌ها یکسان است، پایه‌ای برای تشخیص این گروه‌ها محسوب شده و از پیامدهای بحث

1. HabingAnd Roussos
2. Reckase
3. Reise, Moore, and Maydeu-Olivares
4. Nowakowska

ابعاد بوده و به خودی خود یک فرض محسوب نمی‌شود (لرد[1]، ۱۹۸۰). برخی افراد نظیر مک دونالد (۱۹۸۱) نیز اصل استقلال موضعی را مقدمه تعریف بعد می‌دانند و برخی دیگر مثل گلدشتاین[2] معتقدند لزوما رابطه‌ای بین استقلال شرطی و تک‌بعدی بودن وجود ندارد و برقراری یکی از آن‌ها دلیل بر وجود دیگری نیست (تروب و ولف[3]، ۱۹۸۱). برخلاف جنبه نظری اصل استقلال شرطی، اهمیت تجربی این اصل طی چند دهه گذشته مورد توجه قرار گرفته است (بل، پاتیسون و وایثر[4]؛ ۱۹۸۸). این اصل مبنای تعریف فضای پنهان است، از این رو در اکثر موارد با تعریف بعد همراه است. به همین دلیل تعاریف جدید بعد به دنبال ارایه تعاریف جدید از این اصل صورت گرفته است.

همگام با چالش‌های نظری در خصوص مفهوم بعد در عمل نیز مشخص کردن ابعاد یک آزمون همیشه چالش انگیز است. به دست آوردن داده‌های تک‌بعدی از آزمون‌های چندبعدی و داده‌های چندبعدی از آزمون‌های تک‌بعدی حاکی از پیچیدگی موضوع بعد دارد و نشان می‌دهد که بازنمایی ماتریس نمرات حاصل از اجرای آزمون‌ها و مشخص کردن تعداد ابعاد لازم برای تبیین روابط متقابل در این ماتریس هم به حساسیت سؤال‌های آزمون نسبت به تفاوت‌های فردی در فضای توانایی (تعداد ابعادی که آزمون قادر است افراد را در آن‌ها تفکیک کند) و هم به میزان تغییرپذیری آزمودنی‌های نمونه در سازه‌هایی که آزمون برای اندازه‌گیری آن‌ها ساخته شده (ابعادی که افراد آزمون دهنده در آن متفاوتند) وابسته است. به همین دلیل مشخص کردن تعداد ابعاد برای بازنمایی ماتریس نمره‌های صفر و یک شده سؤال‌ها همیشه چالش انگیز بوده. متأسفانه از دست دادن غنای مهارت‌های لازم برای پاسخ درست به سؤال‌ها هزینه‌ای است که به خاطر صفر و یک کردن عملکرد افراد در سؤال‌ها پرداخت می‌کنیم و به عبارتی ابعاد روان‌شناسی را فدای راحتی نمره‌گذاری می‌نماییم (ریکاس، ۱۹۹۰).

1. Lord
2. Goldstein
3. Traub and Wolfe
4. Bell, Pattison And Withers

به لحاظ نظری در صورتی که جهت بردار تمام سؤال‌های یک آزمون به یک سمت باشد ماتریس داده‌های حاصل از اجرای آن آزمون با یک بعد نیز مدل‌سازی می‌شود، حتی اگر دستیابی به پاسخ درست نیازمند چندین مهارت شناختی باشد. تحت این شرایط تفاوت افراد نمونه در برخی از مهارت‌های شناختی لازم تأثیر کمی بر روی تعداد ابعاد لازم برای مدل‌سازی داده‌ها دارد. این در حالی است که اگر جهت بردار سؤال‌های آزمون یا جهت بردارهای دسته‌ای از آن‌ها نسبت به بقیه متفاوت باشد، مدل‌سازی ماتریس داده‌های حاصل از اجرای این نوع آزمون تابع ابعاد تفکیک کنندگی سؤال‌ها و ابعاد تغییرپذیری افراد نمونه در سازه‌هایی است که به وسیله آزمون اندازه‌گیری می‌شود (ریکاس، اکرمن و کارلسون[1]، ۱۹۸۸). با در نظر گرفتن تعداد ابعاد تفکیک کنندگی آزمون و تعداد ابعاد تغییرپذیری افراد، تعداد ابعاد لازم برای مدل‌سازی داده‌ها برابر با کمترین هر یک از آن‌ها است. چرا که وجود یک بعد در آزمون و نبود آن در نمونه و بالعکس باعث خنثی شدن اثر آن عامل در مدل‌سازی ماتریس داده‌ها می‌شود (اکرمن، ۱۹۹۴؛ ریکاس، ۱۹۹۰). از سوی دیگر عواملی مثل طول آزمون، حجم نمونه، میزان همبستگی بین ابعاد (پاپ[2]، ۲۰۰۴)، جبرانی یا غیرجبرانی بودن ابعاد (دنگ و انسلی[3]، ۲۰۰۰)، ساختار ساده یا پیچیده سؤال‌ها (مک‌دونالد[4]، ۱۹۸۵؛ گرین[5]، ۱۹۸۳) و وجود عامل حدس (یی[6]، ۲۰۰۷) نیز بر تعداد و ماهیت ابعاد آزمون‌ها بی‌تاثیر نیستند. به همین دلیل تعیین تعداد ابعاد یک آزمون هیچ راه حل روشن و مشخصی که مورد توافق همگان باشد ندارد. این در حالی است که اختلاف نظرها و مشکلات این موضوع در بافت مدل‌های دو ارزشی بیشتر نیز می‌شود. ظاهرا روش‌های مبتنی بر رویکرد تحلیل عاملی و بررسی مانده‌ها بیشتر از همه مورد

1. Reckase, Ackerman& Carlson
2. Popp
3. Deng and Ansley
4. McDonald
5. Green
6. Yeh

استفاده قرار می‌گیرند و تا حدی نیز دارای توجیه منطقی هستند (هتی[1]، ۱۹۸۵). از بین روش‌های پارامتریک تعیین ابعاد در متغیرهای دو ارزشی روش‌های تحلیل عاملی غیرخطی، تحلیل عاملی کل اطلاعات و رویکرد SEM بیشتر از همه مورد استفاده قرار می‌گیرند. یکی از روش‌های مناسب در چارچوب SEM در نرم افزار MPLUS موجود است که با توجه به محدودیت‌ها و ویژگی‌های داده‌های دو ارزشی از روش WLSMV برای برآورد استفاده می‌کند (جاسبر[2]، ۲۰۱۰). در بین روش‌های ناپارامتریک نیز DIMTEST و DETECT جزو رایج‌ترین روش‌ها محسوب می‌شوند. روش‌هایی که بر مفهوم استقلال موضعی و رابطه آن با تک‌بعدی بودن استوارند (مثل Q3، $\chi^2$، $G^2$، نسبت بخت و تبدیل‌های لگاریتمی آن‌ها) یا تحلیل موازی (که در حال حاضر بیشتر برای داده‌های پیوسته استفاده می‌شود) به دلیل عدم در دسترس بودن نرم افزار زیاد رایج نیستند (اسوتینا و لوی[3]، ۲۰۱۲). تعدد روش‌های آماری تعیین بعد و نرم نرم افزارهای مبتنی بر آن‌ها گاهی باعث سر درگمی می‌شود. با وجود تعدد روش‌های مختلف اکثر دست‌اندرکاران معتقدند که همراه کردن روش‌های آماری با بررسی محتوایی و تحلیل‌های منطقی لازمه رسیدن به ابعادی است که گمان می‌رود ابعاد واقعی داده‌ها باشند (نانداکومار و اکرمن[4]، ۲۰۰۴؛ اکرمن، ۱۹۹۶، ۱۹۹۴).

با توجه به مباحث فوق و پیچیدگی علل روابط متقابل در ماتریس نمرات سؤال‌ها منطقی نیست که در مورد تعداد ابعاد مورد نیاز برای مدل‌سازی آزمون فکر کنیم بلکه بهتر است به دنبال مدل‌سازی ماتریس نمرات حاصل از اجرای یک دسته سؤال بر روی یک نمونه خاص باشیم. زیرا ابعاد آماری استخراج شده از آزمون‌ها همیشه به نمونه وابسته بوده و قابلیت تعمیم ندارد، حتی اگر یک آزمون برای نمونه‌های مختلف استفاده شود. از سوی دیگر چون پاسخ درست به هر سؤال به سهم خود متفاوت از سایر سؤال‌ها است پس برای بازنمایی دقیق ماتریس

1. Hattie
2. Jasper
3. Svetina and Levy
4. Nandakumar And Ackerman

داده‌ها به ابعادی برابر با تعداد سؤال‌ها نیاز داریم. ولی چون نتایج حاصل از هر یک از سؤال‌ها فاقد ثبات و اعتبار کافی است، برای رسیدن به نتایج دارای اعتبار و ثبات کافی باید راهی پیدا کرد که با ترکیب اطلاعات مشترک تعدادی از سؤال‌ها به مقیاسی با نتایج دارای اعتبار و ثبات کافی برسیم. تحت این شرایط دو هدف وجود دارد. یک هدف آن است که تعداد خوشه‌های سؤالی که جهت بردارهای آن‌ها به اندازه کافی به هم نزدیک و شبیه است مشخص کرده و به این طریق مقیاسی برای گزارش کردن نتایج پیدا کنیم. هدف دوم آن است که برای مدل‌سازی جایگاه افراد و کارکرد سؤال‌های آزمون تعداد محورهای مختصات عمود بر هم را مشخص نماییم. این در حالی است که این دو هدف یکسان نیستند (ریکاس، ۲۰۰۹). پژوهش حاضر به دنبال آن است که نشان دهد کدام یک از روش‌های تعیین ابعاد داده‌های دو ارزشی منعکس کننده تعداد خوشه‌های سوال و کدام یک نشان‌دهنده ابعاد متعامد مناسب برای مدل‌سازی است. این روش‌ها عبارتند از: ۱) روش ناپارامتریک DIMTEST، ۲) روش ناپارامتریک DETECT، ۳) تحلیل محتوای سؤال‌ها براساس قضاوت متخصصان محتوایی و بررسی جواب‌های ارایه شده به آن‌ها، ۴) تحلیل عاملی کل اطلاعات، ۵) تحلیل خوشه بر اساس زاویه بین بردار سؤال‌ها، ۶) تحلیل موازی، ۷) آزمون MAP و ۸) تحلیل عاملی تأییدی. در ادامه هر یک از این روش‌ها به طور مختصر توصیف و سپس نتایج آنها براساس داده‌های تجربی مورد مقایسه و بحث قرار می‌گیرد.

## روش‌ها

DIMTEST: منطق حاکم بر این روش ساده است. اگر سؤال‌های یک آزمون یک صفت را اندازه‌گیری کنند، آنگاه میانگین کواریانس شرطی بین زوج‌های سؤال به سمت صفر میل می‌کند. توجه کنیم که این تعریف از تک‌بعدی بودن براساس فرض استقلال موضعی ضروری[1] است که حالت ضعیف‌تر استقلال شرطی ضعیف است. خود استقلال شرطی ضعیف نیز حالت

1. Essential unidimensionality

تعدیل شده استقلال موضعی قوی است (ین و فیتزپاتریک[1]، ۲۰۰۶). برای اجرای این روش، سؤال‌های آزمون باید به دو دسته تقسیم شوند. یک دسته سؤالاتی هستند که گمان می‌رود بهترین اندازه‌گیری آن‌ها در جهت مشخص شده به وسیله نمره توانایی ترکیبی مرجع[2] است. نمره توانایی ترکیبی مرجع، ترکیب عناصر بردار توانایی است که بهترین جهت اندازه‌گیری آن در جهت بیشترین تشخیص است، این دسته را آزمون تفکیک شده[3] (PT) گویند. سایر سؤال‌های آزمون که گمان می‌رود بهترین جهت اندازه‌گیری آن‌ها بیشترین تفاوت را با سؤال‌های آزمون تفکیک شده دارد و به اصطلاح چیزی متفاوت از آنها را اندازه‌گیری می‌کنند به آزمون ارزیابی[4] (AT) موسومند. انتخاب آزمون AT به روش‌های مختلفی مثل استفاده از نظر متخصصان، استفاده از تحلیل عاملی و خوشه‌بندی سؤال‌ها بر اساس بهترین جهت اندازه‌گیری آن‌ها در فضای توانایی انجام می‌شود. بسته به روش استفاده شده نتایج حاصل متفاوت خواهد بود. راه حل پیشنهاد شده به وسیله استوت[5] آن است که اجازه دهیم خود این نرم افزار با استفاده از الگوریتم خوشه‌بندی سؤال‌های AT را پیدا کند. برای رسیدن به یک نتیجه متعادل بهتر است از یک نمونه برای پیدا کردن سؤال‌های AT و از نمونه دیگر برای محاسبه آزمون آماری استفاده شود. تحت این شرایط میانگین کواریانس زوج‌های سؤال آزمون AT به شرط نمره افراد در آزمون PT محاسبه می‌شود. آماره حاصل که $T_L$ نام دارد دارای سوگیری مثبت است (یعنی آزمون‌های چندبعدی را تک‌بعدی نشان می‌دهد). بنابراین با استفاده از تکرار مجموعه داده‌های شبیه‌سازی شده‌ای که با داده‌های مشاهده شده همخوان هستند ولی با مدل تک‌بعدی ایجاد شده‌اند اصلاح می‌شود. آماره‌ای که از این نوع داده‌های شبیه‌سازی به دست می‌آید $\overline{T}_G$

1. Yen & Fitzpatrick
2. Composite Reference Score
3. Partitioned Test (PT)
4. Assessment Test (AT)
5. Stout

نام دارد. برای اصلاح سوگیری مثبت، $\overline{T}_G$ برای N مجموعه داده شبیه‌سازی شده محاسبه و سپس میانگین آن به دست می‌آید آماره نهایی برابر است با:

$$T = \frac{T_L - \bar{T}_G}{\sqrt{1 + \frac{1}{N}}} \qquad (۲)$$

معنادار شدن این آماره به رد فرض صفر مبنی بر تک‌بعدی بودن ضروری آزمون منجر می‌شود. اجرای این روش اولین گام توصیه شده در تعیین تعداد ابعاد است (ریکاس، ۲۰۰۹).

DETECT: این روش مبتنی بر چند پیش فرض است که به جز یک استثناء همان پیش فرض‌های مربوط به مدل‌های سؤال پاسخ چندبعدی هستند. این فرض‌ها عبارتند از: افزایش یکنواخت توابع سؤال پاسخ همراه با افزایش هر یک از توانایی‌ها و برقراری استقلال موضعی به شرط توانایی‌ها. با این تفاوت که این استقلال از نوع استقلال موضعی زوجی است، که همان فرض مستثنی شده است. بر این اساس، تعریف بعد در روش DETECT عبارت است از: تعداد ابعاد لازم برای برقرار شدن استقلال موضعی زوجی با استفاده از تابع یکنواخت افزایشی مدل‌های سؤال پاسخ است. تعداد ابعادی که به تحقق شدن این ویژگی منجر شوند ابعاد لازم برای بازنمایی ماتریس نمره‌های خام حاصل از تعامل افراد با سؤال‌ها هستند. روش DETECT به دنبال تعیین خوشه‌های سؤال در فضای چندبعدی است. جهت اندازه‌گیری این خوشه‌های همگون، با جهت اندازه‌گیری شده به وسیله کل تست متفاوت است. در این روش سؤال‌های موجود در یک تست به خوشه‌های مختلفی تقسیم می‌شوند.

تقسیم‌بندی که مقدار DETECT[1] را بیشینه کند ملاک دسته‌بندی سؤال‌ها به خوشه‌های مختلف خواهد بود. مقادیر نزدیک به صفر این آماره حاکی از تک‌بعدی بودن و مقادیر نزدیک یک یا بزرگ‌تر از آن به چندبعدی بودن اشاره دارند. همراه با این شاخص آماره‌های

1. DETECT value

شاخص نسبت[1] و شاخص IDN[2] نیز ارایه می‌شوند. شاخص نسبت بر اساس وارسی روایی[3] مقدار DETEC انجام می‌شود. به این صورت که داده‌ها به دو نمونه تصادفی تقسیم می‌شوند. سپس تقسیم‌بندی که مقدار DETEC را برای نمونه اول بیشینه می‌کند مشخص می‌شود ($D_{max}$). به دنبال آن تقسیم‌بندی که مقدار DETEC را برای نمونه دوم بیشینه می‌کند نیز مشخص می‌گردد. سپس مقدار DETEC با استفاده از داده‌های نمونه اول بر اساس تقسیم‌بندی به دست آمده برای نمونه دوم محاسبه می‌شود ($D_{ref}$). نسبت $D_{ref}$ به $D_{max}$ شاخصی برای تعیین تک‌بعدی بودن است. مقدار یک این نسبت نشان دهنده مشابه بودن تقسیم‌بندی‌های به دست آمده در هر دو نمونه و ثبات راه حل‌های به دست آمده است. مقادیر کوچک‌تر از یک نشان دهنده متفاوت بودن تقسیم‌بندی‌ها است. مقادیر بسیار کوچک این نسبت نشان دهنده این است که تقسیم‌بندی‌ها ناشی از شانس است و احتمالاً به تک‌بعدی بودن ضروری اشاره دارد. شاخص IDN بر اساس تقسیم‌بندی سؤال‌های آزمون به دسته‌های مختلف، نسب کواریانس برآورد شده سؤال‌ها که از خوشه‌بندی سؤال‌ها پیروی می‌کنند را مشخص می‌سازد. انتظار می‌رود که سؤال‌های داخل یک خوشه با یکدیگر دارای کوواریانس مثبت و سؤال‌های خوشه‌های مختلف با یکدیگر کوواریانس منفی داشته باشند. مقادیر یک برای این شاخص حاکی از آن است که الگوی کواریانس موجود بین خوشه‌های مختلف از نظر علامت با آنچه مورد انتظار است هم‌خوانی دارد، یعنی از خوشه‌بندی حاصل تبعیت می‌کند؛ به شرط آن که خوشه‌بندی حاصل (چندبعدی بودن) از ساختار ساده یا تقریباً ساده پیروی نماید (رسس و اوزبک[4]، ۲۰۰۶).

تحلیل محتوا: در این روش سوال‌های موجود در آزمون از نظر فرایندهای شناختی درگیر در پاسخ‌گویی تحلیل می‌شوند. در اکثر موارد طبقه‌بندی محتوایی سوال‌ها برای این منظور ناکارآمد بوده و با طبقه‌بندی آنها براساس توانایی‌های شناختی لازم مطابقت ندارند. این

1. Ratio index
2. IDN index
3. Cross-Validation
4. Roussos & Ozbek

موضوع به خصوص در حوزه آزمون‌های پیشرفت تحصیلی اهمیت دارد. از این رو توصیه بر تحلیل پاسخ‌های ارایه شده به سوال‌ها است تا طبقه‌بندی محتوایی آنها. به عبارتی باید تلاش شود سوال‌ها از نظر فرایندهای شناختی لازم برای پاسخ دادن تحلیل شوند هر چند که تحلیل محتوایی نیز می‌تواند مفید باشد.

تحلیل عاملی کل اطلاعات: تعیین تعداد ابعاد براساس این روش که در نرم‌افزار TESTFACT ارایه شده مبتنی بر رویکرد تحلیل عاملی در نظریه سؤال است. این روش براساس فرض استقلال موضعی قوی تعداد ابعاد موجود در آزمون را مشخص می‌کند. در صورتی که همانند شرایط پژوهش فعلی تعداد الگوهای پاسخ ممکن بیشتر از تعداد افراد موجود در نمونه باشد، می‌توان از تفاوت آماره کای دوی حاصل از راه حل‌های عاملی مختلف به عنوان ملاک تعداد ابعاد لازم برای برقرار شدن فرض استقلال موضعی استفاده کرد. تفاوت مقادیر کای دو خود دارای توزیع کای دو با درجات آزادی برابر با تفاوت درجات آزادی کای دوهای مربوطه است. معنادار بودن آن حاکی از ترجیح راه حل دارای عامل‌های بیشتر بر راه حل دارای عامل‌های کمتر است. در سنجش‌های بزرگ (نظیر مطالعه فعلی) تأثیر تفاوت مکانی پاسخ دهنده‌ها باعث بزرگ شدن این آماره به خاطر اثر خوشه‌بندی می‌شود (دوتویت، ۲۰۰۳، ص ۵۸۸). تحت این شرایط برای قابل توجیه بودن عامل‌های حاصل از این روش پیشنهاد شده قبل از بررسی معنادار بودن، آماره کای دوی حاصل بر ۲ یا ۳ تقسیم شود تا میزان آن کوچک‌تر شود. عامل‌هایی که پس از اعمال این ملاک معنادار باشند می‌توانند به عنوان عامل محسوب گردند. این کار به خاطر از بین بردن اثر طرح[1] انجام می‌شود (کیش، ۱۹۶۵). راه کار دیگر برای کنار آمدن با مسئله بزرگ شدن کای دو که به خاطر مخدوش شدن نمونه‌گیری ساده تصادفی است (زیرا کای دوی آماری بر اساس نمونه‌گیری ساده تصادفی استخراج شده) استفاده از رویکردی است که در SEM مورد استفاده قرار می‌گیرد. یعنی کای دو بر درجات آزادی آن تقسیم می‌شود و در صورتی که مقدار به دست آمده کمتر از ۲ یا ۳ باشد، آن راه

1. design effect

حل عاملی پذیرفته شود. البته این راه حل قطعی نیست. بررسی بار عاملی سؤال‌ها بر روی عامل‌های مربوطه، منطقی بودن تفسیر عامل‌های استخراج شده و در نظر گرفتن این که معمولاً روش تحلیل عاملی کل اطلاعات تعداد عامل‌ها را بیش برآورد می‌کند (چون بر اساس فرض استقلال موضعی قوی پایه‌گذاری شده) همگی باید در مشخص کردن تعداد عامل‌ها لحاظ شوند.

تحلیل موازی: اگر چه این روش توسط هورن[1] (۱۹۶۵) مطرح شد، ولی به دلایلی مثل دشواری‌های محاسباتی همراه با توسعه نرم افزارهای مربوطه و کامپیوترها به تازگی برای تعیین تعداد ابعاد استفاده می‌شود (لدسما، والرو-مورا[2]، ۲۰۰۷؛ یو، پاپ، دیگانگی و جانش-پنل[3]، ۲۰۰۷). در این روش مقادیر ویژه حاصل از ماتریس داده‌های واقعی که از مقادیر ویژه ماتریس داده‌های شبیه‌سازی شده بزرگ‌تر هستند به عنوان تعداد عامل‌های واقعی در نظر گرفته می‌شود. حجم نمونه در داده‌های واقعی و شبیه‌سازی شده یکسان است. به علاوه دشواری سؤال‌ها در داده‌های شبیه‌سازی شده برابر با دشواری سؤال‌ها در داده‌های واقعی است ولی بین سؤال‌ها همبستگی وجود ندارد. در تحلیل حاضر از توابع موجود در نرم افزار R برای اجرای تحلیل موازی (ریول[4]، ۲۰۱۳) استفاده شده است.

آزمون متوسط کمینه تفکیکی (MAP) ولیسر[5]: این روش در هر یک از گام‌های متوالی متوسط همبستگی‌های غیرقطری مجذور شده موجود در یک ماتریس را ملاک انتخاب تعداد ابعاد قرار می‌دهد. استخراج عامل‌ها تا جایی ادامه می‌یابد که متوسط همبستگی تفکیکی مجذور شده غیرقطری به کمترین مقدار برسد (ولسیر[6]، ۱۹۷۶). در تحلیل حاضر از توابع موجود در نرم افزار R برای اجرای این آزمون استفاده شده است (پرساقی و دسیمونی، ۲۰۱۳).

---

1. Horn
2. Ledesma And Valero-Mora
3. Yu, Popp, Digangi And Jannasch-Pennell
4. Revelle
5. Velicer's Minimum Average Partial (MAP) Test
6. Velicer

تحلیل خوشه: در این رویکرد بر اساس خوشه‌بندی متغیرها تعداد خوشه‌ها که حد بالایی برای تعداد ابعاد است تعیین می‌شود. در بین روش‌های تحلیل خوشه در حوزه مدل‌های سوال پاسخ روش خوشه‌بندی سلسله مراتبی وارد[1] (۱۹۶۳) نسبت به سایر روش‌های رقیب بهتر ساختار موجود در آزمون را بازنمایی می‌کند. به شرط آن که درون‌داد زاویه بین بردار سوال‌ها باشد، نه ماتریس همبستگی یا ماتریس فاصله اقلیدسی (ریکاس، ۲۰۰۹). روش وارد در هر گام دو خوشه‌ای که مجموع مجذور درون خوشه‌ای برای خوشه حاصل از ترکیب آن‌ها حداقل مقدار ممکن باشد را با هم ترکیب می‌کند. همبستگی بین هر زوج سؤال برابر با کسینوس زاویه بین سؤال‌ها در فضایی به ابعادی معادل با سؤال‌های آزمون است. پس در مقایسه با هر فضایی با ابعاد کمتر (مثلاً دو یا سه بعد) بهترین شرایط را برای بررسی زاویه بین بردار سؤال‌ها را فراهم می‌کند. در این پژوهش پس از تبدیل ماتریس همبستگی تتراکوریک حاصل از TESTFACT به ماتریس زاویه بین بردار سؤال‌ها، تحلیل خوشه با توابع موجود در نرم افزار R صورت گرفت (تیم هسته R[2]). دومین ملاک مناسب برای قضاوت در مورد تعداد ابعاد لازم بر اساس این روش بررسی زاویه بین بردار سؤال‌ها در فضای دو و سه بعدی است. ضرایب تشخیص حاصل از چرخش وریماکس تمایز بین خوشه‌های سؤال را به وضوح نشان می‌دهند. لذا استفاده از آن‌ها برای تحلیل خوشه بهتر از ضرایب تشخیص مایل یا چرخش نیافته است (کیم، ۲۰۰۱). ابتدا بارهای عاملی سؤال‌ها در فضای دو و سه‌بعدی به شیب تبدیل شدند و سپس تحلیل خوشه براساس هر یک از آن‌ها به صورت جداگانه انجام شد. در صورتی که بین مقادیر زوایه بردارهای حاصل از دو روش تفاوت زیادی نداشته باشد راه حل دارای ابعاد کمتر ترجیح داده می‌شود.

تحلیل عاملی تأییدی: یکی از روش‌های پرکاربرد برای تعیین تعداد ابعاد تحلیل عاملی تأییدی است. برخلاف سایر روش‌ها در این روش پژوهش‌گر براساس دانش نظری یا تجربی

1. Ward
2. R Core Team

ساختار عاملی آزمون را ترسیم و سپس با استفاده از روش‌شناسی معادله ساختاری میزان برازش مدل ترسیم شده با داده‌های مشاهده شده را بررسی می‌کند. با توجه به دو ارزشی بودن داده‌ها و تناسب افزار NOHARM برای تحلیل این داده‌ها در فضای چندبعدی اکتشافی و تأییدی از این نرم افزار برای این منظور استفاده شد (فراسر و مک‌دونالد[1]، ۲۰۱۲).

**نمونه**

کلیه شرکت‌کنندگان گروه ریاضی در آزمون سراسری سال ۱۳۹۲-۱۳۹۱ که مجاز به انتخاب رشته بودند (۲۲۹۷۴۱ نفر) جامعه این پژوهش را تشکیل می‌دهند. در پژوهش حاضر از داده‌های آزمون ۵۵ سوالی چهار گزینه‌ای ریاضی برای مقایسه روش‌های تعیین ابعاد استفاده شد. اکثر پژوهش‌گران هنگام انتخاب حجم نمونه بر چند نکته تاکید دارند که عبارتند از: (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. 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. گروه‌های نامتجانس که کل پیوستار صفت مکنون را می‌پوشانند به نمونه بزرگ‌تری نیاز دارند (۵) تعداد سؤال‌ها، گزینه‌ها و رابطه سوال‌ها با سازه زیربنایی (ارلاندو[2]، ۲۰۰۴). شواهد موجود در پیشینه همگی بر نمونه‌های دست کم به حجم ۱۰۰۰ نفر (به خصوص برای مدل‌های حاوی سه پارامتر) توافق دارند (ین[3]، ۱۹۸۷؛ کیم[4]، ۲۰۰۶). اگر چه براساس نظر متخصصان نمونه‌های ۲۰۰۰ تا ۲۵۰۰ به اندازه کافی بزرگ هستند. ولی با توجه به موارد ذکر شده در فوق و تعدد روش‌های مختلف تعیین ابعاد تصمیم گرفته شد از نمونه‌های ۵۰۰۰ تایی

1. Fraser and McDonald
2. Orlando
3. Yen
4. Kim

استفاده شود. در روش‌هایی مثل DETECT که وارسی روایی در آنها توصیه شده، لذا لازم است از دو نمونه مختلف برای اجرای این روش استفاده شود.

**یافته‌ها**

پاسخ‌گویی به سوال‌های این آزمون ۵۵ سوالی چهار گزینه‌ای نیازمند قدرت و تا حدی سرعت است. چون شیوه تحلیل عاملی که یکی از روش‌های تعیین تعداد ابعاد است فقط برای آزمون‌های توانایی کارایی دارد، تاثیر سرعت و محدویت زمانی در آزمون فعلی می‌تواند نتایج آن را تا حدی مخدوش کند. برای کاهش تاثیر سرعت بر تعداد ابعاد حاصل از تحلیل عاملی از دو راه کار استفاده شد. اول آن که از حجم نمونه بزرگی استفاده شد. دوم آن که در هنگام تحلیل عاملی با استفاده از دستورات مربوطه در نرم افزار TESTFACT کلیه سوال‌های بعد از آخرین سوال پاسخ داده شده توسط افراد به عنوان سوال‌هایی که افراد به خاطر محدودیت زمانی به آنها پاسخ نداده‌اند لحاظ شدند و سوال‌های قبل از این سوال به عنوان پاسخ نادرست معرفی شدند. بررسی اولیه سوال‌ها نشان که سوال ۵۳ با نمره کل رابطه منفی پایینی دارد. به علاوه تحلیل محتوایی صورت گرفته توسط معلمان نشان داد که سوال مورد نظر دارای ابهام محتوایی بوده و نادرست است. در سوال ۵۵ نیز جواب با گزینه کلید همخوانی نداشت. پس این دو سوال کنار گذاشته شدند. بر این اساس سوال‌های تحلیل شده به ۵۳ سوال مورد رسید.

DIMTEST: با توجه به نتایج حاصل از این روش که در جدول ۱ ارایه شده می‌توان فرض تک‌بعدی بودن ضروری را رد کرد و آزمون را چندبعدی در نظر گرفت. براساس توصیه متخصصان پس از اجرای این روش در اولین گام باید از روش DETECT برای تعیین تعداد ابعاد استفاده کرد که در ادامه ارایه شده است.

DETECT: اجرای این روش به سه شاخص منجر می‌شود (جدول ۲). مقدار DETECT: برابر با مقدار ۰/۱۸ حاکی از تک‌بعدی بودن آزمون است. این در حالی که نتایج حاصل از DIMTEST به رد فرض تک‌بعدی بودن منجر شد. این گونه تناقض‌ها در نتایج مربوط به

داده‌های واقعی بعید نیست. ظاهرا تحت شرایطی که ساختار ساده به دلایلی مثل همبستگی بین عامل‌ها یا پیچیدگی ساختار نقض شود نتایج حاصل از DETECT و DIMTEST به راحتی تفسیر نمی‌شود (ناناداکومار و اکرمن، ۲۰۰۴). لازم به ذکر است که این شاخص بر اساس وارسی روایی در دو نمونه ۵۰۰۰ به دست آمده است. ۲) شاخص نسبت: برابر با مقدار ۰/۷۴ است که تا حدی نشان دهنده عدم ثبات در خوشه‌بندی حاصل است. ۳) شاخص IDN: برابر با مقدار ۰/۶۵ نشان دهنده عدم برقراری ساختار ساده بین خوشه‌های موجود در آزمون است. زیرا اگر ساختار سه بعدی مبتنی بر خوشه‌بندی به دست آمده ساده باشد میزان این شاخص برابر یک خواهد بود.

**جدول ۱. نتایج حاصل از روش DIMTEST**

| P | T | TG | TL | AT test | PT test |
|---|---|---|---|---|---|
| ۰/۰۰۰۰۱ | ۷/۷۶۱۲ | ۷/۰۴۹۳ | ۱۴/۸۴۹۲ | ۱-۲-۳-۶-۷-۹-۱۳- | ۴-۵-۸-۱۰-۱۱-۱۲-۱۴-۱۸-۱۹-۲۰-۲۴- |
| | | | | ۱۵ | ۲۵ |
| | | | | ۱۶-۱۷-۲۱-۲۲-۲۳ | ۲۶-۲۷-۲۸-۲۹-۳۰-۳۱-۳۲-۳۳-۳۴-۳۵- |
| | | | | | ۳۶ |
| | | | | | ۳۷-۳۸-۳۹-۴۰-۴۱-۴۲-۴۳-۴۴-۴۵-۴۶- |
| | | | | | ۴۷ |
| | | | | | ۴۸-۴۹-۵۰-۵۱-۵۲-۵۳ |

**جدول ۲. نتایج حاصل از روش DETECT**

| Partition | | | Cross Validated DETECT index | Ratio )Dmax/(Dref | IDN Index |
|---|---|---|---|---|---|
| ۳ | ۲ | ۱ | | | |
| ۱۴-۱۲-۱۱-۸-۷-۵ | ۳۲-۲۷-۲۵-۴ | ۱۰-۹-۶-۳-۲-۱ | ۰/۱۸۳۹ | ۰/۷۴۳۱ | ۰/۶۴۶۹ |
| -۲۴-۲۰-۱۹-۱۸-۱۶ | -۴۹-۴۸-۴۷ | -۲۱-۱۷-۱۵-۱۳ | | | |
| ۲۶ | ۵۰ | ۲۲ | | | |
| -۳۴-۳۱-۳۰-۲۹-۲۸ | ۵۳ | ۳۹-۳۳-۲۳ | | | |
| ۳۵ | | | | | |
| -۴۱-۴۰-۳۸-۳۷-۳۶ | | | | | |
| ۴۲ | | | | | |
| -۵۱-۴۶-۴۵-۴۴-۴۳ | | | | | |
| ۵۲ | | | | | |

تحلیل محتوا: ابتدا به کمک سه دبیر تدریس کننده ریاضی سوال‌ها مورد تحلیل محتوایی قرار گرفت. طبقه‌بندی سوالات از نظر محتوایی نشان داد می‌توان تقریباً آنها را در ۱۰ طبقه تقسیم کرد (جدول ۳). با این حال این نتایج در عمل زیاد کارایی نداشت. زیرا معلمان بیشتر تمایل دارند سؤال‌ها را بر اساس طبقه محتوایی تقسیم‌بندی کنند تا صفات بنیادی لازم برای پاسخ‌گویی. در عین حال تقسیم‌بندی برخی از سؤال‌ها از نظر محتوایی در یک طبقه دشوار بود زیرا ساختار محتوایی برخی از آن‌ها حالت سلسله مراتبی داشت. پس این امکان وجود داشت که آن‌ها را به لحاظ محتوایی همزمان در چندین طبقه قرار داد. برخی از اختلاف نظرها بین معلمان از همین موضوع ناشی می‌شد. مثلاً سؤال به مشتق‌گیری از یک معادله مثلثاتی که حاصل ترکیب دو تابع بود ارتباط داشت.

**جدول ۳. نتایج تحلیل محتوا بر اساس نظر معلمان در آزمون ریاضی**

| طبقه یا خوشه | سوال‌ها |
|---|---|
| تابع و محاسبات جبری | ۱۰۱-۱۰۳-۱۰۵-۱۰۶-۱۰۹-۱۱۲-۱۱۵-۱۲۲ |
| مثلثات و هندسه | ۱۰۲-۱۰۷-۱۱۱-۱۲۰-۱۲۵-۱۲۶-۱۲۷-۱۲۸-۱۲۹-۱۳۰-۱۳۱-۱۳۲-۱۳۳-۱۳۴-۱۳۵-۱۳۶ |
| هندسه تحلیلی | ۱۳۷-۱۳۸-۱۳۹-۱۴۰ |
| مشتق و کاربرد ان | ۱۱۰-۱۱۷-۱۱۸-۱۱۹ |
| حد و پیوستگی | ۱۰۸-۱۱۴-۱۱۶-۱۲۱ |
| انتگرال | ۱۲۳-۱۲۴ |
| ریاضیات گسسته | ۱۴۵-۱۰۴-۱۴۹-۱۵۰-۱۵۱-۱۵۲-۱۵۳-۱۵۴-۱۵۵ |
| آمار | ۱۴۱-۱۴۲ |
| سری | ۱۱۳ |
| جبر و احتمال | ۱۴۳-۱۴۴-۱۴۶-۱۴۷-۱۴۸ |

به دنبال ضعف‌های فوق تلاش شد براساس بررسی روند پاسخ ارایه شده به سوال‌ها و ترکیب نتایج حاصل با اطلاعات حاصل از تحلیل محتوا و خوشه‌بندی به دست آمده از روش‌های DETECT و DIMTEST سؤال‌ها بر اساس فرایندهای شناختی که برای رسیدن به پاسخ نهایی لازم‌اند طبقه‌بندی شدند. نتایج حاصله نشان داد که سوال‌ها در سه دسته کلی قرار می‌گیرند. سوالاتی که نیازمند محاسبات جبری بوده و توانایی خواندن تاثیر عمده‌ای در پاسخ گویی به آنها ندارد. این سوال‌ها که بیشترشان در خوشه اول حاصل از DETECT قرار دارند اکثراً جزو سوال‌های ابتدایی بوده و نسبت پاسخ‌های درست آنها تقریبا زیاد است. دسته دوم سوال‌هایی بودند که خواندن و محاسبات جبری تقریباً به یک اندازه در پاسخ‌گویی به آنها سهیم است. اکثر این سوال‌ها به حوزه احتمالات و هندسه مربوط بوده و همان طور که بعدا مشخص می‌شود اکثرا حدس‌پذیری زیادی دارند. بیشتر این سوال‌ها در خوشه دوم DETECT قرار دارند. دسته سوم که اکثر سوال‌ها را شامل می‌شود بیشتر سوال‌های دشوار و انتهایی را در بر می‌گیرد. در این سوال‌ها خواندن و استدلال نسبت به محاسبات نقش مهمتری دارد. این

سوال‌ها اکثرا در خوشه سوم DETECT قرار دارند. برخی از سوال‌های دسته دوم و سوم به تجسم فضایی و استدلال نیاز داشته و شباهت زیادی به هم داشتند. بنابراین احتمال می‌رود که برخی از آنها همزمان بر روی توانایی خواندن یا محاسبات جبری دارای بار عاملی باشند.

TESTFACT: ماتریس تتراکوریک هموار شده[1] آزمون مربوطه دارای سه ریشه بزرگتر از یک بود، که می‌تواند ملاکی سرانگشتی برای تعیین تعداد عامل‌ها باشد. نتایج حاصل از آزمون کای دو (جدول ۴) حاکی از آن است که حتی پس از تقسیم کای دوها بر ۲ باز هم عامل‌های ۱ تا ۴ معنادارند. نسبت کای دو به درجه آزادی نیز برای عامل پنجم کوچکتر از ۳ شده، که حاکی از چهار بعدی بودن ساختار آزمون است. با تقسیم کای دوها بر ۳ باز هم عامل‌های ۱ تا ۴ معنادارند. با این تفاوت که نسبت کای دو به درجه آزادی دو عاملی بودن ساختار آزمون را نشان می‌دهد. چون این نوع تحلیل عاملی بر اساس فرض استقلال موضعی قوی صورت می‌گیرد، پس تعداد عامل‌های استخراج شده از آن برای رسیدن به این فرض بیشتر از هر روش دیگری خواهد بود. توجه کنیم که فرض استقلال موضعی قوی به عواملی مثل ترتیب ارایه سوال‌ها و هر عاملی که بر اجرا تاثیر گذار باشد حساس است لذا تصمیم نهایی در خصوص تعداد عامل‌ها (۲ یا ۴) باید بر اساس ملاحظلات منطقی و سایر شواهد صورت گیرد.

1. Smoothed

**جدول ٤. نتایج روش تحلیل عاملی کل اطلاعات**

| عامل | Chi Square | ۲Chi Square/ | DF | Chi Difference | DF Difference | CHI/DF |
|---|---|---|---|---|---|---|
| یک | ۶۳۱۰۱/۴۴۰ | ۳۱۵۵۰/۷۲۰ | ۴۶۱۶ | ۱۱۲۹/۹۹۰* | ۵۲ | ۲۱/۷۳۱ |
| دو | ۶۰۸۴۱/۴۶۰ | ۳۰۴۲۰/۷۳۰ | ۴۵۶۴ | ۴۰۵/۹۷۰* | ۵۱ | ۷/۹۶۰ |
| سه | ۶۰۰۲۹/۵۲۰ | ۳۰۰۱۴/۷۶۰ | ۴۵۱۳ | ۱۴۶/۰۶* | ۵۰ | ۲/۹۲۱ |
| چهار | ۵۹۷۳۷/۴۰ | ۲۹۸۶۸/۷ | ۴۴۶۳ | ۱۶۲/۲۱* | ۴۹ | ۳/۳۱۰ |
| پنجم | ۵۹۴۱۲/۹۸ | ۲۹۷۰۶/۴۹ | ۴۴۱۴ | ۶۲/۶۸۵* | ۴۸ | ۱/۳۰۶ |
| ششم | ۵۹۲۸۷/۶۱ | ۲۹۶۴۳/۸۱ | ۴۳۶۶ | | | |

*P<۰/۰۵

بارهای عاملی چرخش پروماکس (هنریچسون و وایت،۱۹۶۴) خوشه‌های سوال نامتجانسی که تشکیل یک بعد را می‌دهند نمایان می‌سازد (دوتویت، ۲۰۰۳). بررسی این بارها در جدول ۵ نشان می‌دهد که خوشه‌های سوال از ساختار پیچیده تبعیت می‌کنند نه ساده. از بین ۵۳ سوال فقط ۳ تا ۴ سوال دارای بار خالص بر روی عامل سوم می‌باشند. این سوال‌ها که دارای بیشترین حدس‌پذیری می‌باشند، جزو سوال‌های خوشه دوم حاصل از DETECT بوده و به حوزه هندسه تحلیلی و احتمالات مربوطند که علاوه بر خواندن و محاسبات تا حدی به استدلال و تجسم فضایی نیز وابسته‌اند. سوالاتی که بار بالایی بر روی عامل اول دارند در خوشه اول قرار می‌گیرند. سوالات خوشه سوم نیز بر روی عامل دوم دارای بار هستند. ترتیب خوشه‌های ۲ و ۳ حاصل از DETECT در نتایج حاصل از چرخش پروماکس تست فکت جابه‌جا شده، این در حالی است که دسته‌بندی کلی حفظ گردیده. برخی از سوال‌های نیز روی دو یا سه عامل دارای بار می‌باشند و تعداد معدودی از آنها نیز روی هیچ یک از سه عامل بار قابل قبولی ندارند. بر این اساس می‌توان گفت اولین عامل منعکس‌کننده غالب‌ترین ویژگی اندازه‌گیری شده توسط آزمون است که تغییرپذیری آن نیز در نمونه زیاد است. ولی عامل‌های بعدی بیشتر تابع تعداد سوال‌ها و فاصله جهت‌دار آن‌ها از عامل اول است، مثلا عامل دوم همان خوشه سوم است که

سوال‌های بیشتری دارد. در عین حال این عامل نسبت به عامل سوم (خوشه دوم) فاصله بیشتری نیز با اولین عامل دارد. همبستگی عامل اول با سوم ۰/۶۴، دوم و سوم ۰/۴۹ و اول و دوم ۰/۵۳ است. بر این اساس می‌توان گفت عامل اول و سوم نسبت به دو عامل اول و دوم (خوشه اول و سوم) به هم نزدیک‌ترند. این نتایج قابل توجه است و می‌تواند در تصمیم‌گیری نهایی در مورد تعداد ابعاد موثر باشد.

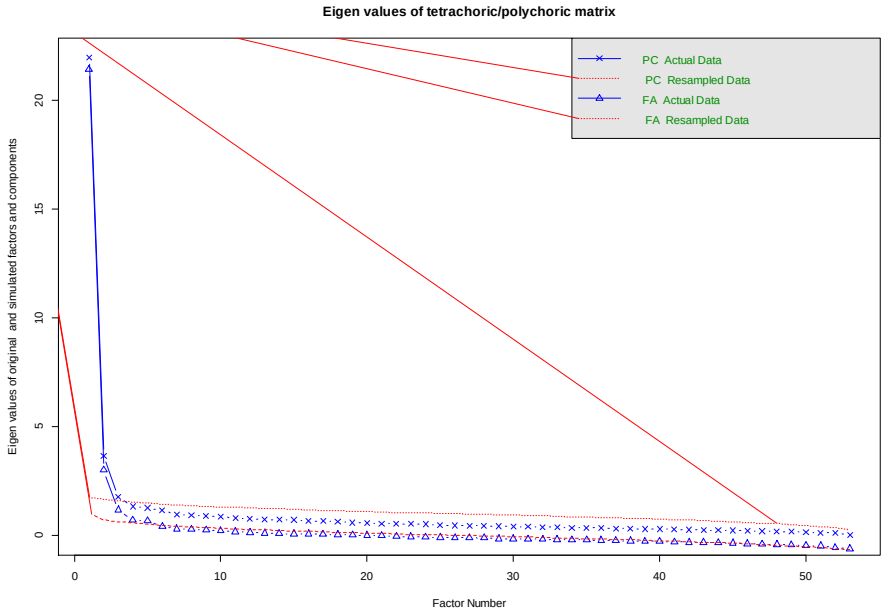
**جدول ۵. بارهای عاملی حاصل از چرخش پروماکس و وریماکس بر روی سه عامل**

| چرخش | پروماکس | | | وریماکس | | | سوال | پروماکس | | | وریماکس | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| سوال | ۱ | ۲ | ۳ | ۱ | ۲ | ۳ | | ۱ | ۲ | ۳ | ۱ | ۲ | ۳ |
| ۱ | ۰/۴۳۴ | -۰/۰۴۹ | ۰/۳۰۷ | ۰/۵۳۶ | ۰/۱۴۰ | ۰/۳۳۳ | ۲۸ | ۰/۰۰۱ | ۰/۴۶۹ | ۰/۳۶۱ | ۰/۲۸۸ | ۰/۵۳۴ | ۰/۳۸۸ |
| ۲ | ۰/۵۹۵ | ۰/۱۳۳ | ۰/۰۱۰ | ۰/۶۰۶ | ۰/۲۸۲ | ۰/۱۳۵ | ۲۹ | -۰/۲۳۳ | ۰/۷۳۷ | ۰/۱۴۲ | ۰/۰۴۰ | ۰/۶۷۵ | ۰/۲۰۲ |
| ۳ | ۰/۷۴۷ | -۰/۳۰۰ | ۰/۲۶۶ | ۰/۷۴۸ | -۰/۰۲۸ | ۰/۳۱۱ | ۳۰ | ۰/۳۱۴ | ۰/۲۲۵ | ۰/۳۱۲ | ۰/۴۹۸ | ۰/۳۷۱ | ۰/۳۶۱ |
| ۴ | -۰/۲۳۶ | ۰/۰۰۹ | ۰/۴۷۱ | -۰/۰۱۲ | ۰/۰۶۳ | ۰/۳۶۵ | ۳۱ | ۰/۰۹۱ | ۰/۳۷۵ | ۰/۳۱۸ | ۰/۳۲۹ | ۰/۴۵۷ | ۰/۳۵۲ |
| ۵ | ۰/۰۷۴ | ۰/۴۸۳ | ۰/۰۲۱ | ۰/۲۰۹ | ۰/۴۸۳ | ۰/۱۱۰ | ۳۲ | ۰/۰۶۲ | ۰/۰۴۳ | ۰/۵۲۶ | ۰/۳۰۵ | ۰/۱۸۴ | ۰/۴۷۱ |
| ۶ | ۰/۵۴۶ | ۰/۰۰۰ | ۰/۲۱۸ | ۰/۶۱۶ | ۰/۱۹۳ | ۰/۲۸۴ | ۰/۳۳ | ۰/۴۰۹ | -۰/۰۹۰ | ۰/۱۳۱ | ۰/۴۲۳ | ۰/۰۵۲ | ۰/۱۷۰ |
| ۷ | ۰/۳۵۶ | ۰/۴۰۴ | -۰/۰۳۶ | ۰/۴۳۰ | ۰/۴۶۷ | ۰/۰۹۸ | ۳۴ | ۰/۴۰۱ | ۰/۳۷۶ | ۰/۰۱۸ | ۰/۴۹۰ | ۰/۴۶۴ | ۰/۱۴۸ |
| ۸ | ۰/۱۳۷ | ۰/۲۶۴ | -۰/۱۰۵ | ۰/۱۵۴ | ۰/۲۶۰ | -۰/۰۲۳ | ۳۵ | ۰/۲۷۸ | ۰/۵۱۳ | -۰/۰۸۳ | ۰/۳۶۵ | ۰/۵۳۸ | ۰/۰۶۲ |
| ۹ | ۰/۷۵۶ | ۰/۰۵۱ | -۰/۰۹۱ | ۰/۶۹۲ | ۰/۲۲۱ | ۰/۰۶۴ | ۳۶ | -۰/۱۷۰ | ۰/۸۱۱ | -۰/۰۷۳ | ۰/۰۲۳ | ۰/۷۰۸ | ۰/۰۴۰ |
| ۱۰ | ۰/۳۰۶ | ۰/۱۶۲ | ۰/۲۱۳ | ۰/۴۲۹ | ۰/۲۸۵ | ۰/۲۶۴ | ۳۷ | ۰/۲۵۴ | ۰/۵۶۸ | ۰/۰۰۶ | ۰/۳۹۶ | ۰/۶۰۶ | ۰/۱۴۳ |
| ۱ | ۰/۱۲۰ | ۰/۴۲۹ | ۰/۱۱۴ | ۰/۲۸۰ | ۰/۴۶۶ | ۰/۱۸۹ | ۳۸ | ۰/۱۳۷ | ۰/۲۲۲ | ۰/۱۳۵ | ۰/۲۵۰ | ۰/۲۷۹ | ۰/۱۷۷ |
| ۱۲ | ۰/۰۲۲ | ۰/۵۴۸ | -۰/۰۱۲ | ۰/۱۶۲ | ۰/۵۲۳ | ۰/۰۸۳ | ۳۹ | ۰/۴۶۰ | ۰/۱۷۷ | ۰/۰۸۳ | ۰/۵۲۱ | ۰/۳۰۷ | ۰/۱۸۲ |
| ۱۳ | ۰/۶۷۴ | ۰/۰۶۸ | ۰/۰۰۵ | ۰/۶۶۱ | ۰/۲۴۰ | ۰/۱۳۵ | ۴۰ | ۰/۰۹۱ | ۰/۵۷۹ | -۰/۰۹۲ | ۰/۲۰۱ | ۰/۵۵۱ | ۰/۰۳۲ |
| ۱۴ | ۰/۲۶۷ | ۰/۳۰۹ | ۰/۰۵۵ | ۰/۳۶۱ | ۰/۳۷۵ | ۰/۱۴۵ | ۴۱ | -۰/۱۵۴ | ۰/۳۸۷ | ۰/۱۷۶ | ۰/۰۳۷ | ۰/۳۷۱ | ۰/۱۸۸ |
| ۱۵ | ۰/۵۴۹ | -۰/۰۹۳ | ۰/۲۴۰ | ۰/۶۰۴ | ۰/۱۱۱ | ۰/۲۸۸ | ۴۲ | -۰/۰۴۱ | ۰/۵۵۸ | ۰/۱۴۰ | ۰/۱۷۳ | ۰/۵۵۳ | ۰/۲۰۵ |
| ۱۶ | ۰/۳۵۰ | ۰/۴۰۵ | -۰/۱۱۶ | ۰/۳۹۰ | ۰/۴۴۷ | ۰/۰۲۹ | ۴۳ | -۰/۴۴۳ | ۰/۶۲۰ | ۰/۰۶۵ | -۰/۲۲۵ | ۰/۴۹۱ | ۰/۰۸۰ |
| ۱۷ | ۰/۸۲۲ | ۰/۰۴۶ | -۰/۱۳۹ | ۰/۷۳۱ | ۰/۲۲۱ | ۰/۰۳۳ | ۴۴ | -۰/۰۵۶ | ۰/۲۲۲ | ۰/۱۵۱ | ۰/۰۷۳ | ۰/۲۳۳ | ۰/۱۵۶ |
| ۱۸ | ۰/۰۸۴ | ۰/۵۲۲ | -۰/۰۳۳ | ۰/۲۰۵ | ۰/۵۰۹ | ۰/۰۷۲ | ۴۵ | ۰/۲۲۰ | ۰/۳۲۲ | ۰/۱۹۳ | ۰/۳۸۱ | ۰/۴۰۹ | ۰/۲۴۷ |
| ۱۹ | -۰/۰۲۶ | ۰/۵۱۲ | ۰/۰۸۳ | ۰/۱۴۹ | ۰/۴۹۹ | ۰/۱۵۰ | ۴۶ | -۰/۰۵۲ | ۰/۳۷۶ | ۰/۳۶۶ | ۰/۲۱۴ | ۰/۴۳۲ | ۰/۳۶۷ |
| ۲۰ | ۰/۳۰۸ | ۰/۳۷۶ | -۰/۰۲۰ | ۰/۳۸۴ | ۰/۴۳۱ | ۰/۰۹۹ | ۴۷ | ۰/۵۷۳ | -۰/۱۴۹ | ۰/۱۰۱ | ۰/۵۴۹ | ۰/۰۳۰ | ۰/۱۶۴ |
| ۲۱ | ۰/۸۹۱ | ۰/۰۱۶ | -۰/۱۹۴ | ۰/۷۶۴ | ۰/۱۹۷ | -۰/۰۰۷ | ۴۸ | ۰/۱۰۰ | ۰/۱۹۷ | ۰/۴۰۵ | ۰/۳۲۹ | ۰/۳۱۱ | ۰/۳۹۹ |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ۲۲ | ۰/۴۳۴ | ۰/۲۸۸ | ۰/۰۷۸ | ۰/۵۲۵ | ۰/۴۰۴ | ۰/۱۹۱ | ۴۹ | -/۰۱۳ | -/۰۴۱ | ۰/۵۰۰ | ۰/۲۰۰ | ۰/۰۷۹ | ۰/۴۲۲ |
| ۲۳ | ۰/۵۷۷ | ۰/۳۵۹ | -/۱۳۷ | ۰/۵۸۳ | ۰/۴۵۶ | ۰/۰۴۳ | ۵۰ | ۰/۲۷۴ | -/۱۱۹ | ۰/۳۳۷ | ۰/۳۷۸ | ۰/۰۳۹ | ۰/۳۱۹ |
| ۲۴ | ۰/۰۷۸ | ۰/۶۱۹ | -/۰۲۲ | ۰/۲۳۰ | ۰/۶۰۳ | ۰/۰۹۶ | ۵۱ | -/۱۲۸ | ۰/۷۰۹ | -/۱۴۱ | ۰/۰۰۵ | ۰/۶۰۶ | -/۰۲۸ |
| ۲۵ | ۰/۰۱۴ | ۰/۰۰۷ | ۰/۶۲۸ | ۰/۲۹۵ | ۰/۱۶۳ | ۰/۵۴۴ | ۵۲ | -/۱۴۴ | ۰/۴۲۴ | ۰/۰۴۶ | -/۰۰۳ | ۰/۳۷۷ | ۰/۰۸۴ |
| ۲۶ | -/۱۹۷ | ۰/۴۴۴ | ۰/۴۵۹ | ۰/۱۳۷ | ۰/۴۸۲ | ۰/۴۳۳ | ۵۳ | ۰/۴۸۷ | -/۴۵۷ | ۰/۳۰۱ | ۰/۴۵۷ | -/۲۳۶ | ۰/۲۷۰ |
| ۲۷ | ۰/۱۴۵ | -/۰۶۱ | ۰/۶۱۵ | ۰/۳۹۶ | ۰/۱۲۹ | ۰/۵۴۵ | | | | | | | |

تحلیل موازی: نتایج حاصل از این روش که در نمودار ۱ ارایه شده نشان می‌دهد که بر اساس تحلیل مولفه‌های اصلی[1] سه عامل و بر اساس تحلیل محورهای[2] اصلی ۵ عامل وجود دارد. ملاک تعیین تعداد عامل‌ها بر اساس این روش تعداد مقادیر ویژه داده‌های واقعی است که از مقادیر ویژه داده‌های شبیه‌سازی شده بزرگترند. این نتایج از ۱۰ بار شبیه‌سازی به دست آمده‌اند.

1. Principal Component Analysis
2. Principal axis Analysis




**نمودار ۱. نتایج حاصل از تحلیل موازی براساس ده بار شبیه‌سازی**

آزمون متوسط کمینه تفکیکی (MAP) ولیسر: نتایج حاصل از آزمون MAP که در جدول ۶ ارایه شده بر اساس همبستگی تتراکوریک که متناسب با شرایط داده‌های این پژوهش است نشان می‌دهد که ساختار آزمون دارای ۵ مولفه اصلی است. این نتایج از تکرار ۱۰ بار شبیه‌سازی به دست آمده‌اند.

**جدول ٦. نتایج آزمون متوسط کمینه تفکیکی ولیسر و توان چهارم آن**

| POLY.MAP.4th | POLY.MAP.squared |
|---|---|
| ۰/۰۳۷۰۹۵ | ۰/۱۶۸۰۸۷ |
| ۰/۰۰۰۸۴۹ | ۰/۰۱۴۹۸۳ |
| ۰/۰۰۰۲۴۷ | ۰/۰۰۸۰۳۸ |
| ۰/۰۰۰۱۷۳ | ۰/۰۰۷۰۰۲ |
| ۰/۰۰۰۱۸۰ | ۰/۰۰۶۹۹۳ |
| ۰/۰۰۰۱۶۱ | ۰/۰۰۶۶۱۴ |
| ۰/۰۰۰۱۹۰ | ۰/۰۰۷۱۳۸ |

تحلیل خوشه: براساس نمودار ۲ تصمیم‌گیری در مورد تعداد خوشه‌ها راحت نیست. محورعمودی به نوعی میزان فاصله بین خوشه‌ها بر اساس درجه را در هر سطح از خوشه‌بندی نشان می‌دهد. تقریبا چهار تا ۵ خوشه کلی وجود دارد. سوالات ۱۵، ۳، ۶، ۱۳، ۲۳، ۹، ۱۷ و ۲۱ که تشکیل دهنده دومین خوشه از سمت چپ هستند همان سوالاتی هستند که در آزمون AT حاصل از DIMTEST و خوشه اول حاصل از DETECT قرار دارند. این سوال‌ها نسبت به بقیه سریعتر تشکیل خوشه داده‌اند. پس می‌توان گفت همگون‌تر از بقیه هستند. بر این اساس می‌توان گفت نتایج روش DETECT بر اساس اصل ساختار ساده، سوال‌های همگونی که نسبت به بقیه زاویه کمتری با هم دارند را مشخص می‌کند. برخی از سوال‌های دومین خوشه از راست نیز تا حدی با سوال‌های دومین خوشه حاصل DETET هم‌خوانی دارند. اکثر سوال‌های اولین خوشه سمت چپ نیز در بین سوال‌های خوشه سوم پیدا می‌شوند. این سوال‌ها فاصله بیشتری با سایر سوال‌ها دارند. به همین دلیل نسبت به خوشه دوم تشکیل عامل مستقل‌تری داده‌اند. در سطح جزیی تقریبا ۹ خوشه در این آزمون وجود دارد. هر چند که تناسبی بین این خوشه‌بندی و طبقه‌بندی سوال‌ها بر اساس نظر معلمان وجود ندارد. در سطح بالاتر می‌توان سوال‌ها را به چهار خوشه تقسیم کرد. بررسی این خوشه‌ها در نمودار دندانه‌ای نشان می‌دهد که دو خوشه سمت راست فاصله کمی از هم دارند و به همین دلیل خیلی سریع با هم ترکیب شده‌اند. این دو خوشه

با کمی فاصله با خوشه سوم از سمت راست ترکیب شده‌اند. این خوشه که همان سوالات AT و اولین خوشه حاصل از DETECT است تفاوت بارزی نسبت به بقیه سوال‌ها دو خوشه سمت راست دارد. ترکیب این سه خوشه با اولین خوشه سمت چپ با فاصله زیادی صورت گرفته. لذا می‌توان ساختار آزمون دو یا سه عاملی در نظر گرفت. در عین حال میانگین زوایه بین بردارها (جدول ۷) برابر با ۶۶/۸۸ است. کمینه زاویه بین بردار سوال‌ها صفر درجه و بیشترین آن ۹۸/۸۱ درجه است که حاکی از پراکندگی نسبتا متوسط در بین آنهاست. انحراف استاندارد زاویه‌ها برابر با ۱۲/۱۴ است. بر این اساس دست کم دوبعدی بودن ساختار آزمون کاملا مشهود است. دومین ملاک مناسب برای قضاوت در مورد تعداد ابعاد لازم بررسی زوایه بین بردار سوال‌ها در فضای دو و سه‌بعدی است. برای این کار از بارهای برآورد شده سوال‌ها با چرخش وریماکس در نرم افزار TESTFACT استفاده شده. ضرایب تشخیص حاصل از چرخش وریماکس تمایز بین خوشه‌های سوال را به وضوح نشان می‌دهند. لذا استفاده از آنها برای تحلیل خوشه بهتر از ضرایب تشخیص مایل یا چرخش نیافته است ( کیم، ۲۰۰۱). ابتدا بارهای عاملی سوال‌ها در فضای دو و سه‌بعدی به شیب تبدیل شدند و سپس تحلیل خوشه بر اساس هر یک از آنها به صورت جداگانه انجام شد. نتایج این تحلیل‌ها حاکی از کاهش میانگین و افزایش پراکندگی بین بردارها است. شاخص‌های توصیفی به دست آمده نشان می‌دهد که به طور متوسط ۸ درجه بین میانگین حاصل از فضای دو و سه بعدی تفاوت وجود دارد (جدول ۷).

**جدول ۷. شاخص‌های توصیفی حاصل از کسینوس همبستگی و فضای دو و سه‌بعدی**

| وضعیت | Mean | SD | Min | Max |
|---|---|---|---|---|
| کسینوس همبستگی | ۶۶/۸۸ | ۱۲/۱۴ | صفر | ۹۸/۸۱ |
| دو بعدی | ۳۰/۵۶ | ۲۲/۷۳ | صفر | ۱۳۱/۹۷ |
| سه بعدی | ۳۸/۶۶ | ۲۱/۳۹ | صفر | ۱۲۸/۴۱ |

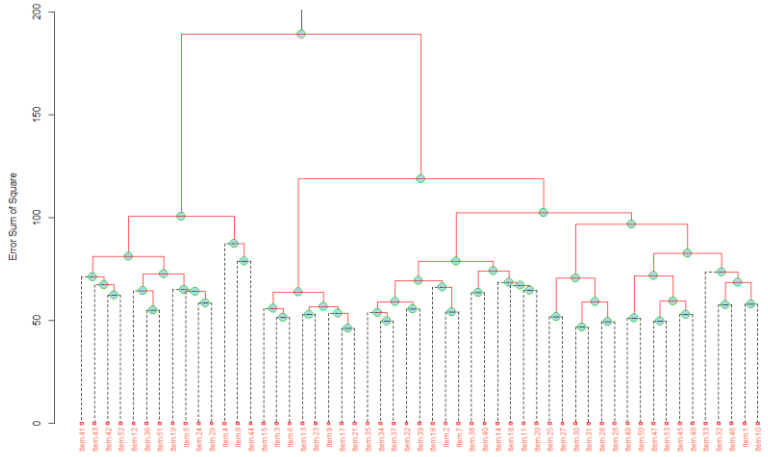
تحلیل عاملی تاییدی با NOHARM: جدول ۸ همبستگی بین نه خوشه حاصل از تحلیل خوشه را نشان می‌دهد. نتایج حاصل از تحلیل تاییدی بر اساس این نوع خوشه‌بندی که در جدول ۹ ارایه شده حاکی از آن است که وقتی ۹ خوشه متمایل باشند عدم برازش به میزان قابل توجهی افزایش می‌یابد. در حالت ۹ خوشه متمایل کاهش قابل قبولی در عدم برازش مدل نسبت به خوشه‌های متمایل رخ داده است. با این حال هنوز خود کای دو و نسبت آن به درجه آزادی نبود برازش را نشان می‌دهند.

**جدول ۸. همبستگی بین نه دسته سوال حاصل از تحلیل خوشه**

| خوشه‌ها | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ |
|---|---|---|---|---|---|---|---|---|---|
| ۱ | ۱ | | | | | | | | |
| ۲ | ۰/۸۶۱ | ۱ | | | | | | | |
| ۳ | ۰/۸۵۷ | ۰/۸۲۵ | ۱ | | | | | | |
| ۴ | ۰/۴۸۰ | ۰/۵۵۴ | ۰/۶۰۳ | ۱ | | | | | |
| ۵ | ۰/۷۴۸ | ۰/۸۲۰ | ۰/۷۴۶ | ۰/۹۰۱ | ۱ | | | | |
| ۶ | ۰/۸۲۰ | ۰/۸۶۵ | ۰/۸۸۰ | ۰/۷۹۸ | ۰/۹۴۸ | ۱ | | | |
| ۷ | ۰/۶۵۶ | ۰/۷۰۶ | ۰/۷۳۱ | ۰/۷۸۱ | ۰/۸۵۹ | ۰/۸۰۰ | ۱ | | |
| ۸ | ۰/۶۸۶ | ۰/۶۲۳ | ۰/۷۹۴ | ۰/۷۴۹ | ۰/۸۴۴ | ۰/۷۸۲ | ۰/۷۸۲ | ۱ | |
| ۹ | ۰/۶۴۵ | ۰/۶۵۶ | ۰/۸۳۸ | ۰/۹۱۶ | ۰/۹۳۶ | ۰/۸۶۸ | ۰/۸۸۴ | ۰/۸۶۸ | ۱ |

در ادامه ترکیب خوشه‌ها براساس نتایج حاصل از تحلیل خوشه صورت می‌گیرد تا روش‌های دیگر مثل همبستگی یا خوشه‌بندی حاصل از نظر ملعمان. همان طور که نمودار دندانه‌ای نیز نشان می‌دهد از سمت چپ خوشه ۵ و ۶ زودتر از سایر خوشه‌ها با هم ترکیب شده‌اند. ترکیب این دو خوشه به کاهش در عدم برازش منجر نشد، بلکه تا حدی نیز آن را افزایش داد. این روند ترکیب خوشه‌ها براساس نزدیکی خوشه‌ها تا سه خوشه نهایی به کاهش

عدم برازش منجر نمی‌شود. لذا بر اساس نتایج تحلیل عاملی تاییدی باید نه عاملی بودن ساختار آزمون را تایید کرد. البته نوع عامل‌ها از نوع متمایل است تا متعامد.




**نمودار ۲. نمودار دندانه‌ای حاصل از تحلیل خوشه مبتنی بر کسینوس زاویه بردارها**

**جدول ۹. آماره‌های برازش مبتنی بر رویکرد SEM**

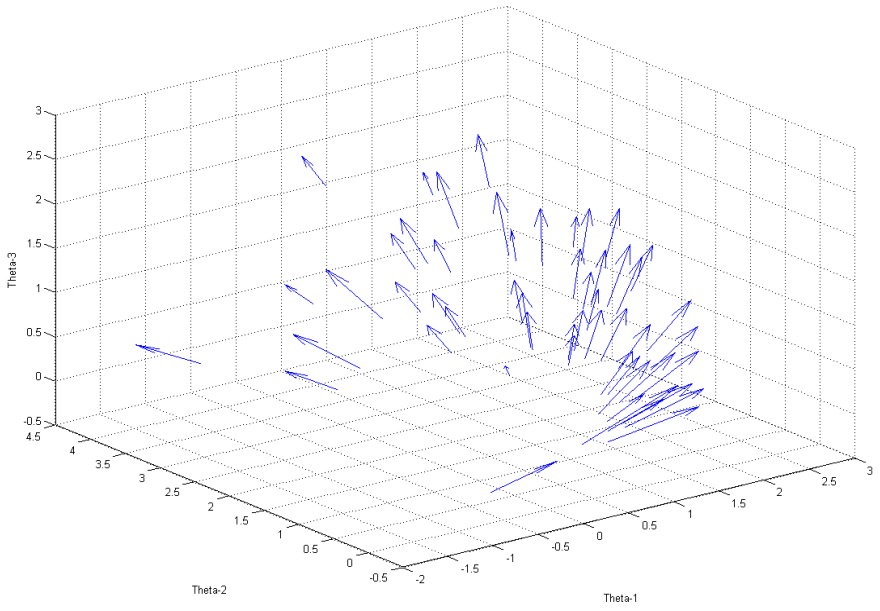
| آماره | ۹متعامد | ۹ مایل | ۸ مایل | ۷ مایل | ۶ مایل | ۵ مایل | ۴ مایل | ۳ مایل |
|---|---|---|---|---|---|---|---|---|
| Chi Square | ۴۳۷۹۸۲/۹۰۴ | ۱۳۱۲۳/۷۹۶ | ۱۳۵۶۱/۰۸۴ | ۱۳۶۱۵/۸۶۹ | ۱۳۹۵۷/۵۷۶ | ۱۴۶۶۱/۷۲۰ | ۱۴۹۵۶/۶۷۰ | ۱۵۹۱۰/۵۴۸ |
| DF | ۱۳۲۵ | ۱۲۸۹ | ۱۲۹۷ | ۱۳۰۴ | ۱۳۱۰ | ۱۳۱۵ | ۱۳۱۹ | ۱۳۲۲ |
| P | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ | ۰/۰۰۰۰۱ |
| Chi/DF | ۳۳۰/۵۵۳ | ۱۰/۱۸۱ | ۱۰/۴۵۶ | ۱۰/۴۴۲ | ۱۰/۶۵۵ | ۱۱/۱۵۰ | ۱۱/۳۳۹ | ۱۲/۰۳۵ |
| RMSEA | ۰/۲۵۷ | ۰/۰۴۳ | ۰/۰۴۳ | ۰/۰۴۳ | ۰/۰۴۴ | ۰/۰۴۵ | ۰/۰۴۵ | ۰/۰۴۷ |
| Tanaka Index | ۰/۵۰۵۳۸ | ۰/۹۸۹۸۱ | ۰/۹۸۹۵۸ | ۰/۹۸۹۵۲ | ۰/۹۸۸۱ | ۰/۹۸۷۱۹ | ۰/۹۸۷۰۶ | ۰/۹۸۶۲۳ |
| RMSR | ۰/۰۲۰۲۳ | ۰/۰۰۲۹ | ۰/۰۰۲۹۴ | ۰/۰۰۲۹۴ | ۰/۰۰۳۱۴ | ۰/۰۰۳۲۶ | ۰/۰۰۳۲۷ | ۰/۰۰۳۳۸ |

## بحث و نتیجه‌گیری

همان طور که بردار سوال‌ها در فضای سه‌بعدی (نمودار ۳) نشان می‌دهد یک دسته از سوال‌ها تقریبا موازی با اولین محور توانایی هستند. اکثر سوال‌های نیازمند محاسبات جبری در این دسته قرار دارند. دومین دسته تقریبا عمود بر صفحه‌اند. اکثر سوال‌های نیازمند توانایی خواندن و محاسبات در این گروه قرار دارند. به علاوه اکثر سوال‌های دارای حدس پذیری در بین دسته اول و به خصوص دوم قرار گرفته‌اند. دسته سوم که جهت بردار آنها به سمت محور سومین توانایی خم شده، اکثرا سوال‌های بسیار دشوار و انتهایی آزمون هستند. احتمالا تاثیر سرعت و دشواری زیاد باعث شده جهت بردار این سوال‌ها تغییر کند. به نظر می‌رسد جهت رو به بالای سوال‌های دسته دوم و به خصوص سوم بیشتر منعکس کننده سرعت و محدودیت زمانی باشد.

جهت متفاوت بردار سوال‌ها در نمودار ۳ حاکی از چندبعدی بودن آزمون است. این موضوع در نتایج DIMTEST، که ضعیف‌ترین شکل تک‌بعدی بودن است نیز منعکس شده است. نتایج حاصل از DETECT که به سه خوشه با ساختار پیچیده و تا حدی بدون ثبات منجر شد بر اساس تحلیل خوشه مبتنی بر کسینوس همبستگی سوال‌ها نیز تایید شد. هر چند که سوال‌های موجود در خوشه‌های دو روش کاملا بر هم منطبق نیست. به علاوه تفاوت بین زاویه بردارها در فضای دو و سه‌بعدی ناچیز بود. زاویه مبتنی بر همبستگی، زاویه بین سوال‌ها در فضایی به ابعاد تعداد سوال‌ها است. بررسی این مقادیر نشان می‌دهد که حداقل زاویه صفر و حداکثر آن ۹۹ درجه است. این موضوع حاکی از تاثیر ناچیز عامل سوم و وجود دو توانایی حتمی است. در کنار این موضوع میانگین زاویه ۶۶/۸۸ نشان می‌دهد که در کل سوال‌ها به دو توانایی وابسته‌اند و همان طور که تحلیل محتوا نیز نشان داد این دو توانایی محاسبات جبری و خواندن هستند. در کنار نتایج دو روش اخیر که تعداد خوشه‌های سوال را نشان می‌دهند. توان اول و چهارم آزمون MAP براساس محورهای اصلی ۵ عامل را نشان داد که بیشتر با تعداد خوشه‌های سطح دوم همخوانی دارند تا عامل‌های مستقل متعامد. نتایج تحلیل موازی نیز بر اساس مولفه‌های اصلی ۳ عامل و بر اساس تحلیل محورهای اصلی ۵ عامل را معرفی کرد.

بر این اساس می‌توان گفت تحلیل موازی براساس مولفه‌های اصلی عامل‌های مستقل بنیادی و بر اساس تحلیل محورهای اصلی با تعداد خوشه‌ها همخوانی دارد. نتایج تحلیل عاملی تاییدی نیز بیشتر تعیین کننده تعداد خوشه‌ها است تا ابعاد مستقل. به علاوه شاخص تاناکه و شاخص‌های مبتنی بر مانده‌ها در اکثر مواقع (به خصوص زمانی که عامل‌ها متعامد هستند) برازش کافی مدل را نشان می‌دهند. این در حالی است که کای دو و نسبت آن به درجه آزادی از عدم برازش حکایت دارند. در حالی که ترکیب خوشه به بدتر شدن برازش منجر می‌شود، شاخص‌های مبتنی بر مانده‌ها این عدم برازش را زیاد منعکس نمی‌کنند. این نتایج نشان می‌دهد استفاده از تحلیل عاملی تاییدی برای تعیین تعداد ابعاد باید با احتیاط به کار گرفته شود و دست کم هر نوع مدل باید حتما بر اساس سایر روش‌ها نیز بررسی گردد.




**نمودار ۳. بردار سوال‌ها در فضای سه‌بعدی**

نتایج حاصل از TESTFACT چهار عاملی بودن را تایید کرد. با توجه به فرض استقلال موضعی قوی این روش به سایر عوامل مثل ترتیب ارایه سوال‌ها، دشواری، سرعت و حدس نیز حساس بوده و تعداد عامل‌ها را بیش برآورد می‌کند. در حالت سه بعدی، عامل اول ۲۹٪، دوم ۶/۵٪ و عامل سوم ۲/۶٪ از واریانس را تبیین می‌کنند. سهم عامل سوم بسیار ناچیز است. با توجه به چندگزینه‌ای بودن آزمون سوال‌های دارای بیشترین حدس عبارتند از: (۰/۰۷۱)۴، (۰/۰۱۴)۱۱، (۰/۰۳۵)۲۵، (۰/۰۱۱)۲۶، (۰/۰۳۱)۳۲، (۰/۰۱۳)۳۳، (۰/۰۱۳)۳۸، (۰/۰۱۷)۴۴، (۰/۰۳)۴۹، (۰/۰۱۳)۵۰ و (۰/۰۱)۵۳. این سوال‌ها علاوه بر این که روی عامل اول یا دوم و یا هر دو دارای بار بالایی هستند بر روی عامل سوم نیز دارای بار می‌باشند. براساس بارهای عاملی چرخش پروماکس سوال‌های ۴، ۲۵ و ۴۹ تنها نشانگرهای خالص عامل سوم محسوب می‌شوند. این در حالی است که بر اساس نتایج چرخش وریماکس که برای انتخاب سوال‌های هر عامل مناسبتر است، فقط دو سوال ۴ و ۴۹ بر روی سومین عامل بار خالص دارند. سوال ۴ که دارای بیشترین میزان حدس‌پذیری است بسیار دشوار نیز می‌باشد. سوال ۴۹ نیز که جزو سوال‌های انتهایی است دارای حدس پذیری زیادی است. لازم به ذکر است که سوال‌های دارای بیشترین حدس پذیری در خوشه دوم و سپس سوم حاصل از DETECT قرار دارند. شش مورد از این سوال‌ها در خوشه دوم، یک مورد در خوشه اول و چهار مورد نیز در خوشه سوم قرار گرفته‌اند. بنابراین تمام خوشه‌ها تا حدی متاثر از حدس هستند. البته این تاثیر در خوشه دوم و سوم بیشتر است. از آنجا که سوال‌های این دو خوشه جزو دشوارترین و انتزاعی‌ترین سوال‌ها هستند، می‌توان گفت افراد در مواجه با سوال‌های سخت دست به حدس می‌زنند. اکثر سوالات دارای حدس به محتوای هندسه و احتمالات مربوط بوده و نیازمند توانایی خواندن و استدلال هستند. میزان حدس‌پذیری سایر سوالات بسیار پایین بود. عواملی مثل حدس پذیری، دشواری و محدودیت زمانی جزو عوامل مؤثر در این آزمون بوده‌اند. دشواری آزمون در کل بالا بوده. اکثر سوالات دشوار و انتهایی آزمون در خوشه سوم یا دوم قرار گرفته‌اند. لذا می‌توان گفت عامل سوم تا حد زیادی ناشی از دشواری، حدس و سرعت (به طور متوسط افراد برای هر سوال

یک دقیقه و پنجا و پنج ثانیه زمان داشته‌اند) است. بر این اساس عامل سوم می‌تواند ناشی از حدس، دشواری و سرعت باشد تا یک ویژگی ذهنی. در نتیجه منطقی است تعداد ابعاد مستقل متعامد موجود در جامعه را دو توانایی در نظر بگیریم تا سه توانایی.

از نظر کاربردی نتایج این پژوهش نشان داد که مشخص کردن تعداد ابعاد بر اساس یک روش بسیار گمراه کننده است. هدف از تحلیل نیز در انتخاب روش مورد نظر موثر است. مثلا در صورتی که هدف پیدا کردن مقیاس برای گزارش کردن نمره‌ها است می‌توان از روش‌های مبتنی بر تحلیل عاملی، MAP، DETECT استفاده کرد. اما اگر هدف بازسازی تعامل افراد و سوال‌ها بر اساس عامل‌های مستقل است می‌توان از روش‌هایی نظیر تحلیل موازی مبتنی بر مولفه‌های اصلی استفاده کرد که در مقایسه با محورهای اصلی عامل‌های مستقل متعامد را مشخص می‌سازد. استفاده از تحلیل خوشه مبتنی بر کسینوس همبستگی و روش‌های گرافیکی و نیز تحلیل محتوایی و شناختی سوال‌ها نیز همیشه باید با سایر روش‌ها همراه شوند، چرا که باعث افزایش بصیرت نسبت به تعداد ابعاد مستقل بنیادی می‌شوند. مقایسه نتایج حاصل از DETECT و تحلیل عاملی حاصل از TESTFACT نشان داد که عامل اول غالب‌ترین عامل در نمونه مورد نظر است که سوال‌های آزمون آن را اندازه‌گیری می‌کنند. این در حالی است که عامل‌های بعدی بیشتر بر اساس تعداد سوال‌ها شکل می‌گیرند. اگر این یافته در مطالعات بعدی نیز اثبات شود حاکی از تاثیر خوشه‌های سوال بر تعداد عامل‌ها است. بر این اساس تعداد خوشه‌های موجود در آزمون‌ها حد بالایی برای ابعاد مستقل واقعی هستند. تعدد روش‌های تعیین بعد باعث گردیده برخی از پژوهش‌گران بر اساس ملاک‌های خاصی الگوریتم‌هایی برای این منظور ارایه دهند (برای مثال جاسبر، ۲۰۱۰). برخی دیگر نیز بر اساس نوع نرم افزار و قابلیت‌های آنها موضوع تعیین تعداد ابعاد را مورد بحث قرار داده‌اند (برای مثال اسوتینا و لوی، ۲۰۱۲). نقطه مشترک همه آنها این است که روش‌های DIMTEST و DETECT به عنوان نقطه شروع مورد استفاده قرار گیرند. سپس بر اساس نتایج آنها در مورد انتخاب‌ها سایر روش‌ها تصمیم‌گیری شود. البته این انتخاب بسته به هدف، امکانات و تمایل پژوهشگر به مشخص کردن دقیق ابعاد

آزمون می‌تواند متفاوت باشد. در حوزه آزمون‌سازی بر اساس این نتایج توصیه می‌شود مشابه این طرح روی سایر آزمون‌های ورود به دانشگاه اجرا شود. به علاوه با عنایت به اهمیت تعداد ابعاد آزمون در تحلیل نتایج، سازمان سنجش دوره‌های آموزشی در این باره برای نویسندگان سوال‌ها طراحی کند، تا از این طریق دست کم تاثیر ابعاد نامربوط بر روی نتایج کاهش یابد.

## منابع لاتین


Ackerman, T. A. (1994). "Using multidimensional item response theory to understand what items and tests are measuring." *Applied Measurement in Education7*(4): 255-278.

Ackerman, T. A. (1996). "Graphical representation of multidmensional item response theory analyses."*Applied Psychological Measurement20*: 311–329.

Deng, H. and T. N. Ansley (2000)."Detecting Compensatory and Noncompensatory Multidimensionality Using DIMTEST."

Bell, R. C., P. E. Pattison, et al. (1988). "Conditional independence in a clustered item test." *Applied Psychological Measurement 12*(1): 15-26.

du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.

Fraser, C., McDonald, R. P. (2012). A Windows program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory.

Green, S. B. (1983). "Identifiability of spurious factors using linear factor analysis with binary items."*Applied Psychological Measurement7*(2): 139-147.

Habing, B. and L. A. Roussos (2003)."On the need for negative local item dependence."*Psychometrika68*(3): 435-451.

Hattie, J. (1985). "Methodology review: assessing unidimensionality of tests and ltenls." *Applied Psychological Measurement9*(2): 139-164.

Hendrickson, A. E. and P. O. White (1964). "Promax: A quick method for rotation to oblique simple structure." *British Journal of Statistical Psychology17*(1): 65-70.

Horn, J. L. (1965). "A rationale and test for the number of factors in factor analysis." *Psychometrika30*(2): 179-185.

Jasper, F. (2010). "Applied dimensionality and test structure assessment with the START-M mathematics test." *International Journal of Educational and Psychological Assessment6*(1): 104-125.

Jiao, H. (2004)."Evaluating the dimensionality of the Michigan English language assessment battery."*Spaan Fellow Working Papers in Second or Foreign Language Assessment Volume 2 20041001*: 27.

Kim, J. P. (2001). *Proximity measures and cluster analyses in multidimensional item response theory*. East Lansing, MI, Unpublished doctoral dissertation, Michigan State University.

Kim, S. (2006). "A comparative study of IRT fixed parameter calibration methods." *Journal of Educational Measurement43*(4): 355-381.

Kish, L. (1965). *SUR VEY SAMPLING*.New York, Wiley.

Ledesma, R. D. and P. Valero-Mora (2007)."Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis."*Practical Assessment, Research & Evaluation12*(2): 1-11.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*, Mahwah, NJ: Erlbaum.

McDonald, R. P. (1981). "The dimensionality of tests and items."*British Journal of mathematical and statistical psychology34*(1): 100-117.

McDonald, R. P. (1985). *Factor analysis and related methods*, Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment*, Lawrence Erlbaum.

McNemar, Q. (1946). "Opinion-attitude methodology."*Psychological Bulletin43*: 289-374.

Nandakumar, R. and T. Ackerman (2004). "Test modeling." *The sage handbook of quantitative methodology for the social sciences*: 93-105.

Nowakowska, M. (1983).*Quantitative psychology: some chosen problems and new ideas.* London: North Holland. .

Orlando, M. (2004). "*Critical issues to address when applying item response theory (IRT) models*."

Popp, E. C. (2004). "*The effects on parameter estimation of sample size ratio, test length and trait correlation in a two-dimensional, two-parameter, compensatory item response model with dichotomous scoring*."

Presaghi, F. and M. Desimoni (2013). *random.polychor*.pa: A Parallel Analysis With Polychoric Correlation Matrices.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria: URL http://www.R-project.org/.

Reckase, M. D. (1985). "The difficulty of test items that measure more than one ability." *Applied Psychological Measurement9*(4): 401-412.

Reckase, M. D. (1990). "*Unidimensional Data from Multidimensional Tests and Multidimensional Data from Unidimensional Tests*."

Reckase, M. D. (2009). *Multidimensional item response theory*, Springer Verlag.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988)."Building a unidimensional test using multidimensional items."*Journal of Educational Measurement25*(3): 193-203.

Reese, L. M. (1999). *A classical test theory perspective on LSAT local item dependence*, Law School Admission Council.

Reise, S., T. Moore, et al. (2011). "Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models." *Educational and Psychological Measurement71*(4): 684-711.

Revelle, W. (2013).psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois. *USA, http://CRAN.R-project.org/package=psych Version = 1.3.9.*

Roussos, L. A., & Ozbek, O. Y. (2006). "Formulation of the DETECT population parameter and evaluation of DETECT estimator bias." *Journal of Educational Measurement43*(3): 215-243.

Svetina, D. and R. Levy (2012)."An Overview of Software for Conducting Dimensionality Assessment in Multidimensional Models."*Applied Psychological Measurement36*(8): 659-669.

Tate, R. (2003). "A comparison of selected empirical methods for assessing the Structure of responses to test items."Applied Psychological Measurement27: 159-203.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, University of Chicago Press.

Traub, R. E. and R. G. Wolfe (1981).Latent Trait Theories and the Assessment of Educational Achievement. *Review of Research of Education*. D. C. Berliner. Washington, DC, American Educational Research Association.9: 377-435.

Vehkalahti, K., Puntanen, S., & Tarkkonen, L. (2009). Implications of Dimensionality on Measurement Reliability.*In Statistical inference, econometric analysis and matrix algebra (pp. 143-160)*, Physica-Verlag HD.

Velicer, W. F. (1976). "Determining the number of components from the matrix of partial correlations."*Psychometrika41*(3): 321-327.

Ward Jr, J. H. (1963). "Hierarchical grouping to optimize an objective function."*Journal of the American statistical association58*(301): 236-244.

Yeh, C.-C. (2007). The effect of guessing on assessing dimensionality in multiple-choice tests: A Monte Carlo study with application, University of Pittsburgh.

Yen, W. M. (1987)."A comparison of the efficiency and accuracy of BILOG and LOGIST."*Psychometrika52*(2): 275-291.

Yen, W. M., & Fitzpatrick, A. R. (2006).Item response theory.*Educational measurement.* R. L. Brennan. Westport, CTPraeger Publishers**:** 111-154.

Yu, C. H., S. O. Popp, et al. (2007). "Assessing unidimensionality: A comparison of Rasch modeling, parallel analysis, and TETRAD."*Practical Assessment, Research & Evaluation12*(14).

Zimmerman, D. W. and R. H. Williams (1980). "Is classical test theory" robust" under violation of the assumption of uncorrelated errors?" *Canadian Journal of Psychology/Revue canadienne de psychologie34*(3): 227.