Educational Measurement and Evaluation Studies

Print ISSN: 2476-2865 Online ISSN: 2783-0942

National Organization for Educational Testing

Comparing Methods of Determining Test Factor Structure Using Empirical data: The Case of National Entrance Exam in 2016

Balal Izanloo

1. Assistant Professor, Faculty of Psychology, Kharazmi University, Tehran, Iran, Corresponding Author izan.b@khu.ac.ir.

Article Info

Article Type:

Research Article

Received:

2020/12/06

Revised:

2021/10/17

Accepted:

2021/11/01

ABSTRACT

Objective: The present study aimed to compare the dimensionality assessment methods using National Entrance Exam data and determine the number of dimensions in the exam's data.

Methods: The data from mathematics (mathematics group), chemistry (experimental sciences group) and Philosophy-logic (humanities group) sub-tests of the National Entrance Exam in 2016 AD (1395 solar) were used for analysis.

Results: Analysis based on 11 methods resulted in 34 related indices and graphical methods, such as hierarchical cluster analysis, exploratory graph analysis and heat map revealed that different methods, depending on their nature, resulted in general factors, specific factors, and a cluster of items. Results showed that the required uni-dimensionality did not exist in most cases, and the structure of the specialized national exam in2016 was bi-factorial. The only difference was that the resulting bi-factor structure did not match the specifications of the previous bi-factor model (i.e., a general factor and several specific factors unrelated to each other and the general factor, so that each item is merely related to one specific factor in addition to the general factor). In other words, besides correlating with the general factor, each item is related to more than one specific factor whose result was a complex or a relatively complex structure. Factor analysis of the total data and nonlinear factor analysis revealed that a gradual increase in lower asymptote reduced the number of dimensions.

Conclusion: It is recommended to apply a combination of methods to find the dimensions of the National Entrance Exam. In addition, the extent of general factor saturation, reflected in item correlations, considering lower asymptote, the way of dealing with omitted responses in analysis, and comparing results of all data with complete data (data without missing values) can be useful for dimensionality assessment. Furthermore, researchers should consider checking the fit of the models extracted from different explanatory methods by confirmatory factor analysis and the interpretation of the extracted model.

Keywords: Dimension, factor, construct, Factor Structure, Multiple Choice Test, binary data.

Cite this article: Izanloo, Balal. (2021). Comparing Methods of Determining Test Factor Structure Using Empirical data: The Case of National Entrance Exam in 2016. *Educational Measurement and Evaluation Studies*, 11 (33): 27-58 pages.

© The Author(s).
Publisher: National Organization of Educational Testing (NOET)

مطالعات اندازه‌گیری و ارزشیابی آموزشی

سازمان سنجش آموزش کشور

شاپا چاپی: ۲۸۶۵-۲۴۷۶ شاپا الکترونیکی: ۰۹۴۲-۲۷۸۳

مقایسهٔ روش‌های تعیین ساختار عاملی آزمون براساس داده‌های تجربی مورد آزمون سراسری ورود به دانشگاه‌ها در سال ۱۳۹۵[۱]

بلال ایزانلو

استادیار دانشکده روانشناسی، دانشگاه خوارزمی، تهران، ایران، نویسنده مسئول izan.b@khu.ac.ir

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

دریافت:
۱۴۰۰/۰۱/۲۲

اصلاح:
۱۴۰۰/۰۷/۱۵

پذیرش:
۱۴۰۰/۰۷/۲۵

چکیده

هدف: هدف این پژوهش مقایسهٔ روش‌های تعیین تعداد ابعاد بر اساس داده‌های آزمون‌های سراسری و تعیین تعداد ابعاد موجود در این داده‌هاست.

روش پژوهش: پس از بررسی پیشینهٔ نظری و تجربی، از داده‌های آزمون‌های ریاضی (گروه ریاضی)؛ شیمی (گروه تجربی)؛ و فلسفه-منطق (گروه انسانی) سال ۱۳۹۵ برای تحلیل استفاده شد.

یافته‌ها: تحلیل داده‌ها با یازده روش تحلیل ابعاد، ۳۴ شاخص مبتنی بر این روش‌ها و نیز روش‌های نموداری تحلیل خوشه سلسله مراتبی، تحلیل شبکهٔ اکتشافی و نقشهٔ حرارتی نشان داد که روش‌های مختلف بسته به ماهیت، عامل یا عوامل کلی، اختصاصی و یا خوشه‌های سؤال موجود در آزمون‌ها را منعکس می‌کنند. نتایج حاصل از تحلیل‌ها نشان داد تک بعدی بودن ضروری (essential unidimensionality) در بیشتر موارد به معنی دقیق آن برقرار نیست و دست‌کم در آزمون‌های تخصصی سراسری سال ۱۳۹۵ ساختار از نوع عامل دوگانه (bifactor) است. با این تفاوت که ساختار دوگانهٔ حاصل با مشخصات مدل دوگانهٔ موجود در پیشینه همخوانی ندارد. یعنی هر سؤال علاوه بر عامل کلی با بیش از یک عامل اختصاصی ارتباط دارد که به ساختار پیچیده یا تا حدودی پیچیده منجر می‌شود.

نتیجه‌گیری: براساس نتایج تحلیل عاملی کل داده‌ها و تحلیل عاملی غیرخطی، افزایش تدریجی مجانب پایین باعث کاهش تعداد ابعاد می‌شود. پیشنهاد می‌شود برای تعیین ابعاد آزمون‌های سراسری از ترکیب چند روش استفاده شود. به علاوه به هنگام تحلیل میزان اشباع عامل کلی، که در میزان همبستگی بین سؤال‌ها منعکس می‌شود، لحاظ کردن مجانب پایین، نوع برخورد با پاسخ‌های سفید و مقایسهٔ نتایج حاصل از کل داده‌ها با داده‌های کامل (داده‌های بدون پاسخ سفید) می‌تواند در این خصوص مفید باشد. به علاوه میزان برازش عامل‌های حاصل از روش‌های مختلف اکتشافی به داده‌ها با استفاده از روش‌های تاییدی نیز بررسی شده و در نهایت به تفسیرپذیری مدل حاصل نیز توجه گردد.

واژه‌های کلیدی: بعد، عامل، سازه، ساختار عاملی، آزمون چندگزینه‌ای، داده‌های دو ارزشی.

استناد: ایزانلو، بلال (۱۴۰۰). مقایسهٔ روش‌های تعیین ساختار عاملی آزمون براساس داده‌های تجربی مورد آزمون سراسری ورود به دانشگاه ۱۳۹۵. مطالعات اندازه‌گیری و ارزشیابی آموزشی، ۱۱ (شماره۳۳)، صفحه ۲۷-۵۸.
ناشر: سازمان سنجش آموزش کشور حق‌مؤلف © نویسندگان.

۱. * این مقاله برگرفته از طرحی پژوهشی با عنوان "روش‌های بهینهٔ تعیین ابعاد زیر بنایی آزمون‌های سراسری سرنوست‌ساز ورود به دانشگاه" به شماره ۱۲/۸۷۳۷۴ است که با حمایت مالی (و از محل اعتبارات پژوهشی) سازمان سنجش آموزش کشور انجام شده است.

مقدمه

مقایسهٔ کیفی و کمی افراد در خصیصهٔ اندازه‌گیری شده با آزمون‌های روان‌شناسی و آموزشی تنها در شرایط تک‌بعدی بودن[1] معنا خواهد داشت (مک‌نمار، ۱۹۴۶). این مسئله از دیرباز مورد توجه بوده و اهمیت آن را افراد زیادی تأکید کرده‌اند. به عنوان مثال، ترســتون[2] (۱۹۳۱) معتقد است «اندازه‌گیری هر چیز، فقط یک ویژگی اندازه‌گیری شدهٔ آن را توصیف می‌کند، این مشخصهٔ کلی تمام اندازه‌گیری‌هاست» (ص ۲۵۷). وجود مزیت‌های ریاضی و آماری در مقیاس‌های تک بعدی باعث شده که مبانی نظری مدل‌های مختلف اندازه‌گیری نیز بر اساس ابزارهایی که خصیصه واحدی را اندازه‌گیری می‌کنند پایه‌گذاری شود. نظریه‌های کلاسیک اندازه‌گیری، سؤال ــ پاسخ (همبلتون[3]، ۱۹۸۹) و مدل‌یابی معادلهٔ ساختاری، هر یک به نوعی بر فرضیهٔ اندازه‌گیری خصیصهٔ واحد تاکید کرده و از روش‌هایی برای درستی این موضوع استفاده می‌کنند (سگرز، ۱۹۹۷؛ جورسکوگ و سوربوم، ۱۹۹۶). بحث تعداد ابعاد[4] آزمون‌ها با مباحث روایی[5]، پایایی[6] (وکالتی، پنتانن و ترکنه[7]، ۲۰۰۹)، کارکرد افتراقی سؤال (تات[8]، ۲۰۰۳) و سنجش بدون سوگیری[9] (جی‌یو[10]، ۲۰۰۴) ارتباط دارد. اگر چه در چهارچوب نظریهٔ کلاسیک اندازه‌گیری موضوع، تعداد ابعاد آزمون‌ها به شــکل امروزی آن مورد توجه نبوده اســت؛ اما به طور غیرمستقیم تأثیر آن بر پارامترهای سؤال‌ها و افراد در قالب اصل استقلال شرطی[11] (ریز[12]، ۱۹۹۹) و همبسته بودن خطاها (زیمرن و ویلیامز[13]، ۱۹۸۰) بررسی شده است. نتایج حاصل از بررسی‌ها در بافت نظریهٔ کلاسیک حاکی از آن است که اگر وابستگی موضوعی در سؤال‌ها زیاد باشد، همبستگی دو رشته‌ای و پایایی، بیشتر از حد واقعی برآورد خواهند شد (ریز، ۱۹۹۹).

مبانی نظری و پیشینهٔ پژوهش

در نظریهٔ صفت پنهان، متغیرهای آشکار یا نشانگرها، تابع صفات پنهان فرض می‌شوند. نشانگرهای مربوط به هر خصیصهٔ پنهان باید دارای ویژگی‌هایی باشند که عبارتند از: الف) هماهنگی نشانگرها با هم که با ضرایب پایایی، به خصوص آلفا، ارزیابی می‌شود. ب) همه نشانگرهای مربوط به یک خصیصه، همان خصیصهٔ پنهان را منعکس کند (روایی هم‌گرا) و در عین حال ج) بتوانند خصیصهٔ مورد نظر را به لحاظ آماری از سایر خصیصه‌های پنهان

1. Unidimensionality
2. Thurstone
3. Hambleton
4. Dimensionality
5. Validity
6. Reliablity
7. Vehkalahti, Puntanen & Tarkkonen
8. Tate
9. Unbiased Testing
10. Jiao
11. Conditional Local Independence
12. Reese
13. Zimmerman and Williams

متمایز کنند (روایی واگرا) و د) واریانس مشترک هر نشانگر با سایر نشانگرهای همان خصیصه، به متغیر پنهان نامشخصی مربوط نباشد (تک‌بعدی بودن) (گفن، ۲۰۰۳). هر یک از این موارد ویژگی‌های متفاوتی هستند که باید با روش‌های مناسبی آنها را بررسی کرد. نادیده گرفتن هر یک از این موارد در ساخت و ارزیابی مقیاس‌ها می‌تواند به نتایج گمراه کننده‌ای منجر شـود. بسیاری از سردرگمی‌های موجود در این خصوص ناشی از عدم تفکیک بین این ویژگی‌ها، به ویژه پایایی، ثبات درونی[1] و همگنی یا تک‌بعدی بودن[2] است (کلارک و واتسون، ۱۹۹۵).

اگر چه تفکیک بین هر یک از این موارد به لحاظ نظری روشـن اسـت، ولی آنچه در عمل بیشتر از همه دیده می‌شود این است که پژوهشگران بیشتر و در بهترین حالت برای ابزار مورد نظر فقط پایایی، روایی هم‌گرا و روایی واگرا را بررسی کرده‌اند و اغلب بدون بررسی ویژگی تک بعدی بودن، نتایج خود را گزارش می‌کنند. گذشته از این، پژوهشگرانی هم که این ویژگی را بررسی کرده‌اند اغلب از روش‌های سنتی نادرستی مانند ضریب آلفا و تحلیل عاملی اکتشافی که برای بررسی پایایی، روایی هم‌گرا و واگرا به کار برده می‌شود، استفاده کرده‌اند و این در حالی است که هیچ یک از این موارد نمی‌توانند تک‌بعدی بودن یا نبودن را مشخص کنند. پژوهشگران زیادی به این موضوع اشاره کرده و بر آن تأکید کرده‌اند(اشمیت، ۱۹۹۶؛ گرین و هرش برگر، ۲۰۰۰؛ ریکوف، ۲۰۰۱). در نظریه‌های جدید اندازه‌گیری (مثل نظریه سؤال-پاسـخ یا مدل‌یابی معادلهٔ سـاختاری) بعدیت[3] آزمون با اصل اسـتقلال موضعی (شـرطی) گره خورده است. این اصل در سه حالت قوی (لرد و نویک، ۱۹۶۸)، ضعیف (مک‌دونالد، ۱۹۸۱) و ضروری (اسـتوت[4]، ۱۹۸۷ و ۱۹۹۰) آن، تعیین کنندهٔ تعاریف مختلف از بعدیت اسـت. بـه همین دلیـل در طی زمان روش‌های زیادی برای بررسـی تعداد ابعاد موجـود در آزمون‌ها به وجود آمده (هتی، ۱۹۸۵)، به تدریج به تعداد این روش‌ها اضافه شده (تات، ۲۰۰۳) و روز به روز نیز در حال افزایش است (استالفسـن و هانوک[5]، ۲۰۰۸؛ لوی و اسـویتنا، ۲۰۱۱، ۲۰۱۲). خلاصه‌ای از روش‌های عمدهٔ تعیین بعدیت، همراه با ویژگی‌های عمده و روش‌های مربوط در جدول (۱) ارایه شده است. تعدد روش‌های آماری تعیین بُعد همراه با آماره‌ها و نرم افزارها مبتنی بر آنها نیز بر سـردرگمی پژوهشگران افزوده است. با وجود تعدد روش‌ها، بیشتر دست‌اندرکاران معتقدند که همراه کردن روش‌های آماری با بررسی محتوایی و تحلیل‌های منطقی، لازمهٔ رسیدن به ابعادی است که گمان می‌رود ابعاد واقعی داده‌ها باشد (اکرمن، ۱۹۹۶، ۱۹۹۴؛ نانداکومار و اکرمن[6]، ۲۰۰۴).

مسـئلهٔ تعیین تعداد ابعاد آزمون از سـه جنبهٔ آماری، روان‌سنجی و ریاضی قابل بررسی است. متأسفانه تمایز

1. Internal consistency Reliability
2. Homogeneity or Unidimensionality
3. Dimensionality(بُعد دار بودن)
4. Stout
5. Stellefson & Hanik
6. Nandakumar And Ackerman

بین این ســه رویکرد حتی در منابع اصلی تحلیل عاملی نیز چندان مورد توجه قرار نمی‌گیرد. تمایز بین این سه رویکرد در خصوص مشکل تعداد عامل‌ها را می‌توان به صورت زیر خلاصه کرد. الف) رویکردهای ریاضی بر ملاحظات وابستگی خطی بین بردارها استوارند و تعداد عناصر مستقلی را بررسی می‌کنند که باید در داده‌ها باشد تا عامل و ضرایب کواریانس_عاملی برآورد شود. چنین ملاحظاتی در کارهای اولیهٔ آلبرت، لیدرمن[1] و در برخوردهای جدید مباحث مربوط به شناسایی نشان داده شده‌اند. ب) در رویکردهای روان‌سنجی روش‌هایی مد نظر است که در آنها خطای اندازه‌گیری و نمونه‌گیری متغیرها بر تعداد عامل‌ها تاثیرگذار است. چنین موضوعاتی در کارهای گاتمن، کایزر و دیکمن[2] که آزمون معروف «ریشهٔ یک» بر آنها استوار است به خوبی نشان داده شده است. ج) آزمون‌های آماری مربوط به شیوه‌های است که در آنها نمونه‌گیری از آزمودنی‌ها بر تعداد عامل‌ها تأثیر دارد. شاید کارهای پیشگامانه لاولی و ماکس‌ول[3] نمونه‌های خوبی از این رویکرد باشند.

امروزه مسئلهٔ تعیین تعداد ابعاد آزمون به طور عمده با کارهای جورسکگ و همکاران[4] در ارتباط است. روش‌های مختلف تعیین تعداد ابعاد را می‌توان بر اساس این سه رویکرد دسته‌بندی کرد. به عنوان نمونه آزمون اسکری و آزمون کای دوی بارتلت[5] که برای تعیین تعداد عامل‌ها استفاده می‌شود منطق یکسانی دارند. چنان که اولی منعکس کنندهٔ تغییرپذیری آماری (نمونه‌گیری آزمودنی) است و دومی معمولاً در برگیرندهٔ تأثیرات روانسنجی (نمونه‌گیری متغیر) است. برای آنکه هر دو روش به نتایج یکسانی منجر شود باید سطح آلفای آزمون بارتلت را به جای سطوح قراردادی مانند ۰/۰۱ به سطوح پایین‌تر (مثلا ۰/۰۰۳ یا پایین‌تر) کشاند. در نمونه‌ای از متغیرها که در یک طرح خاص اســتفاده شده، آزمون اسکری در نمونه‌های کوچک نسبت به نمونه‌های بزرگ معمولا تعداد عامل‌ها را بیش از آنچه هست نشان می‌دهد. در مقابل، آزمون بارتلت در نمونه‌های کوچک، تعداد عامل‌ها را کمتر از آنچه در نمونهٔ بزرگ هست برآورد می‌کند. از این رو، لازم است هنگام استفاده از روش‌های مختلف تعیین تعداد عامل‌ها، به نوع آزمون مورد نظر و اینکه بر منطق ریاضی، آماری یا روانسنجی تعیین تعداد عامل‌ها استوار است، توجه شود (هُرن و اَنگستِرام، ۱۹۷۹).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Albert and Lederman
2. Guttman, Kaiser and Dickman
3. Lawley & Maxwell
4. Jöreskog and His co-workers
5. Bartlett's Chi-Square Test

جدول ۱: چهارچوب تصمیم‌گیری تعیین تعداد ابعاد آزمون براساس ویژگی‌های داده‌ها، فرایند فرضی پاسخ سؤال و رویکردهای ارزیابی بعدیت (اقتباس از لوی و اسوتینا، ۲۰۱۴)

داده‌های دو ارزشی		رویکرد اکتشافی		رویکرد تأییدی	
داده گمشده	مجانب پایین	پارامتری	ناپارامتری	پارامتری	ناپارامتری
بله	بله	●تحلیل عاملی اکتشافی[a] ●تحلیل موازی ●آزمون تفاوت کای دو[a]		●GDDM[f] ●شاخص‌های وابستگی موضعی[b]	
	خیر	●تحلیل عاملی اکتشافی[a,b] ●تحلیل موازی ●آزمون تفاوت کای دو[a]		●کای دوی تعدیل شده[b] ●WRMR[b] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	
خیر	بله	●تحلیل عاملی اکتشافی[a,c] ●تحلیل موازی ●آزمون تفاوت کای دو[a] ●تغییر در RMSR[c]		● تغییر در RMSR[c] ●$\chi^2_{G/D}$[c] ●ALR[c] ●GDDM[f] ●شاخصهای وابستگی موضعی[f]	●آزمون تک‌بعدی بودن برای ساختار ساده[g]
	خیر	●تحلیل عاملی اکتشافی[a,b,c] ●تحلیل موازی ●آزمون تفاوت کای دو[a] ●تغییر در RMSR[c]	●برآورد ساختار ساده چندبعدی[d]	●کای دوی تعدیل شده[b] ●WRMR[b] ● تغییر در RMSR[c] ●$\chi^2_{G/D}$[c] ●ALR[c] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	●آزمون تک‌بعدی بودن ضروری برای ساختار ساده[g] ●ارزیابی ساختار ساده چندبعدی[d]
داده‌های چندارزشی رتبه‌ای		رویکرد اکتشافی		رویکرد تأییدی	
داده گمشده		پارامتری	ناپارامتری	پارامتری	ناپارامتری
بله		●تحلیل عاملی اکتشافی[b] ●تحلیل موازی		●کای دوی تعدیل شده[b] ●WRMR[f] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	
خیر		●تحلیل عاملی اکتشافی ●تحلیل موازی[b]	●برآورد ساختار ساده چندبعدی[e]	●کای دوی اصلاح شده[b] ●WRMR[b] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	●آزمون تک‌بعدی برای ساختار ساده[h] ●ارزیابی ساختار ساده چندبعدی[i]

Note. Entries indicate the examples of such approaches discussed. GDDM=generalized dimensionality discrepancy measure; WRMR=weighted root mean square residual; RMSR=root mean square residual; ALR=approximate likelihood ratio. aTEST-FACT, which supports inputting of lower asymptotes. bMplus. cBased on NOHARM, which supports inputting of lower asymptotes. dDETECT. eNonegligible PolyDETECT. fVia PPMC. gDIMTEST, which supports inputting of a single lower asymptote. hPolyDIMTEST. iPolyDETECT.

از آنجا که ابعاد سؤال و آزمون با مباحث مختلفی مانند همترازسازی[1]، کارکرد افتراقی سؤال[2] و نیز تهیهٔ بانک سؤال در سنجش انطباقی در ارتباط است (رِکیس، ۱۹۹۷)، هر نوع تلاش در جهت استفاده از هر یک از این موارد، به خصوص بانک سؤال، به مشخص شدن تعداد ابعادی که سؤال‌ها اندازه‌گیری می‌کنند وابسته است. مشخص شدن تعداد صفات لازم برای پاسخ دادن به سؤال‌ها، سهم زیادی در دشواری سؤال و سایر آماره‌های سؤال و فرد دارد. در چنین شرایطی آگاهی از ماهیت ابعاد آماری سؤال‌ها و ربط دادن آنها به ابعاد شناختی، هم بینشی برای طراحی سؤال فراهم می‌کند و هم مقدمه‌ای برای سایر کارها از جمله بانک سؤال خواهد بود. اهمیت تک‌بعدی بودن و تعیین تعداد ابعاد آزمون‌ها در موقعیت‌هایی مثل آزمون‌های سراسری ورود به دانشگاه که سازمان سنجش آموزش کشور اجرا می‌کند و هدف اصلی آنها تعیین جایگاه افراد در پیوستار صفت مورد نظر و تصمیم‌گیری در مورد انتخاب آنها است، بارزتر می‌شود. با توجه به تعداد زیاد و روش‌های مختلف تعیین ابعاد، اینکه کدام یک از این روش‌ها، صفات زیربنایی آزمون‌های سراسری را مشخص می‌کنند معلوم نیست. در این باره، پس از بررسی پیشینهٔ نظری و انتخاب روش‌های مناسب تعیین ابعاد، با انتخاب داده‌های تعدادی از آزمون‌های سراسری، تلاش شد تا با روش‌های مختلف تحلیل و مقایسهٔ داده‌ها، کارایی آنها براساس نتایج به دست آمده مشخص شود. با توجه به آنچه ذکر شد این پژوهش به دنبال پاسخگویی به این سؤال‌ها است که چه روش‌هایی برای مشخص کردن تعداد ابعاد داده‌های آزمون‌های چند گزینه‌ای وجود دارد؟ و عملکرد روش‌های مختلف تعیین ابعاد، در داده‌های واقعی چگونه است؟

روش پژوهش

چون هدف روش‌های مختلف تعیین ابعاد، مشخص‌کردن عوامل زیربنایی مؤثر در تبیین کواریانس بین سؤال‌های آزمون است، این پژوهش جزو پژوهش‌های توصیفی- همبستگی محسوب می‌شود و به دلیل استفاده از داده‌های جمع‌آوری شده توسط سازمان سنجش، داده‌ها از نوع ثانویه هستند. از آنجا که داده‌های حاصل از اجرای آزمون سراسری ورود به دانشگاه در گروه‌های سه‌گانهٔ (علوم ریاضی، تجربی و انسانی) در سال ۱۳۹۵برای تحلیل استفاده شد، جامعه آماری این مطالعه از یک طرف شرکت کنندگان در آزمون سراسری سال ۱۳۹۵و از طرف دیگر آزمون‌های اجرا شده در این سال هستند. به دلیل تعداد زیاد آزمون‌های سراسری و ملاحظهٔ محدودیت زمانی و مالی، از بین آنها تعدادی آزمون انتخاب شد[3]. با این دلایل و با وجود تعداد زیاد روش‌های بررسی ابعاد و نیز با توجه به اهمیت درس‌های اختصاصی، از بین درس‌های تخصصی هر یک از سه گروه اصلی انسانی، تجربی و ریاضی، آزمون درس فلسفه- منطق علوم انسانی با ۲۵ سؤال، شیمی تجربی با ۳۵ سؤال و ریاضی گروه ریاضی-فیزیک با ۵۵ سؤال برای تحلیل انتخاب شد. برای تحلیل از داده‌های ده هزار نفری

1. Test Equating
2. Differential Item Function

۳. لازم به یادآوری است که فقط در آزمون سراسری سال ۹۵ برای سه گروه انسانی، تجربی و ریاضی به ترتیب ۹، ۶ و ۴ آزمون اختصاصی و ۴ آزمون عمومی خاص هر گروه اجرا شده است (در کل ۳۱ آزمون). اگر آزمون‌های مربوط به گروه هنر و زبان‌های خارجی را نیز در نظر بگیریم تعداد آزمون‌های سراسری اجرا شده در سال ۹۵-۹۴ بیشتر است. از این رو به دلیل محدودیت زمانی و مالی امکان تحلیل همه این آزمون‌ها برای پژوهش‌گر میسر نبود.

که سازمان سنجش آموزش کشور در اختیار پژوهشگر قرار داد استفاده شد.

در هر آزمون ابتدا افرادی که همهٔ سؤال‌های آزمون را سفید[1] گذاشته بودند حذف شدند. به علاوه با توجه به ماهیت آزمون‌های سراسری، در هر آزمون، تعداد زیادی داده گمشده[2] (M)، یعنی سؤال‌هایی که افراد به آنها پاسخ نداده‌اند، وجود دارد. وجود این نوع داده‌ها به چند دلیل است. اول آنکه در برخی سؤال‌ها افراد از تعیین پاسخ درست ناتوانند، پس سؤال را بدون پاسخ رها می‌کنند. دوم اینکه برخی افراد به دلیل ترس از نمرهٔ منفی، به سؤال‌هایی که از گزینهٔ درست آن مطمئن نیستند پاسخ نمی‌دهند. سوم این که برخی افراد به دلیل محدودیت زمانی برخی سؤال‌ها را (به خصوص سؤال‌های آخر هر خرده آزمون را) بدون پاسخ می‌گذارند. البته بررسی پاسخ‌های سفید نوع اخیر در چند درس نشان داد که احتمال از دست دادن داده‌ها در این حالت سوم بسیار کم است، چرا که توزیع پاسخ‌های سفید بعد از آخرین سؤال پاسخ داده شده تقریباً در بازه صفر تا ۲ در نوسان بود، که نشان می‌دهد دست کم در آزمون‌های استفاده شده در این پژوهش محدودیت زمانی در وقوع پاسخ‌های از دست رفته نقش چندانی نداشته است. به طور منطقی در دو حالت نخست، ترس از نمره منفی باید افراد را از حدس تصادفی پاسخ درست باز دارد، ولی احتمال حدس مبتنی بر نوعی دانش نسبی با حذف یک یا چند گزینه و انتخاب گزینه درست در بین گزینه‌های باقی مانده وجود دارد. در پیشینه به سه نوع داده گمشده اشاره شده است: داده‌های گمشده کاملاً تصادفی[3] (MCAR)، داده‌های گمشدهٔ تصادفی[4] (MAR) و داده‌های گمشدهٔ غیرتصادفی[5] (MNAR). دو مورد اول را می‌توان نادیده گرفت (یعنی می‌توان با آنها طوری برخورد کرد که دارای الگوی تصادفی هستند)، ولی مورد آخر قابل اغماض نیست (روبین و لیتل[6]، ۲۰۰۲). اگر چه تا به امروز مطالعات کافی در مورد نحوه برخورد با داده‌های از دسته رفته آزمون‌های سراسری وجود ندارد، ولی با توجه به اینکه ماهیت این نوع داده‌ها در آزمون‌های سراسری از نوع داده‌های گمشدهٔ غیرتصادفی (MNAR) و در نتیجه قابل چشم پوشی نیست و نتایج برخی مطالعات حاکی از آن است که در نظر گرفتن این نوع داده‌ها به عنوان پاسخ‌های نادرست به سوگیری در برآورد پارمترهای فرد و سؤال منجر می‌شود (رُز، وان‌داویر و زو[7]، ۲۰۱۰)، تصمیم گرفته شده در صورت امکان در تحلیل ابعاد به عنوان پاسخ نادرست کدگذاری نشوند. البته روش‌های تعیین ابعاد، بر اساس لحاظ کردن یا نکردن داده‌های گمشده در تحلیل متفاوتند. مثلا در هر آزمون داده‌های افرادی که به همهٔ سؤال‌ها پاسخ داده بودند (با عنوان داده‌های کامل) برای تحلیل ابعاد با روش‌های خاصی انتخاب شدند که در ادامه به آنها اشاره شده است. روشن است که تعداد افراد دارای داده‌های کامل در هر آزمون کم است. قبل از تحلیل ابعاد، پارامترهای سؤال هر آزمون با مدل سه پارامتری،

1. Omit
2. Missing
3. Missing Completely at Random
4. Missing at Random
5. Missing Not at Random
6. Rubin and Little
7. Rose, Von Davier & Xu

که معمولا در داده‌های آزمون‌های سراســری نسبت به مدل‌های یک و دو پارامتری بیشترین برازش آماری را با داده‌ها دارد، برآورد شد. برآورد پارامترهای سؤال با مدل سه پارامتری به دو صورت انجام شد. اول زمانی که مقادیر گمشده[1] به عنوان داده از دست رفته[2] (MNA) به نرم افزار معرفی شدند. دوم زمانی که داده‌های گمشده به عنوان پاســخ نادرست[3] (MO) کدگذاری شدند. برآورد پارامترهای ســؤال با روش اول در مقایسه با زمانی که داده‌های گمشده به صورت نادرست کدگذاری شده بود، به برآورد مجانب پایین (حدس) بیشتری منجر می‌شود (جدول ۱). به علاوه مقایسهٔ برآورد پارامترهای مدل سه پارامتری در هر دو حالت نیز نشان داد، در کل وقتی پاسخ‌های سفید به عنوان نادرست کدگذاری می‌شوند، در مقایسه با زمانی که پاسخ‌های سفید به عنوان داده از دست رفته کدگذاری می‌شوند، دشواری سؤال‌ها افزایش، شیب کاهش و مجانب پایین بسیار کمتر می‌شود. در نهایت، با توجه به اینکه تعداد ابعاد استخراج شده از داده‌های آزمون‌های سراسری براساس نتایج برخی روش‌ها بسیار بیشتر از حد مورد انتظار اســت، در این پژوهش تعداد خوشه‌های سؤال هر آزمون حد بالای تعداد ابعاد قابل استخراج در نظر گرفته شده است (رِکی‌س[4]، ۲۰۰۹).

با توجه به تعداد زیاد روش‌های بررســی ابعاد در پیشــینه، از روش‌های زیر برای تعیین تعداد ابعاد آزمون‌های سراسری استفاده شد: تحلیل عاملی اکتشافی و تاییدی، تحلیل عاملی غیرخطی، تحلیل عاملی کل اطلاعات، روش‌های ناپارامتریک، تحلیل موازی، روش‌های مبتنی بر اســتقلال موضعی، تحلیل شبکه و سایر روش‌های نموداری. برای اجرای تحلیل عاملی اکتشافی و تأییدی از روش‌های موجود در نرم افزار MPLUS (ماتن و ماتن[5]، ۲۰۱۵)، که مناسب داده‌های دو ارزشی است استفاده شد. برای تحلیل عاملی غیرخطی از NOHARM (فراسر و مک‌دونالد[6]، ۲۰۰۳) برای تحلیل عاملی مبتنی بر کل اطلاعات و روش‌های مبتنی بر استقلال موضعی از بسته mirt (چالمرز[7]، ۲۰۱۲) در R، برای روش‌های مبتنی بر رویکرد ناپارامتریک از DIMPACK (رُسُــس و استوت[8]، ۲۰۰۷)، برای تحلیل موازی از بسته psych (ریول[9]، ۲۰۱۸)، برای تحلیل شبکه از بسته EGA در R (هودسون و گولینو[10]، ۲۰۱۹) و برای سایر روش‌های نموداری از R (تیم هسته R[11]، ۲۰۱۸) استفاده شد. به دلیل گستردگی بیش از حد تحلیل‌ها و کمبود فضا نتایج در جدول‌ها خلاصه شده‌اند و از ارائهٔ اطلاعات جزئی صرف نظر شده است. برای اطلاعات بیشتر در خصوص تحلیل‌ها به ایزانلو[12] (۲۰۱۹) مراجعه شود.

1. Missing
2. Not Available(NA)
3. Uncorract
4. Reckase
5. Muthén & Muthén
6. Fraser, C., & McDonald
7. Chalmers
8. Roussos, L. A., & Stout
9. Revelle
10. Hudson F. Golino
11. R Core Team
12. Izanloo

جدول ۲: آماره‌های توصیفی پارامتر مجانب پایین سؤال‌های آزمون ریاضی، شیمی و فلسفه-منطق در داده‌های حاوی پاسخ‌های گمشده

آماره / آزمون	میانگین	میانه	انحراف استاندارد	کمینه	بیشینه
ریاضی	۰/۱۶۸	۰/۱۸۱	۰/۰۸۵	۰/۰۰۲	۰/۳۲۷
شیمی	۰/۱۷۹	۰/۱۹۳	۰/۰۷۴	۰/۰۱۳	۰/۲۹۶
فلسفه-منطق	۰/۱۶۲	۰/۱۹۲	۰/۰۸۴	۰/۰۰۲	۰/۲۸۰

یافته‌ها

چه روش‌هایی برای مشخص کردن تعداد ابعاد داده‌های آزمون‌های چندگزینه‌ای وجود دارد؟

بررسی پیشینه نشان داد روش‌های زیادی برای بررسی ابعاد آماری داده‌های حاصل از آزمون‌های چند گزینه‌ای، که معمولاً از نوع داده‌های طبقه‌ای و گسسته هستند، وجود دارد. این روش‌ها بر اساس ویژگی‌های زیر طبقه‌بندی می‌شوند و لازم است پژوهشگر قبل از تحلیل با در نظر گرفتن این موارد روش‌های مناسب مربوط به داده‌های خود را انتخاب کند. این ویژگی‌ها عبارتند از: نوع داده‌ها (دو ارزشی یا چندارزشی بودن)، گمشدگی (بودن یا نبودن داده‌های گمشده یا از دست رفته)، مجانب پایین (حدس‌پذیری یا حدس‌ناپذیری در سؤال‌های آزمون)، مدل آماری (پارامتریک در برابر ناپارامتریک) و نوع تحلیل (اکتشافی در برابر تاییدی). برای آگاهی از جزئیات بیشتر در مورد روش‌ها، همراه با شرایط استفاده و نرم افزار لازم برای اجرای هر یک از آنها جدول(۱) را ببینید.

عملکرد روش‌های مختلف تعیین ابعاد در داده‌های واقعی چگونه است؟

در آزمون ریاضی، شیمی و فلسفه-منطق نتایج حاصل از ۱۱ روش تعیین ابعاد و ۳۴ شاخص استفاده شده به نتایج زیر منجر شد که در ادامه به تفکیک در جدول‌های ۳ تا ۵ ارائه شده است. نتایج، بر اساس تعداد ابعاد دسته‌بندی شده‌اند تا تصمیم‌گیری در خصوص روش‌ها راحت شود.

الف) آزمون ریاضی

یک عامل: در تحلیل موازی، شاخص ساختار ساده با پیچیدگی ۱ (VSS.C1)؛ در تحلیل عاملی اکتشافی شاخص‌های CFI، RMSEA و TLI؛ در تحلیل عاملی کل اطلاعات با مجانب پایین ۰/۱ و ۰/۱۵، شاخص BIC و باز در تحلیل عاملی کل اطلاعات با مجانب پایین مدل سه پارامتری، شاخص‌های تفاوت کای دو، SABIC، HQ، BIC، AIC و AICc؛ در تحلیل عاملی غیرخطی روی داده‌های کامل (بدون داده گمشده) با مجانب صفر، ۰/۱، ۰/۱۵ و مجانب پایین مدل سه پارامتری شاخص RMSR؛ در آزمون DIMTEST با مجانب پایین صفر و ۰/۱ (البته سطح معناداری روش ناپارامتری DIMTEST در هر دو مورد برابر ۰/۰۵۰۹ بود که بسیار نزدیک به آلفای ۰/۰۵ است).

دو عامل: در تحلیل موازی، شاخص ساختار ساده با پیچیدگی ۲ (VSS.C2)، مقادیر ویژهٔ تحلیل عاملی مشترک در

داده‌های واقعی و مقادیر ویژهٔ تحلیل مولفه مبتنی بر نمونه‌گیری از داده‌های واقعی؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر، شاخص‌های BIC و SABIC. در تحلیل عاملی کل اطلاعات با مجانب پایین ۰/۱، شاخص SABIC و در تحلیل عاملی کل اطلاعات با مجانب پایین ۰/۱۵، آزمون تفاوت کای دو، AIC، SABIC، HQ و AICc؛ در تحلیل عاملی غیر خطی با مجانب پایین حاصل از مدل سه پارامتری شاخص کاهش RMSR؛ تحلیل خوشه سلسله مراتبی روی ماتریس عدم شباهت مبتنی همبستگی فای و تتراکوریک، هر یک با دو خوشهٔ کلی منحصر به فرد. سه عامل: در تحلیل موازی مقادیر ویژهٔ تحلیل عاملی مشترک مبتنی بر نمونه‌گیری از داده‌های واقعی؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب ۰/۱ و ۰/۱۵ شاخص کاهش RMSR، و نیز شاخص تاناکا در مجانب پایین صفر، ۰/۱، ۰/۱۵ براساس ملاک بزرگ‌تر یا مساوی ۰/۹.

جدول ۱: خلاصه نتایج روش‌های مختلف بررسی ابعاد داده‌های آزمون ریاضی

روش	شاخص		نتیجه	روش	شاخص		نتیجه
تحلیل موازی	SBIC		کم ۱۰ دست عامل	تحلیل عاملی غیرخطی با داده‌های کامل	$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب صفر	۱ عامل
	BIC		کم ۱۰ دست عامل		کاهش در RMSR		۴ عامل
	MAP		۴ عامل		Tanaka Index		۳ عامل قابل قبول و ۷ عامل خوب
	VSS.C1		۱ عامل		$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب ۰/۱	۱ عامل
	VSS.C2		۲ عامل		کاهش در RMSR		۳ عامل
	های مقادیر ویژه داده مشاهده شده	عامل	۲ عامل		Tanaka Index		۳ عامل قابل قبول و ۶ عامل نزدیک به خوب
	های مقادیر ویژه داده شبیه‌سازی		۲۰ عامل		$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب ۰/۱۵	۱ عامل
	گیری مقادیر ویژه نمونه از داده‌ها		۳ عامل		کاهش در RMSR		۳ عامل
	مقادیر ویژه داده‌های مشاهده شده	مولفه	۸ عامل		Tanaka Index		۳ عامل قابل قبول و ۷ عامل نزدیک به مرز خوب
	مقادیر ویژه داده‌های شبیه‌سازی		۷ عامل		$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب مدل 3PL	۱ عامل
	گیری مقادیر ویژه نمونه از داده‌ها		۲ عامل		کاهش در RMSR		۲ عامل
					Tanaka Index		با افزایش عامل‌ها به مرز قابل قبول نزدیک شد

روش	شاخص	نتیجه	روش	شاخص	نتیجه
تحلیل عاملی اکتشافی	آزمون تفاوت کای دو	دست‌کم ۸ عامل	DETECT	$Detect_{cross\ validated}$	۴ خوشه
	RMSEA	۱ عامل			
	فاصله اطمینان ۹۰٪ اطراف RMSEA	۴ عامل		Ratio(R)	بی‌ثباتی ساختار ۴ خوشه‌ای
	CFI	۱ عامل			
	TLI	۱ عامل		IDN	ساختار چندبعدی نسبتا پیچیده
	WRMR	۸ عامل			

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

روش		شاخص	نتیجه	روش	شاخص	نتیجه
تحلیل عاملی کل اطلاعات	مجانب صفر	آزمون تفاوت کای دو	۸ عامل	DIMTEST	مجانب پایین صفر	تایید تک‌بعدی ضروری
		BIC	۲ عامل			
		HQ	۴ عامل		مجانب پایین ۰/۱۰	تایید تک‌بعدی ضروری
		SABIC	۲ عامل			
		AICc	۸ عامل		مجانب پایین ۰/۱۵	رد تک‌بعدی ضروری و تغییر خوشه‌بندی سؤال‌ها
		AIC	۸ عامل			
	مجانب ۰/۱	آزمون تفاوت کای دو	۴ عامل	Q3)استقلال موضعی(	IRTمدل‌های	وابستگی موضعی ضیف
		BIC	۱ عامل			
		HQ	۴ عامل			
		SABIC	۲ عامل		نمره خام	وابستگی موضعی نسبتا بالای برخی سؤال‌ها
		AICc	۴ عامل			
		AIC	۴ عامل			
	مجانب ۰/۱۵	آزمون تفاوت کای دو	۲ عامل	نقشه حرارتی	همبستگی فای	با دامنه ۰/۰۹- تا ۰/۶
		BIC	۱ عامل		همبستگی تتراکوریک	با دامنه ۰/۱- تا ۰/۸
		HQ	۲ عامل	تحلیل خوشه سلسله مراتبی	همبستگی فای	دو خوشه کلی و تقریبا ۸ خوشه جزئی یا بیشتر
		SABIC	۲ عامل		همبستگی تتراکوریک	دو خوشه کلی و تقریبا ۱۱ خوشه جزئی یا بیشتر
		AICc	۲ عامل	تحلیل شبکه	کل داده‌ها	۴ خوشه
		AIC	۲ عامل		داده‌های کامل	۴ خوشه
	مجانب مدل 3PL	آزمون تفاوت کای دو	۱ عامل	تحلیل عاملی تاییدی	ساختار دو عاملی (یک عامل کلی و چهار عامل اختصاصی)	چهار عاملی تحلیل عاملی اکتشافی بهترین برازش را با داده‌ها به دست آورد
		BIC	۱ عامل			
		HQ	۱ عامل			
		SABIC	۱ عامل			
		AICc	۱ عامل			
		AIC	۱ عامل			

چهار عامل (۴ خوشه سؤال): در تحلیل موازی، روش MAP؛ در تحلیل عاملی اکتشافی با فاصله اطمینان ۹۰ درصدی اطراف RMSEA؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر آزمون HQ و با مجانب پایین ۰/۱ آزمون تفاوت کای دو، AIC ،HQ و AICc؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب صفر شاخص کاهش RMSR؛ آزمون DETECT با چهار خوشهٔ بی‌ثبات و ساختار چند بُعدی نسبتاً پیچیده؛ تحلیل شبکهٔ اکتشافی برای داده‌های کامل (بدون داده گمشده) و کل داده‌ها.

شش عامل: در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب پایین ۰/۱، شاخص تاناکا براساس ملاک بزرگ‌تر یا مساوی ۰/۹۵.

هفت عامل: در تحلیل موازی مقادیر ویژه تحلیل مولفه براساس داده‌های شبیه‌سازی؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب پایین صفر، ۰/۱۵ و مجانب پایین مبتنی بر مدل سه پارامتری، شاخص تاناکا براساس ملاک بزرگ‌تر یا مساوی ۰/۹۵.

هشت عامل: در تحلیل موازی مقادیر ویژه تحلیل مولفه در داده‌های واقعی؛ در تحلیل عاملی اکتشافی آزمون تفاوت کای دو؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر، آزمون تفاوت کای دو همراه با AIC و AICc؛ تحلیل خوشهٔ سلسله مراتبی برای ماتریس عدم شباهت مبتنی بر همبستگی فای.

ده عامل و بیشتر: در تحلیل موازی شاخص‌های SBIC و BIC و مقادیر ویژه تحلیل عاملی مشترک داده‌های شبیه‌سازی شده.

آزمون DIMTEST در شرایط مجانب پایین ۰/۱۵ فرض تک بعدی بودن ضروری را با آلفای ۰/۰۵ رد کرد (البته همان طور که یادآوری شد، این آزمون در مجانب پایین صفر و ۰/۱ نیز دارای سطح معناداری ۰/۰۵۰۹ بود که با آلفای ۰/۰۵ در مرز رد شدن قرار داشت). شاخص تاناکا در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل در شرایط مجانب پایین حاصل از مدل سه پارامتری با افزایش تعداد عامل‌ها به مرز قابل قبول نزدیک شد. شاخص‌های استقلال موضعی مبتنی بر مدل‌های یک، دو و سه پارامتری حاکی از وابستگی مثبت ضعیف سؤال‌ها است، ولی براساس نمرهٔ خام وابستگی موضعی مثبت نسبتاً بالایی بین سؤال‌ها وجود دارد. همبستگی فای و تتراکوریک بین سؤال‌ها تقریباً متوسط به بالا است. به جز یک یا دو سؤال که با سؤال‌های دیگر همبستگی نزدیک به صفر دارند، همبستگی سؤال‌ها مثبت و تقریباً متوسط به پایین است. با توجه به تنوع محتوایی گستردهٔ آزمون ریاضی و همبستگی تقریباً متوسط سؤال‌های آن با هم و نیز ماهیت ریاضی که با هوش و استدلال منطقی رابطهٔ بالایی دارد، می‌توان گفت روش‌هایی که در این آزمون منعکس کنندهٔ یک عامل هستند بیشتر عامل غالب آزمون را مشخص می‌کنند و در مقابل، روش‌های منعکس کننده دو عامل، عوامل کلی را منعکس می‌کنند. در نهایت روش‌هایی که از وجود ۳ تا ۴ عامل حمایت می‌کنند، بیشتر خوشه‌های سؤال موجود در آزمون و عوامل اختصاصی را نشان می‌دهند. روش‌ها و شاخص‌هایی که از وجود ۶، ۷ و ۸ عامل

حمایت می‌کنند ممکن است ترکیب عوامل مزاحم و اختصاصی را نشان دهند.
روش‌هایی که بیش از ۸ عامل را نشان می‌دهند احتمالاً منعکس کنندهٔ عوامل آماری موجود در آزمون هستند. در تحلیل عاملی تأییدی، بهترین برازش با ساختاری دو عاملی (یک عامل کلی و چهار عامل اختصاصی) حاصل از چهار عامل، مبتنی بر تحلیل عاملی اکتشافی به دست آمد. لازم به یادآوری است که ساختار دو عاملی بر اساس نتایج سایر روش‌های اکتشافی (مثل DETECT و تحلیل شبکه اکتشافی) نیز بهترین برازش را با داده‌ها داشت. به همین دلیل به نظر می‌رسد آزمون ریاضی متشکل از یک عامل کلی و چهار عامل اختصاصی است. در کنار این موارد باید به نقش مجانب پایین نیز اشاره کرد. مطابق نتایج تحلیل عاملی کل اطلاعات که برای کل داده‌ها (داده‌های حاوی مقادیر گمشده بودند و نیز کامل) صورت گرفت، گنجاندن مجانب پایین در تحلیل، به تدریج به کاهش در برآورد تعداد عامل‌ها بر اساس شاخص‌های موجود در این روش منجر می‌شود. حتی در تحلیل عاملی غیر خطی که در آن داده‌های کامل (بدون داده گمشده) مورد استفاده قرار گرفته‌اند، لحاظ کردن مجانب پایین، منجر به کاهش تعداد عامل‌ها می‌شود (شاخص کاهش RMSR در تحلیل عاملی غیرخطی را ببینید). به نظر می‌رسد مجانب پایین در مدل سه پارامتری، که می‌تواند منعکس کنندهٔ نقش عوامل غیرشناختی و حدس (به خصوص حدس مبتنی بر دانش نسبی) در پاسخگویی به سؤال‌ها باشد، نقش مهمی در کاهش تعداد عامل‌های ناشی از عوامل مزاحم یا بی اهمیت دارد. به علاوه مقایسهٔ نتایج حاصل از تحلیل‌های مبتنی بر کل داده‌ها با تحلیل‌های مبتنی بر داده‌های کامل از این نظر می‌تواند مفید باشد که عدم همخوانی نتایج چه از نظر تعداد عامل‌ها و چه از نظر متفاوت بودن الگوهای به دست آمده، به وجود زیرگروه‌های متفاوتی که تحت تاثیر عوامل مختلف به شیوه گوناگونی به سؤال‌ها پاسخ می‌دهند اشاره دارد. این موضوع برای روش‌های بررسی ابعاد پیامدهای مهمی به همراه دارد.

ب) آزمون شیمی

یک عامل: در تحلیل موازی، آزمون MAP، آزمون ساختار ساده با پیچیدگی ۱ (VSS.C1) و مقادیر ویژهٔ تحلیل عاملی مشترک در داده‌های واقعی؛ در تحلیل عاملی اکتشافی شاخص‌های CFI، RMSEA و TLI؛ در تحلیل عاملی کل اطلاعات با مجانب پایین مبتنی بر مدل سه پارامتری، آزمون تفاوت کای دو، SABIC، HQ، BIC، AIC و AICc؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب صفر و مجانب پایین مربوط به مدل سه پارامتری، شاخص RMSR و شاخص تاناکا براساس ملاک بزرگ‌تر یا مساوی ۰/۹.

دو عاملی: در تحلیل موازی آزمون ساختار ساده با پیچیدگی ۲ (VSS.C2)؛ در تحلیل عاملی اکتشافی با فاصله اطمینان ۹۰ درصدی اطراف RMSEA؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب پایین صفر و مجانب پایین مبتنی بر مدل سه پارامتری، شاخص تاناکا براساس ملاک بزرگ‌تر یا مساوی ۰/۹۵ و شاخص کاهش در RMSR در حالت مجانب پایین مبتنی بر مدل سه پارامتری؛ در تحلیل خوشهٔ سلسله مراتبی برای

ماتریس عدم شباهت، مبتنی بر همبستگی فای و تتراکوریک هر یک با دو خوشهٔ کلی خاص.
سه عاملی: فقط تحلیل خوشهٔ سلسله مراتبی برای ماتریس عدم شباهت مبتنی بر همبستگی تتراکوریک با دست کم سه خوشهٔ جزئی.
چهار عامل (۴ خوشه سؤال): در تحلیل عاملی کل اطلاعات با مجانب پایین صفر، شاخص BIC و SBIC؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب پایین صفر، شاخص کاهش در RMSR؛ روش Detect با چهار خوشه دارای ساختار پیچیدهٔ نسبتاً بی‌ثبات- خوشه‌های جزئی در تحلیل خوشهٔ سلسله مراتبی برای ماتریس عدم شباهت مبتنی بر همبستگی فای؛ تحلیل شبکهٔ اکتشافی مبتنی بر کل داده‌ها.
پنج عامل: در تحلیل موازی مقادیر ویژه مبتنی بر نمونه‌گیری از داده‌های واقعی در تحلیل مولفه.
شش عامل: در تحلیل موازی مقادیر ویژه مبتنی بر نمونه‌گیری از داده‌های واقعی در تحلیل عاملی مشترک، مقادیر ویژه داده‌های شبیه‌سازی در تحلیل مولفه.
هفت عاملی: در تحلیل موازی مقادیر ویژه داده‌های واقعی در تحلیل مولفه- تحلیل شبکه اکتشافی مبتنی بر داده‌های کامل. البته در تحلیل موازی نتایج آزمون SBIC و BIC با دست کم ۷ عامل نیز هم در قسمت هفت عاملی و هم هشت عاملی قرار می‌گیرد.
هشت عامل و بالاتر: در تحلیل موازی نتایج آزمون BIC، SBIC و مقادیر ویژهٔ تحلیل عاملی مشترک در داده‌های شبیه‌سازی شده؛ در تحلیل عاملی اکتشافی آزمون تفاوت کای دو و WRMR؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر، آزمون تفاوت کای دو همراه با AICc، HQ و AIC.
آزمون DIMTEST در شرایط مجانب پایین صفر، ۰/۱ و ۰/۱۵ فرض تک بعدی بودن ضروری را با آلفای ۰/۰۵ رد کرد. اگر چه نتایج بررسی استقلال موضعی براساس شاخص Q3 مبتنی بر مدل‌های یک، دو و سه پارامتری نقض ضعیف این پیش فرض را نشان نداد، ولی نتایج این شاخص بر اساس نمره‌های خام حاکی از نقض نسبتاً بالای این پیش فرض است. توزیع مقادیر شاخص Q3 مبتنی بر مدل سه پارامتری حاکی از کجی منفی این آماره است که نشان می‌دهد سؤال‌های این آزمون نسبت به هم صفات متفاوتی را اندازه‌گیری می‌کنند. این موضوع در همبستگی پایین بین سؤال‌های آن نیز منعکس شده است.
در آزمون شیمی نتایج مربوط به نقشهٔ حرارتی همبستگی بین متغیرها، حاکی از همبستگی کم آنها با هم است. بر اساس نتایج تحلیل شبکه اکتشافی، چهار خوشه وجود دارد که خوشهٔ چهارم فقط شامل دو سؤال است و از این رو به لحاظ عملی نمی‌تواند عاملی تشکیل دهند. نتایج تحلیل خوشهٔ سلسله مراتبی به روش **وارد** روی ماتریس عدم شباهت مبتنی بر همبستگی‌های تتراکوریک و فای نشان داد که دو خوشه کلی در این آزمون وجود دارد. سؤال‌های آزمون AT (سؤال‌هایی که نسبت به سؤال‌های PT، که بیشتر سؤال‌های آزمون را تشکیل می‌دهند، بُعد متفاوتی را اندازه‌گیری می‌کنند) در دو خوشهٔ کلی حاصل از هر دو نوع همبستگی پخش شده‌اند. خوشه‌بندی

حاصل از تحلیل خوشـــه سلســـله مراتبی روی ماتریس عدم شـــباهت مبتنی بر همبستگی تتراکوریک و فای متفاوت است. به این معنی که خوشه‌بندی مبتنی بر همبستگی تتراکوریک از دو خوشه کلی و سه خوشه جزئی حمایت می‌کند ولی خوشه‌بندی مبتنی بر همبستگی فای از دو خوشه کلی و چهار خوشه جزئی حکایت دارد.

جدول ۲: خلاصه نتایج روش‌های مختلف بررسی ابعاد داده‌های آزمون شیمی

روش	شاخص		نتیجه	روش	شاخص		نتیجه
تحلیل موازی	SBIC		کم ۷دست عامل	های کامل تحلیل عاملی غیرخطی با داده	$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب صفر	۱ عامل
	BIC		کم ۷دست عامل		کاهش در RMSR		۴ عامل
	MAP		۱ عامل		Tanaka Index		۱ عامل قابل قبول و ۲ عامل خوب
	VSS.C1		۱ عامل		تحلیل نشد	مجانب ۰/۱	-
	VSS.C2		۲ عامل		تحلیل نشد		-
	عامل	های مقادیر ویژه داده مشاهده شده	۱ عامل		تحلیل نشد		-
		های مقادیر ویژه داده شبیه‌سازی	۱۵ عامل		تحلیل نشد		-
		گیری مقادیر ویژه نمونه از داده‌ها	۶ عامل		تحلیل نشد	مجانب ۰/۱۵	-
	مولفه	مقادیر ویژه داده‌های مشاهده شده	۷ عامل		تحلیل نشد		-
		مقادیر ویژه داده‌های شبیه‌سازی	۶ عامل		$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب مدل 3PL	۱ عامل
		گیری مقادیر ویژه نمونه از داده‌ها	۵ عامل		کاهش در RMSR		۲ عامل
					Tanaka Index		۱ عامل قابل قبول و ۲ عامل خوب
تحلیل عاملی اکتشافی	آزمون تفاوت کای دو		۸ عامل	DETECT	$Detect_{cross\ validated}$		۴ خوشه
	RMSEA		۱ عامل				
	فاصله اطمینان ۹۰٪ اطراف RMSEA		۲ عامل		Ratio(R)		بی‌ثباتی ساختار ۴ خوشه‌ای
	CFI		۱ عامل				
	TLI		۱ عامل		IDN		ساختار چندبعدی نسبتا پیچیده
	WRMR		بیش از ۸ عامل				

روش		شاخص	نتیجه	روش	شاخص	نتیجه
تحلیل عاملی کل اطلاعات	مجانب صفر	آزمون تفاوت کای دو	۸ عامل	DIMTEST	مجانب پایین صفر	رد تک‌بعدی ضروری
		BIC	۴ عامل			
		HQ	۸ عامل		مجانب پایین ۰/۱۰	رد تک‌بعدی ضروری
		SABIC	۴ عامل			
		AICc	۸ عامل		مجانب پایین ۰/۱۵	رد تک‌بعدی ضروری
		AIC	۸ عامل			
	مجانب ۰/۱	تحلیل نشد	–	Q3)استقلال موضعی(	IRTمدل‌های	وابستگی موضعی منفی ضعیف
		تحلیل نشد	–			
		تحلیل نشد	–			
		تحلیل نشد	–		نمره خام	وابستگی موضعی مثبت بالای برخی سؤال‌ها
		تحلیل نشد	–			
		تحلیل نشد	–			
	مجانب ۰/۱۵	تحلیل نشد	–	نقشه حرارتی	همبستگی فای	با دامنه ۰/۰۳- تا ۰/۵
		تحلیل نشد	–		همبستگی تتراکوریک	با دامنه ۰/۱- تا ۰/۷
		تحلیل نشد	–	تحلیل خوشه سلسله مراتبی	همبستگی فای	دو خوشه کلی و دست‌کم ۴ خوشه جزئی
		تحلیل نشد	–		همبستگی تتراکوریک	دو خوشه کلی و دست‌کم ۳ خوشه جزئی
		تحلیل نشد	–	تحلیل شبکه	کل داده‌ها	۴ خوشه
		تحلیل نشد	–		داده‌های کامل	۷ خوشه
	مجانب مدل 3PL	آزمون تفاوت کای دو	۱ عامل	تحلیل عاملی تاییدی	ساختار دو عاملی (یک عامل کلی و چهار عامل اختصاصی)	چهار خوشه حاصل از بهترینDETECT برازش را با داده‌ها به دست آورد
		BIC	۱ عامل			
		HQ	۱ عامل			
		SABIC	۱ عامل			
		AICc	۱ عامل			
		AIC	۱ عامل			

توجه: تحلیل عاملی کل اطلاعات و تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل در حالت مجانب پایین ۰/۱ و ۰/۱۵ به همگرایی نرسید.

برخلاف آزمون ریاضی که خوشه‌بندی سؤال‌ها براساس تحلیل شبکه اکتشافی برای کل داده‌ها و داده‌های کامل از نظر تعداد خوشه یکسان بود، در آزمون شیمی تحلیل کل داده‌ها و داده‌های کامل به ترتیب به چهار و هفت خوشه منجر شد. خوشه شش و هفت مبتنی بر داده‌های کامل فقط دارای دو سؤال‌اند، که در نتیجه می‌توان گفت عامل مربوط به آنها بسیار ضعیف است و خوشه چهار مربوط به کل داده‌ها نیز فقط حاوی دو سؤال است و در نتیجه تشکیل یک عامل توسط آنها از نظر عملی ارزشمند نیست.

اگرچه در آزمون شیمی تنوع محتوایی سؤال‌ها، به اندازه آزمون ریاضی نیست، ولی همبستگی سؤال‌های آن نسبت به سؤال‌های آزمون ریاضی تقریباً پایین است. بر اساس نتایج به دست آمده می‌توان گفت در این آزمون نیز تقریبا همان روش‌هایی که آزمون ریاضی را یک عاملی نشان می‌دادند، در آزمون شیمی نیز عامل کلی و غالب آزمون را مشخص کرده‌اند. با کمی اختلاف جزئی همان روش‌های منعکس کنندهٔ دو عامل در ریاضی، در اینجا نیز ساختار دو عاملی را پیشنهاد داده‌اند. روش‌هایی که از وجود سه و به خصوص چهار عامل حمایت می‌کنند بیشتر خوشه‌های سؤال موجود در آزمون و یا عوامل اختصاصی را نشان می‌دهند. روش‌ها و شاخص‌هایی که از وجود پنج، شش و به خصوص هفت عامل حمایت می‌کنند ممکن است ترکیب عوامل مزاحم و اختصاصی را نشان دهند. روش‌هایی که هشت و بیش از هشت عامل را نشان می‌دهند احتمالاً منعکس کنندهٔ عوامل آماری موجود در آزمون هستند.

نتایج تحلیل عاملی تأییدی بر اساس روش‌های تحلیل عاملی اکتشافی چهارتایی، روش ناپارامتریک DETECT و تحلیل شبکهٔ اکتشافی نشان داد که در این آزمون هم ساختار دو عاملی (یک عامل کلی و چهار عامل اختصاصی) نسبت به بقیهٔ مدل‌ها برازش بهتری با داده‌ها دارد، ولی برخلاف نتایج آزمون ریاضی، ساختار چهار خوشه‌ای حاصل از DETECT نسبت به ساختار حاصل از تحلیل عاملی اکتشافی چهارتایی و یا چهار خوشهٔ حاصل از تحلیل شبکهٔ اکتشافی برازش بهتری با داده‌ها دارد. این نتیجه دور از انتظار نیست، زیرا بر اساس نتایج تحلیل شبکهٔ اکتشافی همبستگی عامل چهار با سه عامل دیگر پایین است. شاخص‌های مربوط به اشباع عامل کلی تقریباً خوب است ولی به اندازهٔ آزمون ریاضی چندان بالا نیست. نتایج تبدیل اشمیت-لایمن بر اساس سه تا هشت عامل جزئی و یک عاملی کلی نشان داد که سؤال ۲۳۹ دفترچه روی هیچ یک از عوامل جزئی و عامل کلی قرار نگرفت و سؤال‌های ۲۴۹، ۲۵۱، ۲۶۵، ۲۶۸ و ۲۶۹ روی عاملی کلی قرار نگرفتند. این نتایج ساختار عاملی خاص این آزمون را نشان می‌دهد. در این آزمون سؤال‌هایی وجود دارند که به تنهایی صفت خاصی را اندازه‌گیری می‌کنند. سه ویژگی منحصر به فرد آزمون شیمی نسبت به ریاضی بدین قرار است: الف) همبستگی کم بین سؤال‌ها، ب) تفاوت تعداد خوشه‌ها در داده‌های کامل و کل داده‌ها و ج) وجود سؤال‌هایی

که با بقیهٔ سـؤال‌ها از نظر محتوایی همخوانی چندانی نداشـتند. به همین سبب ساختار عاملی این آزمون، به خصوص با روش‌های مبتنی بر تحلیل عاملی، چندان مشخص نیست. به نظر می‌رسد یکی از دلایل این موضوع همبسـتگی کم سـؤال‌ها با هم و ترکیب خاص داده‌ها براساس پاسخ‌دهندگان باشد. به این معنی که احتمالاً افراد راهبردهای گوناگونی را برای پاسـخ به سـؤال‌های این آزمون انتخاب کرده‌اند. یعنی جامعهٔ افراد شرکت کننده از نظر سـبک پاسـخ‌دهی تنوع زیادی دارند، به طوری که وقتی داده‌های افراد فاقد پاسخ سفید تحلیل می‌شوند، تعداد خوشه‌های سؤال و به تبع آن عامل‌ها بیشتر از زمانی است که داده‌های کل افراد مورد تحلیل قرار می‌گیرند. آیا ترکیب نمونه‌هایی که راهبردهای متفاوتی برای پاسخ دادن به سؤال‌ها دارند می‌تواند عوامل موجود در آزمون را مشتبه کرده به طوری که ساختار عاملی به خوبی مشخص نشود؟ کارهای رکیس (۲۰۰۹) در خصوص ابعاد آزمون و جامعه و همخوان بودن یا نبودن آنها با هم مؤید این مطلب است ولی برای اطمینان بیشتر باید پژوهش‌های بیشتری در این خصوص انجام داد.

ج) آزمون فلسفه-منطق

یک عامل: در تحلیل عاملی اکتشافی، شاخص RMSEA؛ در تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل با مجانب پایین صفر و مجانب پایین مبتنی بر مدل سه پارامتری، شاخص‌های کاهش RMSR و تاناکا براساس ملاک ۰/۹۵ به بالا.

دو عامل: در تحلیل موازی آزمون MAP و شاخص ساختار ساده با پیچیدگی ۱ (VSS.C1)، مقادیر ویژهٔ تحلیل عاملی مشترک در داده‌های واقعی؛ در تحلیل عاملی اکتشافی فاصلهٔ اطمینان ۹۰درصدی اطراف RMSEA همراه با شاخص‌های CFI و TLI؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر شاخص‌های BIC و SBIC؛ در تحلیل عاملی کل اطلاعات با مجانب پایین مبتنی بر مدل سه پارامتری آزمون تفاوت کای دو، HQ، SBIC، BIC، AICC و AIC؛ تحلیل شبکه اکتشافی مبتنی بر کل داده‌ها.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول ۳: خلاصه نتایج روش‌های مختلف بررسی ابعاد در داده‌های آزمون فلسفه-منطق

روش	شاخص		نتیجه	روش	شاخص		نتیجه
تحلیل موازی	SBIC		کم ۸دست عامل	روش‌های کامل تحلیل عاملی غیرخطی با داده	$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب صفر	۱ عامل
	BIC		کم ۸دست عامل		RMSRکاهش در		۱ عامل
	MAP		۲ عامل		Tanaka Index		۱ عامل عامل خوب
	VSS.C1		۲ عامل		تحلیل نشد		-
	VSS.C2		۴ عامل		تحلیل نشد		-
	عامل	های مشاهده شده‌مقادیر ویژه داده	۲ عامل		تحلیل نشد		-
		های شبیه‌سازی‌مقادیر ویژه داده	۱۰ عامل		تحلیل نشد		-
		گیری از داده‌هامقادیر ویژه نمونه	۴عامل		تحلیل نشد		-
	مولفه	مقادیر ویژه داده‌های مشاهده شده	۷ عامل		تحلیل نشد		-
		مقادیر ویژه داده‌های شبیه‌سازی	۶ عامل		$RMSR \leq 4 \times 1/\sqrt{n}$	مجانب مدل 3PL	۱ عامل
		گیری از داده‌هامقادیر ویژه نمونه	۳ عامل		RMSRکاهش در		۱ عامل
تحلیل عاملی اکتشافی	آزمون تفاوت کای دو		۷ عامل		Tanaka Index		۱ عامل خوب
	RMSEA		۱ عامل	DETECT	$Detect_{\text{cross validated}}$		۴ خوشه
	RMSEAفاصله اطمینان ۹۰٪ اطراف		۲ عامل		Ratio(R)		بی‌ثباتی ساختار ۴ خوشه‌ای
	CFI		۲ عامل		IDN		ساختار چندبعدی نسبتا پیچیده
	TLI		۲ عامل				
	WRMR		۴ عامل				

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

روش	شاخص		نتیجه	روش	شاخص	نتیجه
تحلیل عاملی کل اطلاعات	مجانب صفر	آزمون تفاوت کای دو	۷ عامل	DIMTEST	مجانب پایین صفر	رد تک‌بعدی ضروری
		BIC	۲ عامل			
		HQ	۳ عامل		مجانب پایین ۰/۱۰	رد تک‌بعدی ضروری
		SABIC	۲ عامل			
		AICc	۴ عامل		مجانب پایین ۰/۱۵	رد تک‌بعدی ضروری و تغییر خوشه‌بندی
		AIC	۴ عامل			
	مجانب ۰/۱	تحلیل نشد	-	استقلال موضعی (Q3)	IRTمدل‌های	وابستگی موضعی ضعیف
		تحلیل نشد	-			
		تحلیل نشد	-			
		تحلیل نشد	-		نمره خام	وابستگی موضعی مثبت بالا بین برخی سؤال‌ها
		تحلیل نشد	-			
		تحلیل نشد	-			
	مجانب ۰/۱۵	تحلیل نشد	-	نقشه حرارتی	همبستگی فای	با دامنه ۰/۱۱- تا ۰/۲۹
		تحلیل نشد	-		همبستگی تتراکوریک	با دامنه ۰/۲- تا ۰/۴۵
		تحلیل نشد	-	تحلیل خوشه سلسله مراتبی	همبستگی فای	۳ خوشه کلی و دست‌کم ۵ خوشه جزئی
		تحلیل نشد	-		همبستگی تتراکوریک	۳ خوشه کلی و دست‌کم ۵ خوشه جزئی
		تحلیل نشد	-	تحلیل شبکه	کل داده‌ها	۲ خوشه
		تحلیل نشد	-		داده‌های کامل	۳ خوشه
	مجانب مدل 3PL	آزمون تفاوت کای دو	۲عامل	تحلیل تاییدی	ساختار دو عاملی (یک عامل کلی و ۲ عامل اختصاصی)	۲ خوشه حاصل از تحلیل شبکه بهترین برازش را با داده‌ها به دست آورد
		BIC	۲ عامل			
		HQ	۲ عامل			
		SABIC	۲ عامل			
		AICc	۲ عامل			
		AIC	۲ عامل			

توجه: تحلیل عاملی کل اطلاعات و تحلیل عاملی غیرخطی مبتنی بر داده‌های کامل در حالت مجانب پایین ۰/۱ و ۰/۱۵ به همگرایی نرسید.

سه عامل: در تحلیل موازی مقادیر ویژه مبتنی بر نمونه‌گیری از داده‌های واقعی؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر شاخص HQ؛ تحلیل خوشهٔ سلسله مراتبی برای ماتریس عدم شباهت مبتنی بر همبستگی فای و تتراکوریک بر اساس سه خوشهٔ کلی؛ تحلیل شبکهٔ اکتشافی داده‌های کامل.

چهار عامل (۴ خوشه سؤال): در تحلیل موازی شاخص ساختار ساده با پیچیدگی ۲ (VSS.C2)، در تحلیل عاملی مشترک مقادیر ویژه مبتنی بر نمونه‌گیری از داده‌های واقعی؛ در تحلیل عاملی اکتشافی شاخص WRMR؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر شـاخص‌های AIC و AICc؛ روش DETECT با چهار خوشـه بی‌ثبات و ساختار چند بعدی نسبتاً پیچیده.

پنج عامل: به جز تحلیل خوشهٔ سلسله مراتبی برای ماتریس عدم شباهت مبتنی بر همبستگی فای و تتراکوریک که از وجود دست کم پنج خوشه حکایت داشت، هیچ روش دیگری وجود ساختار پنج عاملی را نشان نداد.

شش عامل: در تحلیل موازی مقادیر ویژه تحلیل مولفه در داده‌های شبیه‌سازی.

هفت عامل و بیشتر (عامل‌های آماری): در تحلیل موازی شاخص‌های BIC، SBIC، مقادیر ویژهٔ تحلیل عاملی مشترک در داده‌های شبیه‌سازی شده و مقادیر ویژهٔ تحلیل مولفه در داده‌های واقعی؛ در تحلیل عاملی اکتشافی آزمون تفاوت کای دو؛ در تحلیل عاملی کل اطلاعات با مجانب پایین صفر آزمون تفاوت کای دو.

آزمون DIMTEST در شرایط مجانب پایین صفر، ۰/۱ و ۰/۱۵ فرض تک بُعدی بودن ضروری را با آلفای ۰/۰۵ رد کرد. تحلیل استقلال موضعی براساس شاخص *Q3* در مدل سه پارامتری نشان داد هیچ یک از زوج‌های سؤال دارای وابستگی بزرگتر از ۰/۲ نیستند ولی توزیع آماره *Q3* نامتقارن و تاحدی دارای کجی منفی است. این در حالی است که بر اساس نتایج مدل یک پارامتری و به خصوص نمره‌های خام، وابستگی مثبت بین سؤال‌ها تا حدی بالا است. در این آزمون نیز با کمی اختلاف عملکرد روش‌های تعیین ابعاد با دو آزمون دیگر تقریباً یکسان است. نتایج نقشهٔ حرارتی همبستگی بین سؤال‌های آزمون فلسفه-منطق نشان داد که همبستگی بین متغیرها تقریباً پایین است. تحلیل شبکهٔ اکتشافی از وجود دو عامل در داده‌ها حکایت دارد و تحلیل‌های مبتنی بر تحلیل خوشهٔ سلسله مراتبی نیز از وجود دو عامل کلی حکایت دارد. این نتایج با توجه به وجود دو حوزهٔ محتوایی فلسفه و منطق در این آزمون جالب است.

سـؤال‌های آزمون AT در دو خوشـهٔ کلی نمودارهای تحلیل خوشهٔ سلسه مراتبی مبتنی بر همبستگی فای و تتراکوریک پخش شده‌اند و خوشه‌بندی سلسله مراتبی مبتنی بر همبستگی فای و تتراکوریک نیز تقریباً یکسان است. اگرچه خوشه‌بندی حاصل از تحلیل شبکهٔ اکتشافی و تحلیل خوشه سلسله مراتبی برای ماتریس عدم شباهت، مبتنی بر همبستگی فای و تتراکوریک چندان مطابقت ندارد، ولی خوشهٔ اول تحلیل شبکه بیشتر در سؤال‌های سمت چپ نمودار تحلیل خوشهٔ سلسله مراتبی و خوشهٔ دوم بیشتر در سؤال‌های سمت راست تحلیل خوشـهٔ سلسله مراتبی نمایان شده است. نتایج روش‌های تحلیل عاملی تأییدی مبتنی بر روش‌های اکتشافی

DETECT، تحلیل عاملی و تحلیل شبکه نشان داد، بهترین برازش با داده‌ها به دو خوشهٔ سؤال حاصل از تحلیل شـــبکه (یعنی یک عامل کلی و دو عامل جزئی) مربوط است. مدل عامل مرتبه دوم در تحلیل تأییدی مبتنی بر روش‌های DETECT، تحلیل خوشـــه و تحلیل عاملی اکتشافی به دلیل همبستگی کامل عامل کلی با یکی از عامل‌های جزئی قابل اجرا نبود. بر اســـاس نتایج تحلیل عاملی اکتشـــافی، ساختار عاملی این آزمون چندان مشخص نبود که یکی از دلایل آن پایین بودن میزان اشباع عامل کلی در این آزمون است که در همبستگی پایین بین سؤال‌ها متبلور شده است. در آزمون فلسفه و منطق نیز تعداد خوشه‌ها در داده‌های کامل و کل داده‌ها متفاوت بود، ولی این اختلاف به شدت آزمون شیمی نیست.

بحث

مطابق با نتایج حاصل از روش‌های مختلف تحلیل ابعاد، ساختار عاملی حاکم بر آزمون‌های تخصصی سراسری از نوع عامل دوگانه[1] (یک عامل کلی و چند عامل اختصاصی) است. به این معنی که در آزمون‌های تخصصی تحلیل شـــده یک عامل غالب و چند عامل اختصاصی وجود دارد. بر اساس مبانی نظری مدل دوگانه، همه سؤال‌های آزمون علاوه بر اینکه یک عامل اصلی یا کلی را اندازه‌گیری می‌کنند، دسته‌هایی از آنها با هم در داخل آزمون تشکیل خرده آزمون‌هایی را داده که هر کدام از این دسته‌ها یک عامل اختصاصی را اندازه‌گیری می‌نمایند. این عوامل اختصاصی که واریانس ماندهٔ مشترک بین خرده آزمون‌ها را تبیین می‌کنند، با عامل کلی و با یکدیگر متعامد بوده و همبستگی ندارند (دمارز[2]، ۲۰۱۳). برازش نتایج حاصل از تحلیل عاملی اکتشافی، تحلیل شبکهٔ اکتشافی و روش ناپارامتریک Detect با استفاده از تحلیل عاملی تأییدی نشان داد که در هر سه مورد، همیشه برازش مدل دو عاملی نسبت به مدل مرتبهٔ دوم[3] (یک عامل کلی مرتبه دوم و چهار عامل مرتبه اول)، یا نسبت به مدل حاوی چهار عامل همبسته و یا نسبت به مدل حاوی چهار عامل مستقل با داده‌ها بهتر است. با این تفاوت که مدل دو عاملی موجود در آزمون‌های تخصصی سراسری تحلیل شده، ویژگی‌های خاصی دارد که با مبانی نظری مدل دو عاملی همخوانی چندانی ندارد. به این معنی که عوامل اختصاصی موجود در آزمون‌های تخصصی یا توسط یک سؤال مشخص می‌شوند که ارزش عملیاتی ندارد (مثل سؤال ۱۲۱ریاضی در بخش تحلیل عاملی تاییدی) و یا ســـؤال یا ســـؤال‌هایی که یک عامل اختصاصی را اندازه‌گیری می‌کنند با عامل کلی ارتباط ندارند (بخش تحلیل عاملی تأییدی در آزمون‌های شیمی و فلسفه-منطق را ببینید). به علاوه در آزمون‌های تخصصی سراســـری ســـؤال‌ها هم زمان با چند عامل اختصاصی ارتباط دارند. اگر همبستگی بین سؤال‌ها متوسط به بالا باشـــد (همانند سؤال‌های آزمون ریاضی با متوسط همبستگی تتراکویک ۰/۴ و نمای ۰/۵۶ و همبستگی فای ۰/۳ و نمای ۰/۳۸)، که حاکی از غلبهٔ عامل کلی است، برازش مدل دو عاملی برای عامل‌های حاصل از تحلیل

1. Bifactor
2. DeMars
3. Second Order Model

عاملی اکتشـــافی نســـبت به مدل دو عاملی مربوط به نتایج تحلیل شبکهٔ اکتشافی و روش ناپارامتریک Detect برازش بهتری با داده‌ها دارد. در مقابل وقتی همبستگی بین سؤال‌ها پایین باشد (مثل سؤال‌های آزمون شیمی با متوسط همبستگی تتراکوریک ۰/۲ و نمای ۰/۳۵ و همبستگی فای ۰/۱۵ با نمای ۰/۲۲)، که حاکی از کاهش غلبهٔ عامل کلی است، نتایج حاصل از تحلیل شبکهٔ اکتشافی نسبت به نتایج حاصل از تحلیل عاملی اکتشافی و روش ناپارامتریک Detect براساس مدل دو عاملی برازش بهتری با داده‌ها دارد. در صورتی که همبستگی بین ســؤال‌ها خیلی پایین باشد (مثل آزمون فلسفه-منطق با متوسط همبستگی تتراکوریک ۰/۰۶ و نمای ۰/۱۶و همبستگی فای ۰/۰۴ با نمای ۰/۱) نتایج حاصل از روش ناپارامتریک Detect نسبت به نتایج حاصل از تحلیل عاملی اکتشافی و تحلیل شبکه اکتشافی براساس مدل دو عاملی برازش بهتری با داده‌ها دارد. این نتیجه دور از انتظار نیســت، چرا که آزمون فلســفه-منطق از نظر محتوایی از دو کتاب طراحی شده و دست کم محتوای این دو درس تا حدی متفاوت است. به همین دلیل در این حالت برازش نتایج روش ناپارامتریک Detect که به خوشه‌های سؤال نامتجانس حساس است برازش بهتری با داده‌ها دارد.

در کنار این موارد باید به نقش مجانب پایین در تعیین تعداد عامل‌ها اشــاره کرد. همان طور که شــاخص‌های مختلف مربوط به روش تحلیل عاملی کل اطلاعات در آزمون ریاضی نشان داد، همراه با افزایش مجانب پایین، از تعداد عامل‌های اســتخراج شــده نیز کاسته می‌شــود و وقتی مجانب پایین مبتنی بر مدل سه پارامتری در تحلیل عاملی کل اطلاعات لحاظ می‌شود، بر اساس تمام شاخص‌های موجود در این روش یک عامل پیشنهاد می‌گردد. در آزمون شــیمی و فلســفه-منطق نیز این روند تأیید شد، با این تفاوت که در آزمون فلسفه-منطق به دلیل وجود دو محتوای مختلف در آزمون، بر اساس شاخص‌های موجود در تحلیل عاملی، کل اطلاعات دو عامل پیشنهاد شده است.

در نهایــت اینکه تحلیل داده‌های کامل (بدون داده گمشــده) با تحلیل عاملی غیرخطــی، روش ناپارامتریک DETECT و DIMTEST و مقایسهٔ نتایج آنها با نتایج حاصل از تحلیل کل داده‌ها در سایر روش‌ها، می‌تواند تأثیر حدس و الگوها و روش‌های مختلف پاسخگویی را نمایان سازد. همان طور که مشخص شد، در آزمون شیمی و فلسفه-منطق تحلیل داده‌های کامل و کل داده‌ها با تحلیل شبکهٔ اکتشافی به خوشه‌های سؤال متفاوتی منجر شد که حاکی از تأثیر الگوهای مختلف پاسخگویی و شاید راهبردهای افراد در تشکیل آزمون‌های خاص خود است. به این معنی که هر فرد بسته به دانش و توانایی خود از بین همهٔ سؤال‌های آزمون، آزمون خاص خود را انتخاب و پاسخ می‌دهد. البته این موضوع نیازمند بررسی بیشتر است.

نتیجه‌گیری

استخراج عامل‌های تفسیرپذیر، که مهمترین هدف در بررسی تعداد عامل‌ها و ابعاد موجود در داده‌ها است، به این معنی است که تعداد عامل‌ها بیشتر منعکس کنندهٔ خلاقیت پژوهشگر است تا داده‌ها. روش‌های مختلف

استخراج عامل فقط جنبهٔ یافتن سرنخ دارد و هیچ وقت نباید صرفاً به آنها تکیه کرد. وجود روش‌های متعدد برای بررسی و مشخص کردن ابعاد داده‌ها حاکی از پیچیدگی این موضوع است. مشخص کردن ابعاد داده‌های حاصل از اجرای آزمون‌ها (به خصوص داده‌های دو ارزشی) روش منحصر به فردی ندارد. کاربرد روش‌های مختلف بررسی ابعاد، به نتایج متفاوتی منجر می‌شود. هیچ ملاک قطعی برای این منظور وجود ندارد. به همین دلیل مشخص کردن درست ابعاد داده‌ها، نیازمند ترکیب روش‌های مختلف آماری، نموداری، محتوایی و در نهایت خلاقیت پژوهشگر است (تات، ۲۰۰۳، لوی و اسویتنا، ۲۰۱۴). همگام با گسترش فناوری و توسعه نرم افزارها روش‌های مختلفی برای این منظور پیشنهاد شده است (هتی، ۱۹۸۵؛ تات، ۲۰۰۳؛ استالفسن و هانوک[1]، ۲۰۰۸؛ لوی و اسویتنا، ۲۰۱۱، ۲۰۱۲). پژوهش‌گران مختلفی (از جمله لوی و اسویتنا، ۲۰۱۴) روش‌هایی را که تا به امروز برای این منظور استفاده شده خلاصه کرده‌اند. هر یک از روش‌های تعیین ابعاد در شرایط خاصی کارایی لازم را دارد که در ادامه به مهم‌ترین این موارد اشاره شده است.

آزمون تفاوت کای‌دو به تعداد آزمودنی‌ها حساس است و به تعداد عواملی منجر می‌شود که به لحاظ عملی اهمیت ندارند. به این معنی که اگر فردی به دنبال عامل‌های بیشتری باشد می‌تواند با اجرای تحلیل با آزمودنی‌های بیشتری به هدف مد نظر برسد. تحلیل موازی مبتنی بر داده‌های شبیه‌سازی شده نیز تا حدودی به اندازهٔ نمونه حساس است و از این نظر در نمونه‌های بزرگ (N>۱۰۰۰) تمام مقادیر ویژهٔ عامل‌های تصادفی بسیار به یک نزدیک می‌شوند. آزمون اسکری این قابلیت را دارد که به اختلاف‌های تفسیری در خصوص موقعیت شکست نمودار این روش منجر شود. اگر داده‌ها به لحاظ عاملی خیلی پیچیده باشند روش ساختار بسیار ساده چندان خوب کار نمی‌کند (نباید پیچیدگی برخی سؤال‌ها بیشتر از ۲ باشد). قاعدهٔ مقدار ویژهٔ بزرگ‌تر از یک، با اصل تقسیم تعداد متغیرها بر سه همخوانی دارد و در بیشتر موارد بدترین همهٔ ملاک‌ها است (ریول، ۲۰۱۷). در بیشتر موارد تحلیل موازی مبتنی بر تحلیل عاملی مشترک، جزو روش‌هایی است که در شرایط مختلف، با ثبات رفتاری خوب برای این منظور دقت کافی دارد. هنگام استفاده از این روش، در صورتی که پژوهشگر نگران باشد که عامل‌ها با سؤال‌هایی کم (پنج یا کمتر) بازنمایی شوند، می‌تواند مدل حاوی یک عامل بیشتر را نیز بررسی کند. شاخص AIC، یا تعداد عامل‌ها را درست مشخص می‌کند و یا برآوردی در حد یک عامل بیشتر به دست می‌دهد. به همین دلیل، با ملاحظهٔ این شرط که برازش مدل حاوی یک عامل کمتر نیز با این شاخص بررسی شود، جایگزین خوبی برای تحلیل موازی محسوب می‌شود. در صورت نداشتن دسترسی به هیچ یک از این روش‌ها و در حالت تعداد کم متغیرها و بالا بودن همه اشتراک‌ها، می‌توان از قاعدهٔ مقادیر ویژه مبتنی بر تحلیل عاملی مشترک که مقدارشان بزرگ‌تر از میانگین مقادیر ویژه است استفاده کرد. قاعدهٔ سنتی مقادیر ویژهٔ بزرگ‌تر از یک به هیچ وجه کارایی خوبی ندارد (پیرسون، ماندفرام و پِکُنی[2]، ۲۰۱۳).

1. Stellefson & Hanik
2. Pearson, Mundfrom, & Piccone

مقایسهٔ روش‌های ارزیابی ساختار آزمون‌های دو ارزشی، شامل روش‌های تحلیل عاملی اکتشافی خاص داده‌های دو ارزشی (مثلا با MPLUS)، تحلیل عاملی غیرخطی با NOHARM، تحلیل عاملی کل اطلاعات، آزمون بعدیت ضروری در DIMTEST و شاخص DETECT نشان می‌دهد که در شرایط حاکم بر آزمون‌های با مقیاس بزرگ که تعداد سؤال‌ها و آزمون دهنده‌های آنها زیاد است و نیز در داده‌های شبیه‌سازی شده تک بُعدی و چند بُعدی، تمام این روش‌ها در شرایطی بسیار متنوع، تا حد زیادی عملکرد معقول و خوبی دارند. تنها مسئله زمانی است که داده‌های آزمون از پیش فرض‌های روش مورد استفاده، انحرافی جدی داشته باشند. به عنوان نمونه زمانی که حدس وجود دارد اما در تحلیل لحاظ نمی‌شود یا زمانی که ساختار چند بُعدی آزمون ساده نیست ولی هدف روش برآورد، میزان ساختار چند بُعدی ساده است (تات، ۲۰۰۳). بررسی تأثیر ساختار پیچیده بر ارزیابی بُعدیت در مدل‌های چند بُعدی جبرانی نظریهٔ سؤال پاسخ با استفاده از DETECT و NOHARM براساس دقت شناسایی تعداد درست ابعاد و توانایی بازیابی درست دسته‌های سؤال نشان می‌دهد که روش‌های مبتنی بر DETECT نسبت به روش‌های مبتنی بر NOHARM در شرایط دو بعدی و سه بعدی به نسبت بالاتری از دفعات، تعداد ابعاد را درست بازیابی می‌کند، به ویژه زمانی که همبستگی بین ابعاد کوچک‌تر یا مساوی ۰/۶ است، پیچیدگی داده‌ها کوچک‌تر یا مساوی ۳۰ درصد و حجم نمونه هزار باشد. در صورت افزایش پیچیدگی و کاهش حجم نمونه کارآمدی هر دو روش به طور معمول کاهش می‌یابد. موفقیت روش‌های موجود در NOHARM از نظر بازیابی دسته‌های سؤال نسبت به روش‌های موجود در DETECT بهتر یا برابر آن است و از سطوح پیچیدگی بیشتر متاثر می‌شوند. روش‌های مبتنی بر DETECT بیشتر متأثر از طول آزمون است به طوری که با افزایش تعداد سؤال‌ها ضرایب شباهت تطبیق ساده به طور اساسی کاهش می‌یابد (اسویتنا و لوی، ۲۰۱۶).

کاربرد روش‌های تعیین بُعد در داده‌های آزمون‌های سراسری نشان داد هنگامی که در یک آزمون همبستگی بین سؤال‌ها پایین است و داده‌ها در حجم وسیع از مکان‌های مختلف جمع‌آوری می‌شوند، پیدا کردن ساختار عاملی بر اساس روش‌های مبتنی بر تحلیل همبستگی (مثل تحلیل عاملی) گمراه کننده و تقریباً ناموفق است. در چنین شرایطی بهتر است از روش‌های اکتشاف خوشه‌های سؤال (مثل DETECT و یا تحلیل شبکه) نیز استفاده شود. دست کم در پژوهش حاضر نتایج مربوط به تحلیل آزمون ریاضی، شیمی و فلسفه-منطق این ایده تأیید شد. همبستگی بین سؤال‌ها در آزمون ریاضی بیشتر از همبستگی سؤال‌ها در آزمون شیمی و فلسفه-منطق بود. در واقع در آزمون‌های شیمی و فلسفه-منطق همبستگی سؤال‌ها بسیار پایین بود. بنابر این نتایج روش‌های مبتنی بر تحلیل خوشه، نسبت به نتایج روش تحلیل عاملی دست کم در این پژوهش برازش بهتری با داده‌ها داشت. به علاوه وقتی همبستگی بین سؤال‌ها پایین است پیدا کردن ساختار عاملی با دشواری‌هایی همراه است، به طوری که در نهایت پیدا کردن تعداد عامل‌ها دشوار خواهد بود و انتساب سؤال‌ها به عامل‌ها به دشواری‌های زیادی منجر می‌شود. البته در این باره باید ناهمگونی افراد از نظر نحوهٔ پاسخگویی به سؤال‌ها را

نیز در نظر گرفت. در شرایط همبستگی پایین، ساختار عاملی به دست آمده فاقد ثبات است و به نظر می‌رسد که عواملی غیر از توانایی زیربنایی در خوشه‌بندی سؤال‌ها موثر باشد. بر اساس نتایج حاصل از دو آزمون شیمی و فلسفه-منطق وقتی غلبه عامل کلی پایین است (که توسط امگای سلسله مراتبی[1] ω_h قابل ارزیابی است و مقادیر زیر ۰/۸ آن به کاهش غلبه عامل کلی و همسانی درونی قابل تامل اشاره دارد، کراتزن و پیتز[2]، ۲۰۱۷ و فایست و همکاران[3]، ۲۰۱۹) ساختار عاملی چندان مشخص نیست. به علاوه در این شرایط تحلیل داده‌های کامل و کل داده‌ها از نظر خوشه‌بندی به نتایج متفاوتی منجر می‌شود. ظاهراً در این حالت راهبرد افراد برای پاسخ دادن با هم متفاوت است و همین خود باعث نامشخص شدن ساختار عاملی می‌شود. به عبارت دیگر ترکیب چند زیر نمونه که هر یک راهبردهای خاص خود را برای پاسخگویی به سؤال‌ها دارند (یا زیر گروه‌هایی که هر یک بر بخشی از محتوای موجود در آزمون تسلط کافی یا نسبی دارند) باعث مشتبه شدن نتایج روش‌های برآورد تعداد ابعاد می‌گردد. در کنار این موضوع باید به وجود همبستگی منفی برخی سؤال‌ها با سایر سؤال‌ها نیز اشاره کرد. وجود همبستگی منفی برخی سؤال‌ها با سایر سؤال‌های آزمون باعث کاهش واریانس مشترک و در نتیجه دشواری در پیدا کردن عامل‌ها خواهد شد. در این پژوهش چون هیچ یک از سؤال‌های آزمون‌های ریاضی، شیمی و فلسفه-منطق را سازمان سنجش حذف نکرده بود، پس تصمیم به بررسی ساختار به همان صورت گرفته شد و سؤالی حذف نشد. با این حال می‌توان این موضوع را در پژوهش دیگری به صورت نظری و تجربی بررسی کرد.

گنجاندن مجانب پایین در تحلیل، با استفاده از تحلیل عاملی کل اطلاعات، تحلیل عاملی غیرخطی، *DETECT* و *DIMTEST* می‌تواند نقش حدس یا سایر عوامل مزاحم و غیر مهم را مشخص کند. مقایسهٔ نتایج حاصل از تحلیل داده‌های کامل (نبود داده‌های گمشده) با نتایج حاصل از کل داده‌ها نیز می‌تواند راهنمای خوبی برای تعیین تعداد ابعاد باشد. باید توجه شود که روش‌های تحلیل عاملی غیرخطی، *DETECT* و *DIMTEST* نه تنها برای تحلیل داده‌های کامل مناسب هستند بلکه می‌توانند مجانب پایین را نیز در تحلیل لحاظ کنند. بررسی میزان اشباع عامل کلی با روش‌های نموداری نیز در این خصوص مفید است. در نهایت برازش مدل‌های اکتشافی با روش‌های تأییدی همراه با تفسیرپذیری عوامل باید حتماً در تحلیل گنجانده شود. در خصوص برازش بهتر مدل دو عاملی نسبت به مدل مرتبهٔ دوم در داده‌های واقعی (کوچینا و بایل[4]، ۲۰۱۷) نیز می‌توان گفت وجود ساختار دو عاملی نشان می‌دهد که آزمون‌های سراسری عموماً یک عامل کلی و چند عامل اختصاصی دارند. عوامل اختصاصی می‌توانند به دلیل محتوای مشترک، عوامل مزاحمی مانند وابستگی به متن مشترک، یا فعل منفی در تنهٔ سؤال‌ها و سایر موارد باشد (دمارز، ۲۰۱۳). چنین به نظر می‌رسد که در آزمون‌های سراسری، عامل‌های اختصاصی لزوماً

1. Omega Hierarchical
2. Crutzen & Peters
3. Feißt and et al
4. Cucina & Byle

منعکس کنندهٔ عوامل شناختی یا محتوایی نیستند و امکان اینکه این عوامل منعکس کنندهٔ عوامل غیرشناختی باشـد نیز وجود دارد. در کنار این موضوع باید به پیامدهای نمره کل‌سـازی بر اسـاس این ساختار عاملی نیز توجه کرد. با توجه به پایایی کم عوامل اختصاصی، لازم اسـت که به مقایسـهٔ افراد در این عوامل توجه شـود. برای بررسی بُعدیت آزمون‌های سراسری ضروری است حتماً از ترکیب روش‌های مختلف استفاده شود. به عنوان مثال، در کنار تحلیل عاملی با اسـتفاده از روش‌های مناسـب داده‌های دو ارزشی، از سایر روش‌ها مثل تحلیل موازی و یا روش‌های گرافیکی مثل نقشهٔ حرارتی و تحلیل شبکه نیز استفاده شود. بهتر است قبل از گزارش نمره‌ها و مقایسهٔ افراد، ساختار عاملی آزمون‌ها با روش‌هایی مثل تحلیل خوشه و نقشهٔ حرارتی بررسی شود، زیرا آزمون‌های دارای همبستگی کم بین سؤال‌ها، بیشتر متشکل از عوامل اختصاصی هستند تا عامل کلی و در این حالت مقایسهٔ افراد ممکن است نادرست باشد. دلیل این امر این است که عوامل اختصاصی پایایی کمی دارند و این موضوع مقایسه پذیری افراد را با دشواری‌هایی همراه خواهد ساخت. تحلیل داده‌های کامل و کل داده‌ها این فرضیه را قوت می‌بخشد که در آزمون‌های سراسری ورود به دانشگاه بیشتر افراد از میان کل سؤال‌های یک آزمون، آزمون خاص خود را انتخاب و پاسخ می‌دهند که می‌تواند پیامدهای نامطلوبی داشته باشد. استفاده از نتایج مدل دو عاملی (یک عامل کلی و چند عامل اختصاصی) برای نمره‌گذاری و مقایسهٔ افراد. از آنجا که در مدل دو عاملی، عامل کلی و عوامل اختصاصی قابل تفکیک هستند ولی در مدل یک عاملی (نمره‌گذاری کلی) این دو عامل ترکیب می‌شوند و با این ملاحظه که پایایی عامل کلی بیشتر از عوامل اختصاصی است، لازم است برای مقایسهٔ افراد از نمره‌های عامل کلی استفاده شود، چرا که پایایی عوامل اختصاصی پایین است و مقایسهٔ افراد بر اساس نمرهٔ کلی می‌تواند نادرست باشد. مشخص کردن ساختار محتوایی آزمون توسط طراحان سوال برای اجرای تحلیل‌های بعد از اجرای آزمون می‌تواند نقشـهٔ راهی برای مطابقت سـاختار محتوایی آزمون با نتایج حاصل از روش‌های آماری باشد. وجود ساختار محتوایی آزمون می‌تواند منجر به افزایش روایی سازه آزمون‌های سراسری می‌شود. نداشـتن دسترسی به طراحان سؤال برای تحلیل کیفی سؤال‌ها و تعیین تعداد ابعاد شناختی بر اساس نظر آنها، محدودیـت عمـدهٔ این پژوهش بود. البته تجربهٔ کاری با افراد متخصص در حوزهٔ محتوایی (معلمانی که محتوای دروس مورد نظر را تدریس می‌کنند) نیز نشان داده است که معمولاً این افراد توانایی این کار را ندارند. حتی پس از توضیح مطالب، آنها بیشتر تمایل داشتند سؤال‌ها را بر اساس طبقه‌بندی بلوم رده‌بندی کنند که این کار نتایج خوبی نداشت و مشکل را حل نمی‌کرد. برای انجام درست این کار لازم بود برای آنها آموزش‌های خاصی فراهم شود تا پس از درک هدف، این کار را انجام دهند. متأسفانه به دلیل محدودیت‌های مالی و زمانی چنین کاری امکان پذیر نبود. به همین دلیل امکان استفاده از روش‌های تأییدی ارزیابی بُعدیت بر اساس محتوا و نظر افراد متخصص امکان پذیر نبود. البته بررسی محتوای آزمون‌های تحلیل شده این احتمال را قوت می‌بخشد که ساختار محتوایی چندان مشخصی که از قبل در طراحی این آزمون‌ها لحاظ شده باشد وجود ندارد. به دلیل محدودیت زمانی و مالی امکان اجرای تحلیل‌های استفاده شده برای تمام درس‌های تخصصی و عمومی آزمون سراسری امکان پذیر نبود، از این

رو لازم است در تعمیم نتایج پژوهش به سایر درس‌ها به خصوص آزمون‌های عمومی احتیاط کرد. مثلاً آزمون‌های عمومی مثل زبان انگلیسی و عربی و یا سایر درس‌های تخصصی ممکن است از ساختار دو عاملی پیروی نکنند.

تقدیر و تشکر

بدینوسیله از دست‌اندرکاران محترم سازمان سنجش آموزش کشور به خاطر حمایت مالی، معنوی و همکاری در اجرای پژوهش حاضر سپاس‌گزاری می‌شود.

References

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *7*(4), 255-278.
- Ackerman, T. A. (1996). Graphical representation of multidmensional item response theory analyses. *Applied Psychological Measurement*, 20, 311–329.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, 7(3), 309.
- Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health psychology review*, *11*(3), 242-247.
- Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence*, 5(3), 27.
- DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, 13(4), 354-378.
- Feißt, M., Hennigs, A., Heil, J., Moosbrugger, H., Kelava, A., Stolpner, I., ... & Rauch, G. (2019). Refining scores based on patient reported outcomes–statistical and medical perspectives. *BMC medical research methodology*, *19*(1), 1-9.
- Fraser, C., & McDonald, R. P. (2003). *NOHARM version 3.0*: A windows program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory computer program.
- Gefen, D. (2003). Assessing unidimensionality through LISREL: An explanation and an example. *Communications of the Association for Information Systems*, 12(1), 2.
- Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural equation modeling*, 7(2), 251-270.
- Hambleton, R.K.(1989). Principles and selected applications of item response theory. In R. Linn(editor). *Educational Measurement (3rd edition)* (pp. 147-200) New York: Macmillan.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltenls. *Applied Psychological Measurement*, 9(2), 139-164.

- Horn,J.L.,&Engstrom,R.(1979).Cattell'sscreetestinrelationtoBartlett'schi-squaretestandother observationsonthenumberoffactorsproblem. *MultivariateBehavioralResearch*, 14(3),283-300.
- Hudson F. Golino (2019). EGA: Exploratory Graph Analysis: Estimating the number of dimensions in psychological data. *R package version 0.4*.
- Izanloo, B.(2019). Optimum methods of Determining underlying dimensions of high-stakes' university entrance exams. *Research project reporte approved by National Organization of Education Testing* (NOET) (in persion).
- 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*. University of Michigan, Ann Arbor, MI.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8*: User's reference guide. Scientific Software International.
- Levy, R., & Svetina, D. (2011). A generalized dimensionality discrepancy measure for dimensionality assessment in multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*, 64(2), 208-232.
- Lord, F. M., & Novick, M. R. (1968). *statistical theories of mental test scores*: Addison Wesley.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of mathematical and statistical psychology*, 34(1), 100-117.
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*,43, 289-374.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus (Version 7.4)*. Los Angeles, CA.
- Nandakumar, R., & Ackerman, T. A. (2004). Test modeling. *The sage handbook of quantitative methodology for the social sciences*, 93-105.
- Pearson, R., Mundfrom, D., & Piccone, A. (2013). A comparison of ten methods for determining the number of factors in exploratory factor analysis. *Multiple Linear Regression Viewpoints*, 39(1), 1-15.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- Raykov, T. (2001). Bias of coefficient afor fixed congeneric measures with correlated errors. *Applied psychological measurement*, 25(1), 69-76.
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1), 25-36.
- Reese, L, M. (1999)A Classical Test Theory Perspective on LSAT Local Item Dependence, *LSAC Research Report Series, Statistical Report.*
- Revelle, W. (2017). *How To Use the psych package for Factor Analysis and data reduction*.

- Revelle, W. (2018) *psych*: Procedures for Personality and Psychological Research, Northwestern University,Evanston,Illinois,USA,https://CRAN.R-project.org/package=psychVersion=1.8.12.
- Rose, N., von Davier, M., & Xu, X. (2010). Modeling nonignorable missing data with item response theory (IRT). *ETS Research Report Series*, 2010(1), i-53.
- Roussos, L. A., & Stout, W. (2007). *Dimpack. Version 1.0*. The Roussos-Stout Software.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350.
- Segars, A. H. (1997). Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. *Omega*, 25(1), 107-121.
- Stellefson, M., & Hanik, B. (2008). Strategies for Determining the Number of Factors to Retain in Exploratory Factor Analysis. *Online Submission*.
- Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52(4), 589-617.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325.
- Svetina, D., & Levy, R. (2012). An Overview of Software for Conducting Dimensionality Assessment in Multidimensional Models. *Applied Psychological Measurement*, 36(8), 659-669. doi: 10.1177/0146621612454593
- Svetina, D., & Levy, R. (2012). An overview of software for conducting dimensionality assessment in multidimensional models. *Applied Psychological Measurement*, 36(8), 659-669.
- Svetina, D., & Levy, R. (2014). A framework for dimensionality assessment for multidimensional item response models. *Educational Assessment*, 19(1), 35-57.
- Svetina, D., & Levy, R. (2016). Dimensionality in compensatory MIRT when complex structure exists: Evaluation of DETECT and NOHARM. *The Journal of Experimental Education*, 84(2), 398-420.
- Tate, R. (2003). A comparison of selected empirical methods for assessing theStructure of responses to test items. *Applied Psychological Measurement*,27, 159-203.
- Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, 26(3), 249.
- Vehkalahti,K;Puntanen,S&Tarkkonen,L.(2009).Implicationsofdimensionalityonmeasurement reliability. *Statistical Inference, Econometric Analysis and Matrix Algebra*. Festschrift in Honour of Götz Trenkler (Bernhard Schipp, Walter Krämer, eds.), Physica-Verlag/Springer, 143-160.
- Zimmerman, W.D & William, R. H (1980). is Classical test theory "Robust" Under Violation of the Assumption of Uncorrelated Errors? *Canadian Journal of Psychology*, Vol 34(3).