Educational Measurement and Evaluation Studies

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

National Organization for Educational Testing

Introduction and Application of Moken Scale Analysis (MSA) in Educational and psychological Measurements

Jalil Younesi[1], Hengameh Aminaee[2] , Reyhaneh Reza zadeh[3]

1. Associate Professor of Educational Measurement and evaluation at Allameh Tabataba'i University, Tehran, Iran; (Corresponding Author), Email: jalilyounesi@gmail.com
2. PhD student of Educational Measurement and Evaluation at at Allameh Tabataba'i University, Tehran, Iran. Email: h_aminaee@atu.ac.ir
3. PhD Student of Educational Measurement and Evaluation at Allameh Tabataba'i University, Tehran, Iran. Email: rey_rezazadeh@atu.ac.ir

Article Info

Article Type:

Research Article

Received:

2021.06.07

Received in revised form:

2021.11.16

Accepted:

2021.12.10

Published online:

2021.12.27

ABSTRACT

Objective: Besides introducing the MSA[1] scales, this article shows its application in educational evaluation by providing evidence of Mocken pattern fitting in the study of knowledge amount of Kerman dental students (2016-2017) about the correct way to measure blood pressure according to the American Heart Association.

Methods: Mokken scale analysis (MSA) is a probabilistic-nonparametric approach of Item Response Theory (IRT) that can be used in educational and psychological evaluations and measurements with more lenient assumptions than the parametric models of Item Response Theory (IRT).

Results: 138 fifth and sixth year dental students of Kerman University answered to a 14-item test that was done in a census method, and according to the results of the secondary analysis of the test data (using R software) the scaling coefficients of individual questions and the scaling coefficients of pairs of questions were positive with a low standard error was obtained. The total scalability coefficient was 0.490 (medium), and omission of item 5, increased the total scalability index to 0.54 (high). Therefore, the results under the Mocken's monotone homogeneity Model (MH)[2] showed that there is an assumption of uniformity for the test and one-dimensionality has been achieved. The assumption of invariant ranking of the questions was checked by the number of violation times of the assumption of non-intersection of the response functions to the question, and the question selection method was confirmed by the bottom-up method of omitting question five. Obtaining the scaling factor H^{T3} at the rate of 0.45, showed that there is an evidence of fitting the double uniformity mode (DM)[4]. Four reliability coefficients obtained for the set of 13 items ranged from 0.86 to 0.91.

Conclusion: The evidence showed that this scale can be used both for ranking people and for ranking questions.

Keywords: non-parametric item- response theory, Moken scale analysis, monotone homogeneity (MH), double monotonicity model (DM).

Cite this article: Younesi, Jalil; Aminaee, Hengameh; Rezazadeh, Reyhaneh (2021). Introduction and Application of Moken Scale Analysis (MSA) in Educational and Psychological Measurements. *Educational Measurement and Evaluation Studies*, 11 (36): 143-176 pages. DOI:10.22034/EMES.2021.252028

© The Author(s).

Publisher: National Organization of Educational Testing (NOET)

مطالعات اندازه‌گیری و ارزشیابی آموزشی

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

سازمان سنجش آموزش کشور

معرفی و کاربرد مقیاس دو ارزشی موکن در اندازه‌گیری‌های آموزشی و روانشناختی

جلیل یونسی[۱]، هنگامه امینائی[۲]، ریحانه رضازاده[۳]

۱. دانشیار سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران، (نویسنده مسئول)، رایانامه: younesi@atu.ac.ir
۲. دانشجوی دکتری سنجش و اندازه گیری دانشگاه علامه طباطبائی، تهران، ایران، رایانامه: h_aminaee@atu.ac.ir
۳. دانشجوی دکتری سنجش و اندازه گیری دانشگاه علامه طباطبائی، تهران، ایران، رایانامه: rey_rezazadeh@atu.ac.ir

اطلاعات مقاله

نوع مقاله:
مقاله مروری

دریافت:
۱۴۰۰/۰۴/۱۵

اصلاح:
۱۴۰۰/۰۸/۲۵

پذیرش:
۱۴۰۰/۰۹/۱۹

انتشار:
۱۴۰۰/۱۰/۰۶

چکیده

هدف: این مقاله ضمن معرفی مقیاس موکن[۱]«MSA»، کاربرد آن در زمینهٔ ارزیابی آموزشی را با ارائهٔ شواهد برازش الگوهای موکن در مطالعهٔ سنجش میزان آگاهی از روش صحیح اندازه‌گیری فشارخون دانشجویان دندانپزشکی کرمان (سال تحصیلی ۹۶-۹۵) بر اساس دستورالعمل انجمن قلب آمریکا نشان می‌دهد.

روش پژوهش: تحلیل مقیاس موکن، رویکردی غیرپارامتری از نظریهٔ سؤال-پاسخ است که می‌تواند در ارزیابی و اندازه‌گیری‌های آموزشی و روان‌شناختی با فرضیاتی سهل‌گیرانه‌تر از الگوهای پارامتری نظریهٔ سؤال- پاسخ استفاده شود.

یافته‌ها: در نتایج تحلیل ثانویه داده‌های آزمونی ۱۴ سؤالی که به شیوهٔ سرشماری از ۱۳۸ نفر دانشجوی سال پنجم و ششم دندانپزشکی دانشگاه کرمان به‌عمل‌آمده بود (با استفاده از نرم‌افزار R)، ضرایب مقیاس‌پذیری تک‌تک سؤال‌ها و ضرایب مقیاس‌پذیری جفت سؤال‌ها، مثبت با خطای معیار کم به دست آمد. ضریب مقیاس‌پذیری کل ۰/۴۹۰ (در حد متوسط) بود که حذف سؤال پنج، شاخص مقیاس‌پذیری کل را به ۰/۵۴ (در حد قوی) افزایش داد. لذا نتایج تحت الگوی همگن یکنواخت[۲] (MH) موکن نشان داد که فرض یکنواختی برای آزمون وجود دارد و تک بُعدی بودن حاصل شده است. فرض رتبه‌بندی نامتغیر سؤال‌ها از راه تعداد دفعات نقض مفروضهٔ عدم تقاطع توابع پاسخ به سؤال، بررسی شد و روش انتخاب سؤال به روش پس‌رو، حذف سؤال پنج را تأیید کرد. به‌دست آمدن ضریب مقیاس‌پذیری[۳] H^T به میزان ۰/۴۵، نشان داد که شواهد برازش الگوی یکنواختی مضاعف[۴] (DM) وجود دارد. چهار ضریب پایایی به‌دست آمده برای مجموعهٔ ۱۳ سؤالی از ۰/۸۶ تا ۰/۹۱ متغیر بود.

نتیجه‌گیری: شواهد نشان داد که از این مقیاس می‌توان هم برای رتبه‌بندی افراد و هم برای رتبه‌بندی سؤال‌ها استفاده کرد.

واژه‌های کلیدی: نظریهٔ سؤال- پاسخ غیرپارامتری، تحلیل مقیاس موکن، الگوی همگنی یکنواخت، الگوی یکنواختی مضاعف.

استناد: یونسی، جلیل؛ امینائی، هنگامه؛ رضازاده، ریحانه (۱۴۰۰). معرفی و کاربرد مقیاس دوارزشی موکن در اندازه‌گیری های آموزشی و روانشناختی. مطالعات اندازه گیری و ارزشیابی آموزشی، ۱۱ (شماره۳۶)، صفحه ۱۴۳-۱۷۶. DOI:10.22034/EMES.2021.252028

ناشر: سازمان سنجش آموزش کشور حق‌مؤلف© نویسندگان.

1. Mokken scale analysis
2. Monotone Homogeneity
3. Scalability
4. Double Monotonicity

مقدمه (مبانی نظری و پیشینه پژوهش)

استفاده از آزمون‌ها و پرسشنامه‌ها در آموزش و روانشناسی بسیار رایج است. در واقع، آنها بخشی جدایی‌ناپذیر از اقدامات روزمره مربیان، معلمان، پزشکان، دانشجویان یا فارغ‌التحصیلان در این زمینه‌ها هستند. تصمیمات و تشخیص‌های بالینی و همچنین یافته‌های تحقیقات به‌شدت به کیفیت اندازه‌گیری‌های مورد استفاده بستگی دارد. علاوه بر این، در آزمون‌های سرنوشت‌ساز که در آن داوطلبان بر اساس نتایج آزمون به شغل یا تحصیل دسترسی پیدا می‌کنند، دقت در رتبه‌بندی داوطلبان اهمیت زیادی دارد. بنابر این تحقیق دربارهٔ رواسازی و اطمینان از اینکه آزمون‌ها به طور دقیق سطوح سازهٔ مربوط را اندازه‌گیری می‌کنند، ضروری است (بقایی، ۲۰۲۱).

در اندازه‌گیری‌های آموزشی، الگوهای سؤال- پاسخی[1] که بیشتر به کار برده می‌شود، بر اساس رابطهٔ پارامتری ساخته می‌شود که در آن، تابع جبری خاصی، شکل تابع سؤال- پاسخ[2] IRF“» را مشخص می‌کند و با استفاده از تبدیل‌هایی منجر به اندازه‌گیری در یک مقیاس با فاصلهٔ برابر می‌شود. در حالی‌که می‌توان خصوصیات اندازه‌گیری را با استفاده از یکی از روش‌های غیرپارامتری نظریهٔ سؤال-پاسخ مانند تحلیل مقیاس موکن بررسی کرد (ویند[3]، ۲۰۱۷).

موکن از دو الگوی تابع سؤال-پاسخ (IRT) غیرپارامتری تشکیل شده که بر این فرض استوار است که رابطهٔ بین θ و احتمال پاسخ صحیح با محدودیت‌های ترتیب کنترل می‌شود. یعنی برای دو نفر m و n با $\theta_m < \theta_n$ داریم:

$$P(X_i = 1 \mid \theta_m) \leq P(X_i = 1 \mid \theta_n)$$

و این احتمالِ پاسخ‌های صحیح، تابع سؤال-پاسخ را تشکیل می‌دهد. احتمال اینکه فرد m با مقدار θ کمتر به‌درستی به سؤال i پاسخ دهد باید همیشه کوچک‌تر (مساوی) شخص n با مقدار θ بالاتر باشد. معادلهٔ بالا بیان می‌کند که تابع سؤال-پاسخ تابعی غیر کاهشی از θ است و این تنها محدودیت دراین تابع است. بنابر این در الگوهای سؤال-پاسخ غیرپارامتری، توابع سؤال-پاسخ می‌توانند اشکالی مختلف و بسیار نامنظم داشته باشند. تا زمانی که رابطهٔ بین θ و احتمال پاسخ صحیح غیرکاهشی باشد، هر شکلی برای این تابع مجاز است (سیجتسما[4]، مولنار[5]، ۲۰۰۲).

شکل (۱)، نمونه‌ای از تابع سؤال-پاسخ را نشان می‌دهد. در الگوهای پارامتری سؤال-پاسخ مانند الگوی راش[6]، شکل این تابع با استفاده از یک تابع (مثلاً تابع لجستیک) مشخص می‌شود.

1. Item Response Models
2. Item Response Function (IRF)
3. Wind
4. Sijtsma
5. Molenaar
6. Rasch

بیش از پنجاه سال پیش، لوییز گاتمن[1] روش اسکلوگرام[2] خود را معرفی کرد. با این الگوی جبرگرایانه[3] می‌توان افراد را به صورت «خطی» (با توجه به توانایی‌های آنها) و سؤال‌ها را (بر اساس دشواری) رتبه‌بندی کرد. از آن زمان به بعد، الگوی گاتمن حداقل در دو جهت، یکی از نظر آمار پارامتری بر اساس الگوی راش (۱۹۶۰) و دیگری با الگوی همگنی یکنواخت موکن (۱۹۷۱) خانواده‌ای از الگوهای احتمالی «خطی» تعمیم داده شده است (آنلو[4]، ۲۰۰۷).

در تجزیه و تحلیل داده‌ها بر اساس نظریهٔ کلاسیک آزمون[5]، تأکید بر سؤال‌ها و آماره‌ها نمونه‌ای است که بر حسب نمرات خام محاسبه می‌شود و شاید شناخته‌شده‌ترین این آماره‌ها شاخص همسانی درونی کودر-ریچادسون[6] ۲۰ (KR20) باشد. اما در نظریهٔ صفت پنهان بر برآورد پارامتر سؤال‌ها و افراد و واریانس برآوردگرها که تبدیل غیرخطی از نمرات خام است تأکید می‌شود (اندریچ[7]، ۱۹۸۲).

همسانی درونی این است که سؤال‌های آزمون، همگی «یک‌چیز» را منعکس کنند. «چیز»، همان‌طور که توسط کرونباخ[8] (۱۹۵۱) و لومزدن[9] (۱۹۵۷) توضیح داده شده است، لزوماً به معنای یک عامل یا ویژگی نیست. هر سؤال ممکن است شامل بیش از یک ویژگی یا ترکیبی از ویژگی‌ها باشد، اما اگر همهٔ سؤال‌ها، این ویژگی‌ها را به نسبت یکسان داشته باشند، آنگاه به نظر می‌رسد که آزمون از نظر درونی همسان است. ایدهٔ همسانی درونی در نظریهٔ صفت پنهان، ایدهٔ تک بُعدی بودن است. به این معنا که صفت یک فرد ممکن است با یک مقدار واحد در یک پیوستار پنهان نشان داده شود. تک بُعدی بودن در نظریه صفت پنهان، معادل همسانی درونی در نظریه کلاسیک آزمون، امری نسبی است و نه مطلق.

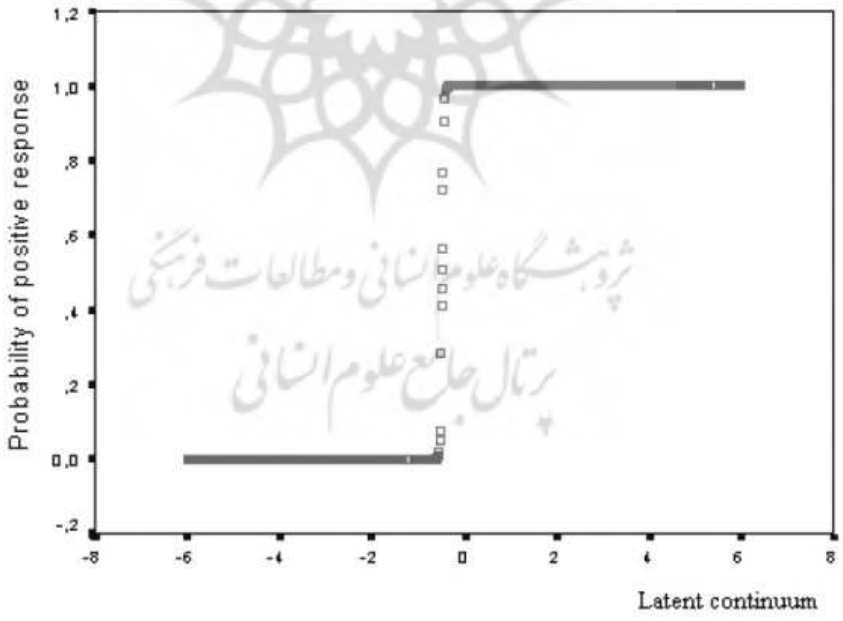
از دیدگاه عملیاتی مقیاس گاتمن از نظر شواهد تک بُعدی بودن، ایده‌آل تلقی می‌شود. ارتباط آن با مفاهیم همسانی درونی نظریهٔ کلاسیک این است که هرچه پاسخ‌ها به این ایده‌آل نزدیک‌تر باشد، ارزش شاخص KR20 بیشتر است. دستیابی به ایده‌آل مقیاس گاتمن در آزمون واقعی دشوار است و مانع اصلی، الزاماتی است که پاسخ‌های هر فرد به یک سؤال را به شیوهٔ معین محدود می‌کند، بنابر این در صورتی‌که پراکندگی دشواری سؤال‌ها زیاد باشد و نیز هیچ دو سؤالی در مقیاس نزدیک به هم نباشد، تحقق مقیاس گاتمن افزایش می‌یابد. همچنین اگر دامنهٔ توانایی‌های افراد نسبتاً زیاد باشد و حتی‌الامکان طیف گسترده‌ای از دشواری سؤال‌ها را پوشش دهد، کمک‌کننده است. میزان انطباق مجموعه‌ای از پاسخ‌ها با مقیاس گاتمن

1. Louis Guttman
2. Scalogram
3. Deterministic
4. Ünlü
5. Classical test theory (CCT)
6. Kuder – Richardson20
7. Andrich
8. Cronbach
9. Lumsden

را می‌توان از شاخص‌های مختلف قابلیت بازسازی[1] نشان داد (گاتمن، ۱۹۴۷).
در الگوهای تابع سؤال-پاسخ، احتمال پاسخ صحیح به سؤال دوارزشی به یک یا چند پارامتر فرد و یک یا چند پارامتر سؤال بستگی دارد. یکی از الگوهای بسیار محبوب تابع سؤال-پاسخ الگوی لجستیک تک پارامتری (الگوی ۱PL) است که به نام مخترع آن، آماردان دانمارکی، جورج راش به‌عنوان الگوی راش نیز شناخته می‌شود. در این الگو احتمال پاسخ صحیح به هر سؤال به‌صورت زیر تعریف می‌شود (که θ_s پارامتر شخص s است، δ_i پارامتر سؤال i است و α را می‌توان به‌عنوان ثابتی برای یک مقیاس خاص در نظر گرفت):

$$P(X_{is}=1|\theta_s,\delta_i)=\frac{e^{\alpha(\theta_s-\delta_i)}}{1+e^{\alpha(\theta_s-\delta_i)}}$$

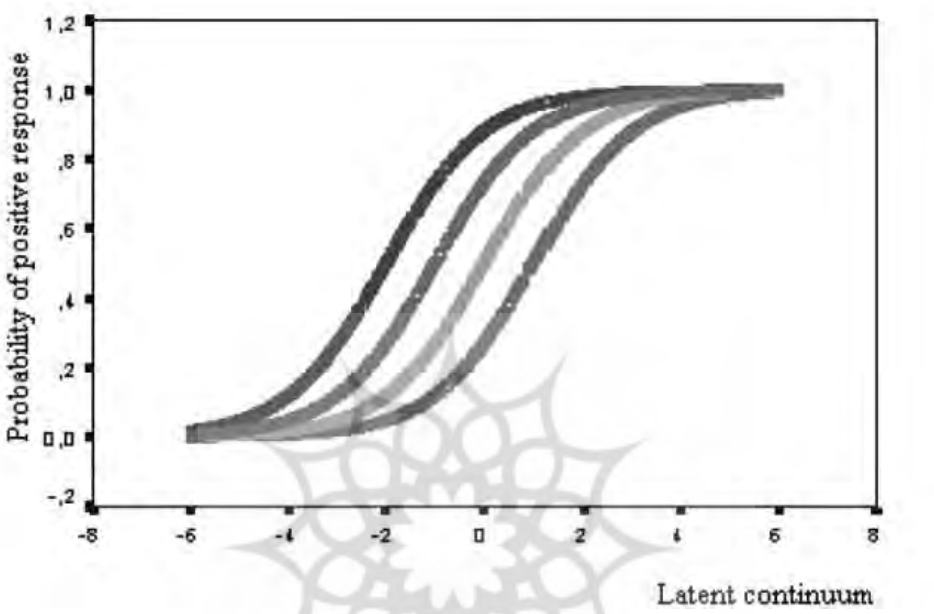
همان‌طور که در شکل (۱) نشان داده شده است، برای $\alpha=+\infty$، در صورتی الگو به مقیاس گاتمن کاهش می‌یابد که در آن اگر $\theta_s>\delta_i$، احتمال پاسخ صحیح به یک نزدیک می‌شود و اگر $\theta_s<\delta_i$، احتمال پاسخ صحیح به صفر نزدیک می‌شود و تابع سؤال پاسخ که احتمال پاسخ صحیح به سؤال i را توسط آزمودنی‌ها با پارامترهای مختلفِ شخص را نشان می‌دهد، تابعی پلکانی[2] به نظر می‌رسد (ون شوور[3]، ۲۰۰۳).



شکل ۱. تابع پاسخ پلکانی

1. Reproducibility
2. Step function
3. Van Schuur

برای مقادیر ثابت α (به‌عنوان‌مثال ۱ =α)، همان‌طور که شکل (۲) نشان می‌دهد، تابع سؤال-پاسخ "IRF" یک سؤال خاص تابعی لجستیک (S شکل) است و توابع "IRF" برای سؤال‌های دیگرِ یک مقیاس، با فرض برابر بودن α های‌شان، همگی موازی هستند (ون شوور،۲۰۰۳).



شکل ۲. توابع پاسخ موازی

مقیاس موکن رویکرد احتمالی غیرپارامتری از نظریهٔ سؤال-پاسخ است که چهارچوبی منظم برای ارزیابی کیفیت اندازه‌گیری‌ها فراهم می‌کند. الگوهای مبتنی بر مقیاس موکن، غیرپارامتری در نظر گرفته می‌شود و رابطهٔ بین متغیر پنهان و احتمال پاسخ (تابع سؤال-پاسخ) مادامی‌که مفروضه‌های اصلی ترتیب برآورده شود، لازم نیست با شکل خاصی مطابقت کند. اگر الگویی از تابع "IRT" را برای ساختن آزمونی به کار ببریم و اندازه‌گیری پاسخ‌دهندگان بر اساس مقیاس ترتیبی برای کارکردهای پیش‌بینی‌شده کافی باشد، استفاده از الگوهای پارامتری "IRT" برای چنین اهداف پیش‌بینی‌شده‌ای می‌تواند بی‌جهت ایجاد محدودیت کند. برازاندن الگویی پارامتری با توابع "IRF" لجستیک بر داده‌ها می‌تواند بیشتر سؤال‌ها را به این دلیل که نمی‌توانند ارتباط کارکردی با توانایی را با یک تابع لجستیک الگوسازی کنند، رد کند، در حالی‌که گنجاندن سؤال‌ها با توابع "IRF" غیرکاهشی، دقت اندازه‌گیری را افزایش می‌دهد (موکن، ۱۹۷۱).

مقیاس موکن به محققان کمک می‌کند تا سؤال‌هایی را که طبق الگوهای محدود کنندهٔ تابع "IRT"

پارامتری بایـد حذف کرد، حفظ کرد. همچنین مقیاس موکن"MSA" مجموعه‌ای از ابزارهای اکتشـافی را بـرای تجزیـه و تحلیل ابعاد فراهم می‌کند کـه در الگوهای پارامتری "IRT" امکان‌پذیر نیسـت (امانز[1]، سیجتسـما و پدرسـن[2]، ۲۰۱۲). بعضی از متخصصان آزمون‌سـازی اسـتدلال می‌کنند که در عمل اکثر توابع "IRF"، لجسـتیک یا چیزی نزدیک به آن هستند. مخالفان این بحث می‌گویند که استفاده از الگوی تک‌پارامتری لجسـتیک باعث رد همهٔ سـؤال‌هایی می‌شـود که شـیب توابع "IRF" آنها متفاوت با الگوی تک‌پارامتری است. استفاده از الگوهای دوپارامتری هم باعث رد سؤال‌هایی می‌شود که مجانب پایین توابع "IRF" آنها بزرگ‌تر از صفر است (به‌عنوان نمونه سؤال‌های چندگزینه‌ای) و استفاده از الگوی سه‌پارامتری لجسـتیک باعث می‌شود که سـؤال‌هایی که مجانب بالای آنها کمتر از یک است و همچنین سؤال‌هایی که توابع "IRF" با شـکل نامنظم دارنـد و یک یا بیش از یک خمیدگی تیز دارند از آزمون یا پرسشـنامه حذف شوند (سیجتسما، مولنار، ۲۰۰۲). هدف اصلی الگوی موکن، رواسازی مقیاس رتبه‌ای متغیر پنهان است. برای سـؤال‌هایی که پیش‌فرض‌های الگوی موکن را برآورده می‌کنند، می‌توان از مجموع پاسخ‌های موجود به سـؤال‌ها برای رتبه‌بندی پاسـخ‌دهندگان در ویژگی پنهان اسـتفاده کرد (همکر[3]، سیجتسـما، مولنار و جانکر[4]، ۱۹۹۷). در مقایسـه بـا الگوهای تابع "IRT" پارامتری، الگوی موکن بـا توجه به رابطهٔ بین صفت پنهان و پاسـخ به سـؤال‌ها به مفروضه‌های کمتری نیاز دارد. بنابر این امکان نگه‌داشتن سؤال‌های مهم‌تر را فراهم می‌کند. در نتیجه ترتیب افراد دقیق‌تر اسـت (سیجتسـما و مولنار، ۲۰۰۲). بسـیاری از محققان به‌ویژه چرنیسـنکو[5]، اسـتارک[6]، چان[7]، دراسـگو[8] و ویلیامز[9]، ۲۰۰۱؛ و همچنین میجر[10] و بنکر[11]، ۲۰۰۴، میجر، تندرو[12] و ندرز[13]، ۲۰۱۴ و رایس[14] و والر[15]، ۲۰۰۹ اشاره کرده‌اند که رویکرد تابع سؤال-پاسخ "IRT" غیرپارامتـری به طورکلی و به طور خاص مقیاس موکن در زمینه‌هایی که در آن فرآیندهای پاسـخ زیربنایی (مانند متغیرهای عاطفی[16]) به‌خوبی درک نمی‌شوند، مفید است. اگرچه مقیاس موکن به طور گسترده در انواع حوزه‌های عاطفی اسـتفاده شده است، اسـتفاده از این روش در ارزیابی‌های آموزشی کمتر متداول اسـت. در آزمون‌های پیشـرفت تحصیلی نیز فرایندهای پاسخگویی آزمون‌شوندگان به‌خوبی مطالعه نشده

1. Emons
2. Pedersen
3. Hemker
4. Junker
5. Chernyshenko
6. Stark
7. Chan
8. Drasgow
9. Williams
10. Meijer
11. Baneke
12. Tendeiro
13. Wanders
14. Reise
15. Waller
16. Affective

است و مقیاس موکن چهارچوبی مفید برای کاوش در خصوصیات اساسی اندازه‌گیری این ارزیابی‌ها از جمله درجهٔ تغییرناپذیری دانش‌آموزان و ترتیب سؤال‌ها، فراهم می‌کند. به عنوان مثال سنجش پیشرفت تحصیلی فرایندی پیچیده است که از داوران خواسته می‌شود تا درک خود از پیشرفت دانشجو را در یک مقیاس رتبه‌بندی منعکس کنند. داوری با استفاده از متغیرهایی مانند کارهای دانشجویی، مقاله‌ها، مقیاس‌های رتبه‌بندی و ویژگی‌های رقابتی فرد صورت می‌پذیرد. به همین ترتیب فرآیندهای پاسخگویی برای سؤال‌های چندانتخابی[1] شامل تعامل بین مکان فرد در یک متغیر پنهان، قالب‌های سؤال، محتوای مورد ارزیابی و سایر ویژگی‌های فرد است. در این حالت به‌کاربردن مقیاس موکن در ارزیابی‌های آموزشی مناسب به نظر می‌رسد. در واقع هنگامی‌که محققان چه در سنجش‌های تشخیصی و چه در سنجش‌های تکوینی علاقه‌مندند که میزان مطابقت ارزیابی با خصوصیات اساسی اندازه‌گیری شده را بررسی یا پیگیری کنند، استفاده از مقیاس موکن مفید است. همچنین در زمینه‌هایی که در آن اطلاعاتی در مورد کیفیت اندازه‌گیری و مرتب کردن افراد یا سؤال‌ها نیاز است، اما حجم نمونه برای دستیابی به برآوردهای باثبات بر اساس الگوهای پارامتری نظریهٔ سؤال-پاسخ پارامتری کافی نیست، این رویکرد کاربرد دارد (ویند، ۲۰۱۷). تحقیقات بر پایه الگوهای مقیاس موکن در مواردی که شواهد پیروی از مفروضه‌های الگوهای آن مهیاست، از تفسیر نمره‌های کل (جمع نمره‌ها) به منظور مرتب کردن معنادار افراد و سؤال‌ها روی متغیر پنهان و بدون هرگونه تبدیل پارامتری، حمایت می‌کنند. به‌علاوه در مواردی که نتیجه‌گیری مطلوب از روش ارزیابی به سطح مقیاس فاصله‌ای نیاز نداشته باشد (سنجش انطباقی با رایانه یا بعضی از روش‌های همترازسازی)، اما بتوان از اطلاعات رتبه‌ای مرتبط با دانش‌آموزان و سؤال‌ها مطلع بود؛ مقیاس موکن مناسب است (سیجتسما و میجر، ۲۰۰۷).

تحلیل مقیاس موکن مجموعه‌ای از ابزارهای آماری را برای ساخت مقیاس، اندازه‌گیری توانایی افراد و بررسی سؤال‌ها در حوزه‌های شخصیتی، شناختی، کیفیت زندگی مرتبط با سلامتی، جامعه‌شناسی، بازاریابی و زمینه‌های دیگری که ابزارهایی با مجموعه‌ای از سؤال‌ها است را فراهم می‌کند. این روش از مفروضه‌های نسبتاً سهل‌گیرانه‌ای استفاده می‌کند که حکایت از ترتیب افراد در یک مقیاس با استفاده از نمره‌های کل افراد در مجموعه سؤال‌ها دارد (سیجتسما و مولنار، ۲۰۰۲).

مقیاس موکن همچنین در مواردی که نتیجه‌گیری مطلوب از یک روش ارزیابی به سطح مقیاس فاصله‌ای نیاز نداشته باشد اما بتوان از اطلاعات رتبه‌ای مرتبط با آزمون‌شوندگان و سؤال‌ها مطلع بود؛ مناسب است (ویند، ۲۰۱۷). هنگامی‌که سطح مقیاس فاصله‌ای مورد نیاز است (سنجش انطباقی با رایانه یا بعضی از روش‌های همترازسازی) رویه‌های مبتنی بر مقیاس موکن می‌تواند یک مرور اولیهٔ اکتشافی در مورد

1. Multiple choice

درجه‌ای که طبق آن روش اندازه‌گیری به الزامات اصلی اندازه‌گیری پایبند است و می‌تواند به‌عنوان لنزی عمل کند تا بتوان داده‌ها را با استفاده از نتایج عددی و نمایشگرهای گرافیکی مورد بررسی قرارداد، فراهم کند (سیجتسما و میجر، ۲۰۰۷). مقیاس موکن به‌سرعت در بین محققان کاربردی محبوبیت پیداکرده است. جستجو در ادبیات نشان می‌دهد که از سال ۲۰۱۵ پژوهش‌های بسیاری با استفاده از این مقیاس انجام شده، اما تاکنون پژوهش‌های چندانی با استفاده از این مقیاس در کشور ما صورت نگرفته است. هدف اصلی این مقاله، ارائه مقدمه‌ای مختصر از ویژگی‌های اصلی رویکرد مقیاس موکن و کاربردهای آن در ارزیابی‌های آموزشی و روان‌شناختی است.

مقیاس بندی الگوهای موکن برای پاسخ‌های دوارزشی

ریشه مقیاس موکن را می‌توان در مقیاس گاتمن جستجو کرد. این مقیاس توسط لوئیس گاتمن (۱۹۵۰) برای سؤال‌های دوارزشی با هدف استخراج یک بُعدِ واحد از داده‌ها و قرار دادن افراد و سؤال‌ها در این بُعد با مقدار عددی ایجاد شد (وبند، ۲۰۱۷). جدول (۱) نمونه‌ای از مقیاس کامل گاتمن را برای آزمون زبان نشان می‌دهد. سؤال‌ها اول تا پنجم از ساده‌ترین به سخت‌ترین و آزمون‌شوندگان A تا E از تواناترین به کم‌توان‌ترین مرتب شده‌اند. انتظار داریم فردی که به سؤال دوم پاسخ صحیح داده است به سؤال اول نیز پاسخ صحیح داده باشد. به طور مثال فردی که نحوهٔ تشکیل سؤال‌ها را در زمان گذشته یاد گرفته است (سؤال دوم) باید زمان حال ساده را نیز یاد گرفته باشد (سؤال اول). زبان‌آموزی که به حد معینی از توانایی دستورزبانی رسیده باشد می‌تواند به تمام سؤال‌های زیر این سطح به‌درستی پاسخ دهد و هیچ یک از سؤال‌های بالای آن را نتواند پاسخ دهد. بر اساس نمرهٔ کل، بازتولید بُردار پاسخ هر فرد نیز باید امکان‌پذیر باشد. نمره کل دو به معنی الگوی پاسخ ۱۱۰۰۰ و نمره کل چهار به معنی الگوی پاسخ ۱۱۱۱۰ است. اگر مقیاس گاتمن برقرار باشد، گواه این است که سؤال‌ها با یک مجموعه تک‌بعدی مطابقت دارند و افراد در یک پیوستار مرتب شده‌اند. پاسخ‌های غیرمنتظره به‌عنوان خطاهای گاتمن نامیده می‌شود (بقایی، ۲۰۲۱).

جدول ۱. نمونه‌ای از مقیاس کامل گاتمن

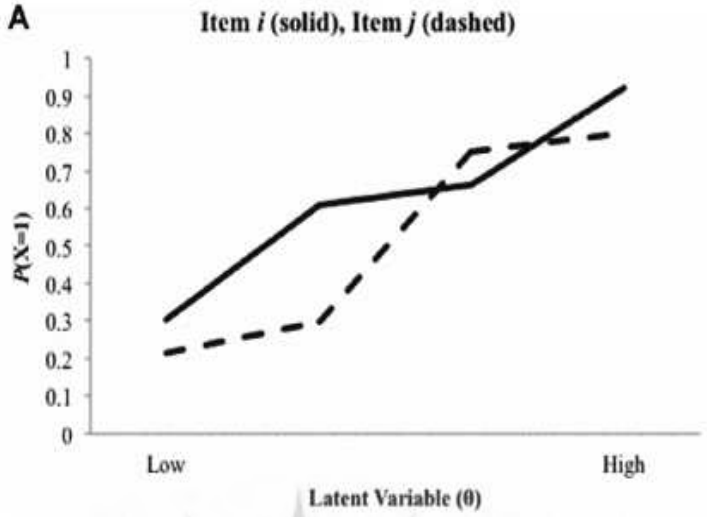
	سؤال‌ها				
آزمون‌شوندگان	۱	۲	۳	۴	۵
A	۱	۱	۱	۱	۱
B	۱	۱	۱	۱	۰
C	۱	۱	۱	۰	۰
D	۱	۱	۰	۰	۰
E	۱	۰	۰	۰	۰

در ارائهٔ اصلی مقیاس موکن (۱۹۷۱) برای ارزیابی کیفیت مقیاس‌های ساخته شده از سؤال‌های دو ارزشی ($X = 0$ یا 1)، دو الگوی ارائه شـد. برای اینکه پاسـخ‌های دوارزشـی برای مقیاس موکن مناسب باشند، نمـره یک بایـد موقعیت مکانی بالاتـر در متغیر پنهان (توانایـی بالاتر در آزمون‌های پیشـرفت تحصیلی) نسـبت به نمره صفر را نشـان دهد. الگوی اول، الگوی همگنی یکنواخت است که از بین دو الگوی اصلی مقیاس موکن، عمومی‌تر اسـت (ویند، ۲۰۱۷). الگو همگن یکنواخت مبتنی بر سه فرض اساسی به شرح زیـر اسـت: (۱) یکنواختی[1]: با افزایش مـکان در متغیر پنهان، احتمال پاسـخ صحیح ($X = 1$) کاهش نمی‌یابد. (۲) تک بُعدی بودن[2]: پاسـخ‌های سـؤال، نشانگر شـواهدی از تنها یک متغیر پنهان است. (۳) استقلال موضعی[3]: پاسخ به یک سؤال پس از کنترل متغیر پنهان، تحت تأثیر پاسخ به سؤال‌ها دیگر نیست (سیجتسما و وندرارک، ۲۰۱۷).

چندین ویژگی در مورد الگوی همگن یکنواخت قابل‌توجه اسـت: ۱) الگوی همگن یکنواخت تابع پاسـخ سؤال را به‌جز فرض یکنواختی بر پیش‌فرض خاصی محدود نمی‌کند. در نتیجه تابع پاسخ سؤال‌هایی که به الگوی همگن یکنواخت پایبند هسـتند ممکن اسـت اشکال مختلفی به خود بگیرد که لزوماً با شکل اجایو لجستیک که معمولاً با الگوهای پارامتریک نظریهٔ سؤال-پاسخ همراه است، مطابقت ندارد (ویند، ۲۰۱۷). شـکل (۳) نشان‌دهندهٔ یک جفت تابع پاسخ سؤال برای دو سؤال است که با مفروضه‌های الگوهای همگن یکنواخت مطابقت دارند. محور Y، احتمال پاسـخ صحیح (تابع پاسخ سؤال) و محور X متغیر پنهان (θ) را نشـان می‌دهد. اگرچه توابع پاسخ سؤال‌های i و j متقاطع اند، اما آنها به الزامات الگوی همگن یکنواخت پایبند هستند، زیرا در اثر افزایش مکان در متغیر پنهان، کاهش نمی‌یابند (سیجتسما و وندرارک، ۲۰۱۷).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

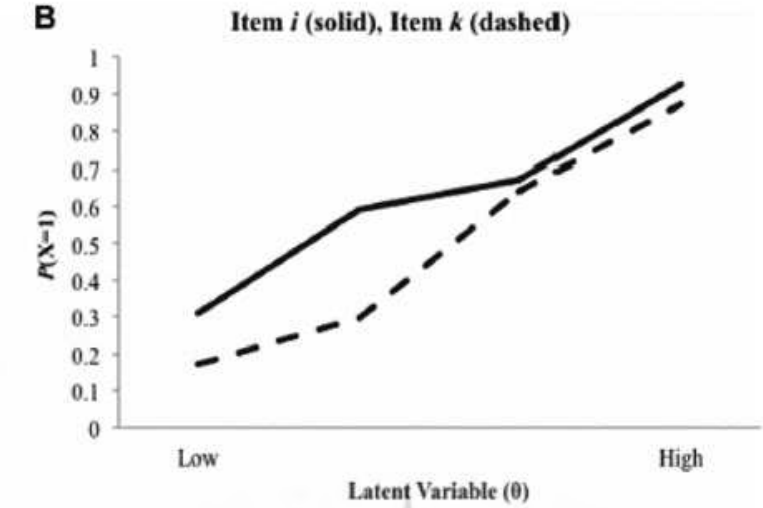
1. Monotoniciy
2. Unidimensionality
3. Local independence



شکل ۳. توابع پاسخ غیرکاهشی و متقاطع

هنگامی که داده‌ها متناسب با مفروضه‌های الگوی همگن یکنواخت است، ترتیب نسبی آزمون‌شوندگان در متغیر پنهان در تمام سؤال‌ها، ثابت است. از آنجا که متناسب با مفروضه‌ها، الگوی همگن یکنواخت شواهدی برای ترتیب ثابت افراد در سؤال‌ها فراهم می‌کند، این الگو می‌تواند مشابه الگو لجستیک دو پارامتری در نظریهٔ سؤال-پاسخ پارامتری باشد (وینـد، ۲۰۱۷). ۲) الگوی دوم موکن (۱۹۷۱) الگوی یکنواختی مضاعف است که حالتی خاص از الگوی همگن یکنواخت به حساب می‌آید. الگوی همگن مضاعف در سه فرض الگوی همگن یکنواخت مشترک است، اما فرض چهارمی را هم شامل می‌شود و آن اینکه ترتیب‌بندی نامتغیر سؤال‌ها یا نامتغیر بودن ترتیب‌بندی سؤال‌ها[1] (IIO): توابع پاسخ برای سؤال‌های خاص با توابع پاسخ برای هر سؤال دیگر تلاقی ندارند (سیجتسما و وندرارک، ۲۰۱۷). شکل (۴) نشانگر یک جفت تابع برای دو سؤال دو ارزشی است که با مفروضه‌های الگوی همگن مضاعف مطابقت دارند زیرا توابع پاسخ سؤال‌های i و k هر دو یکنواخت (غیر کاهشی در سرتاسر متغیر پنهان) و غیرمتقاطع هستند (وینـد، ۲۰۱۷).

1. Invariant item ordering (IIO)



شکل ٤. توابع پاسخ غیرکاهشی و غیرمتقاطع

نتیجهٔ مهم این پیش‌فرض‌ها برای سؤال‌های دو ارزشی این است که وقتی داده‌ها با مفروضه‌های الگوی همگن مضاعف متناسب است، سؤال‌ها به روش یکسان در بین دانش‌آموزان مرتب می‌شود، زیرا تناسب با مفروضه‌های الگوی همگن مضاعف، شواهدی برای ترتیب‌بندی ثابت، هم برای آزمون‌شوندگان و هم برای سؤال‌ها فراهم می‌کند. این الگو به‌عنوان نسخهٔ رتبه‌ای الگوی راش دو ارزشی یا الگوی لجستیک تک پارامتری در نظریهٔ سؤال- پاسخ پارامتری توصیف شده است (انگلهارد[1]، ۲۰۰۸).

مقیاس‌پذیری

موکن (۱۹۷۱) بسط‌هایی از ضرایب مقیاس‌پذیری که ابتدا توسط لووینگر[2] (۱۹۴۸) ارائه شده است را برای توصیف درجه‌ای که هر سؤال و مجموعه‌های کلی از سؤال‌های مقیاس می‌تواند برای مرتب کردن افراد در یک سازه استفاده شود را توسعه داد. شاخص‌های مقیاس‌پذیری خلاصه‌ای از تأثیر خطاهای گاتمن بر کیفیت یک روش اندازه‌گیری را ارائه می‌دهد. خطاهای کمتر گاتمن تفسیر نمره‌های کل افراد را به‌عنوان شاخص‌های مرتب کردن فرد در سازه تسهیل می‌کند (ویند، ۲۰۱۷).

موکن از سه ضریب مقیاس‌پذیری استفاده می‌کند. ضریب مقیاس‌پذیری سؤال (H_i)، ضریب مقیاس‌پذیری جفت سؤال (H_{ij}) و ضریب مقیاس‌پذیری کل (H^T). ضریب مقیاس‌پذیری با استفاده از نسبت فراوانی‌های خطای گاتمن مشاهده شده و مورد انتظار محاسبه می‌شود. ضرایب مقیاس‌پذیری به‌عنوان یک منهای

1. Engelhard
2. Loevinger

نسبت خطاهای گاتمن مشاهده‌شده (F_{ij}) و مورد انتظار (E_{ij}) برای هر سؤال، برای جفت سؤال‌ها و کل آزمون محاسبه می‌شود. ضریب مقیاس‌پذیری برای جفت سؤال‌ها:

$$H_{ij} = 1 - \frac{F_{ij}}{E_{ij}}$$

که در آن F_{ij} تعداد خطاهای گاتمن برای جفت سؤال i و j است و E_{ij} تعداد مورد انتظار خطاهای گاتمن تحت استقلال حاشیه‌ای ($\frac{\text{بیشتر فراوانی} \times \text{کمتر فراوانی}}{\text{کل فراوانی}}$) است. این روش معادلِ محاسبهٔ کواریانس دو سؤال است (سیجتسما و مولنار، ۲۰۰۲). ضریب مقیاس‌پذیری جفت سؤال (H_{ij}) به‌عنوان نسبت کواریانس بین هر جفت از سؤال‌های i و j حداکثر کواریانس ممکن آنها با توجه به توزیع حاشیه‌ای نمره‌های دو سؤال تعریف می‌شود و منعکس‌کنندهٔ سازگاری درونی هر جفت سؤال است (موویج[1]، ۲۰۱۲).

به همین ترتیب ضرایب مقیاس‌پذیری برای هر یک از سؤال‌ها H_i با استفاده از رابطهٔ بالا با در نظر گرفتن تعداد خطاهای گاتمن مشاهده شده و مورد انتظار برای همهٔ جفت سؤال‌های مرتبط با سؤال مورد نظر محاسبه می‌شود. دو سؤال i و j را طوری در نظر گرفته می‌شود که i سخت‌تر از j باشد. جدول دو در دو زیر نتایج احتمالی را نشان می‌دهد.

جدول ۲. نتایج احتمالی برای دو سؤال

	سؤال j	
سؤال i	۱	۰
۱	A(۱,۱)	B(۱,۰)
۰	C(۰,۱)	D(۰,۰)

سلول‌های A ، B ، C و D به ترتیب بیانگر تعداد پاسخ‌دهندگانی است که به هر دو سؤال به‌درستی پاسخ داده‌اند، به سؤال i درست پاسخ داده‌اند اما سؤال j را اشتباه گفته‌اند، به سؤال i اشتباه اما به سؤال j صحیح پاسخ داده‌اند و آنها که به هر دو سؤال اشتباه پاسخ داده‌اند. ازآنجایی‌که سؤال j ساده‌تر از سؤال i است، انتظار داریم سلول C بزرگ‌تر از سلول B باشد. ارزیابی این فراوانی‌ها نشان می‌دهد که آیا محدودیت‌های ترتیب برای سؤال‌ها حفظ شده است یا خیر. این دقیقاً شبیه به بررسی این است که آیا خطاهای گاتمن رخ داده است یا خیر. هرچه تعداد خطاها کمتر باشد، پاسخ‌ها بیشتر از الگوی گاتمن پیروی می‌کند و سلسله مراتبی که توسط سؤال‌ها تشکیل می‌شود قوی‌تر است (بقایی، ۲۰۲۱).

ضریب مقیاس‌پذیری سؤال (H_i) به‌صورت نسبت جمع کل کواریانس‌های جفتی با توجه به هر سؤال i و

1. Mooij

مجموع حداکثر کواریانس‌های دوتایی این سؤال بیان می‌شود. به طور خلاصه دقت تمییز سؤال و قدرت رابطه بین سؤال و مجموعه کل سؤال‌ها است (امانز[1] و همکاران، ۲۰۱۲). مقادیر بالاتر H_i نشان‌دهندهٔ قدرت تشخیص بالاتر است (موویج، ۲۰۱۲).

وقتی داده‌ها با الگوی "MH" مطابقت دارند، مقادیر ضرایب مقیاس‌پذیری مثبت و از صفر تا یک متغیرند، مقدار یک، نشان دهندهٔ نبود خطاهای گاتمن و مقادیر نزدیک به صفر، نشان‌دهنده بسیاریِ خطاهای گاتمن است (وینـد، ۲۰۱۷). در نهایت ضریب مقیاس‌پذیری برای کل آزمون، H^T، با شمارش تعداد خطاهای گاتمن مشاهده شده و مورد انتظار برای کل داده‌ها با فرض حفظ ساختار کامل گاتمن محاسبه می‌شود. ضریب H^T برای کل مقیاس، نقش محوری در الگو همگنی یکنواخت دارد و نشان می‌دهد که ساختار داده تا چه حد به الگوی گاتمن کامل نزدیک است (بقایی، ۲۰۲۱). ضریب H^T شاخصی دقیق از تک‌بعدی بودن و معنی‌داری نمرهٔ کل برای مرتب کردن افراد است (سیجتسما و مولنار، ۲۰۰۲). موکن (۱۹۷۱) اظهار داشت که «ما ترجیح می‌دهیم از ضریب همگنی H^T، به‌عنوان معیار مقیاس‌پذیری به معنای همگنی یکنواخت استفاده کنیم. ضریب مقیاس‌پذیری، به‌عنوان تنها معیار مقیاس‌پذیری ما خواهد بود». اگر ضریب مقیاس‌پذیری کل بین ۰/۳ و ۰/۴ باشد ($0/3 \leq H^T < 0/4$) مقیاس، ضعیف در نظر گرفته می‌شود، ضریب بین ۰/۴ تا ۰/۵، ($0/4 \leq H^T < 0/5$) مقیاس متوسط و اگر بزرگ‌تر از ۰/۵، ($0/5 \leq H^T$) باشد، قوی در نظر گرفته می‌شود. مقادیر H_{ij} باید بیشتر از صفر باشد و سرانجام اگر ضریب ($H_j < 0/3$) باشد، سؤال j باید بررسی یا حذف شود، اما اگر ($0/3 \leq H_j$) باشد، سؤال می‌تواند برای تشکیل مقیاس موکن انتخاب شود (عبدالحمید[2] و همکاران، ۲۰۲۰).

ضریب مقیاس‌پذیری کل H^T نسبت مجموع تمام کواریانس‌های جفتی و جمع کل حداکثر کواریانس‌های جفتی است و مقادیر بالاتر H^T نشان می‌دهد که از میانگین نمرهٔ کل می‌توان برای مرتب کردن افراد با دقتی بالا استفاده کرد (موویج، ۲۰۱۲).

یکنواختی

همان طور که قبلاً توضیح داده شد، یکنواختی به این معنی است که با افزایش مکان افراد در متغیر پنهان، احتمال پاسخ صحیح یا باید افزایش یابد یا ثابت بماند. چندین روش برای ارزیابی یکنواختی از جمله بررسی نمره‌های ماندهٔ گروهی[3] وجود دارد. نمرهٔ مانده[4] عبارت از امتیاز یا نمرهٔ خام کل بدون احتساب نمرهٔ سؤال مورد بررسی است. هر نمرهٔ مانده باید حداقل تعداد اعضای ثابتی داشته باشد. سپس با استفاده از روش‌های گرافیکی و روش‌های آماری می‌توان یکنواختی را ارزیابی کرد. نمرهٔ ماندهٔ گروهی و

1. Emons
2. Abdelhamid
3. .Restscore groups
4. ·Restscore

نسبت پاسخ‌دهندگانی است که به یک سؤال به‌درستی پاسخ داده‌اند (احتمال پاسخ صحیح) در محورهای x و y رسم می‌شوند. با افزایش نمرهٔ مانده، احتمال پاسخ صحیح باید افزایش یابد یا ثابت بماند. به این معنی که توابع "IRF" باید نسبت به مقادیر فزایندهٔ نمرهٔ مانده که مقادیر θ بالاتر را نشان می‌دهند، غیر کاهشی باشند (وینـد، ۲۰۱۷).

آزمونی را با ده سؤال در نظر بگیرید. فرض کنید می‌خواهیم یکنواختی را برای سؤال ده ارزیابی کنیم. به‌استثنای سؤال ده، ده گروه نمره مانده وجود دارد، یعنی دانش‌آموزانی که در کل آزمون نمره صفر گرفتند، آنهایی که نمرهٔ ۱ گرفتند، آنهایی که نمرهٔ ۲ تا کسانی که نمرهٔ ۹ گرفتند. جدول (۳) نشان می‌دهد که در گروهی که در کلّ آزمون نمره صفر کسب کرده‌اند، هیچ‌کس به سؤال ۱۰ پاسخ صحیح نداده است. در گروهی که در کل آزمون نمره ۱ را کسب کرده‌اند، پنج درصد به سؤال ۱۰ پاسخ صحیح داده‌اند و ... (بقایی، ۲۰۲۱).

جدول ۳. نمرات باقی‌مانده گروهی و نسبت‌های مرتبط با آنها

نمرات باقی‌مانده	۰	۱	۲	۳	٤	٥	٦	۷	۸	۹
نسبت پاسخ‌های صحیح	۰/۰	۰/۰۵	۰/۰۹	۰/۱٤	۰/۲٥	۰/٤۰	۰/٤۹	۰/٦٥	۰/۷۲	۰/۸٤

اگر گروه‌های باقی‌مانده کوچک باشند، نسبت پاسخ‌های صحیح، برآورد قابل‌اعتمادی از احتمال پاسخ صحیح نیست. اگر این اتفاق بیفتد، گروه‌های باقی‌مانده با گروه‌های مجاور ترکیب می‌شوند. اگر ۵۰۰ N> باشد، حداقل تعداد برای گروه نمرات باقی‌مانده $\frac{N}{10}$ است. اگر ۵۰۰>N>۲۵۰ باشد، حداقل تعداد افراد برای گروه‌های باقی‌مانده $\frac{N}{5}$ است؛ اگر ۲۵۰N> باشد، حداقل اندازه گروه‌های باقی‌مانده $\frac{N}{3}$ است (سیجتسما و مولنار، ۲۰۰۲). این فرمول‌ها تضمین می‌کنند که حداقل تعداد در گروه‌های باقی‌مانده ۵۰ باشد.

رتبه‌بندی نامتغیر سؤال‌ها[1]

در آزمون‌های روان‌شناختی و آموزشی، سؤال‌ها برای نشان دادن سطوح مختلف سازه نوشته می‌شوند. در یک آزمون گرامر انگلیسی، سؤال شرطی نوع سوم[2] باید دشوارتر از سؤال‌های مربوط به مبحث مجهول[3] باشد. فردی که به اولی به‌درستی پاسخ می‌دهد، نسبت به کسی که دومی را درست پاسخ می‌دهد، در گرامر انگلیسی مهارت بیشتری دارد. تفاوت در دشواری سؤال‌ها، با میانگین سؤال‌ها در جمعیت موردنظر اندازه‌گیری می‌شود. بااین‌حال، این ترتیب دشواری که از پاسخ‌های جمعیت، کمی سازی می‌شود ممکن

1. Invariant item ordering
2. .The conditional type III
3. Passive voice

است برای همه افراد صادق نباشد. باید به‌صورت تجربی بررسی شود که سؤال شرطی نوع سوم درواقع دشوارتر از سؤال مجهول برای همه امتحان شوندگان بوده است؛ یعنی ترتیب سؤال‌ها به‌لحاظ دشواری، در بین افراد با سطوح متفاوت توانایی گرامری ثابت می‌ماند. سازندگان و کاربران آزمون باید بتوانند فرض کنند که فردی با نمره کل بالاتر به تمام سؤال‌هایی که فردی با نمره کل پایین‌تر پاسخ داده است، به‌علاوه چند سؤال دیگر به‌درستی پاسخ داده است. اگر مجموع نمرات، ترکیبی تصادفی از پاسخ‌های صحیح یا به عبارتی فاقد الگوهای قابل پیش‌بینی باشد، معنی‌دار بودن نمرات کل زیر سؤال می‌رود (بقایی ۲۰۲۱).

اگر در یک پرسشنامه افسردگی یا اضطراب، IIO وجود داشته باشد، نشان می‌دهد که فردی با نمره کل بالاتر، تمام علائم فردی با نمره پایین‌تر و برخی علائم اضافی را دارد. علاوه بر این نمره کل را می‌توان هم به‌عنوان شاخصی از سطح صفت و هم به‌عنوان فهرستی از علائمی که بیمار از آن رنج می‌برد به‌طور معناداری تفسیر کرد؛ زیرا سلسله‌مراتب، علائم را جذب می‌کند (لیگتوت[1] و همکاران، ۲۰۱۱).

علاوه بر این تحلیل الگوهای پاسخ افراد که به‌عنوان برازش فرد[2] در ادبیات IRT شناخته می‌شود، تنها در صورتی منطقی است که IIO برقرار باشد. اگر الگوی پاسخ فردی به ۱۰ سؤال که از ساده‌ترین تا سخت‌ترین مرتب‌شده‌اند، ۱۱۱۱۰۰۰۱۰۱ باشد، بدیهی است که پاسخ صحیح او به سؤال‌های هشتم و دهم غیرمنتظره است و این به معنای عدم برازش فرد است زیرا این دو سؤال سخت را در حالی به‌درستی پاسخ داده است کهسه سؤال ساده‌تر را ازدست‌داده است. این نتیجه‌گیری تنها در صورتی معتبر است که بتوانیم فرض کنیم که ترتیب دشواری سؤال‌ها برای همه امتحانشوندگان یکسان است، یعنی IIO برقرار است. اگر IIO پابرجا نباشد، ممکن است سؤال‌های هشتم و دهم برای این آزمودنی خاص سخت نباشد و نتیجه‌گیری عدم برازش برای این فرد اشتباه باشد. وقتی یک سؤال برای یک آزمودنی آسان و برای دیگری سخت باشد، تفسیر نمره کل مشکل خواهد بود. ازاین‌رو، IIO تفسیر نمره آزمون را تسهیل می‌کند (بقایی ۲۰۲۱). IIO را می‌توان با استفاده از چندین روش ازجمله روش مانده[3]، تقسیم سؤال[4]، روش ماتریس نسبت[5] و روش تقسیم نمرات مانده[6] ارزیابی کرد (ون شوور، ۲۰۱۱). در روش مانده‌ها، IIO برای همه جفت سؤال‌ها به‌طور جداگانه ارزیابی می‌شود و عدم تلاقی IRF های آن‌ها بررسی می‌شود. مانده عبارت از نمره کل منهای نمره جفت سؤال‌ها است. نمرات مانده بر روی محور x و احتمال پاسخ صحیح متناظرش (نسبت پاسخ‌های صحیح) در محور y رسم می‌شوند. به‌این‌ترتیب IRF ها ترسیم‌شده و عدم تقاطع آن‌ها بررسی می‌شود. اگر IRF ها ترتیب نسبی یکسانی را در محدوده نمرات مانده داشته باشند، یعنی ترتیب

1. Ligtvoet
2. Person fit
3. Restscore method
4. .Item-splitting method
5. Proportion matrices
6. .Restscore splitting method

آن‌ها معکوس نشـود، خاصیت IIO برای جفت سـؤال‌ها باقی می‌ماند. علاوه بر بررسـی گرافیکی، آزمون فرضیه‌های آماری نیز وجود دارد (بقایی، ۲۰۲۱).

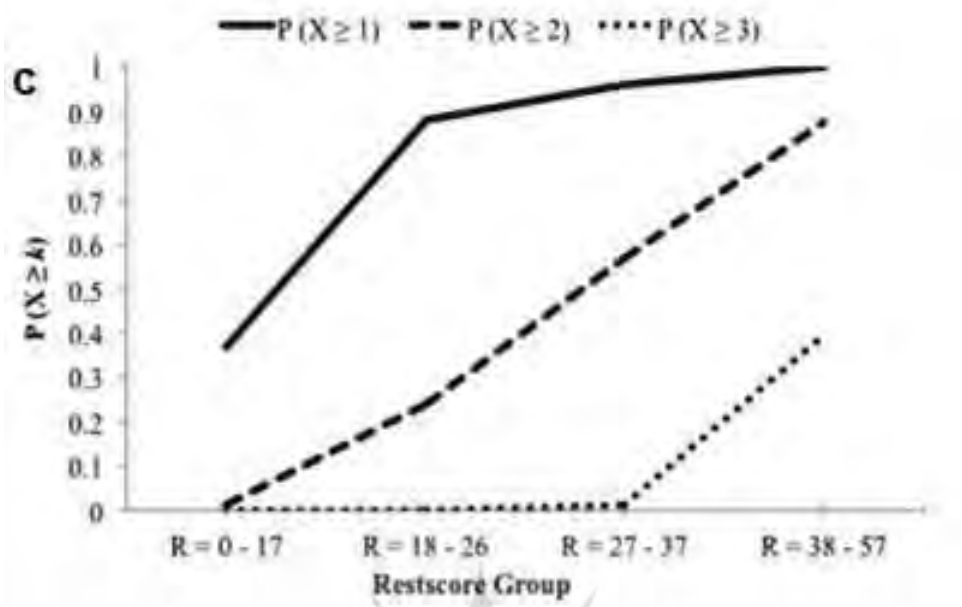
مقیاس بندی الگوهای موکن برای پاسخ‌های چند ارزشی

مولنار (۱۹۹۷)، تحلیل مقیاس موکن را برای سـؤال‌های چند ارزشـی[1] گسـترش داد. اصول MSA برای سـؤال‌های چندارزشـی مانند سـؤال‌های دوارزشی اسـت با این تفاوت که برای سـؤال‌های چندارزشی، مفروضات الگو هم در سطح سؤال و هم در سطح گزینه‌های سؤال بررسی می‌شود. یک سؤال لیکرت با پنج رده پاسـخ «کاملاً موافق»، «موافق»، «بی‌نظر»، «مخالف» و «کاملاً مخالف» دارای چهار مرحله[2] است. در مورد سـؤال‌های چند ارزشـی، هر مرحله دارای یک تابع پاسخ اسـت که به‌عنوان تابع پاسخ مرحله سؤال[3] (ISRF) شناخته می‌شود. ISRF ها احتمال تأیید یک رده‌ی خاص را در صفت نهفته θ نشان می‌دهند.

تعداد ISRF ها همیشه $m-1$ در هر سؤال است، m تعداد رده‌های پاسخ در پرسشنامه است؛ بنابراین در یک مقیاس لیکرت ۵ درجه‌ای، هر سـؤال دارای چهار ISRF اسـت. فرض کنید x_i امتیاز سؤال i باشد، $i = 0,1,\dots,m$، بنابرایـن بـرای یک مقیـاس پنج‌درجه‌ای، این به معنای $x_i = 0,1,2,3,4$ اسـت. احتمال $x_i=0$ (کمترین مقدار مقیاس) نادیده گرفته می‌شـود زیرا طبق تعریف، این برابر با یک است. الگو همگنی یکنواخت برای سـؤالات چندارزشی باید برای مراحل سـؤال رعایت شود؛ یعنی احتمال تأیید رده k یا بالاتر، با افزایش مقادیرθ، کاهشـی نداشته باشـد. رده‌های پاسخ باید نظم معنی‌داری داشته باشند و سطوح بالاتر، مقدار بیشتری از صفت پنهان را نشان دهند (بقایی، ۲۰۲۱).

فـرض یکنواختـی برای سـؤال‌های چندارزشـی حاکی از آن اسـت کـه ISRF ها باید در θ غیرکاهشـی باشـند. یکنواختی پنهان[4] یک فرض غیرقابل مشـاهده اسـت، یعنی $P(X_i \geq x|\theta)$در θ غیرکاهشـی اسـت. یکنواختی آشـکار[5] یک ویژگی قابل‌مشاهده اسـت به‌عنوان‌مثال $P(X_i \geq x|R=r)$؛ r نمره مانده‌ایسـت که نماینده θ برای بررسی یکنواختی پنهان قابل‌مشاهده است (شبیه به‌IIO آشکار که ویژگی قابل‌مشاهده IIO است). برای پاسخ‌های دوارزشی سؤال‌های دارای یکنواختی آشکار دلالت بر یکنواختی پنهان دارد اما برای سؤال‌های چند ارزشی این‌طور نیست و این رابطه پیچیده‌تر است. شکل ۵ (c) تبعیت از فرض یکنواختی را نشان می‌دهد زیرا ISRF ها با افزایش نمرات مانده، کاهش نمی‌یابند (مولنار، ۱۹۹۷).

1. polytomous items
2. Steps
3. Item Step Response Function
4. Latent monotonicity
5. Manifest monotonicity



شکل ۵. توابع ISRF برای یک سؤال ٤ گزینه‌ای

الگو یکنواختی مضاعف برای سؤال‌های چندارزشی شامل سه فرض: تک‌بعدی بودن، یکنواختی و استقلال موضعی به همراه ISRF های غیرمتقاطع است. به این معنی که احتمال تأیید یک رده معین یا بالاتر روی یک سؤال باید دارای نظم نسبی یکسانی در تمام مقادیر θ باشد. ISRF ها طبق تعریف در درون یک سؤال متقاطع نیستند. مولنار (۱۹۹۷) بیان کرد که ISRFهای سؤال‌های مختلف نیز باید غیر متقاطع باشند. بااین‌حال برآوردن این شرایط در داده‌های واقعی بسیار دشوار است. علاوه بر این ISRF های ثابت برای سؤال‌های مختلف به معنای IRFهای ثابت برای سؤال‌ها نیست (سیجتسما، میر و وندرارک ۲۰۱۰). ازآنجا که ISRF های ثابت برای سؤال‌های مختلف به ایجاد امتیازهای ترتیبی مفید کمک نمی‌کنند و در سؤال‌های چندارزشی تمرکز بر IRF های کلی ثابت است (سیجتسما، میر و وندرارک ۲۰۱۰). توجه به این نکته مهم است که برای MSA چندارزشی، نسخه ضعیف‌تری از ترتیب تصادفی بر روی متغیر پنهان[1] (SOL) وجود دارد. برای سؤال‌های دوارزشی، اگر الگو MH برقرار باشد از نمره کل X^{+} می‌توان برای ترتیب تصادفی افراد روی θ استفاده کرد (سیجتسما و مولنار ،۲۰۰۲). برای سؤال‌های چند ارزشی، از دیدگاه نظری، X^{+} ممکن است در مرتب کردن افراد روی θ شکست بخورد (همکر، سجیتسما، مولنار و یونکر، ۱۹۹۷)؛ اما وندرآرک (۲۰۰۵) نشان داد که اگر تعداد سؤال‌ها بیش از پنج باشد، نقض ترتیب تصادفی به‌ندرت رخ می‌دهد.

1. Stochastic Ordering on the Latent variable

وندرآرک و برگسما[1] (۲۰۱۰) ثابت کردند که برای سؤال‌های چند ارزشی شکل ضعیف‌تری از ترتیب تصادفی برقرار است. در سؤال‌های چند ارزشی اگر الگو همگنی یکنواخت برقرار باشد، به این معنی است که با استفاده از نمره کل، نمونه را می‌توان به دو گروه تقسیم کرد، به‌طوری‌که افراد دارای بالاترین و کمترین مقدار θ را شناسایی کرد. همکر، سیتسما، مولنار و یونکر (۱۹۹۷) استدلال کردند که در مورد سؤال‌های چند ارزشی، نمره کل X^+ پاسخ‌دهندگان را دقیقاً بر روی θ مرتب نمی‌کند. بااین‌وجود، وندرآرک (۲۰۰۰) با استفاده از مطالعات شبیه‌سازی نشان داد که برای اهداف عملی، X^+ به‌درستی افراد را مرتب می‌کند و تحریف‌ها فقط برای پاسخ‌دهندگان با نمره کلی بسیار نزدیک (به‌اندازه یک یا دو امتیاز از هم) رخ می‌دهد.

ده گام برای انجام تجزیه تحلیل مقیاس موکن

مرحلهٔ اول: بررسی داده‌ها. داده‌ها بررسی‌شده و در صورت بروز مشکلات خاص، اقدامات مناسب انجام می‌شود (گام‌های ۱ تا ۳).

گام (۱) کدگذاری مجدد. باید نمرات سؤال‌های که به‌صورت منفی[2] نوشته‌شده‌اند بر اساس ویژگی موردانداز‌ه‌گیری مجدداً کدگذاری شوند. به‌طوری‌که برای همه سؤال‌ها، نمرات بالاتر به معنای موقعیت بالاتر در مقیاس اندازه‌گیری باشد (سیجتسما و وندرآرک، ۲۰۱۷).

گام (۲) نمرات غیرقابل‌قبول[3] و داده‌های بدون پاسخ[4]. با نمرات غیرقابل‌قبول به‌عنوان مقادیر بدون پاسخ (گمشده) برخورد می‌شود. باید درصد کلی سؤال‌ها بدون پاسخ را در مجموعه داده‌ها و پاسخ‌دهندگانی که تعدادی سؤال را بی‌پاسخ گذاشته یا تعدادی پاسخ نامعقول داده‌اند به دست آورد و از خود پرسید که چرا این حجم از داده‌ها ازدست‌رفته‌اند و چرا برخی پاسخ‌دهندگان خاص تعداد زیادی از سؤال‌ها را بی‌پاسخ رها کرده‌اند. اگر به‌عنوان‌مثال بیش از ۱۰٪ از کل داده‌ها ازدست‌رفته‌اند، آیا در طرح مطالعه[5] یا عبارت[6] برخی سؤال‌ها مشکلی وجود دارد؟ اگر به‌عنوان‌مثال آزمودنی حداقل ۳۰٪ از سؤال‌ها را بی‌پاسخ رها کرده است، آیا آن‌ها انجام آزمون را جدی تلقی کرده بودند؟ می‌توان نمرات سؤال‌ها بدون پاسخ را با استفاده از روش جاگذاری دوراهه[7] محاسبه کرد (ونگینکل[8]، وندرآرک[9] و سیجتسما، ۲۰۰۷) اما روش‌های دیگری همچون جاگذاری چندگانه (وندرآرک و سیجتسما، ۲۰۰۵) نیز امکان‌پذیر است.

گام (۳) داده‌های پرت[10]. باید مشخص شود که آیا نمره‌های سؤال‌ها واجد شرایط برای داده‌های پرت،

1. Bergsma
2. Negatively worded
3. Inadmissible scores
4. Missing data
5. Study design
6. Wording
7. Two-way imputation
8. Van Ginkel
9. Van der Ark
10. Outliers

دارای الگوی خاصی هستند یا خیر؟ چراکه می‌توان شاهد الگوهای پاسخ غیرمنتظره بود. به‌عنوان نمونه بسیار مشاهده‌شده است که افراد به سؤال‌ها دشوار، پاسخ صحیح و به سؤال‌ها آسان، پاسخ نادرست می‌دهند. در تجزیه‌وتحلیل داده‌ها توصیه می‌شود که تحلیل یک‌بار بر روی داده‌های کامل و بار دیگر بدون داده‌های پرت انجام گردد. اگر حذف تعداد کمی از داده‌های پرت تا حد زیادی بر نتایج مقیاس‌گذاری تأثیر می‌گذارد، حذف موجه خواهد بود، زیرا نمی‌توان پذیرفت که تنها تعداد اندکی مشاهدات، نتایج مقیاس بندی را تا حد زیادی تحت تأثیر قرار دهند. می‌توان از تعداد خطاهای گاتمن با شاخص +G و ترکیب آن با روش‌های توکی[1] برای تشخیص داده‌های پرت استفاده کرد (زیجلسترا[2]، وندرآرک و سیجستسما، ۲۰۱۱).

مرحلهٔ دوم: شناسایی مقیاس. باید یک یا چند مقیاس را شناسایی کرد که هر دو الگو همگنی یکنواخت و در صورت نیاز الگو رضایت‌بخش‌تر یکنواختی مضاعف را نیز بررسی کند (گام‌های ۴ تا ۷).

گام (۴) مقیاس‌پذیری. اگر هدف بررسی مقیاس از پیش[3] تعیین‌شده است، ضرایب H_j، H_{jk} و H^T را برای مقیاس کامل محاسبه می‌شود. اگر هدف بررسی ابعاد[4] مجموعه داده‌ها است، باید AISP[5] را c برابر ۰، ۰/۰۵، ۰/۱۰ و ۰/۵۵ اجرا کرد. می‌توان از همکر[6] و همکاران (۱۹۹۵) برای جستجوی الگوهای مرتبط استفاده کرد. بسته به داده‌ها، الگوی نتایج ممکن باید قبل از رسیدن به بالاترین مقدار c مشخص گردد. AISP به‌صورت تقریبی مقیاسی را مرتب می‌کند که افراد را با استفاده از نمرهٔ کل با خطای نسبتاً کم رتبه‌بندی می‌کند، اما نیاز است تعدادی از سؤال‌ها مهم برای تصمیم‌گیری‌های دقیق در نظر قرار گیرند: (مراجعه به گام‌های ۵ تا ۷).

هنگامی‌که تمام مقادیر ضریب مقیاس‌پذیری جفت سؤال‌ها مثبت باشد و همه ضرایب مقیاس‌پذیری تک‌تک سؤال‌ها بیشتر از ۰/۳۰ باشد، این تعداد سؤال، به‌عنوان مقیاس موکن در نظر گرفته می‌شوند (واتسون[7] و همکاران، ۲۰۱۲).

علاوه بر ضرایب فوق، ابعاد آزمون می‌تواند با استفاده از مقادیر حد پایین الگوریتم ژنتیک[8] مورد بررسی قرار گیرد (استرات[9]، وندرآرک و سیجتسما، ۲۰۱۳). حد پایین c حداقل مقدارتمیز(H_j)، برای سؤال‌های که مقیاس موکن را تشکیل می‌دهند نشان می‌دهد. تقسیم مجموعه‌ای از سؤال‌ها به مقیاس موکن با استفاده از تابع AISP یک روش اکتشافی برای به‌دست آوردن مجموعه‌ای از سؤال‌ها است که برخی از انتظارات قابل‌مشاهده اساسی را که با الگو همگنی یکنواخت و قدرت تمیز منطقی همخوانی دارد، برآورده

1. Tukey's fences
2. Zijlstra
3. Priori
4. Dimensionality
5. Automated item selection procedure
6. Hemker
7. Watson
8. Genetic algorithm
9. Straat

می‌کند (وندرآرک، ۲۰۱۲). به گفته سیجتسما و مولنار (۲۰۰۲)، چنانچه همه سؤال‌ها انتخاب‌شده در یک مقیاس (آزمون) دارای (c ۰/۳≤) باشند، مقیاس یک‌بعدی است. درصورتی‌که مقادیر c افزایش یابد، این سؤال‌ها برای قرار گرفتن در مجموعه سؤال‌ها مقیاس انتخاب نمی‌شوند.

گام (۵) استقلال موضعی. فرض استقلال موضعی بیان می‌کند که اگر پاسخ‌های شخص به سؤال‌ها مقیاس، تنها به سطح وی در ویژگی نهفته‌ای که اندازه‌گیری می‌شود، وابسته باشد، پاسخ به یک سؤال تحت تأثیر نمرهٔ سایر سؤال‌ها قرار نمی‌گیرد. باید تأکید کرد که ردیابی این ویژگی تا حد زیادی حدسی است زیرا استقلال موضعی، پدیده‌ای غیرقابل‌کشف و عملاً غیرقابل تحقق است (پالمگرن[1] و همکاران، ۲۰۱۸). لرد (۱۹۸۰) می‌گوید وقتی‌که مفروضه تک‌بعدی بودن برقرار است، استقلال موضعی نیز حاصل می‌شود و از این حیث دو مفهوم باهم برابرند (همبلتون، سوامیناتان و راجرز، ۱۹۹۰). بااین‌وجود استرات[2]، وندرآرک و سیجتسما (۲۰۱۳)، معتقدند که استقلال موضعی را می‌شود با استفاده از روش وابستگی شرطی[3] بررسی کرد. روش وابستگی شرطی شامل سه شاخص W_1 ، W_2 و W_3 است که جفت سؤال‌های دارای وابستگی (همبستگی) موضعی را علامت‌گذاری می‌کند (سیجتسما و وندرآرک، ۲۰۱۷).

گام (۶) یکنواختی. یکنواختی توابع "IRF" را با استفاده از رگرسیون غیرپارامتری نمرهٔ سؤال بر روی نمرهٔ کل در $J-1$ سؤال دیگر در همان مقیاس بررسی می‌شود (جانکر[4] و سیجتسما، ۲۰۰۰). تحلیل گرافیکی، این امکان را فراهم می‌آورد که نشان دهد تا چه حد می‌توانیم منحنی مشاهده شده را یکنواخت در نظر بگیریم و انحراف موضعی (مکانی) از یکنوایی را می‌توان با استفاده از آزمون آماری، مورد آزمون قرار داد (سیجتسما و مولنار، ۲۰۰۲).

گام (۷) رتبه‌بندی نامتغیر سؤال‌ها. محققانی که قصد ساخت مقیاسی را دارند که افراد را در یک بُعد رتبه‌بندی کند، می‌توانند از این مرحله صرف‌نظر کنند. با این حال اگر به رتبه‌بندی نامتغیر سؤال‌ها نیاز باشد، ضریب مقیاس‌پذیری H^T بیانگر درجه‌ای است که پاسخ‌دهندگان به‌صورت نامتغیر به سؤال‌ها پاسخ داده‌اند (لیگتوت[5]، وندرآرک، تی مارولد[6] و سیجتسما، ۲۰۱۰). فرض IIO از راه تعداد دفعات نقض مفروضه‌ها (عدم تقاطع توابع IRF) و روش انتخاب سؤال به روش پس‌رو[7] (سؤال‌هایی را که باید با نقض IIO حذف شوند حذف می‌کند)، بررسی می‌کنند. علاوه بر این شاخص (H^T) فاصلهٔ بین توابع "IRF" را تخمین می‌زند. هر چه فاصلهٔ بین توابع "IRF" « بیشتر باشد، IIO بیشتر خواهد بود؛ بنابر این آمارهٔ H^T به عنوان شاخصی از دقت IIO بر اساس معیارهای زیر گزارش می‌شود: ($H^T < ۰/۳$) به معنای نادرست بودن

1. Palmgren
2. Straat
3. Conditional association procedure
4. Junker
5. Ligtvoet
6. Te Marvelde
7. Backward

ترتیب سؤال است، ($0/3 \leq H^T < 0/4$) به معنی دقت کم، ($0/4 \leq H^T < 0/5$) نشانگر دقت متوسط و ($H^T \geq 0/5$) نشانگر دقت بالا است (سیجتسما، میر[1] و وندرآرک، ۲۰۱۰). سرانجام مقدار ملاک[2] پیشنهادی توسط سیجتسما و مولنار (۲۰۰۲) برای بررسی اندازهٔ اثر یک نقض[3] (تخلف) معنی‌دار با معیارهای زیر به کار می‌رود. اگر ($Crit < 40$) تخلف جزئی است. اگر ($40 \leq Crit < 80$)، نقض جدّی نیست، اما باید توسط محقق بررسی شود؛ اگر ($Crit \geq 80$) باشد، نقض جدی است (ون شوور، ۲۰۱۱).

مرحلهٔ سوم: خصوصیات (ویژگی‌های) مقیاس. باید ویژگی‌های مقیاس مشخص شده در مرحلهٔ دوم را معلوم کرد (مراحل ۸ تا ۱۰).

گام (۸) پایایی[4]. برای ارزیابی میزان پایایی مقیاس، چهار ضریب پایایی قابل برآورد است. اولین ضریب آلفا است. دومین ضریب، آماره لامبدا-۲[5] (سیجتسما، ۲۰۰۹) است که به‌عنوان جایگزین آلفای کرونباخ معرفی شده است. سومین ضریب، آمارهٔ سیجتسما- مولنار (MS) است که به‌عنوان یک برآورد کنندهٔ پایایی با اریبی جزئی (سیجتسما و مولنار، ۲۰۰۲)؛ برای الگوهای همگنی مضاعف (سیجتسما و مولنار، ۱۹۸۷) کاربرد دارد؛ و چهارمین ضریب، پایایی طبقهٔ پنهان[6] (LCRC) است که آماره‌ای نااُریب از پایایی نمرهٔ آزمون (وندرآرک، وندرپالم[7]، سیجتسما، ۲۰۱۱) محسوب می‌شود.

گام (۹) هنجارها. محققانی که برای اندازه‌گیری افراد به جداول هنجار نیاز دارند، هنجارها و فواصل اطمینان را برای هنجار با استفاده از روش رگرسیون تخمین می‌زنند (اوسترهوز[8]، وندرآرک و سیجتسما، ۲۰۱۶). در تحقیقات علمی، برای تفسیر عملکرد افراد در آزمون نیازی به هنجاریابی نیست. در این مورد، می‌توان از مرحله ۹ عبور کرد.

گام (۱۰) مقایسهٔ گروه. اگر نمونه شامل زیرگروه‌های معنی‌دار باشد، بررسی ترکیب مقیاس‌ها و ویژگی‌های مقیاس (گام‌های ۴ تا ۹) را می‌توان در بین زیرگروه‌ها نیز تعمیم داد. اگر ترکیب مقیاس و ویژگی‌های مقیاس در گروه‌ها متفاوت است، آگاهی از این تنوع هنگام تفسیر عملکرد آزمون افراد و در پژوهش‌هایی که ویژگی‌های گروه مورد مطالعه مانند میانگین نمرات مقیاس و همبستگی نمرات مقیاس با سایر متغیرها مورد توجه است مفید خواهد بود (سیجتسما و وندرآرک، ۲۰۱۷).

روش پژوهش

از آنجا که برای نشان دادن کاربرد مقیاس موکن "MSA"در ارزیابی عملکرد آموزشی نیاز به داده‌های

1. Meijer
2. Criterion
3. Violation
4. Reliability
5. Lambda-2
6. Latent class reliability coefficient
7. Van der Palm
8. Oosterhuis

دو ارزشی است، داده‌های ثانویه پژوهش حاضر، از مطالعهٔ سلطانی‌نژاد (۱۳۹۶) که هدف آن سنجش میزان آگاهی از روش صحیح اندازه‌گیری فشارخون دانشجویان دو سال آخر دندانپزشکی دانشگاه کرمان (سال تحصیلی ۹۵-۹۶) بر اساس دستورالعمل انجمن قلب آمریکا بوده است، انتخاب شد. جمعیت مورد مطالعه در این پژوهش، دانشجویان دورهٔ عمومی دو سال آخر دانشکدهٔ دندانپزشکی کرمان در سال تحصیلی ۹۵-۹۶ بوده‌اند. روش جمع‌آوری اطلاعات این مطالعه، سرشماری بوده است. بدین نحو که از همهٔ دانشجویانی که در سال تحصیلی ۹۶-۹۵ دانشجوی سال پنجم یا ششم دندانپزشکی بودند، (مجموعاً ۱۳۸ نفر) خواسته شد تا در موعد مقرر، به سؤال‌های این آزمون، پاسخ دهند. ۴۱/۴ درصد از دانشجویان مذکر و ۵۸/۶ درصد مؤنث بودند. ۵۲/۵ درصد دانشجوی سال پنجم و ۴۷/۵ درصد دانشجوی سال ششم دندانپزشکی بودند. ۲۶/۳ درصد از دانشجویان شهریه‌پرداز و ۷۳/۷ درصد غیرشهریه‌پرداز بودند. ابزار اندازه‌گیری در این پژوهش شامل آزمونی است که ضمن ثبت خصوصیات جمعیت‌شناختی (جنسیت، سال تحصیلی، نحوهٔ ورود به دانشگاه) شامل ۱۴ سؤال مربوط به روش صحیح اندازه‌گیری فشارخون بیمار (به عنوان نمونه روی‌هم انداختن پاها هنگام اندازه گیری فشارخون ممکن است منجر به کاهش فشارخون شود) است و بر اساس دستورالعمل انجمن قلب آمریکا طراحی گردیده است. برای هر سؤال سه پاسخ (صحیح- غلط و نمی‌دانم) در نظر گرفته‌شده تا دانشجویان پاسخ صحیح را انتخاب کنند. هر جواب صحیح یک نمرهٔ مثبت داشت و برای پاسخ غلط و نمی‌دانم امتیازی در نظر گرفته نشده است. بدین ترتیب مجموع نمرهٔ آزمون ۱۴ است. بر اساس مطالعهٔ سلطانی‌نژاد (۱۳۹۶) سنجش روایی محتوایی آزمون با نظرخواهی از ده نفر از صاحب‌نظران رشته‌های بیماری‌های دهان و جراحی انجام شده و روایی آن ۰/۹۹ گزارش شده است. برای ارزیابی پایایی آزمون از روش بازآزمایی استفاده شد، بدین نحو که سؤال‌های آزمون در اختیار بیست نفر از دانشجویان قرارگرفت و پس از دو هفته مجدداً همان افراد به سؤال‌ها پاسخ دادند. ضریب پایایی ۰/۸ محاسبه شد.

یافته‌ها

برای تحلیل داده‌های آزمون از بسته موکن[1] در نرم‌افزار R (وندرآرک، ۲۰۱۸) استفاده شد. گام اول، دوم و سوم نتایج بررسی داده‌ها نشان می‌دهد که نمره‌گذاری همهٔ سؤال‌ها مثبت و در یک جهت است و نیاز به کدگذاری معکوس ندارد. در ضمن دادهٔ بی‌پاسخ و نیز دادهٔ پرت در آنها وجود ندارد و همان‌طور که در جدول (۴) نشان داده شده است، مقادیر کجی و کشیدگی در محدودهٔ توزیع نرمال است.

1. .Mokken Package

جدول ٤. آماره‌های توصیفی

تعداد سؤال‌ها	چولگی	خطای معیار چولگی	کشیدگی	خطای معیار کشیدگی
١٤	-٠/٠٦	٠/٢٤٣	٠/٠٥٥	٠/٤٨١

گام چهارم: برآورد ضرایب مقیاس‌پذیری

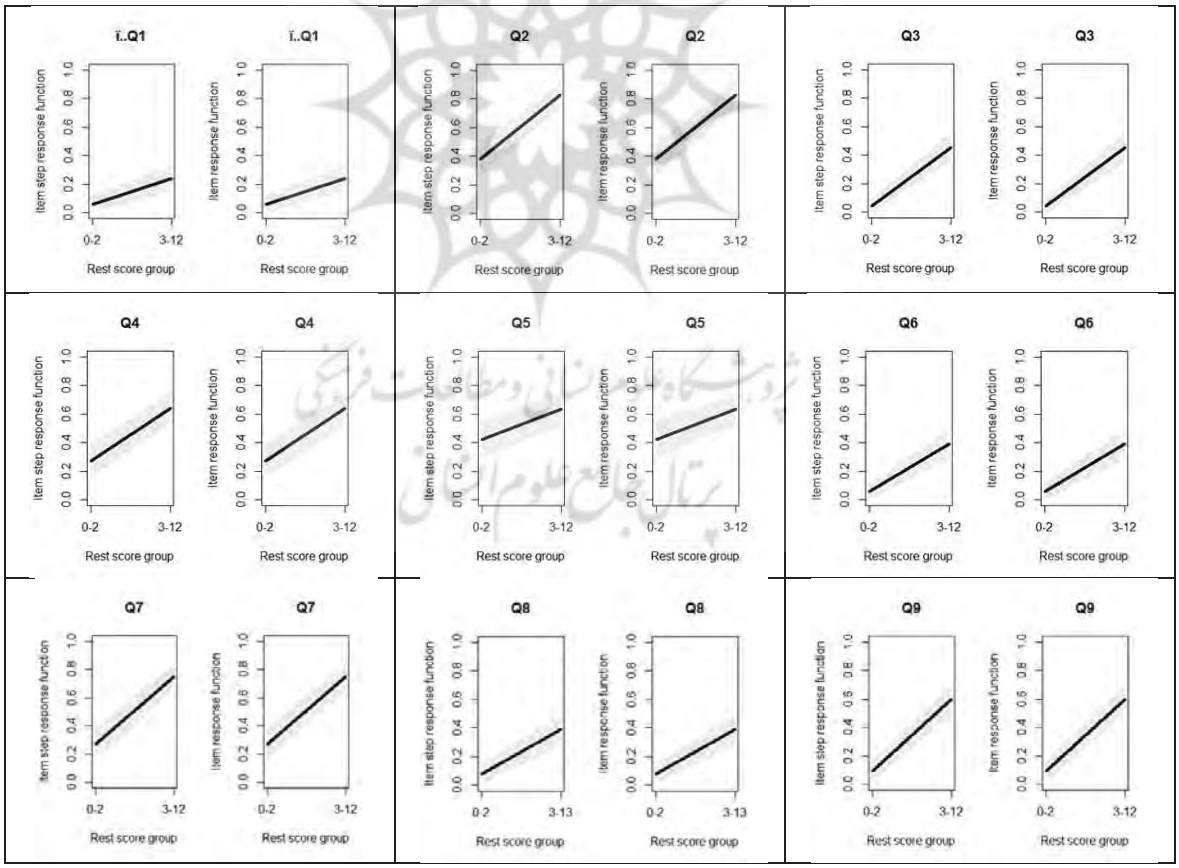
ضرایب مقیاس‌پذیری موکن شامل H_j، H_{ij} و H^T با توجه به معیارهای پیشنهادی موکن (۲۰۱۱) بررسی شد. طبق نتایج مندرج در جدول (۵)، ضرایب مقیاس‌پذیری تک‌تک سؤال‌ها و نیز ضرایب مقیاس‌پذیری جفت سؤال‌ها (به‌جز در یک مورد) مثبت با خطای معیار کم به‌دست‌آمد. ضریب مقیاس‌پذیری کل (۰/۴۹۰) نشان می‌دهد که مجموعهٔ ۱۴ سؤالی در تشکیل مقیاس موکن در حد متوسط عمل می‌کند. علاوه بر ضرایب فوق، ابعاد آزمون با استفاده از مقادیر حد پایین c=۰/۳ الگوریتم ژنتیک بررسی شد. تحلیل داده‌ها (جدول ۵) نشان می‌دهد که به‌جز سؤال پنج، سایر سؤال‌های آزمون دارای c<۰/۳ هستند و بنابر این می‌توانند در مجموعه سؤال‌های مقیاس موکن برای سنجش یک صفت تک بُعدی قرار گیرند. لذا سؤال پنج به‌عنوان کاندیدای حذف در نظر گرفته می‌شود.

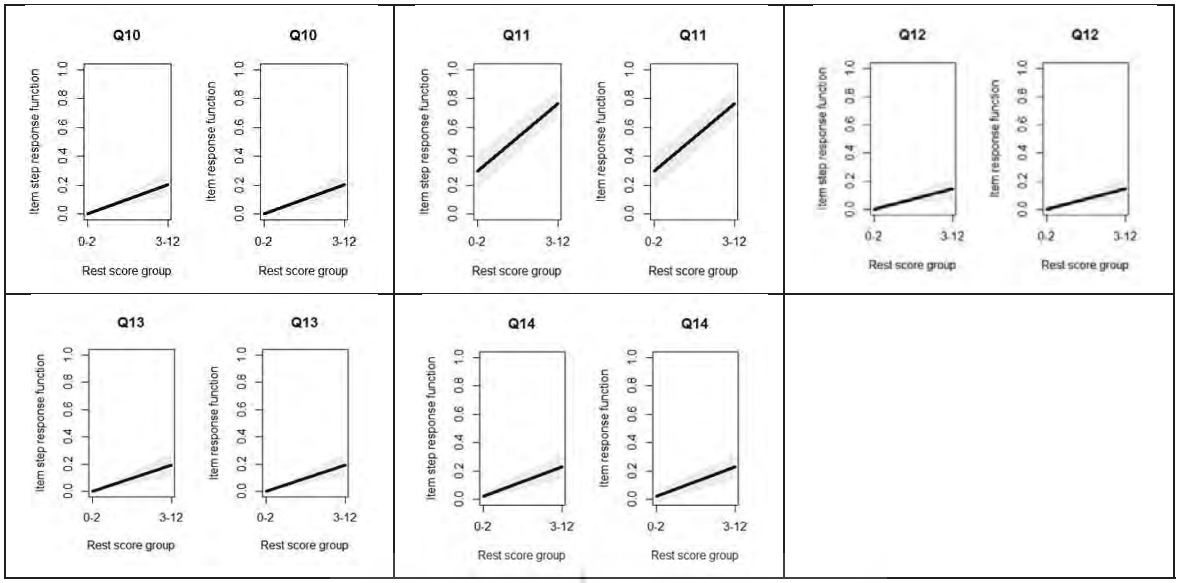
جدول ۵. ضرایب مقیاس‌پذیری تک سؤال‌ها

شماره سؤال	ضریب مقیاس‌پذیری برآورد شده هر سؤال (H_j)	خطای معیار برآورد	ضریب مقیاس‌پذیری کل (H)	الگوریتم ژنتیک تعیین تعداد زیر مقیاس با c= ۰/۳
١	٠/٤٣٢	٠/٠٨٣	٠/٤٩٠	١
٢	٠/٥١٥	٠/٠٥٧		١
٣	٠/٦٢٠	٠/٠٥٤		١
٤	٠/٤٧٢	٠/٠٥٨		١
٥	٠/٢٣٤	٠/٠٧٥		.
٦	٠/٥٠٢	٠/٠٦٥		١
٧	٠/٤٨٨	٠/٠٥٤		١
٨	٠/٤٩٢	٠/٠٦٨		١
٩	٠/٦١٣	٠/٠٤٧		١
١٠	٠/٤٥٧	٠/٠٦٣		١
١١	٠/٥٠٢	٠/٠٥٢		١
١٢	٠/٥٨٧	٠/٠٥٧		١
١٣	٠/٤٣٢	٠/٠٨١		١
١٤	٠/٥٣٢	٠/٠٧٤		١

گام پنجم: استقلال موضعی. از آنجا که مفروضهٔ تک بُعدی بودن برقرار است، استقلال موضعی نیز حاصل شده است. چون استقلال موضعی به طور خودکار از تک بُعدی بودن پیروی می‌کند و مفروضهٔ جداگانه‌ای به حساب نمی‌آید. البته برخی از متخصصان نظیر همبلتون و سوامیناتان معتقدند استقلال موضعی به‌معنای تک بُعدی بودن نیست (لرد، ۱۹۸۱؛ ترجمه دلاور و یونسی، ۱۳۹۱). با این وجود با توجه به اینکه ردیابی استقلال موضعی تا حد زیادی حدسی و عملاً غیرقابل تحقق است (پالمگرن و همکاران، ۲۰۱۸) و وابستگی موضعی در موقعیت‌های سنجش پیشرفت تحصیلی و استعداد ایجاد مشکل نمی‌کند (امبریتسون و رایس، ۲۰۱۳) در اکثر تحلیل‌ها این مفروضه به‌طور فنی وارسی نمی‌شود.

گام ششم: یکنواختی. بررسی گرافیکی فرض یکنواختی توابع پاسخ سؤال‌ها (شکل ۶) تأیید می‌کند که در همهٔ سؤال‌های آزمون، فرض یکنواختی (غیر کاهشی بودن) برقرار است. لذا نتایج تحلیل تا به اینجا نشان می‌دهد که الگوی همگن یکنواخت موکن با داده‌ها برازش دارد. در شکل (۳) نمودارهای غیر کاهشی توابع پاسخ سؤال‌های آزمون آورده شده است.





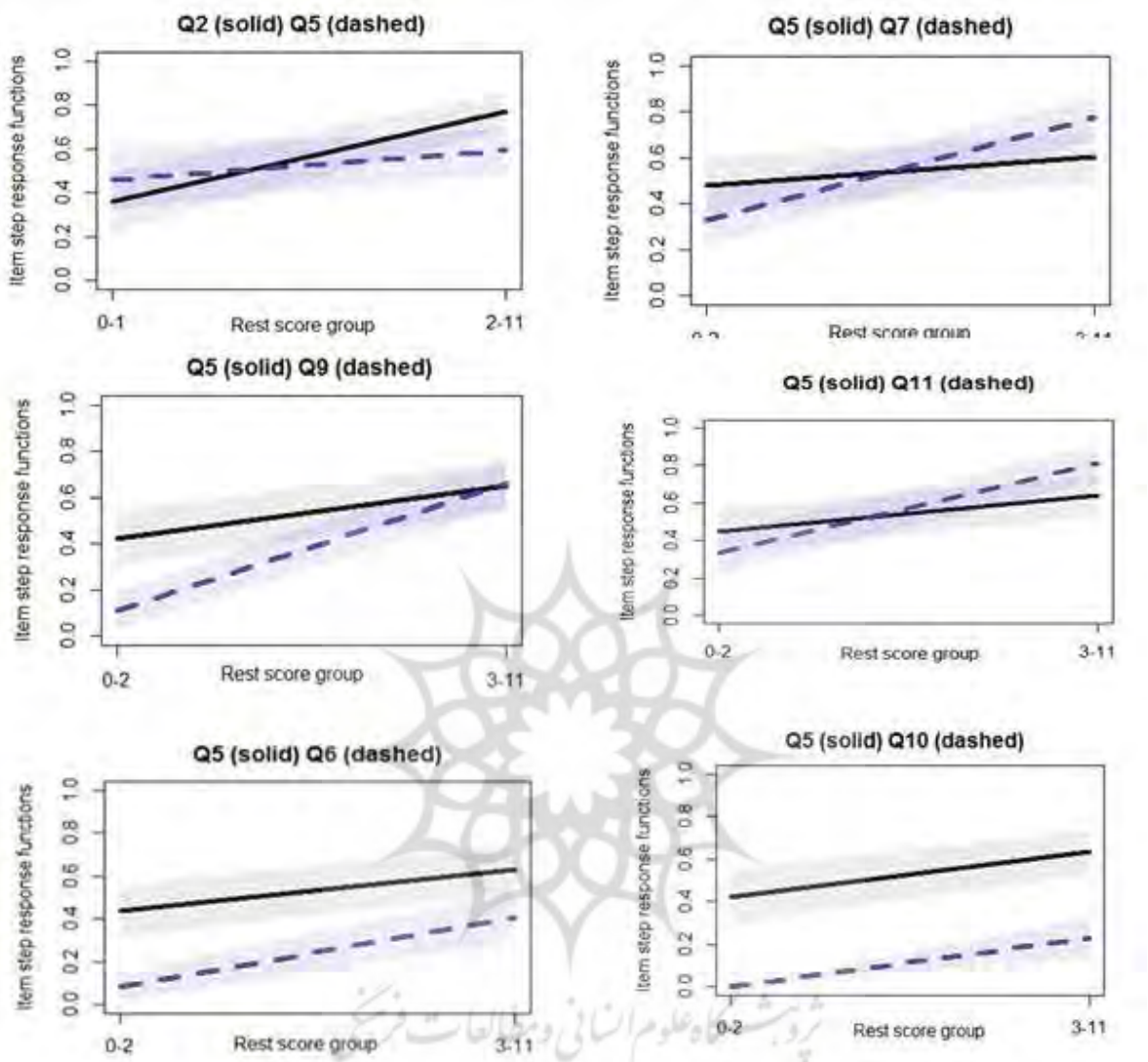
شکل ٦. نمودارهای غیرکاهشی توابع پاسخ سؤال‌های آزمون

گام هفتم: رتبه‌بندی نامتغیر سؤال‌ها: فرض IIO را با در نظر گرفتن تعداد دفعات نقض مفروضه‌های (عدم تقاطع توابع IRF) و روش انتخاب سؤال به روش پس‌رو بررسی شد. برای بررسی اندازهٔ اثر یک نقض (تخلف) معنی‌داری آن با شاخص مقدار ملاک بررسی گردید. نتایج جدول (۶)، نشان می‌دهد که به‌عنوان مثال تابع پاسخ سؤال دو، با ۱۳ سؤال دیگر مقایسه شده است که از این تعداد، تنها در یک مورد تابع «IRF" سؤال ۲ با سؤال دیگری تلاقی دارد که آن مقدار تلاقی معنی‌دار نیست. برای سؤال هفت یک مورد نقض معنی‌دار وجود دارد، اما چون شاخص ملاک آن زیر ۸۰ به‌دست آمده است، لذا سؤال هفت هم در مرحلهٔ انتخاب پس‌رو، کاندیدای حذف معرفی نشده است. اما برای سؤال پنج، چهار مورد نقض مفروضه وجود دارد (شکل ۷) که از این تعداد، یک مورد معنی‌دار است. مقدار ملاک هم تنها برای سؤال پنج بیشتر از ۸۰ به دست آمده است که نشان می‌دهد نقض جدّی مفروضه در سؤال پنج روی داده است و لذا نتایج انتخاب سؤال به روش پس‌رو نشان می‌دهد که برای داشتن مجموعه سؤال‌هایی با الگوی همگنی مضاعف، سؤال پنج باید حذف شود. ضریب H^T با مقدار (۰/۴۵) نشان‌دهندهٔ ترتیبی قابل قبول است. هر چند که نتایج تکرار تحلیل با حذف سؤال پنج، نشان داد که ضریب H^T برای اطلاعات باقیمانده همچنان ۰/۴۵ است اما حذف سؤال پنج، شاخص مقیاس‌پذیری کل را به ۰/۵۴ (در حد قوی) افزایش داد. همچنین ضرایب پایایی برای مجموعهٔ ۱۳سؤالی از ۰/۸۶ تا ۰/۹۱ متغیر بود (جدول شماره ۷). در کل نتایج نشان می‌دهد که شواهد برازش الگوی یکنواختی مضاعف نیز وجود دارد. لذا از مجموعه سؤال‌ها (با حذف سؤال پنج) می‌توان هم برای رتبه‌بندی افراد و هم برای رتبه‌بندی سؤال‌ها استفاده کرد.

جدول ٦. شاخص‌های بررسی مفروضه تغییرناپذیری ترتیب سؤال‌ها

شمارهٔ سؤال	تعداد مقایسه‌های جفتی برای هر سؤال	تعداد موارد نقض مفروضه	تعداد نقیضه‌های معنی‌دار	crit	انتخاب پس رو		H^T
					مرحلهٔ ۱	مرحلهٔ ۲	
۲	۱۳	۱	۰	۲۳	۰	۰	۰/٤٤۹۹
۱۱	۱۳	۱	۰	۲۹	۰	۰	
۷	۱۳	۱	۱	٥٦	۰	۰	
٥	۱۳	٤	۱	۱۱٥	۱	NA	
٤	۱۳	۱	۰	۱۲	۰	۰	
۹	۱۳	۰	۰	۰	۰	۰	
۳	۱۳	۱	۰	٥	۰	۰	
۸	۱۳	۱	۰	۱۲	۰	۰	
٦	۱۳	۰	۰	۰	۰	۰	
۱	۱۳	۰	۰	۰	۰	۰	
۱٤	۱۳	۰	۰	۰	۰	۰	
۱۰	۱۳	۰	۰	۰	۰	۰	
۱۳	۱۳	۰	۰	۰	۰	۰	
۱۲	۱۳	۰	۰	۰	۰	۰	

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



شکل ۷. نمونه‌ای از نمودارهای جفتی تابع پاسخ سؤال ۵ با سؤال‌های دیگر

گام هشتم: پایایی

بـرای ارزیابی میزان پایایی مقیاس، چهار ضریب پایایی برآورد شـد. جـدول (۷)، مقادیر ضرایب پایایی را، هم برای کل سـؤال‌ها (۱۴ سؤال) و هم بعد از حذف سـؤال پنج، برای مجموعهٔ ۱۳ سؤالی نشان می‌دهد. مقایسه ضرایب پایایی نشان می‌دهد که همهٔ شاخص‌های پایایی پس از حذف سؤال پنج، افزایش یافته‌اند.

جدول ۷. مقادیر ضرایب پایایی

پایایی مجموعه ۱۴ سؤالی				پایایی مجموعه ۱۳ سؤالی (با حذف سؤال ۵)			
MS	λ_2	alpha	LCRC	MS	λ_2	alpha	LCRC
۰/۸۵۶	۰/۸۶۱	۰/۸۵۳	۰/۹۰۶	۰/۸۰۶	۰/۸۶۷	۰/۸۶۱	۰/۹۱۲

شکل (۶) نشان می‌دهد که به‌عنوان مثال تابع پاسخ سؤال پنج با سؤال‌های دو، هفت، نُه و یازده تلاقی دارد که از این تعداد، تنها در یک مورد مقدار تلاقی معنی‌دار است و برای مثال تابع "IRF" سؤال پنج با سؤال‌های شش و ده تلاقی ندارد. از آنجا که در پژوهش حاضر برای تفسیر عملکرد افراد در آزمون نیاز به هنجاریابی نداریم و مقایسهٔ گروه‌ها جزو اهداف نبوده است، گزارش گام‌های نهم و دهم ضروری نیست.

بحث و نتیجه‌گیری

امروزه آزمون‌ها، مقیاس‌های درجه‌بندی، پرسشنامه‌ها و تمامی ابزارهای اندازه‌گیری جزئی جداناپذیر از زندگی بشر شده‌اند. این بحث تا حدی حائز اهمیت است که افراد از بدو تولد با استفاده از ابزارهای مختلف سنجش می‌شوند و نتایج حاصل از این سنجش‌ها در تصمیم‌گیری‌های متعددی مورد استفاده قرار می‌گیرد. با احتساب این نقش برای آزمون‌ها در زندگی فردی و اجتماعی افراد جامعه، درک درستی از نمره‌های حاصل از این ابزار در طول تاریخ همیشه مورد توجه بوده است و نظریه‌های مختلفی ازجمله نظریهٔ کلاسیک و نظریه سؤال-پاسخ هر یک به نوبه خود درصدد بهبود کیفیت اندازه‌گیری‌ها بوده‌اند (تیلور، ۲۰۱۳ ترجمه یونسی، ۱۳۹۸). هر چند که با روی کار آمدن برنامه‌های متعدد و رایانه‌های پیشرفته، گرایش متخصصان اندازه‌گیری، در تحلیل سؤال و ساخت ابزار به سمت الگوهای سؤال -پاسخ بیشتر شده است، اما پیچیدگی‌های محاسباتی زیربنایی آن هنوز به قوت خود باقی است. نیرومندی و اهمیت نظریهٔ سؤال-پاسخ و اینکه روش‌های اندازه‌گیری "IRT" از جمله ابزارهای قدرتمند در ساخت و تجزیه تحلیل آزمون‌ها و مقیاس‌ها به شمار می‌آید، بر هیچ‌کس پوشیده نیست. شاید با توجه به پیشرفت در حوزه‌های مختلف و روی کار آمدن رایانه‌های پیچیده بتوان اذعان کرد که عصر نظریهٔ کلاسیک به پایان رسیده است. امروزه بیشتر تحقیقات موجود با استفاده از روش‌های روان‌سنجی جدید مبتنی بر نظریهٔ سؤال-پاسخ پارامتری است. "IRT" پارامتری شامل تعدادی از روش‌های اندازه‌گیری قدرتمند است، اما برای استفاده از این روش‌ها چندین معیار باید برآورده شود و لذا استفاده از آنها بسیار پیچیده است؛ اما روش‌های نظریه سؤال -پاسخ غیرپارامتری "NIRT" محدودیت کمتری نسبت به الگوهای "IRT" پارامتری دارد و بعضی از روش‌های "NIRT" ممکن است ساده‌تر و بصری‌تر از روش‌های پارامتری "IRT" باشند؛ مانند نظریه کلاسیک آزمون،

معیار اصلی در "NIRT" نمره جمع اســت، اما روش‌های "NIRT" در روش انتخاب ســؤال‌ها، نحوهٔ ارزیابی کیفیت مقیاس و سایر اندازه‌ها از روش‌های کلاسیک متفاوت است (سیجتسما و مولنار، ۲۰۰۲). الگوهای اندازه‌گیری مانند "CTT" یا تحلیل عاملی بر این فرض اســت که همهٔ ســؤال‌ها وزن یکسانی دارند (توزیع فراوانی یکســانی دارند). هر زمان که این فرض نقض شــود مانند سؤال‌هایی که مقیاس تجمعی را تشکیل می‌دهند، نتیجه غیرواقعی که می‌تواند بروز کند، این است که سؤال‌ها به‌اندازه کافی همگن نیستند تا یک متغیر پنهان واحد را اندازه بگیرند. در این شرایط، یک مزیت عمدهٔ الگوهای "IRT"، ازجمله تحلیل مقیاس موکن نســبت به الگوهای "CTT" این اســت که الگوهای "IRT" در معرفی پارامترهای الگو برای سؤال‌ها، صریحاً در نظر می‌گیرند که سؤال‌ها وزن‌ها یا (اهمیت[1]) متفاوت دارند. دومین مزیت تجزیه‌وتحلیل مقیاس موکن نســبت به الگوهای "CTT" در تأکید دقیق بر تناســب الگو نهفته است. تمام ضرایب H_{ij} (و بنابر این تمام همبســتگی‌های زوجی) باید مثبت باشــند و هر ســؤال باید به‌اندازه کافی با بقیه همگن باشــد. این الزامات منجر به ابزار اندازه‌گیری می‌شــود که با اســتانداردهای بالاتر قابلیت اطمینان (پایایی) و همگنی مطابقت دارند تا ابزارهایی که فقط در یک تحلیل پایایی استاندارد بررسی شده‌اند. سومین و کاربردی‌ترین مزیت روش مقیاس‌گذاری موکن نســبت به سایر روش‌های اندازه‌گیری، روش جستجوی خوشه‌ای سلسله مراتبی[2] «از پایین به بالا» است که برای یافتن زیرمجموعه‌ای با حداکثری از سؤال‌های همگن بکار می‌رود. این روش به‌ویژه در تحقیقات اکتشــافی با هدف ایجاد ابــزار اندازه‌گیری جدید به محقق کمک می‌کند که حتی در صورت وجود تعداد محدود ســؤال‌ها، کاندیداهای جدید متغیرهای پنهان را کشف کند. سرانجام تجزیه‌وتحلیل مقیاس موکن یک الگوی «IRT" اســت که می‌تواند با موفقیت برای تعداد اندکی از ســؤال‌ها اســتفاده شــود. مولنار (۱۹۹۷) نشــان داد که وقتی تعداد سؤال‌ها نسبتاً کم باشــد، نتایج تحلیل مقیاس موکن و تحلیل مقیاس راش، اغلب به نتایج یکسانی منجر می‌شود (ون شوور، ۲۰۰۳).

References

Abdelhafez, A. M. (2007). Postgraduate Research Students' Knowledge and Attitudes towards Good Supervisory Practice at the University of Exeter. *Online Submission.*

Abdelhamid, G. S., Gómez-Benito, J., Abdeltawwab, A. T., Abu Bakr, M. H., & Kazem, A. M. (2020). A Demonstration of Mokken Scale Analysis Methods Applied to Cognitive Test Validation Using the Egyptian WAIS-IV. *Journal of Psychoeducational Assessment, 38*(4), 493-506.

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR.

1. Popularity
2. Hierarchical clustering search

20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9*(1), 95-104.

Baghaei, P. (2021). *Mokken Scale Analysis in Language Assessment.* Germany: Münster.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). (2001). Fitting item response theory models to two personality inventories: Issues and insights. Multivariate Behavioral Research, 36(4), 523-562.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

Emons, W. H., Sijtsma, K., & Pedersen, S. S. (2012). Dimensionality of the Hospital Anxiety and Depression Scale (HADS) in cardiac patients: comparison of Mokken scale analysis and factor analysis. *Assessment, 19(3), 337-353., 19*(3), 337-353.

Engelhard Jr, G. (2008). Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement, 6*(3), 155-189.

Guttman, L. (1947). On Festinger's evaluation of scale analysis. *Psychological Bulletin, 44*(5), 451.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Hemker, B. T., K. Sijtsma, I. W. Molenaar, and B. W. Junker. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, 331-347.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken I RT model. *Applied Psychological Measurement, 19*(4), 337-352.

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika, 62*(3), 331-347.

Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement, 24*(1), 65-81.

Ligtvoet, R. V. (2011). Polytomous latent scales for the investigation of the ordering of items. *Psychometrika*(2), 200-216.

Ligtvoet, R., Van der Ark, L. A., Te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement, 70*(4), 578-595.

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of" scale analysis" and factor analysis. *Psychological bulletin, 45*(6), 507.

Lord, F. M. (1980). Applications of item response theory to practical testing problems.(A. Delavar، J. Younesi, Trans.). Roshd Press.

Lumsden, J. (1957). A factorial approach to unidimensionality. *Australian Journal of Psychology, 9*(2), 105-111.

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychological methods, 9*(3), 354.

Meijer, R. R., Tendeiro, J. N., & Wanders, R. B. (2014). *The use of nonparametric item response theory to explore data quality.* Routledge.

Mokken, R. J. (1971). A theory and procedure of scale analysis. The Hague, The Netherlands: Mouton. *MokkenA Theory and Procedure of Scale Analysis1971, 62*(3), 331-347.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research* (Vol. (Vol. 1)). Walter de Gruyter.

Molenaar, I. W. (1997). *Nonparametric models for polytomous responses. In Handbook of modern item response theory.* New York, NY: Springer.

Mooij, T. (2012). A Mokken scale to assess secondary pupils' experience of violence in terms of severity. *Journal of Psychoeducational Assessment,, 30*(5), 496-508.

Oosterhuis, H. E., van der Ark, L. A., & Sijtsma, K. (2017). Standard errors and confidence intervals of norm statistics for educational and psychological Tests. *psychometrika, 82*(3), 559-588.

Palmgren, P. J., Brodin, U., Nilsson, G. H., Watson, R., & Stenfors, T. (2018). Investigating psychometric properties and dimensional structure of an educational environment measure (DREEM) using Mokken scale analysis–a pragmatic approach. *BMC medical education, 18*(1), 1-16.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology, 5*, 27-48.

Sijtsma, K. (2009). Correcting fallacies in validity, reliability, and classification. *International Journal of Testing, 9*(3), 167-194.

Sijtsma, K., & Meijer, R. R. (2007). 22 Nonparametric Item Response Theory and Special Topics. *Handbook of statistics, 26*, 719-746.

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item re-

sponse theory. *Psychometrika, 52*(1), 79-97.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (Vol. Vol. 5). sage.

Sijtsma, K., & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology, 70*(1), 137-158.م

Sijtsma, K., and I. W. Molenaar. (2002). Introduction to Nonparametric Item Response. *Thousand Oaks, CA: Sage*.

Sijtsma, K., Meijer, R. R., & van der Ark, L. A. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences, 50*(1), 31-37.

Soltaninejad, Fatemeh. (1396). Assessing the knowledge of Kerman dental students about the correct way to measure the patient's blood pressure according to the standard of the American Heart Association in the 95-96 academic year. (Unpublished doctoral dissertation). Kerman University of Medical Sciences and Health Services, School of Dentistry.

Straat, J. H., Van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification, 30*(1), 75-99.

Taylor, C. S. (2013). Validity and validation.(J. Younesi, Trans.). Allameh Tabatabai University Press.

Ünlü, A. (2007). Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory. *Journal of Mathematical Psychology, 51*(6), 383-400.

Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software, 48*(5), 1-27.

Van der Ark, L. A. (2005). Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika, 70*(2), 283-304.

Van der Ark, L. A. (2010). A note on stochastic ordering of the latent trait using the sum of polytomous item scores. *Psychometrika, 75*(2), 272-279.

Van Der Ark, L. A., & Sijtsma, K. (2005). The effect of missing data imputation on Mokken scale analysis. *New developments in categorical data analysis for the social and behavioral sciences*, 147-166.

Van der Ark, L. A., van der Ark, M. L. A., & Suggests, M. A. S. S. (2018). Package 'mokken'.

Van der Ark, L. A., van der Palm, D. W., & Sijtsma, K. (2011). A latent class approach to estimating test-score reliability. *Applied Psychological Measurement, 35*(5), , 380-392.

Van Ginkel, J. R., Van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivariate Behavioral Research, 42*(2), 387-414.

Van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis, 11*(2), 139-163.

Van Schuur, W. H. (2011). *Ordinal item response theory: Mokken scale analysis.* Thousand Oaks: SAGE.

Watson, R. v. (2012). Item response theory: how Mokken scaling can be used in clinical practice. *Journal of clinical nursing, 21*(19pt20), 2736-2746.

Wind, Stefanie A. (2017). An Instructional Module on Mokken Scale Analysis. *Educational Measurement*, 1–17.

Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2011). Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics, 36*(2), 186-212.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی