فصلنامه پژوهش‌های ارتباطی/ سال بیست‌وششم/ شماره ۲ (پیاپی ۹۸) / تابستان ۱۳۹۸/ ۱۶۵ـ ۱۳۹
Quarterly Journal of Communication Research, 2019, Vol. 26, No. 2 (98), 139-165

تحلیل داده‌های نگرش‌سنجی: تفاوت نظریه کلاسیک و سؤال ـ پاسخ

دکتر عنایت‌اله زمانپور✉، دکتر جلیل یونسی*، مصطفی رستگارآگاه، مهسا مهرابی*****

چکیده

سنجش نگرش با تحلیل داده‌ها به پایان نمی‌رسد بلکه آغاز راهی است که بتوان کیفیت گویه‌های نگرش‌سنجی را افزایش داد و فرصتی فراهم کرد تا خطای اندازه‌گیری را کاهش داد. در این پژوهش تفاوت بین دو نظریه کلاسیک اندازه‌گیری (CTT) و نظریه سؤال- پاسخ (IRT) در تعیین ویژگی‌های روان‌سنجی داده‌های حاصل از افکارسنجی مطالعه شده است؛ قبل از ارائه نتایج تحلیل های مربوط به این دو نظریه به منظور بررسی روایی و نیز ساخت و تأیید ابزار از تحلیل عامل اکتشافی (EFA) و تحلیل عامل تأییدی (CFA) نیز بهره برده شده است. موضوع مورد مطالعه، داده‌های واقعی مربوط به افکارسنجی ۳٦۸٤ نفر از اقشار مختلف مردم درخصوص برنامه‌های تلویزیونی بوده است. یافته‌ها حاکی از آن است که هر دو نظریه کلاسیک و سؤال- پاسخ می‌توانند اطلاعات مشابهی درباره گویه‌های مقیاس در ارتباط با خصیصه مکنون مورد اندازه‌گیری ارائه دهند، اما دو نظریه اندازه‌گیری تصاویر متفاوتی از دقت مقیاس‌ها به‌دست می‌دهند. به‌طور کلی، نظریه سؤال- پاسخ، هم اطلاعات غنی‌تری درباره دقت اندازه‌گیری و هم پیشنهادهای روشن‌تری برای بهبود مقیاس ارائه می‌دهد.

کلید واژه‌ها: نگرش‌سنجی، نظریه کلاسیک اندازه‌گیری، نظریه سؤال- پاسخ، ویژگی‌های روان‌سنجی

✉ نویسنده مسئول: دکترای سنجش و اندازه‌گیری، استادیار دانشگاه علامه طباطبائی، تهران، ایران
Email: enayat_zamanpour@yahoo.com

* دکترای سنجش و اندازه‌گیری، دانشیار دانشگاه علامه طباطبائی، تهران، ایران
** دانشجوی دکترای سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران
*** کارشناس ارشد تحقیقات آموزشی، دانشگاه تهران، ایران

تاریخ دریافت: ۹۷/۱/۳۰ پذیرش نهایی: ۹۸/۵/۲

DOI: 10.22082/CR.2019.113772.1917

مقدمه

با اینکه تعداد بسیار زیادی از مقیاس‌های نگرش با استفاده از نظریه کلاسیک اندازه‌گیری ساخته شده‌اند و این مدل‌ها را می‌توان به صورت وسیع در ساخت و گسترش آزمون‌ها و مسائل مربوط به تحلیل نمرات مقیاس‌های نگرش به‌کار گرفت، ولی چون نظریه کلاسیک آزمون[۱]، مبتنی بر مفروضه‌های ضعیفی[۲] است، یعنی برای اکثر مجموعه داده‌های حاصل از اجرای آزمون‌ها این مفروضه‌ها به آسانی برقرار می‌شوند و به دلیل ضعف‌هایی در نظریه کلاسیک اندازه‌گیری، تلاش متخصصان روان‌سنجی معطوف به ساخت و گسترش نظریه‌های مناسب‌تر اندازه‌گیری‌های روانی[۳] شده است (همبلتون[۴] و سوامیناتان[۵]، ۲۰۱۳؛ لرد[۶]، ترجمه دلاور و یونسی، ۱۳۹۱ و دی آیالا[۷]، ۲۰۰۹). در حال حاضر، متداول‌ترین مجموعه سازه‌ها، مدل‌ها، و مفروضه‌ها برای استنباط صفات یا ویژگی‌ها، با تکیه به نظریه صفت مکنون[۸] تعریف و سازماندهی شده است؛ بنابراین، در حال حاضر، توجه عمده به حوزه نظریه صفت مکنون، نظریه منحنی ویژگی سؤال[۹] یا همان‌گونه که لرد (ترجمه دلاور و یونسی، ۱۳۹۱) این نظریه را نظریه سؤال ـ پاسخ[۱۰] (IRT) می‌نامد، معطوف شده است. تاکنون به‌کارگیری نظریه جدید برای ساخت آزمون‌ها در ابعاد بزرگ و مورد استفاده عمومی مورد غفلت بوده اما در پژوهش حاضر بر آنیم که با مقایسه تأثیر دو نظریه کلاسیک آزمون و سؤال ـ پاسخ بر روی نگرش به پیچیدگی برنامه‌های تلویزیونی به رهیافت‌های کاربردی در ساخت آزمون‌های سنجش نگرش دست یابیم.

1 . Classical Test Theory (CTT)
2 . weak assumptions
3 . mental measurement
4. Hambleton
5. Swaminathan
6. Lord
7. De Ayala
8. latent trait theory
9. item characteristic curve theory
10. Item Response Theory (IRT)

دو نظریه کلاسیک آزمون و سؤال ـ پاسخ در اندازه‌گیری کاربرد زیادی داشته و از اهمیت زیادی در اندازه‌گیری سازه‌های مکنون برخوردارند (وندر لیندن[1] و همبلتون، ۲۰۱۳). بیکر[2] و کیم[3] (۲۰۰٤) بر این باور هستند که لازارسفلد[4] احتمالاً اولین فردی است که اصطلاح «صفت مکنون[5]» را معرفی کرده است ولی بیان دارند که با این حال، کارهای لرد و ناویک[6]، (۱۹٦۸) معمولاً به عنوان پیدایش نظریه سؤال ــ پاسخ تلقی می‌شود. اگرچه هیچ اندازه‌گیری دقیقی از یک متغیر مکنون وجود ندارد، به‌واسطه آزمون و چگونگی پاسخ‌دهی فرد به تعدادی گویه در ارتباط با یک تک‌بعدی زیربنایی، پژوهشگران می‌توانند نمراتی را به‌عنوان تخمینی از سطح فرد در آن صفت مکنون بسازند. نظریه کلاسیک و جدید اندازه‌گیری هر دو ابزاری برای این منظور هستند اما فراتر از این هدف مشترک، سازوکار این دو نظریه، بی‌شباهت با یکدیگر است. این دو نظریه به‌طور آشکار در فرایند مدل‌سازی نمرات، با یکدیگر تفاوت دارند علاوه بر این، در مفروضات بنیادی، ماهیت سازه مورد اندازه‌گیری و اینکه چطور افراد به گویه‌ها پاسخ می‌دهند متفاوت هستند. گسترش نظریه سؤال ـ پاسخ از تحلیل داده‌های دوارزشی به داده‌های پاسخ چندارزشی و پیوسته و از مدل‌های تک‌بعدی به مدل‌های چندبعدی توسط سامی‌جیما[7] (۱۹٦۹) منجر به پیشرفت‌های نظری مهمی در IRT شده است که از نظر عملی بسیار مفید است. اطلاعات عمیق‌تر درباره نظریه کلاسیک اندازه‌گیری را می‌توان در آثار لرد و ناویک (۱۹٦۸) و آلن و ین (ترجمه دلاور، ۱۳۹٦) و جزییات بیشتر درباره نظریه سؤال ـ پاسخ را در اثر امبرتسون[8] و ریس[9] (۲۰۰۰)

1. Van der Linden
2. Baker
3. Kim
4. Lazarsfeld
5. latent trait
6. Novick
7. Samejima
8. Embretson
9. Reise

مطالعه کرد. در ادامه، یک شمای کلی از هر نظریه به‌منظور مقایسه آن در ارتباط با اندازه‌گیری سازه‌ها ارائه شده است:

شاید مهم‌ترین فرضیه بنیادی نظریه کلاسیک این باشد که نمره مشاهده شده هر پاسخ‌دهنده در یک مقیاس یا آزمون بازنمایی از نمره واقعی (حقیقی) به علاوه خطای تصادفی آن فرد است. در مدل کلاسیک آزمون، دو سازه نمره واقعی و نمره خطا هر دو غیرقابل مشاهده هستند. نمره واقعی یک آزمودنی را می‌توان به عنوان نمره مورد انتظار آزمون وی در اجراهای مکرّر (فرم‌های موازی) آزمون تعریف کرد. نمره خطا را هم می‌توان به عنوان تفاوت نمره واقعی و نمره مشاهده شده تعریف کرد. همچنین در مدل کلاسیک آزمون اصل بر این است که ۱) نمرات خطا نمراتی تصادفی با میانگین صفر هستند و با نمرات خطا و نمرات واقعی در یک آزمون موازی همبستگی ندارند و ۲) نمرات واقعی، نمرات مشاهده شده و نمرات خطا رابطه‌ای خطی با یکدیگر دارند. برآوردهای نظریه کلاسیک وابسته به آزمون هستند و هر آزمون و مقیاس، ویژگی‌های روان‌سنجی متفاوتی دارد. علاوه بر این، برآوردهای نظریه کلاسیک از خطای اندازه‌گیری، وابسته به نمونه است چراکه تنها راه برآورد خطای استاندارد اندازه‌گیری[1] استفاده از اطلاعات یک گروه از پاسخ‌دهنده‌هاست. آماره‌های متداول و مشهور مدل کلاسیک آزمون، مانند خطای استاندارد اندازه‌گیری[2]، فرمول‌های جبران کاهش[3]، فرمول اسپیرمن- براون[4]، و فرمول کودر- ریچاردسون[5] ۲۰، تنها تعدادی از بسیاری از آماره‌های مهمی هستند که بخشی از مدل کلاسیک آزمون و تکنیک‌های وابسته به آن را تشکیل می‌دهند (گالیکسن[6]، ۱۹۵۰؛ لرد و ناویک، ۱۹۶۸ و دی آیالا، ۲۰۰۹).

در نظریه سؤال ـ پاسخ چنین فرض می‌شود که در موقعیت‌های سنجش، می‌توان از

1. SEM
2. standard error of measurement
3. disattenuation formula
4. Spearman-Brown formula
5. Kuder-Richardson 20
6. Gulliksen

طریق تعریف ویژگی‌های آزمودنی، که به صفات یا توانایی‌های وی اشاره دارد، عملکرد آزمودنی در یک آزمون را پیش‌بینی یا تبیین کرد؛ سپس نمرات آزمودنی‌ها در این صفات (که نمرات توانایی[1] نامیده می‌شوند) را برآورد کرد؛ و در نهایت از این نمرات برای پیش‌بینی یا تبیین عملکرد سؤال و آزمون[2] استفاده کرد (لرد و ناویک، ۱۹٦۸). چون صفات به طور مستقیم قابل اندازه‌گیری نیستند، به آنها صفات یا توانایی‌های مکنون می‌گویند. یک مدل سؤال ـ پاسخ رابطه‌ای بین عملکرد قابل مشاهده آزمودنی و صفات یا توانایی‌های غیر قابل مشاهده وی که فرض می‌شود زیربنای عملکرد فرد در آزمون است، برقرار می‌کند. در چارچوب گسترده نظریه سؤال- پاسخ می‌توان بسیاری از مدل‌ها را عملیاتی کرد، چون انتخاب‌های زیادی برای فرم ریاضی منحنی‌های ویژگی سؤال وجود دارد. اما، با اینکه نمی‌توان درست یا نادرست بودن نظریه سؤال ـ پاسخ را نشان داد، ولی می‌توان مناسب بودن مدل‌های خاص برای هر مجموعه از داده‌های آزمون را با استفاده از اجرای یک آزمون مناسب نیکویی برازش نشان داد.

چندین تفاوت بین IRT و CCT وجود دارد که پژوهشگران در اندازه‌گیری سنجش نگرش با آن مواجه‌اند (اوستینی[3] و نرینگ[4]، ۲۰۰٦). اول اینکه، مقادیر بسیار متداول آماره‌های سؤال همانند ضریب دشواری و ضریب تشخیص سؤال که در ساخت و توسعه آزمون به‌کار گرفته می‌شوند، به نمونه‌های خاص آزمودنی‌ها ــ که از طریق آن‌ها این آماره‌ها محاسبه می‌شوند ـ وابسته است. میانگین سطح توانایی و دامنه نمرات توانایی در یک نمونه از آزمودنی‌ها بر مقادیر آماره‌های سؤال، اغلب تأثیر زیادی می‌گذارد (لرد و ناویک، ۱۹٦۸). نتیجه نهایی اینکه آماره‌های سؤال تنها زمانی برای انتخاب و گزینش سؤال مفید هستند که آزمون برای جامعه‌هایی ساخته شود که به نمونه آزمودنی‌های مورد استفاده در محاسبه آماره‌های سؤال شباهت بسیاری داشته

1. ability scores
2. item and test performance
3. Ostini
4. Nering

باشد. نقطه ضعف دیگر نظریه کلاسیک آزمون، از منظر روش‌ها و تکنیک‌های متداول در طراحی و تحلیل آزمون، این است که مقایسه آزمودنی‌ها بر اساس یک توانایی ـ که با مجموعه‌ای از سؤالات آزمون اندازه‌گیری می‌شودـ محدود به موقعیت‌هایی است که در آن بر روی آزمودنی‌ها یک آزمون یکسان یا آزمون‌های موازی اجرا می‌شود. سومین نقطه‌ضعف نظریه کلاسیک آزمون این است که یکی از مفاهیم بنیادی این نظریه، یعنی پایایی آزمون[1]، بر حسب فرم‌های موازی[2] تعریف شده است. در عمل تحقق یافتن مفهوم اندازه‌های موازی[3] دشوار است: افراد هرگز در اجرای دوم آزمون، دقیقاً همانند دفعه اول عمل نمی‌کنند. آنها مسائل را فراموش می‌کنند، مهارت‌های جدیدی را یاد می‌گیرند، سطح اضطراب یا انگیزشی آنها ممکن است تغییر کند و غیره (وندر لیندن و همبلتون، ۲۰۱۳). چهارمین نقطه‌ضعف نظریه کلاسیک آزمون این است که در این نظریه مبنایی برای تعیین این که یک آزمودنی زمانی‌که با یک سؤال آزمون روبه‌رو می‌شود چگونه عمل خواهد کرد، وجود ندارد. به هنگام منطبق کردن یک آزمون با سطح توانایی آزمودنی، داشتن برآوردی از احتمال پاسخ صحیح آزمودنی به یک سؤال خاص از ارزش قابل ملاحظه‌ای برخوردار است. پنجمین نقطه ضعف نظریه کلاسیک آزمون این است که در این نظریه چنین فرض می‌شود که واریانس خطاهای اندازه‌گیری برای تمامی آزمودنی‌ها یکسان است. به هر حال مشاهده اینکه بعضی از افراد در انجام تکلیفی نسبت به دیگران از ثبات بیشتری برخوردارند و این میزان ثبات با سطح توانایی تغییر می‌کند، غیر معمول نیست. با توجه به این مطلب، انتظار می‌رود که عملکرد آزمودنی‌های با توانایی بالا نسبت به عملکرد آزمودنی‌های با توانایی متوسط در چندین فرم موازی یک آزمون دارای ثبات بیشتری باشد. علاوه بر پنج نقطه ضعف نظریه کلاسیک آزمون، این نظریه و روش‌های مرتبط با آن در فراهم کردن راه‌حل‌های رضایت‌بخش برای بسیاری از مسائل و مشکلات سنجش، مثل طراحی آزمون‌ها،

1. test reliability
2. parallel forms
3. parallel measures

تشخیص سؤالات دارای سوگیری[1] و همترازسازی نمرات آزمون[2] با شکست مواجه شده است (همبلتون و سوامیناتان، ۲۰۱۳ و دی آیالا، ۲۰۰۹).

هدف پژوهش

هدف از این پژوهش، بررسی گویه‌های سنجش نگرش مخاطبان نسبت به پیچیدگی برنامه‌های تلویزیونی است که با دو نظریه اندازه‌گیری یعنی CCT و IRT مطالعه شده است. از سوی دیگر، تاکنون هر مطالعه‌ای که بر روی ساخت مقیاس‌ها ازجمله نگرش‌سنج‌ها صورت گرفته، به‌طور صریح و یا ضمنی، براساس CCT ساخته شده است و کمتر مطالعه‌ای براساس نظریه IRT انجام شده است. در ضمن، هیچ مطالعه‌ای تاکنون به مقایسه CCT و IRT در حوزه نگرش به برنامه‌های تلویزیونی پرداخته و مزایا و نقاط ضعف و قوت این دو نظریه را در ساخت نگرش‌سنج‌ها مقایسه نکرده است. ازاین‌رو، هدف پژوهش حاضر، مقایسه دو نظریه CTT و IRT در سنجش نگرش مخاطبان نسبت به پیچیدگی برنامه‌های تلویزیونی و بررسی ویژگی‌های روان‌سنجی گویه‌های نگرش است. ویژگی‌های روان‌سنجی به‌معنای بررسی اعتبار (روایی) و اعتمادپذیری (پایایی) است. اعتبار به معنای آن است که آزمون، آنچه را مدنظر دارد، اندازه بگیرد. اعتمادپذیری نیز اشاره به همبستگی بین نمرات واقعی و مشاهده شده دارد (آلن و ین، ترجمه دلاور، ۱۳۹۶). در این پژوهش، به دنبال پاسخ به پرسش‌های زیر هستیم:

۱) ویژگی‌های روان‌سنجی گویه‌های نگرش به بعد پیچیدگی برنامه‌های تلویزیونی براساس CCT و IRT چگونه است؟

۲) آیا بین اطلاعات روان‌سنجی این دو نظریه تشابه و تناقضی وجود دارد؟

روش داده‌ها

داده‌های این مطالعه، حاصل افکارسنجی صداوسیما درخصوص پیچیدگی

1. biased Items
2. equating of test scores

برنامه‌های تلویزیونی است. این داده‌ها براساس بخشی از داده‌های مربوط به پژوهش افکارسنجی مرکز صداوسیما درباره برنامه‌های تلویزیونی به‌دست آمده است. گفته می‌شود که گویه‌های خاصی برای سنجش عامل پیچیدگی در نظر گرفته شده است ولی این ادعا از طریق تحلیل عامل اکتشافی بررسی نشده؛ البته با توجه به اینکه یکی از مفروضات اصلی استفاده از نظریه سؤال- پاسخ، تک‌بعدی بودن داده‌هاست، لذا با شیوه‌های مختلف آماری تک‌بعدی بودن داده‌ها بررسی شده است.

نمونه

داده‌های پژوهش حاضر، حاصل از پژوهش مرکز افکارسنجی صداوسیما در خصوص برنامه‌های تلویزیونی بوده، در مجموع ۳٦۸٤ مخاطب تلویزیون به گویه‌های پژوهش پاسخ داده‌اند و میانگین سنی نمونه مورد مطالعه برابر با ۳۷/۷۲ سال با انحراف استاندارد ۱۵/۳۱ سال گزارش شده است به نحوی که دامنه سنی از ۱۵ تا بالای ۹۰ سال را شامل می‌شود. بیش از نیمی (۵۱/۵ درصد / ۱۹۸۹ نفر) از شرکت‌کنندگان در این پژوهش زن و ٤۸/۵ درصد (۱۷۸٦ نفر) مرد بوده‌اند.

انتخاب گویه‌ها و تحلیل مقیاس

در بانک اولیه، ۱۹ گویه وظیفه سنجش سه عامل «انتخابی بودن» به معنای قرار گرفتن در معرض رسانه به‌صورت آگاهانه؛ «توجه» به مفهوم درگیر شدن با محتوا به لحاظ شناختی و عاطفی و «پیچیدگی» با شاخص فکر کردن در مورد برنامه و ایجاد ارتباط بین محتوای برنامه و تجربیات قبلی فرد را بر عهده داشتند. برای عامل پیچیدگی، پرسش‌های ۱٤ تا ۱۹ این مجموعه در نظر گرفته شده و همین گویه‌ها مورد واکاوی قرار گرفته است. دلیل انتخاب تنها یک عامل، امکان‌پذیر شدن بررسی مفروضات IRT، ازجمله مفروضه تک بعدی بودن بود ازاین‌رو؛ صرفاً یک بعد از پرسشنامه برای تحلیل در نظر گرفته شد. در جدول ۱، شاخص‌های توصیفی گویه‌های عامل پیچیدگی ارائه شده است.

جدول ۱. گویه‌های سنجش نگرش به همراه شاخص‌های توصیفی عامل پیچیدگی

شماره	گویه‌ها	میانگین	انحراف استاندارد
q.14	به آنچه در برنامه‌ها گفته می‌شود فکر می‌کنم.	۲,۲۲	۱,۰۶
q.15	هنگام تماشای برنامه‌ای در تلویزیون به اینکه این برنامه چه اثراتی روی من دارد فکر می‌کنم.	۲,۱۵	۱,۱۱
q.16	هنگام تماشای برنامه‌ای در تلویزیون به اینکه این برنامه چه اثراتی روی دیگران می‌گذارد فکر می‌کنم.	۲,۰۱	۱,۱۶
q.17	هنگام تماشای برنامه‌ای در تلویزیون درباره اینکه این برنامه با اطلاعات و دانسته‌های قبلی من چه ارتباطی دارد فکر می‌کنم.	۲,۰۶	۱,۱۰
q.18	معمولاً اطلاعاتی را که از برنامه‌های تلویزیون دریافت می‌کنم با سایر رسانه‌ها مانند سایت‌های اینترنتی و شبکه‌های اجتماعی و ماهواره‌ای مقایسه می‌کنم.	۱,۶۶	۱,۳۰
q.19	به درست بودن اطلاعاتی که در برنامه‌ها ارائه می‌شود شک می‌کنم.	۱,۷۱	۱,۲۱

تقسیم داده‌ها برای ساخت و تأیید مقیاس

در بخش نخست، داده‌های موجود به‌طور تصادفی، به دو نمونه «الف» و «ب» با تعداد یکسان تقسیم شده است. یکی، با هدف ساخت (بهبود) و پایش مقیاس و دیگری، با هدف ارزیابی و تأیید مقیاسی که در نهایت تولید شده است. نخستین گام برای بهبود و پایش استفاده از تحلیل عاملی اکتشافی (EFA) است که به پژوهشگران توصیه شده در طول ساخت اولیه مقیاس یا ابزار مورد استفاده قرار گیرد (کان[1]، ۲۰۰۶؛ ریس[2] و

1. Kahn
2. Reise

همکاران، ۲۰۰۰؛ ماریزوت[1] و همکاران، ۲۰۰۷ و ورتینگتون[2] و ویتاکر[3]، ۲۰۰۶). فرایند تحلیل عاملی اکتشافی در این پژوهش، با هدف شناسایی تنها یک صفت (عامل) مکنون زیربنایی، دنبال و پس از شناسایی مجموعه‌ای از گویه‌ها، گام ارزیابی و تأیید با استفاده از تحلیل عاملی تأییدی انجام شده است. انجام تحلیل عاملی تأییدی به‌منظور آزمون فرضیه‌های مربوط به ساختار زیربنایی است که حاصل از تحلیل عاملی اکتشافی است. به‌طور ویژه، آنچه در تحلیل عاملی تأییدی آزمون شده، این است که عامل زیربنایی به‌طور مناسب کوواریانس‌های بین هر مجموعه آزمون را توصیف می‌کند. در فرایند CFA تمرکز بیشتر بر روی شاخص‌های برازش مدل است که منطق ساختار عاملی تک بعدی را ارزیابی می‌کند. تعداد نمونه «الف» که برای تحلیل عامل اکتشافی استفاده شده است ۱۷۸۹ نفر (تقریباً ۵۰ درصد از کل داده‌ها) و نمونه «ب» که برای تحلیل عامل تأییدی استفاده شده، ۱۸۵٤ نفر بوده است. پس از انجام این دو تحلیل و انتخاب گویه‌های مناسب، نمونه‌ها بار دیگر با یکدیگر ترکیب شدند و تحلیل‌های بعدی روی یک نمونه بزرگ انجام گرفت.

تحلیل عامل اکتشافی با هدف انتخاب گویه‌ها (نمونه الف)

در تحلیل عامل اکتشافی مفروضاتی باید برقرار باشد که ازجمله آنها، شاخص کفایت نمونه‌گیری (KMO) کیسر[4]، است؛ سرنی[5] و کیسر (۱۹۷۷) معتقدند که وقتی مقدار KMO بزرگ‌تر از ۰/٦ باشد به راحتی می‌توان تحلیل عاملی را اجرا کرد؛ هر چه این مقدار بزرگ‌تر باشد، مناسب‌تر و کفایت نمونه‌گیری بیشتر است (هومن، ۱۳۸۵). در پژوهش حاضر، این مقدار برابر با ۰/۷۹۱ گزارش شده که در حد مطلوب بوده است.

1. Morizot
2. Worthington
3. Whittaker
4. Kaiser
5. Cerny

از دیگر شاخص‌ها آزمون کرویت بارتلت است که مقصود از اجرای آن، رد فرضیه صفر مبنی بر درست بودن ماتریس واحد، یعنی ماتریسی است که عناصر قطری آن یک و تمام عناصر غیرقطری در جامعه برابر صفر باشد. آزمون کرویت بارتلت این فرضیه را که ماتریس همبستگی مشاهده شده، متعلق به جامعه‌ای با متغیرهای ناهمبسته است، می‌آزماید. برای آنکه یک مدل عاملی مفید باشد، لازم است متغیرها همبسته باشند، در غیر این صورت، دلیلی برای تبیین مدل عاملی وجود ندارد. مقدار آماره خی‌دو (کای دو) برای این آزمون معنادار بوده است؛ از این رو، می‌توان ادعا کرد که شرایط این مفروضه نیز رعایت شده است (χ^2=۳۰۸۹/۸۹ ، df = ۱۵، p <۰۰۰۱).

علاوه بر دو مفروضه فوق میزان همبستگی ضد تصویر[1] نیز که شاخص کفایت گویه برای استخراج عامل است بررسی شد. مقادیر محاسبه شده، همگی بیشتر از ۰/۷ و حاکی از کفایت گویه‌هاست؛ همچنین مقادیر مربوط به ماتریس همبستگی بین گویه‌ها[2] نیز بررسی شده که نشان می‌دهد به جز مقادیر مربوط به دو گویه ۱۸ و ۱۹ که با یکدیگر هبستگی بالای ۰/۳۸ ولی با دیگر گویه‌ها، به جز یکی از آنها (همبستگی بین دو گویه ۱۸ و ۱۷ که برابر با ۰/۳۱ بوده) همبستگی کمتر از ۰/۳ داشته‌اند، همبستگی دیگر گویه‌ها همگی بالای ۰/٤ گزارش شده است. همچنین میزان همبستگی گویه با کل[3] نیز محاسبه گردید؛ مقادیر محاسبه شده جز در دو گویه ۱۸ و ۱۹ که به ترتیب، برابر با ۰/۳۱ و ۰/۳۰ بوده همگی بالای ۰/۵۰ گزارش شده است. شواهد بالا نشان از نبود تجانس دو گویه ۱۸ و ۱۹ با دیگر گویه‌ها دارد اما برای حذف نهایی دو گویه از تحلیل‌ها، به ماتریس باقی‌مانده پس از تحلیل عامل اکتشافی به‌عنوان مرجع نهایی رجوع شده است.

اگر همبستگی بین دو متغیر، به نسبت بالا باشد (بالاتر از ۰/۳) و در ماتریس همبستگی باقی‌مانده آنها نیز این مقدار قابل توجه باشد، برای مثال، ۰/۲، این موضوع (کوواریانس تبیین نشده براساس تحلیل عاملی اکتشافی تک عاملی) حاکی از وابستگی

1. Anti-image Correlation
2. inter-Item correlation matrix
3. corrected item-total correlation

مکانی[1] بین دو متغیر است. به نظر می‌رسد که این همبستگی بین دو متغیر، نه تنها به‌دلیل عامل مکنون زیربنایی «پیچیدگی» بلکه به‌دلیل بعد محتوایی دوم دیگری هم باشد. این در حالی است که در CCT موضوع وابستگی مکانی لزوماً مایه نگرانی نیست اما در IRT وجود وابستگی مکانی یک مسئله جدی است چراکه می‌تواند سبب تحریف در برآورد پارامترهای گویه‌ها شود. چنانچه مجموعه‌ای از گویه‌ها دارای استقلال مکانی باشند، پس از کنترل همبستگی‌های درونی بین متغیرها برای مثال با کنترل عامل زیربنایی مشترک، متغیرها مستقل از یکدیگر خواهند بود یا همبستگی بزرگی نخواهند داشت. ازاین‌رو دو گویه ۱۸ و ۱۹ که گفته شد همبستگی آنها با دیگر گویه‌ها نیز کمتر و در حدود ۰/۳ گزارش شده است، با داشتن میزان باقی‌مانده ۰/۲۹۶ به‌دلیل نقص استقلال مکانی کنار گذاشته شدند. در ادامه، مقادیر مربوط به همبستگی و کوواریانس گویه‌های باقی‌مانده ارائه شده است به نحوی که قطر پایینی، شامل میزان همبستگی، قطر بالایی مقادیر کوواریانس‌ها و قطر اصلی شامل واریانس‌هاست.

جدول ۲. مقادیر مربوط به همبستگی، واریانس و کوواریانس بین گویه‌ها

گویه	q.۱٤	q.۱٥	q.۱٦	q.۱۷
q.۱٤	۱,۱۳۳	۰,۵۶۴**	۰,۴۶۸**	۰,۴۵۷**
q.۱٥	۰,۶۶۶	۱,۲۳۱	۰,۶۴۶ **	۰,۵۴۳**
q.۱٦	۰,۵۷۶	۰,۸۲۸	۱,۳۳۷	۰,۵۵۵**
q.۱۷	۰,۵۳۶	۰,۶۶۴	۰,۷۰۷	۱,۲۱۴

** P < ۰/۰۱

از شیوه استخراج تحلیل مؤلفه‌های اصلی[2] (PCA) به‌منظور انجام تحلیل عامل اکتشافی بهره برده شده است. هدف تحلیل عامل اکتشافی این است که آیا واریانس مشترک بین گویه‌ها (پرسش‌ها) از طریق تعدادی متغیر (عامل) مکنون تقلیل‌یافته قابل

1. iocal dependence
2. Principal Components Analysis (PCA)

تبیین و توضیح است. در بافت ساخت مقیاس، پژوهشگرانی که از تحلیل عاملی استفاده می‌کنند، به دنبال این هستند که آیا روابط درونی بین متغیرها در یک مقیاس می‌تواند از طریق تنها یک عامل زیربنایی به بهترین نحو تبیین شود (ریس و همکاران، ۲۰۰۰ و راسل، ۲۰۰۲). اگر متغیرهای مقیاس به وضوح تحت تأثیر تنها یک عامل قرار گیرند، آنچه «اعتبار عاملی» در CTT نامیده می‌شود، به‌دست می‌آید (آلن و ین، ترجمه دلاور، ۱۳۹۶) و نشانه‌ای برای «تک‌بعدیت» در IRT می‌تواند باشد (امبرتسون و ریس، ۲۰۰۰). روایی (اعتبار) عاملی و تک‌بعدی بودن، هر دو مطلوب نظر است چراکه پژوهشگران نوعاً به دنبال سنجش یک بعد یا همان خصیصه مکنون هستند. اگر با استفاده از راه‌حل تک عاملی نتوان به طور کامل همبستگی درونی بین گویه های مقیاس را تبیین کرد یا اگر مجموعه‌ای از گویه‌ها وجود داشته باشند که همبستگی‌هایی به اشتراک بگذارند که نتوان با یک عامل زیربنایی تبیین کرد، نمی توان ادعا کرد که مقیاس مورد نظر تنها چیزی را اندازه گیری می کند که برای آن طراحی شده است (هتی[1]، ۱۹۸۵و ریس و همکاران، ۲۰۰۰).

در ادامه، مقادیر مربوط به بارعاملی در تحلیل عامل اکتشافی برای کل نمونه با در نظر گرفتن یک بعد زیربنایی (تک عامل) آمده است. بنابر آنچه در جدول ۳ ارائه شده، مجموعه گویه‌های حاضر هم دارای اعتبار عاملی بوده و هم مفروضه تک‌بعدی بودن برای آنها برقرار است. مقادیر ویژه[2] نخست به مراتب از دیگر مقدار ویژه بزرگ‌تر است به نحوی که نسبت اولی به دومی برابر با ٤/٤٧ گزارش شده است. علاوه بر این، مقدار ویژه نخست تنها مقدار ویژه‌ای است که بزرگ‌تر از یک گزارش شده است. نسبت تقریباً چشمگیر مقادیر ویژه و میزان کوچک دیگر مقادیر ویژه به جز مقدار نخست، همگی شواهدی هستند بر این مدعا که راه‌حل تک‌عاملی مناسب‌ترین راه‌حل است (لرد، ترجمه دلاور و یونسی، ۱۳۹۱).

1. Hattie
2. eigenvalues

جدول ۳. مقادیر مربوط اشتراکات استخراجی اولیه و بارهای عاملی هر گویه در تحلیل عامل اکتشافی

گویه‌ها	اشتراکات استخراجی	بارهای عاملی
q.۱۴	۰,۵۷۸	۰,۷۶۰
q.۱۵	۰,۷۴۳	۰,۸۶۲
q.۱۶	۰,۶۹۰	۰,۸۳۱
q.۱۷	۰,۶۱۱	۰,۷۸۲

جدول ٤. مقادیر ویژه هر عامل به همراه نسبت آنها در تحلیل عامل اکتشافی

عامل‌ها	مقدار ویژه	درصد واریانس تبیینی	نسبت مقادیر ویژه
۱	۲,۶۲	۶۵,۵۵	۴,۴۷
۲	۰,۵۹	۱۴,۶۵	۱,۲۹
۳	۰,۴۵	۱۱,۳۴	۱,۳۴
٤	۰,۳۴	۸,۴۶	--

تحلیل عامل تأییدی با هدف تأیید گویه‌ها (نمونه ب)

به منظور ارزیابی ساختارهای عاملی که تحلیل عامل اکتشافی بر روی نمونه «الف» پیشنهاد داده، از نمونه «ب» استفاده شده است تا تحلیل عاملی تأییدی این مهم انجام پذیرد. جدا کردن دو نمونه به تصادف و انجام تحلیل عاملی به این نحو که روی یک نمونه، تحلیل عامل اکتشافی اجرا شود و تحلیل عامل تأییدی آن روی نمونه دوم به اجرا درآید، وارسی اعتبار گویند که برای اطمینان بیشتر به یافته‌های مربوط به اعتبار است (هومن، ۱۳۸۵). برای این منظور، از نرم‌افزار آی‌آرتی پرو ۲,۱ (تسیوت[1] و تویت[2]، ۲۰۱۱) در برآورد پارامترها از روش بیشینه

1. Thissen
2. Toit

بزرگ‌نمایی[1]، استفاده شده است. در ادامه، برخی از مهم‌ترین شاخص‌های برازش[2] که نشان‌دهنده میزان برازش مدل مفهومی (نظری) با مدل تجربی هستند، ارائه شده است.

جدول ۵. شاخص‌های برازش تحلیل عامل تأییدی عامل پیچیدگی نگرش به برنامه‌های تلویزیونی

شاخص	X^2	(X^2/ df)	(RMSEA)	(NFI)	(NNFI)	(CFI)	(GFI)	(AGFI)
آماره	۵/۴۹ (p<۰/۰۰۱)	۲/۷۴	۰/۰۷۰	۰/۹۹	۰/۹۸	۰/۹۹	۰/۹۹	۰/۹۷

1. maximum likelihood

۲. شاخص نسبت مجذور خی‌دو به درجه آزادی($\chi 2/df$): بنا بر عقیده کلاین نسبت مجذور خی‌دو به درجه آزادی بین ۲و ۳ نسبت قابل قبولی است و هرچه این نسبت کوچک‌تر باشد، برازندگی مدل بهتر است (کلاین/ترجمه صدرالسادات و مینایی، ۱۳۸۲: ۲۰۷)؛ ولی نسبت مجذور کای به درجه آزادی بسیار به حجم نمونه وابسته است و نمونه بزرگ کمیت خی دو را بیش ازآنچه بتوان آن را به‌غلط بودن مدل نسبت داد، افزایش می‌دهد (هومن، ۱۳۸٤: ٤۲۲).

شاخص تقریب ریشه میانگین مجذورات خطا Root Mean Square Error of Approximation (RMSEA): این شاخص، تقریب ریشه میانگین مجذورات خطا است. این شاخص برای مدل‌های بسیار خوب برابر ۰/۰۵ یا کمتر است. مدل‌هایی که RMSEA آن‌ها ۰/۰۵ تا ۰/۰۸ باشند برازش مناسبی دارند و اگر تا ۰/۱۰ باشد برازش ضعیفی دارند (هومن، ۱۳۸۵: ٤۲ و مارویاما، ۱۹۹۷: ۲٤۳).

شاخص‌های GFI و AGFI: شاخص نیکویی برازش Goodness of fit index (GFI) مقدار نسبی واریانس‌ها و کوواریانس‌ها را به گونه مشترک از طریق مدل ارزیابی می‌کند. دامنه تغییرات GFI بین صفر و یک است. مقدار GFI باید برابر یا بزرگ‌تر از ۰/۹۰باشد. شاخص نیکویی برازش تعدیل‌شده Adjusted Goodness of Fit Index (AGFI) یا همان مقدار تعدیل‌یافته شاخص GFI برای درجه آزادی است. این مشخصه معادل با کاربرد میانگین مجذورات به‌جای مجموع مجذورات در صورت و مخرج (۱- GFI) است. مقدار این شاخص نیز بین صفر و یک است. شاخص‌های GFI و AGFI را که جارزکاگ و سوربوم (۱۹۸۹) پیشنهاد کرده‌اند بستگی به حجم نمونه ندارد.

شاخص بنتلر- بونت NFI، شاخص تاکر –لویز (NNFI) و شاخص CFI: براونی و کادک (۱۹۹۳) معتقدند که مقادیر بالاتر از ۰/۹۰ برای شاخص‌های بنتلر-بونت یا شاخص نرم شده برازش NFI و شاخص تاکر – لویز یا شاخص نرم نشده برازش (NNFI) و شاخص CFI حاکی از برازش خوب مدل است.

با توجه به شاخص‌های برازش به جز شاخص مربوط به خی دو که به‌دلیل تعداد بالای نمونه، معنادار گزارش شده است، بقیه شاخص‌ها در حد مطلوب گزارش شده‌اند و ازاین‌رو، می‌توان استنباط کرد که مدل تجربی با مدل مفهومی برازش مطلوبی دارد. تنها یک بعد زیربنایی برای گویه‌های موجود مفروض است. در ادامه بتای استاندارد، آماره تی به همراه مجذور همبستگی چندگانه برای هر گویه در تحلیل عامل تأییدی ارائه شده است.

جدول ۶. پارامترهای تحلیل عامل تأییدی عامل پیچیدگی نگرش به برنامه‌های تلویزیونی

گویه	β استاندارد	t^{**}	مجذور همبستگی چندگانه
q8.14	۰٫۶۶	۲۹٫۶۳	۰٫۴۴
q8.15	۰٫۸۴	۴۰٫۵۷	۰٫۷۱
q8.16	۰٫۷۷	۳۵٫۹۵	۰٫۵۹
q8.17	۰٫۶۷	۳۰٫۴۵	۰٫۴۵

** P < ۰/۰۱

در جدول ٦ میزان بتای استاندارد شده به همراه معناداری ضرایب مسیر از طریق آزمون t ارائه شده، علاوه بر این، میزان تبیین هر یک از گویه‌ها از طریق مجذور همبستگی چندگانه نیز قابل مشاهده است که همگی حکایت از تأثیر بسزا و تقریباً یکسان گویه‌ها دارند. با توجه به همخوان بودن نتایج تحلیل عامل اکتشافی و تأییدی این موضوع روشن شده است که گویه‌های منتخب برای انجام تحلیل بر اساس نظریه‌های CCT و IRT مناسب و دارای اعتبار (روایی) هستند. در ادامه تحلیل‌های روان‌سنجی بر روی ترکیب دو نمونه «الف» و «ب» انجام شده است.

نتایج تحلیل CCT

در جدول ۷ آماره‌های مرسوم تحلیل CCT برای مقیاس سنجش نگرش ارائه شده است. این موارد شامل آلفای کرونباخ، همبستگی گویه با نمره کل، آلفا با حذف گویه و میانگین همبستگی‌های درونی گویه می‌شود (آلن و ین، ترجمه دلاور، ۱۳۹٦و راسل، ۲۰۰۲).

جدول ۷. خلاصه نتایج تحلیل CCT

گویه	همبستگی گویه با کل	میانگین همبستگی با گویه‌ها	آلفا با حذف گویه
q.14	۰٫۵۸۳	۰٫۵۰۱	۰٫۸۰۷
q.15	۰٫۷۲۵	۰٫۵۸۸	۰٫۷۴۳
q.16	۰٫۶۷۷	۰٫۵۵۷	۰٫۷۶۵
q.17	۰٫۶۱۲	۰٫۵۱۴	۰٫۷۹۴

آلفای کرونباخ

اگرچه آلفا به‌صورت مرسوم و گسترده مورد استفاده قرار می‌گیرد و باید تنها به‌عنوان بخشی از سنجش تک‌بعدی بودن و نه تعیین قطعی و منحصربه فرد این مفروضه به‌کار گرفته شود، همگونی و سازگاری درونی یا میانگین کلی از میزان همبستگی گویه‌ها با یکدیگر را در یک مقیاس نشان می‌دهد. براساس نظر کورتینا[1] «آلفا تابعی از روابط درونی بین گویه‌هاست، لیکن چیزی که باید به خاطر داشته باشیم آن است که آلفا دلالتی بر تک بعدی بودن ندارد و مجموعه‌ای از گویه‌ها می‌توانند با یکدیکر به‌طور نسبی رابطه داشته باشند و چندبعدی باشند». بنابراین تنها اگر نتایج تحلیل عاملی انحرافی تک بعدی بودن را نشان ندهد، می‌توان از آلفا نیز تک بعدی بودن را نتیجه گرفت. در مطالعه پیش رو، تحلیل عاملی شواهد کافی را برای این موضوع که تنها یک عامل مکنون باعث ایجاد همبستگی بین گویه‌ها شده، فراهم ساخته است. اگرچه برای چند گویه نهایی مقدار آلفا ۰/۸۲٤ گزارش شده است، با توجه به جدول ۷، با حذف هیچ گویه‌ای مقدار آلفا افزایش نخواهد داشت؛ همچنین در جدول ۷، میانگین همبستگی بین هر گویه با همه گویه‌های دیگر و نیز همبستگی‌های بین گویه با کل مقیاس که حاصل مجموع تمام گویه‌هاست، ارائه شده است. همه آنها نیز بالاتر از ۰/۵۰ گزارش شده‌اند. لازم است توضیح داده شود که همبستگی‌های گویه با مقیاس

1. Cortina

سنجه‌ای از رابطه بین هر گویه و کل مقیاس است و درکل نشان می‌دهد که آیا گویه‌ها در یک مقیاس، با کل سازه به‌صورت کلی، رابطه خوبی دارند و آیا متعلق به آن سازه هستند یا خیر، همچنین هر آیتم تا چه اندازه در اندازه‌گیری آن سازه سهیم است؛ نتایج از تأثیر چشمگیر و تقریباً یکسان گویه‌ها حکایت دارد.

نتایج تحلیل CCT شواهدی ارائه دادند که به‌خوبی نشان داد مجموعه گویه‌های در نظر گرفته شده برای سنجش نگرش پیچیدگی برنامه‌های تلویزیونی مناسب هستند. همگونی (سازگاری) درونی خوبی دارند و همبستگی گویه با کل نیز بالاست؛ همچنین بارعاملی برای هر گویه براساس تحلیل تک عاملی، بالا و قابل قبول گزارش شده است. بنابراین منطقی به نظر می‌رسد که نتیجه گرفته شود مقیاس سنجش پیچیدگی، تنها یک خصیصه زیربنایی را اندازه‌گیری می‌کند.

درواقع، تحلیل انجام شده برای پژوهشگر، این توجیه منطقی را ارائه می‌دهد که مجموع یا میانگین گویه‌های مقیاس، تک نمره‌ای ایجاد می‌کند که نمایانگر سطح یا میزان نگرش به پیچیدگی برنامه‌های تلویزیونی در میان جامعه هدف است. این در حالی است که تحلیل‌هایی اضافی براساس IRT وجود دارد که می‌تواند اطلاعات بیشتر و به احتمال، متفاوتی درباره برازش گویه‌ها و مقیاس‌ها ارائه کند.

تحلیل IRT مدل پاسخ مدرج[1]

از آنجا که گویه‌های مقیاس سنجش نگرش به‌صورت مقیاس لیکرت درجه‌بندی شده‌اند مدل مناسب IRT برای آن مدل پاسخ مدرج سامی‌جیما (۱۹۶۹) در نظر گرفته شده است. چند نکته درباره IRT و مدل پاسخ مدرج (GRM) قبل از تحلیل و ارائه نتایج باید ذکر شود. برازش GRM با یک سری گویه چندگزینه‌ای منجر به برآورد دو نوع پارامتر می‌شود. هر گویه یک ضریب تمییز[2] یا شیب دارد که با α نمایش داده می‌شود و شاخصی است که نشان می‌دهد تا چه اندازه گویه با خصیصه زیربنایی مورد

1. The Graded Response Model (GRM)
2. discriminant

نظر (در اینجا نگرش به پیچیدگی برنامه‌های تلویزیونی) ارتباط دارد. گویه‌هایی که مقدار تمییز بالایی دارند، اطلاعات بیشتری درباره خصیصه مکنون ارائه می‌دهند. هر گویه همچنین تعدادی پارامتر دشواری دارد که به آنها پارامتر آستانه[1] و یا پارامترهای مکانی[2] نیز گفته می‌شود. تعداد پارامترهای آستانه/دشواری برای یک گویه برابر با تعداد طبقات (سطوح) پاسخ منهای یک است (K-1)، آستانه‌ها با β ارائه شده‌اند. مقیاس پارامترهای دشواری (βها) دقیقاً با مقیاس صفت زیربنایی (θ) یکسان است و به‌منظور شناسایی مدل فرض می‌شود که توزیعی نرمال با میانگین صفر و انحراف استاندارد یک دارند (امبریتسون و ریس، ۲۰۰۰). ازاین‌رو، پارامترهای $\beta_{i,k-1}$ اساساً بر مبنای مقیاس نمرات z تفسیر پذیرند. پارامترهای α_i و $\beta_{i,k-1}$ برای ترسیم نمودارهایی که به منحنی‌های پاسخ طبقات[3] (CRC) معروف‌اند، استفاده می‌شوند که به‌صورت دیداری، احتمال پاسخ پاسخ‌دهنده‌ها را در یک پرسش به هر طبقه به‌عنوان تابعی از سطح نگرش آنها نمایش می‌دهند. پارامترهای یادشده، همچنین برای ترسیم نمودارهای تابع آگاهی گویه/پرسش[4] (IIF) به‌کار می‌روند که نمایانگر میزان آگاهی روان‌سنجی هر گویه در سطوح مختلف نگرش‌اند. در مجموع نمودارهای تابع آگاهی، تابع آگاهی مقیاس[5] (SIF) را می‌سازند که میزان آگاهی کل مقیاس را به‌عنوان تابعی از نگرش به پیچیدگی فراهم می‌سازد. به‌طورکلی، پارامتر تمییز بیشتر هر گویه (α_i)، منجر به منحنی‌های پاسخ طبقات (CRS های) باریک‌تر و شیب‌دارتر و در نتیجه تابع، آگاهی با برجستگی بیشتر می‌شود. گستره بیشتر پارامترهای آستانه $\beta_{i,k-1}$ منجر به یکنواختی و تختی CRS ها و IIF ها می‌شود. پارامترهای تمییز (α) و دشواری/آستانه (β) در IRT بدون توجه به نمودار گرافیکی آنها نیز تفسیر پذیرند. در کل پارامترهای تمییز، نشانه‌ای از قوت ارتباط

1. Threshold Parameter
2. local parameter
3. Category Response Curves
4. Item Information Functions (IIF)
5. Scale Information Fuction(SIF)

و رابطه بین گویه و صفت زیربنایی آن هستند. در بسیاری از جوانب این پارامترها شبیه بارهای عاملی یا همبستگی گویه با کل هستند. پارامتر تمییز بیشتر از ۱/۷ بسیار بالا، بین ۱/۳۵ و ۱/۷ در حد بالا، و بین ۰/٦۵ تا ۱/۳٤ متوسط هستند (بیکر[1]، ۲۰۰۱). پارامترهای آستانه به‌عنوان نقاطی بر روی پیوستار صفت مکنون (برای مثال، سطح نگرش) مشخص می‌شوند که در آن، نقاط پاسخ‌دهنده با احتمال ۵۰ درصد به آن طبقه خاص و یا طبقات بالاتر و با احتمال ۵۰ درصد به هر طبقه پایین‌تر پاسخ می‌دهد (امبرتسون و ریس، ۲۰۰۰). به‌منظور توضیح بیشتر فرض کنید یک گویه i با سه طبقه احتمال پاسخ با طبقات هرگز، گه‌گاه و اکثر اوقات داشته باشیم که دارای مقادیر $\beta_{i,1}$ برابر با ۲- و $\beta_{i,2}$ برابر با صفر است؛ این بدین معنی است که مدل پیش‌بینی می‌کند؛ یک مشارکت‌کننده با سطح صفت مکنون برابر با دو انحراف استاندارد زیر میانگین (۲-=θ)، ۵۰ درصد شانس پاسخ به طبقه نخست (هرگز) و ۵۰ درصد شانس پاسخ به دو طبقه دوم و سوم (گه‌گاه و اکثر اوقات) دارد؛ در صورتی که پاسخ‌دهنده‌ای با سطح صفت مکنون در حد میانگین (۰=θ)، ۵۰ درصد شانس پاسخ به طبقه نخست و دوم (هرگز و گه‌گاه) و ۵۰ درصد شانس پاسخ به طبقه سوم (اکثر اوقات) دارد. پاسخ‌دهندگانی که خصیصه مکنون آنها زیر ۲- است، بیشتر احتمال دارد پاسخ «هرگز» را بدهند، آنهایی که بین ۲- و ۰ هستند، بیشتر محتمل است پاسخ «گه‌گاه» و آنهایی که بالای ۰ هستند بیشتر احتمال دارد پاسخ «اکثر اوقات» را انتخاب کنند. میزان آگاهی بخشی هر گویه در هر ناحیه، بستگی به میزان $\beta_{i,k-1}$ آن و اینکه چگونه نمودار آن گسترده یا جمع شده، دارد. طبق دیدگاه دی‌مارس[2] (۲۰۱۰) گویه‌هایی که ضریب $\beta_{i,k-1}$ آنها بین ۲- تا ۲ است، از ویژگی های ضریب دشواری خوبی برخوردار هستند. در ادامه، به تحلیل گویه‌های عامل پیچیدگی پرداخته می‌شود. شاخص برازش RMSEA برای برازش کل مدل با داده‌ها برابر با ۰/۰۵ بود که نشان از برازش کلی مدل با داده‌ها داشت. برای

1. Baker
2. Demars

تحلیل مدل پاسخ مدرج سیمجیما به‌کـار گرفتـه شـده و تحلیـل‌هـا از طریـق نـرم‌افـزار آی‌آرتی پرو[1] انجام شده است.

جدول ۸. پارامترهای تمییز و آستانه/دشواری برآورد شده

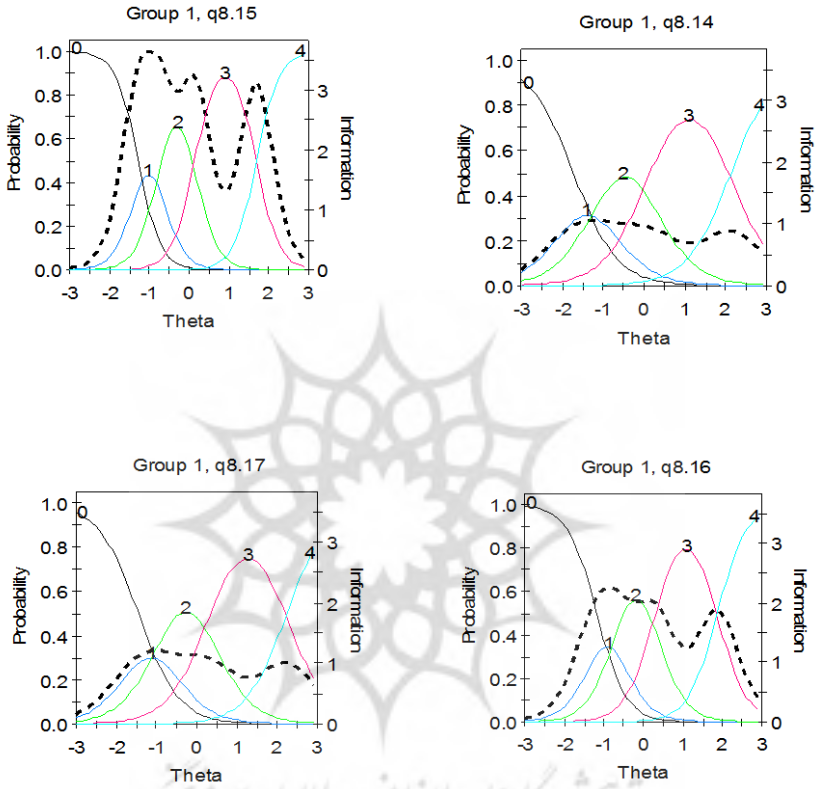
گویه	ضریب تمییز(α)	s.e.	β_1	s.e.	β_2	s.e.	β_3	s.e.	β_4	s.e.
q.14	۱,۷	۰,۰۶	−۱,۷۳	۰,۰۵	−۱,۰۳	۰,۰۴	۰,۰۹	۰,۰۳	۲,۱۲	۰,۰۶
q.15	۳,۵۳	۰,۱۵	−۱,۲	۰,۰۳	−۰,۷۵	۰,۰۲	۰,۱۳	۰,۰۲	۱,۶۹	۰,۰۴
q.16	۲,۷۳	۰,۰۹	−۱,۱	۰,۰۳	−۰,۶۵	۰,۰۳	۰,۲	۰,۰۲	۱,۷	۰,۰۵
q.17	۲,۰۰	۰,۰۶	−۱,۴۵	۰,۰۴	−۰,۳	۰,۰۳	۰,۲۹	۰,۰۳	۲,۲۳	۰,۰۶

براساس دیدگاه بیکر، ضریب تمییز هر چهار گویه در سـطح بسـیار بـالا (α=۱/۷) گزارش شده است و این نشان از سنجش مطلوب عامل مد نظـر دارد. عـلاوه بـر ایـن، براساس دیدگاه دی‌مارس، ضرایب آستانه/دشواری نیز همگـی بـین ۲- تـا ۲ هسـتند و حکایت از مطلوبیت ضرایب دشواری برای طبقات مختلف گویـه‌هـا دارنـد. عـلاوه بـر پارامترهای برآورد شده از دیگر ویژگـی‌هـای تحلیـل IRT میـزان اطلاعـات و آگـاهی بخشی است که هر پرسش در دامنه تتا و نیز کل آزمـون ارائـه مـی‌کنـد، در ادامـه و در بخش نمودارهای گرافیکی تحلیل IRT خروجـی‌هـای مربـوط بـه منحنـی‌هـای پاسـخ طبقات (CRS ها) و تابع آگاهی گویه‌ها (IIF ها) بـه همـراه تـابع آگـاهی کـل آزمـون (TIF) ارائه شده است.

1. IRT-PRO

نمودارهای گرافیکی تحلیل IRT

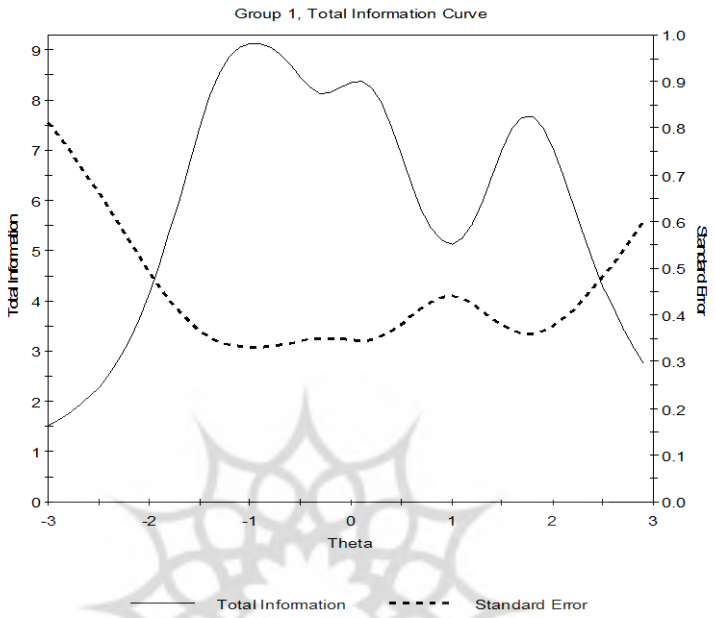
شکل۱. نمودارهای گرافیکی تحلیلIRT گویه‌های سنجش عامل پیچیدگی



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

با توجه به نمودارهای بالا، بیشترین میزان آگاهی بخشی مربوط به گویه ۱۵ بوده که بالاترین ضریب تمییز را نیز داشته است. پایین‌ترین آگاهی بخشی نیز مربوط به دو گویه ١٤ و ١٧ بوده که به‌طور مشابه دارای کمترین مقادیر ضریب تمییز بوده‌اند. میزان آگاهی بخشی همه گویه‌ها در قالب یک آزمون نیز در نمودار ۲ ارائه شده است.

شکل۲. تابع آگاهی عامل پیچیدگی



همان‌طور که ملاحظه می‌شود، بیشترین سنجش آگاهی آزمون مربوط به تتای ۱- بوده و کمترین میزان نیز به جز در کرانها در تتای ۱ گزارش شده است. این بدین معناست که مجموعه گویه‌ها برای افرادی که مقدار تتای آنها بین ۲- تا ۰ بوده بهترین عملکرد را داشته و بهتر است گویه‌های دیگری طراحی شوند که بتوانند افراد را با توانایی در حدود ۱ نیز به خوبی تفکیک کنند. این توانایی IRT است که می‌تواند اطلاعات دقیق‌تری نسبت به آگاهی بخشی هر گویه ارائه کند همچنین می‌تواند میزان خطای استاندارد کل آزمون (در اینجا گویه‌های عامل پیچیدگی) و به‌تبع آن، آگاهی کل آزمون را ارائه کند.

بحث و نتیجه‌گیری

مقایسه نتایج CCT و IRT نشان می‌دهد چنانچه مفروضات تک‌بعدی بودن به‌صورت سختگیرانه رعایت شود، انتخاب گویه‌های مناسب در هر دو تحلیل به نتایج یکسانی می‌انجامد و این کار به علت انجام تحلیل عامل اکتشافی و تأییدی بر روی داده

است، حتی بارهای عاملی در تحلیل عاملی اکتشافی و تحلیل عـاملی تأییـدی و یـا میـزان همبستگی گویه با کل و نیز همگنی میـزان آلفـا بـه ازای حـذف گویـه در تحلیـل CCT همگی حاکی از تناسب و مطلوبیت بیشتر دو گویه ۱۵ و ۱٦ نسبت بـه دو گویـه ۱٤ و ۱۷ بوده است که این نتایج با میزان آگاهی بخشـی و ضـریب تمییـز محاسـبه شـده در IRT یکسان است زیرا بیشترین مقادیر تمییز (جدول ۸) و آگاهی‌بخشی (شـکل ۱) مربـوط بـه گویه‌های ۱۵ و ۱٦ بوده است؛ این موضوع به لحاظ نظـری دور از انتظـار نیسـت چراکـه مقادیر ضرایب تمییز با بارهای عاملی در تحلیل عاملی تشابه دارد. در تحلیـل عـاملی بـار عاملی نشان از میزان همبستگی گویه با صفت مکنون دارد که هرچه میزان آن بالاتر باشـد نشان می‌دهد گویه تناسب بیشتری با عامل دارد، در IRT ضریب تمییز نیز حاکی از میزان تناسب گویه با صفت زیربنایی مورد سنجش است. اگرچه در کل نتـایج CCT و IRT در انتخاب گویه منجر به گویه‌های یکسانی شده است، در IRT این امکان فراهم است کـه از بین بانک گویه‌ها/پرسش‌ها انتخاب به نحـوی صـورت پـذیرد کـه در دامنـه ۳- تـا ۳+ از ضرایب دشواری/آستانه گویه‌های مناسب برای سنجش دامنه ۳- تا ۳+ تتـا وجـود داشـته باشد. بر اساس نتایج این پژوهش پیشنهاد می‌شود که ساخت پرسشنامه‌های مربوط بـه بـه پیماش‌ها و سنجش افکار بر اساس نظریه سؤال پاسخ انجام شود چرا کـه شـاهد افـزایش دقت اندازه‌گیری صفات مکنون مورد نظر بوده و در این رهگذار مـی‌تـوان میـزان آگـاهی بخشی هر گویه و یا کل پرسشنامه تعیین کرد. علاوه بر ایـن، اسـتفاده از IRT در تحلیـل داده‌های سنجش نگرش، راه را فراهم می‌سازد تا از کاربردهای دیگر ایـن نظریـه ازجملـه شناسایی سؤال ها و گویه‌هایی که نسبت به یک گروه اجتمـاعی خـاص مـثلا زنـان و یـا اقوام مختلف عملکرد متفاوتی دارند بهره برد.

اگرچه تعداد کم گویه برای سنجش بعد پیچیدگی یکـی از محـدودیت‌هـای پـژوهش حاضر، بوده، لیکن سعی شده در تحلیل‌ها کمال دقت رعایت شود، با این حـال مـی‌تـوان برای انجام تحلیل عاملی از ماتریس همبستگی پلی‌کوریک بهره برده شود و نتـایج بـا ایـن پژوهش مقایسه شود. علاوه بر این، با توجه به اینکه اغلـب آزمـون‌هـا و پرسشـنامه‌هـای سنجش نگرش چند بعدی بوده‌اند، بهتر است از مدل‌های سؤال- پاسخ چند بعـدی بـرای ساخت آزمون‌های سنجش نگرش در آینده بهره برد.

منابع

آلن مرجی، و ین، وندی‌ام. (۱۳۹۶). **مقدمه‌ای بر نظریه‌های انـدازه‌گیـری (روان‌سـنجی)** (ترجمه علی دلاور). تهران: سمت.

لرد، فردریک. ام. (۱۳۹۱). **کاربردهای نظریه سؤال – پاسخ** (ترجمـه علـی دلاور و جلیل یونسی). تهران: رشد

هومن، حیدرعلی. (۱۳۸۵). **تحلیل داده‌های چنـد متغیـری در پـژوهش رفتـاری**. تهران: پیک فرهنگ.

Baker, F. (2001). **The Basics of Item Response Theory** (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved November 14, 2009, from, http://edres.org/irt/baker/

Baker, F. B. & Kim, S. H. (2004). **Item response Theory: Parameter Estimation Techniques**. CRC Press.

Cerny, B. A. & Kaiser, H. F. (1977). A Study of A Measure of Sampling Adequacy For Factor-Analytic Correlation Matrices. **Multivariate Behavioral Research**, 12(1), 43-47.

De Ayala, R. J. (2009). **The Theory and Practice of Item Response Theory**. Guilford Publications.

DeMars, C. (2010). **Item Response theory**. Oxford University Press.

Embretson, S. E. & Reise, S. P. (2000). Multivariate Applications Books Series**. Item Response Theory for Psychologists**. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Gulliksen, H. (1950). Intrinsic validity. American Psychologist, *5*(10), 511.

Hambleton, R. K. & Swaminathan, H. (2013). **Item Response Theory: Principles and Applications.** Springer Science & Business Media.

Van der Linden, W. J. & Hambleton, R. K. (Eds.). (2013). **Handbook of Modern Item Response Theory**. Springer Science & Business Media.

Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and items. **Applied Psychological Measuremen**t, 9(2), 139–164.

Kahn, J. (2006). Factor analysis in Counseling Psychology Research, Training and Practice: Principles, Advances, and Applications. **The Counseling Psychologist**, 34(5), 684–718.

Lord, F. M.y. & Novick, MR. (1968). **Statistical Theory of Mental Test Scores**. Addison-Wesley Publishing Company.

Morizot, J.; Ainsworth, A. & Reise, S. (2007). **Toward Modern Psychometrics: Application of Item Response Theory Models in Personality Research.** In R. Robins, R. Fraley, & R. Krueger (Eds.), Handbook of Research Methods in Personality Psychology (407–423). New York, NY: Guilford Press.

Ostini, R. & Nering, M. L. (2006). **Polytomous Item Response Theory Models** (No. 144). Sage.

Reise, S.; Waller, N. & Comrey, A. (2000). Factor Analysis and Scale Revision. **Psychological Assessment**, 12(3), 287–297.

Russell, D. (2002). In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis In Personality and Social Psychology Bulletin. **Personality And Social Psychology Bulletin**, 28(12), 1629–1646.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. **Psychometrika Monograph Supplement**, 17.

Sharkness, J. & DeAngelo, L. (2011). Measuring Student Involvement: A Comparison of Classical Test Theory and Item Response Theory In The Construction of Scales From Student Surveys. **Research In Higher Education**, 52(5), 480-507.

Thissen, D.; Cai, L. & du Toit, S. H. C. (2011). **IRTPRO 2.1 for Windows**. Chicago, IL: Scientific Software International.

Worthington, R. & Whittaker, T. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. **The Counseling Psychologist**, 34(6), 806–838.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی