مجلهٔ روانشناسی 53 / سال چهاردهم، شمارهٔ 1 / بهار 1389
Journal of Psychology, 2010, 14 (Spring) 39-61

مقایسهٔ ضریب آلفا و روش‌های برگرفته از مدل معادلات ساختاری در برآورد اعتبار[à]

A Comparison between Alpha Coefficient and Structural Equation Modeling Methods to Estimation of Reliability[à]

مسعود کبیری [à à]

Masoud Kabiri, [*]

چکیده

هدف این مطالعه، مقایسه بین ضریب آلفا و روش‌های برگرفته از مدل معادلات ساختاری در برآورد اعتبار آزمون است. انطباق با مدل اندازه‌گیری اساساً معادل-تاو و عدم همبستگی بین خطاهای اندازه‌گیری مهم‌ترین پیش‌فرض‌های ضریب آلفا است. در مقابل، روش‌های برگرفته از مدل معادلات ساختاری علاوه بر عدم الزام در رعایت چنین پیش‌فرض‌هایی امکان بررسی و آزمون هریک از آن‌ها و قابلیت وزن‌بندی گویه‌ها جهت تشکیل ترکیب بهینه برای تشکیل مقیاس را نیز به‌دست می‌دهد. در این مقاله چهار روش برآورد اعتبار در مدل معادلات ساختاری، شامل «روش‌های مجذور همبستگی چندگانه»، «اعتبار مرکب»، استفاده از «متغیرهای خیالی» و «اعتبار بیشینه‌ای» مورد بحث قرار گرفت. هر یک از این روش‌ها بر روی داده‌های حاصل از مقیاس ادراک نسبت به نگرش معلم ریاضی از مجموعه مقیاس‌های نگرش نسبت به ریاضی فنما و شرمن به‌کار برده شد. داده‌های این مقیاس از نمونه 340 نفری دانش‌آموزان سال اول متوسطه به‌دست آمد. نتایج کاربرد هر یک از این روش‌های برگرفته از مدل معادلات ساختاری نشان دادکه روش‌های معرفی شده مقادیر بالاتری از اعتبار مقیاس نسبت به ضریب‌آلفا را نشان می‌دهند. کاهش ضریب آلفا ممکن است به علت عدم انطباق مقیاس با مدل اساساً معادل-تاو باشد. به‌طور کلی بحث پیشینه‌ای به‌همراه نتایج مطالعه نشان داد که روش‌های برگرفته از مدل معادلات ساختاری برآوردهای دقیق‌تری برای اندازه‌گیری اعتبار آزمون هستند.

کلیدواژه‌ها: *مدل معادلات ساختاری، ضریب آلفا، اعتبار مرکب، اعتبار بیشینه‌ای، متغیر خیالی، مجذور همبستگی چندگان*

Abstract

The purpose of this study is comparison between alpha coefficient and given methods of structural equation modeling to estimate reliability of a scale. Due to necessity of restricted assumptions, that is, fitting to essentially τ-equivalence measurement model and uncorrelated measurement errors. Alpha coefficient may be an inaccurate estimator, In contrast, methods of structural equation modeling does not require considering this assumptions, also there are some possibilities to test assumptions and weighting items to construct an optimal scale. In this article, we discuss four estimators of reliability in structural equation modeling: square multiple coefficient, composite reliability, phantom variable and maximal reliability. Each of them was analyzed on perception of math teacher' attitude scale- a scale of Modified Fennema and Sherman's mathematics attitude scales. Data gathered from 340 ninth grad students. The results reveal these methods have higher values of reliability scale rather than alpha. The reduction of alpha may be due to miss-fit of scale to essentially τ-equivalent model. In general, the results of this study along with literature review show that given methods of structural equation modeling are more accurate estimators to measurement of reliability.

Keywords: alpha coefficient, reliability, structural equation modeling,estimate reliability, square multiple coefficient, composite phantom variable, maximal reliability, phantom variable, maximal reliability.

à دریافت مقاله :1387/3/4تصویب نهایی :1388/6/1
à à پژوهشگاه مطالعات وزارت آموزش و پرورش

à Rresearch Institute for Education of Education ministry, I,R.Iran

* Email: mkabiri@rie, ir, Tel: (+98)21 88 95 2003, 09102123787

Archive of SID

www.SID.ir

– مقدمه

در نظریهٔ کلاسیک آزمون، «*اعتبار*» به عنوان رابطه واریانس واقعی به واریانس مشاهده شده (رایکوف، 2004) و ضریب اعتبار به عنوان سهم واریانس نمره مشاهده شده‌ای که توسط واریانس نمره واقعی تبیین می‌شود (گراهام، 2006) درنظر گرفته می‌شود. بر این اساس ضریب اعتبار به عنوان شاخص کلی دقت اندازه‌گیری ملاحظه می‌شود (رایکوف، 2004) که در نقطهٔ مقابل خطای اندازه‌گیری است. علاوه‌براین، می‌توان «*اعتبار*» را به عنوان تجانس نمرات یا ثبات درونی نیز تعریف نمود (آناستازی، 1379). این تعریف که یکی از رایج‌ترین مبانی محاسبات اعتبار را تشکیل می‌دهد، بر روش به‌دست آوردن اعتبار بر اساس یک‌بار اجرای آزمون استوار است که در مقابل سایر روش‌های محاسبهٔ اعتبار مثل بازآزمایی و فرم‌های همتا قرار دارد. معروف‌ترین روش محاسبه اعتبار در این مبنا، ضریب آلفا است که توسط *کرونباخ* پیشنهاد شده است. بحث این مقاله بر روی این روش از محاسبه اعتبار متمرکز شده است.

بررسی پیشینه پژوهشی نشان می‌دهد که ضریب آلفا دارای پیش‌فرض‌های مهم و جدی است که عدم توجه به آن‌ها می‌تواند بر برآورد اعتبار اثر داشته باشد. این پیش‌فرض‌ها شامل اساساً «*معادل-تاو* (τ) بودن»[1] گویه‌ها و عدم همبستگی بین خطاهای اندازه‌گیری گویه‌ها می‌باشند(بیکون، سائور و یونگ، 1995؛ رایکوف، 2001؛ 2004؛ رایکوف و شراوت، 2002؛ کوماروف، 1997؛ گراهام، 2006؛ گرین و هرشبرگر، 2000). اساساً معادل-تاو بودن گویه‌ها به یکسانی نمرات مشاهده شدهٔ گویه‌ها یا تساوی بار گویه‌ها مرتبط است. عدم رعایت این پیش‌فرض باعث می‌شود که آلفا دارای اعتبار کم‌تری باشد. به این دلیل در برخی مواقع اعتبار به عنوان مرز پائین اعتبار شناخته می‌شود (رایکوف و شراوت، 2002؛ گراهام، 2006).

برای بررسی «*معادل-تاو بودن گویه‌ها*» (و یا اندازه‌گیری در مقیاس یکسان) می‌توان انحراف معیار گویه‌ها را ملاحظه نمود. در صورتی‌که انحراف معیار نمرات گویه‌های تشکیل دهندهٔ یک آزمون از یکدیگر تفاوت زیادی داشته باشند، می‌توان چنین نتیجه‌گیری کرد که این گویه‌ها احتمالاً در مقیاس متفاوتی اندازه‌گیری می‌شوند، در غیر این‌صورت، وجود مقیاس یکسان محتمل است. چنین بررسی می‌تواند توسط

Archive of SID

www.SID.ir

ایجاد فواصل اطمینان پیرامون انحراف معیار انجام شود. به این‌صورت که با ایجاد فواصل اطمینان حدود انحراف معیار، می‌توان معادل بودن آن‌ها را از نظر دیداری بررسی کرد (گراهام، 2006). با این‌حال، رعایت نکات دیگری می‌تواند اطمینان از معادل-تاو بودن گویه‌ها را حاصل کند. به‌طور مثال، در صورتی‌که از اشکال متفاوتی برای دریافت پاسخ استفاده کنیم، احتمال معادل-تاو نبودن گویه‌ها زیاد خواهد بود (گراهام، 2006). به عبارت دیگر، اگر در مقیاسی هم گویه‌های صحیح و غلط و هم گویه‌های 5 درجه‌ای طیف لیکرت وجود داشته باشد، احتمالاً نمی‌توان این گویه‌ها را گویه‌هایی معادل دانست. وزن‌بندی یکسان گویه‌ها نیز از عوامل به‌دست آوردن گویه‌های معادل-تاو است (بیکون، سائور و یونگ، 1995). در بیشتر مطالعات، گویه‌های آزمون با وزن یکسانی با هم‌دیگر ترکیب می‌شوند. اگر گویه‌ها وزن‌های یکسانی برای ترکیب و ایجاد یک متغیر ترکیبی نداشته باشند، احتمالاً مقیاس معادل-تاو را نخواهیم داشت. از عوامل دیگری که بر معادل-تاو بودن اثر می‌گذارد، تعداد گویه‌هاست. به‌طوری که، «*کم برآورده شدن*»[۲] ضریب آلفا در آزمون‌هایی که تعداد گویه‌های بیشتری دارند، کم‌تر اتفاق می‌افتد (گراهام، 2006). به این دلیل که با معادل-تاو نبودن یک گویه مقدار کمی از سهم واریانس نمره واقعی تغییر می‌کند و بنابراین کم‌تر در معادل-تاو بودن مقیاس تأثیر می‌گذارد. از طرف دیگر، در مقیاس‌های کوچک‌تر، معادل-تاو نبودن یک گویه سهم بیشتری بر معادل-تاو بودن مقیاس دارد. با این‌حال، نتایج برخی از مطالعات نشان می‌دهد که وجود حتی یک گویه غیر معادل-تاو می‌تواند بر دقت ضریب آلفا موثر باشد (رایکوف، 1997). با وجود اهمیت این نکات هیچ روش دقیق و مشخصی بر اساس روش‌های مرسوم اندازه‌گیری برای بررسی معادل-تاو بودن گویه‌ها وجود ندارد.

به‌طور کلی چهار نوع «*مدل موازی*»[۳]، معادل-تاو، اساساً معادل-تاو و «متجانس»[٤] وجود دارند که بسته به مفاهیم و عناصر محدود شده آن تعریف می‌گردند.

- «*مدل موازی*» محدودترین مدل اندازه‌گیری در تعریف نمرات واقعی است. در این مدل علاوه بر آن‌که همه گویه‌های آزمون متغیر نهفتهٔ یکسانی را می‌سنجند (مفهوم تک‌وجهی بودن که در تمام آزمون‌های ثبات درونی فرض می‌شود)، فرض می‌شود که همه گویه‌های آزمون دقیقاً معادل یکدیگر هستند. علاوه‌برآن، همه گویه‌ها باید متغیر

Archive of SID

www.SID.ir

نهفته را در مقیاس مشابهی اندازه‌گیری کرده که با درجه یکسانی از دقت و مقدار یکسانی از خطا همراه باشند (گراهام، 2006). به عبارت دیگر، مقیاس اندازه‌گیری، دقت و خطای اندازه‌گیری همه گویه‌ها یکسان در نظر گرفته می‌شود. این مطلب را می‌توان با استفاده از معادل‌های ریاضی آن به ترتیب به این صورت نشان داد:

$$(1)\ b_i = \ldots = b_k = ۱$$

$$(2)\ a_i = \ldots = a_k = ۰$$

$$(3)\ \mathrm{Var}(E_i) = \ldots = \mathrm{Var}(E_k)$$

در این عبارت‌ها b_i ضریب بین متغیرهای مشاهده و متغیر نهفته (بارهای عاملی)، $\mathrm{Var}(E_i)$ واریانس خطای اندازه‌گیری، و a_i ضریب ثابت درمعادله $T_i = a_{ij} + b_{ij} T_j$ (4) است. این معادله رابطه خطی بین نمرات T_i و T_j را نشان می‌دهد (رایکوف، 1997).

¨ «*مدل معادل- تاو*» نیز ساختاری همانند «*مدل موازی*» دارد به جز آن‌که در این مدل، خطاهای واریانس گویه‌ها می‌توانند از همدیگر متفاوت باشند. این موضوع دلالت بر این دارد که تک‌تک گویه‌ها، متغیر نهفته یکسانی را در مقیاس یکسان و با درجه دقت یکسانی اندازه‌گیری می‌کنند ولی احتمالاً میزان خطای متفاوتی وجود دارد (رایکوف، 1998a؛ گراهام، 2006). معنای مفهومی این مدل آن است که اگرچه همه نمرات واقعیِ گویه‌ها معادل هستند ولی هر گویه عبارت‌های خطای منحصر به فردی دارد. بر حسب عبارات ریاضی پیشین می‌توان گفت که در مدل معادل- تاو عبارت (3) درنظر گرفته نمی‌شود.

¨ «*مدل اساساً معادل- تاو*» فرض می‌کند که هر گویه، متغیر نهفته یکسانی را در یک مقیاس یکسان ولی با دقت احتمالاً متفاوتی اندازه‌گیری می‌کند. علاوه‌براین، در این مدل همانند مدل معادل-تاو احتمالاً خطاهای اندازه‌گیری نیز متفاوت هستند (رایکوف، 1997). در این تعریف تفاوت بین دقت و مقیاس گویه برجسته شده است. در مدل معادل- تاو فرض می‌شود که نمرات واقعی گویه‌ها معادل هستند، ولی در مدل اساساً معادل- تاو نمره واقعی گویه می‌تواند توسط ترکیب با یک ضریب یگانه اضافی به هر زوج از متغیرها متفاوت باشد. این پیش‌فرض به‌شکل ریاضی این‌گونه نوشته می‌شود:

$$(۵)\ X_{ik} = (\alpha_k + T_i) + E_{ik}$$

Archive of SID

www.SID.ir

عبارت فوق این حقیقت را منعکس می‌کند که اگرچه نمرات واقعی گویه‌ها در مقیاس یکسانی اندازه‌گیری می‌شوند (یعنی واریانس مشابهی دارند)، ولی ممکن است دقت متفاوتی (میانگین‌های متفاوت) داشته باشند. گنجانیدن ضریب اضافی تنها بر میانگین گویه‌ها تأثیر می‌گذارد، ولی بر واریانس و کوواریانس آن‌ها متفاوت نیست (گراهام، 2006). بر حسب عبارت‌های مشخص شده در مدل موازی، می‌توان مدلی را «*مدل اساساً معاول-تاو*» دانست که عبارت‌های (2) و (3) را رعایت نکرده باشد. برای استفاده از ضریب آلفای کرونباخ لازم است که اندازه‌گیری حداقل بر مبنای مدل اساساً معادل-تاو استوار شده باشد. به این معنا که مقیاس و نمرات واقعی گویه‌ها یکسان باشد.

- «*مدل متجانس*» در نهایت کم‌ترین محدودیت در آن وجود دارد و به همین دلیل عمومی‌ترین مدل برای استفاده در برآورد «*اعتبار*» است. در این مدل فرض می‌شود که تک‌تک گویه‌ها متغیر نهفته یکسانی را با مقیاس، دقت و میزان خطای احتمالاً متفاوت اندازه‌گیری می‌کنند (آدامسون، شولین، لوید و لوئیس، 2000؛ گراهام، 2006). در «مدل متجانس» تفاوت این «*مدل با مدل اساساً معادل-تاو*» در این است که در «*مدل اساساً معادل-تاو*» نمرات واقعی گویه‌ها توسط ضریب «*جمع‌پذیری*»[۵] با همدیگر متفاوت هستند، در حالی‌که در مدل متجانس بین نمرات واقعی گویه‌ها رابطه خطی فرض شده و بین هر زوج از آن‌ها، ضریب جمع‌پذیر یگانه‌ای وجود دارد که می‌توان آن‌را به این صورت نشان داد (رایکوف، 1998a):

$$(۶)\ X_j = a_j + b_j T_1$$

به‌طوری‌که در عبارت (6) ملاحظه می‌شود «*مدل متجانس*» عمومی‌ترین مدل اندازه‌گیری است، زیرا می‌توان با رعایت هریک از پیش‌فرض‌های مدل‌های قبلی، مدل متجانس را به یکی از مدل‌های محدودتر تبدیل کرد. بنابراین، سه مدل مورد اشاره در بالا زیر مجموعه‌های «*مدل متجانس*» هستند و به همین دلیل است که برخی از صاحب‌نظران سایر مدل‌های اندازه‌گیری را یکی از حالت‌های مدل متجانس می‌دانند (رایکوف و هانکوک، 2005).

- برای تطبیق داده‌های حاصل از اندازه‌گیری با هر یک از مدل‌ها از ویژگی سلسله مراتبی مدل‌های اندازه‌گیری استفاده می‌شود. به عبارت دیگر، برای انتخاب بهترین

Archive of SID

www.SID.ir

مدل، از نامحدودترین و غیراقتصادی‌ترین مدل به‌سوی محدودترین و اقتصادی‌ترین مدل حرکت می‌شود تا بهینه‌ترین مدل انتخاب گردد (گراهام، 2006). بدین منظور از «*مدل‌های آشیانه‌ای*»[6] استفاده می‌شود، یعنی ابتدا از مدل متجانس شروع شده و برازش این مدل با داده‌ها آزمون می‌شود. در صورت معنی‌داری، مدل معادل-تاو و یا اساساً معادل-تاو آزمون می‌شود، یعنی مدلی که بارهای عاملی آن برابر با یک درنظر گرفته می‌شود. در صورتی‌که تفاوت در مقادیر آمارهٔ خی‌دو بین این دو مدل از نظر آماری معنی‌دار نبود مدل محدودتر انتخاب شده و در غیر این‌صورت مدل اولی و محدودتر به عنوان مدل مناسب برگزیده می‌شود. در صورتی‌که مدل معادل-تاو پذیرفته شود، مدل موازی آزمون می‌گردد و خطاهای اندازه‌گیری گویه‌ها برابر فرض می‌گردند. پس از آزمودن مدل، تفاوت مقادیر خی‌دو (آماره تفاوت خی‌دو) بررسی می‌گردند. همانند مرحله قبلی، در صورت معنی‌دار نبودن آماره تفاوت خی‌دو مدل محدودتر و در غیر این‌صورت مدل نامحدودتر انتخاب می‌گردد (رایکوف، 1997). بنابراین، با توجه به اصل «*بهینه بودن*»[7] در اندازه‌گیری، اگر دو مدل به‌طور منطقی برازش پیدا کنند، برآورد مدل محدودتر از مدل با محدودیت کم‌تر ترجیح داده خواهد شد. به این دلیل توصیه شده است که قبل از انجام هرگونه عمل اندازه‌گیری، پیش‌فرض‌های برآورد اعتبار آزمون شده و مناسب‌ترین مدل در رابطه با داده‌ها انتخاب گردد.

پیش‌فرض «*خطای اندازه‌گیری ناهمبسته*» نیز از جمله شرایط استفاده از ضریب آلفا است. مطالعات نشان داده‌اند که همبستگی بین خطاهای اندازه‌گیری بر روی برآورد ضریب آلفا موثر است (رایکوف، 2001؛ 2004؛ کوماروف، 1997؛ گرین و هرشبرگر، 2000؛ لوک، 2005). عمده مطالعات در این زمینه نشان می‌دهند که انحراف از پیش‌فرض خطاهای مستقل می‌تواند ضریب آلفا را متورم ساخته و در نتیجه ضرایب آلفا بیش از مقدار واقعی آن برآورد گردند (کوماروف، 1997؛ گرین و هرشبرگر، 2000). با این‌حال، برخی از نوشته‌ها حاکی از کم‌برآورد کردن اعتبار در هنگام همبسته بودن خطاهای اندازه‌گیری دارد (رایکوف، 2001؛ 2004). ولی به‌طور کلی با توجه به پژوهش‌هایی که به‌طور منظم این موضوع را بررسی کرده‌اند، می‌توان چنین برداشت کرد که چنان‌چه بین خطاهای اندازه‌گیری در یک آزمون، همبستگی و به خصوص همبستگی مثبت وجود داشته باشد، آلفا اعتبار را «*بیش‌برآورد*»[8] می‌کند (زیمرمن، زومبو

Archive of SID

www.SID.ir

و لالونده، 1993). این موضوع در مطالعه دیگری نیز تأئید شد و بر اساس آن مشخص گردید که وجود خطای همبسته می‌تواند تا اندازه‌ای، مقداری از کم‌برآورد شدن اعتبار به علت معادل-تاو نبودن را کاهش دهد (کوماروف، 1997) که معرف تعامل رعایت نکردن هر دو پیش‌فرض است. مقدار سوگیری آلفا در هنگام بروز خطاهای همبسته قابل محاسبه خواهد بود (رایکوف، 2001). ناهمبستگی بین خطاهای اندازه‌گیری از طریق اجرای «*استقلال آماری*»[9] به‌دست می‌آید. استقلال آماری به دو موضوع اشاره دارد: اول آن‌که مقدار مشاهده شدهٔ افراد در یک گویه مستقل از مقادیر آن‌ها در گویه‌های دیگر باشد و دوم این‌که مقدار مشاهده شده‌ی افراد در گویه به‌طور آزمایشی از مقدار مشاهده شده‌ی هر فرد دیگر در آن گویه یا گویه‌های دیگر متفاوت باشد (کوماروف، 1997). بنابراین پیش‌فرض استقلال آماری در هنگامی‌که آزمون مشابه باشند یا گویه‌های مقیاس، تقریب هم‌زمان نزدیکی داشته باشند، غیر قابل دفاع خواهد بود. همبستگی بین خطاهای اندازه‌گیری به دلایل متفاوتی رخ می‌دهد که از آن جمله وجود یک نظم مشخص بین گویه‌ها خواهد بود. در نتیجه برای پرهیز از بروز چنین مشکلی می‌توان برای به حداقل رساندن ثبات تصنعی، نظم گویه‌ها را به‌صورت تصادفی تنظیم کرد. به عبارت دیگر با این‌کار می‌توان امیدوار بود که پاسخ افراد به نمره واقعی آن‌ها وابسته است و نه به پاسخ‌های آنان در گویه‌های قبلی. علاوه‌براین، مواردی چون اجرا شدن همه گویه‌ها در یک مقیاس به‌طور متوالی و در یک وضعیت یکسان، داشتن عبارت‌پردازی یکسان و نمره‌گذاری در یک جهت نیز از مواردی است که همبستگی بین گویه‌ها و خطاها را افزایش داده و در نتیجه ضریب آلفا را بیش‌برآورد خواهد کرد (گرین و هرشبرگر، 2000).

" دشواری رعایت پیش‌فرض‌ها به همراه فقدان قابلیت بررسی و آزمون آن‌ها در روش‌های مرسوم اندازه‌گیری، ابهاماتی را در دقیق بودن ضریب آلفا به‌وجود می‌آورد. به همین خاطر معمولاً از ضریب آلفا به عنوان برآورد غیردقیق اعتبار یاد می‌شود. با این‌حال، در مواقعی که تعداد نسبتاً زیادی از مولفه‌ها وجود داشته باشند (بیشتر از 6)، یا بارها به‌طور بالایی در یک سازهٔ نهفتهٔ مشترک بار داشته باشند (حداقل0/6) و خطاهای اندازه‌گیری ناهمبسته‌ای وجود داشته باشند استفاده از آلفا مجاز خواهد بود (رایکوف، 2001؛ 2004).

Archive of SID

www.SID.ir

" با توجه به مطالبی که گفته شد می‌توان نتیجه گرفت کـه در اسـتفاده از ضـریب آلفا به عنوان «*شاخص اعتبار*» حداقل دو مشکل وجود دارد. مشکل اول برآورده کـردن پیش‌فرض‌های لازم برای استفاده از ضریب آلفا و مشکل دوم آزمون کردن هر یک این پیش‌فرض‌ها در رابطه با داده‌های در دسترس می‌باشد. برای حل مشکل اول روش‌های جای‌گزین ضریب آلفا پیشنهاد شده است. به عنوان نمونه در پیش‌فرض ناهمبسته بودن خطاهای اندازه‌گیری گویه‌ها، در مدل معادلات ساختاری خطاهای انـدازه‌گیـری را بـه خطای همبستهٔ تولید شده توسط عامل و خطای همبسـته تولیـد شـده توسـط خطـای اندازه‌گیری تفکیک می‌کنند. خطای همبسته‌ای که توسط عامل‌ها تولید می‌شود واریانس معتبر را مورد توجه قرار می‌دهد، در حالی‌که خطاهای همبسته تولید شده توسط خطای اندازه‌گیری تصادفی واریانس «*نامعتبر*» را درنظر می‌گیرند (گـرین و هرشـبرگر، 2000). برای مواقعی‌که مدل اساساً معادل-تاو نیست نیز روش‌های دیگری جای‌گـزین ضـریب آلفا شده‌اند که از آن جمله روش امگا یا معادل آن روش اعتبار سازه (بیکون، سـائور و یونگ، 1995؛ رایکوف و شراوت، 2002؛ زینبرگ، رول و یوول، 2007)، روش امگـای وزن‌بندی شده (بیکون، سـائور و یونـگ، 1995) و روش‌هـای موجـود در روش مـدل معادلات ساختاری هستند. برای حل مشـکل دوم (عـدم توانـایی بـرای آزمـون کـردن پیش‌فرض‌ها) مناسب‌ترین و رایج‌ترین روش، استفاده از مدل معادلات ساختاری است. در این مقاله سعی می‌شود که جنبه‌های استفاده از شاخص‌ها و روش‌های ویژه موجـود در مدل معادلات ساختاری برای برآورد اعتبار و همچنین آزمون پـیش‌فـرض‌هـا مـورد بررسی قرار گیرند. هم‌چنین برخی مطالعات نشان داده‌اند کـه ضـریب آلفـا نسـبت بـه وجود نقاط پرت تأثیرپذیر است (لیو و زومبو، 2007)

" کاربرد روش مدل معادلات ساختاری در برآورد اعتبار ابزار به علت قابلیت‌ها و مزایای استفاده از آن رواج یافته است. مهم‌ترین دلیل استفاده از این روش عـدم اجبـار برای رعایت‌کردن پیش‌فرض‌های سخت و جدی است که در روش آلفا وجود دارد. در این روش امکان آن وجود دارد که از مدل اندازه‌گیری کم‌تر محدودشده‌ای همانند مدل متجانس استفاده شود و در نتیجه الزام استفاده از مدل اساساً معادل-تاو از بین مـی‌رود (آدامسون و همکاران، 2000؛ رایکوف، 1997؛ رایکـوف و شـراوت، 2002). خطاهـای ناهمبسته نیز از جمله پیش‌فرض‌های اساسی در ضریب آلفاست که به عنوان همبستگی

Archive of SID

www.SID.ir

خطای اندازه‌گیری درنظر گرفته می‌شود، ولی در روش مدل معادلات ساختاری کاملاً به‌صورت خطای اندازه‌گیری «خالص یا ناب»[10] درنظر گرفته نمی‌شود بلکه به عنوان عوامل باقیمانده مشخص‌نشده مورد توجه قرار می‌گیرند. این نمرات باقیمانده هم مولفه‌های تصادفی و هم مولفه‌های غیرتصادفی را دربر می‌گیرند. خطاهای همبسته به‌صورت هم‌پراکندگی تولید شده توسط مولفه‌های غیرتصادفی یا عواملی هستند که لزوماً به‌صورت نااعتبار ملاحظه نمی‌شوند (گرین و هرشبرگر، 2000)، در نتیجه می‌توان بین خطاهای اندازه‌گیری گویه‌ها همبستگی درنظر گرفت. به عبارت دیگر، دو گویه می‌توانند تأثیرات تصادفی مشترکی را دربرگیرند که در نمرات خطای مرتبط با آن‌ها مشترک بوده و به همبستگی بین این خطاها منجر شوند (رایکوف، 2001؛ کوماروف، 1997). علاوه بر مزیت آزاد بودن از رعایت پیش‌فرض‌ها، در روش‌های برگرفته از مدل معادلات ساختاری امکان محاسبه اعتبار با استفاده از عملکرد وزن‌بندی کردن گویه‌ها نیز وجود دارد. استفاده از بارهای غیرمساوی گویه‌ها برای برآورد اعتبار در زمانی‌که گویه‌ها به‌طور مساوی بر برآورد سازه‌ها تأثیر نمی‌گذارند، مفید است (بیکون، سائور و یونگ، 1995؛ رایکوف، 1998a). هم‌چنین به علت انعطاف بالای روش مدل معادلات ساختاری امکان بررسی تفاوت‌های گروهی در اعتبار مرکب، بررسی تفاوت اعتبار در نسخه‌های متفاوت ابزار اندازه‌گیری (مثلاً نسخه قلم و کاغذی در مقابل نسخه رایانه‌ای) (رایکوف، 2004)، استفاده از داده‌های مقوله‌ای (رایکوف، 1998a)، به‌دست آوردن خطاهای معیار و فواصل اطمینان برای اعتبار مرکب در حالتی که به پیش‌فرض نرمال بودن مولفه‌ها وابسته نباشند (رایکوف، 1998a؛ رایکوف و شراوت، 2004)، فراهم می‌گردد. امکان محاسبه میزان سوگیری و کم‌برآوردی ضریب آلفا نیز از دیگر قابلیت‌های روش مدل معادلات است (رایکوف، 1997؛ 1998b؛ 2001). به تمامی مزایای اشاره شده می‌توان امکان آزمون کردن پیش‌فرض‌ها را اضافه نمود که با سایر روش‌ها قابل انجام نیست.

با وجود مزایای مدل معادلات ساختاری در برآورد اعتبار، ملاحظاتی در مورد استفاده از این روش باید انجام پذیرد که مهم‌ترین آن نیاز به نمونه‌های بزرگ مشارکت‌کنندگان است. این امر به‌ویژه در مورد داده‌های طبقه‌ای که تعداد محدودی گزینه برای پاسخ دارد، از اهمیت خاصی برخوردار است زیرا در این‌گونه مواقع روش

Archive of SID

www.SID.ir

برآورد حداقل «*مجذورات وزن‌بندی شده*»[11] برای آزمون مدل مورد نیاز بوده که این روش حساسیت زیادی در مورد استفاده از نمونه‌هایی دارد که به‌طور بهینه بزرگ شناخته می‌شوند (رایکوف، 1997؛ 2001؛ 2004). لذا در این موقعیت‌ها استفاده از نمونه‌های کوچک برای برآورد اعتبار مرکب گمراه‌کننده است. علاوه‌براین، تعداد گویه‌ها ملاحظه قابل توجهی در آزمون مدل‌های اندازه‌گیری به‌شمار می‌رود. در مواقعی که تنها دو گویه وجود داشته باشد، روش‌های محاسبه اعتبار تنها برای مدل‌های موازی یا معادل-تاو قابل کاربرد هستند زیرا برای حالت آزمون موازی، مدل دو پارامتر آزاد دارد و بنابراین «*بیش‌شناسایی‌شده*»[12] است (درجه آزادی برابر با یک است). در این موقعیت با افزودن محدودیت‌های اضافی هم‌چون برابری بارهای شاخص (مدل معادل-تاو) و یا برابری واریانس خطا (مدل موازی) مدل در حالت قابل برازشی قرار می‌گیرد (رایکوف و هنکوک، 2005). برای مدل معادل-تاو، این مدل سه پارامتر آزاد دارد و «*کاملاً شناسایی‌شده*»[13] است (درجه آزادی برابر با صفر است). مدل متجانس در این وضعیت کم‌شناسایی‌شده بوده (درجه آزادی آن منفی است) و بنابراین کاربرد عملی ندارد (یورسکاگ و سوربوم، 1993). در زمانی‌که سه گویه وجود دارد، مدل متجانس کاملاً شناسایی‌شده و قابل برآورد می‌شود ولی نمی‌توان آن‌را با «*مدل اشباع‌شده*»[14] مقایسه کرد زیرا هر دوی این مدل‌ها دارای درجه آزادی صفر هستند (رایکوف، 1997). با گویه‌های بالاتر از چهار، تمامی مدل‌ها کم‌شناسایی‌شده بوده و در مقابل مدل اشباع‌شده قابل دفاع خواهند بود.

¨ در استفاده از مدل معادلات ساختاری برای برآورد اعتبار روش‌های مختلفی پیشنهاد شده است. به‌طور کلی می‌توان این روش‌ها را به دو دسته‌ی روش‌های مبتنی بر تک‌تک گویه‌ها یا متغیرهای مشاهده شده و روش‌های مبتنی بر ترکیب متغیرهای مشاهده شده تقسیم نمود. گرایش بیشتر پژوهشگران و صاحب‌نظران بر استفاده از روش‌های مبتنی بر ترکیب گویه‌هاست تا به این وسیله بتوانند جایگزین‌های مناسبی را برای روش‌های مرسومی هم‌چون آلفا پیشنهاد دهند. با این وجود در ابتدا روش‌های مبتنی بر تک‌تک گویه‌ها معرفی می‌گردند.

¨ در مفهوم «*اعتبار*» مشخص شد که اعتبار قسمتی از اندازه‌گیری است که مشمول خطای تصادفی نمی‌شود. بنابه تعریف در «*مدل معادلات ساختاری*»، اعتبار به عنوان

Archive of SID

www.SID.ir

«واریانسی که به‌وسیله خطای اندازه‌گیری تبیین نمی‌شود» تعریف می‌شود (راینس- اودی، 2000، 128). مشهورترین شاخصی که بر مبنای بررسی تک‌تک گویه‌ها معرفی شده است، «*ضریب مجذور همبستگی چندگانه*»[15] نام دارد که نشان‌دهنده مقدار واریانس تبیین شده در متغیر x توسط متغیر نهفته است (راینس- اودی، 2000؛ قاضی طباطبایی، 1377؛ یورسکاگ و سوربوم، 1993). این شاخص به عنوان حد پائین اعتبار برای متغیرهای x مورد توجه قرار می‌گیرد، به این معنی که اعتبار هر گویه حداقل برابر با مجذور همبستگی‌های چندگانه است (قاضی طباطبایی، 1377؛ یورسکاگ و سوربوم، 1993). مقادیر این شاخص از طریق خروجی‌های برنامه‌های رایانه‌ای مدل معادلات ساختاری هم‌چون LISREL به‌دست می‌آیند و مقدار آن بین صفر و یک است. جهت تفسیر این شاخص‌ها، مقادیر بالاتر از 0/5 به عنوان ضرایب مناسب درنظر گرفته می‌شوند (راینس- اودی، 2000). هم‌چنین جدای از این شاخص، همبستگی هر گویه با متغیر نهفته در هر زیرمقیاس نیز به عنوان ضریبی برای ثبات درونی هر گویه معرفی شده است (عابدی، 2002).

راهبرد دوم ترکیب متغیرها که رواج زیادی پیدا کرده است، ترکیب همه گویه‌های یک عامل با همدیگر است که از آن روش به عنوان «*اعتبار مرکب*»[16] یاد می‌شود. این روش بر اساس مفهوم اصلی اعتبار در نظریه کلاسیک آزمون بنا شده است که اعتبار را به عنوان سهم واریانس نمره واقعی به واریانس نمره مشاهده شده درنظر می‌گیرد. بدین منظور لازم است که برآوردهایی برای واریانس نمره واقعی و واریانس نمره مشاهده شده داشته باشیم. واریانس نمره مشاهده شده‌ی یک اندازه می‌تواند توسط ایجاد یک متغیر مشاهده شده ترکیبی در دسترس قرار گیرد که توسط جمع واریانس‌های تک‌تک متغیرهای مشاهده شده به‌وجود می‌آید (گراهام، 2006). در این روش یک شاخص آماری برای برآورد اعتبار مجموعه‌ای از گویه‌ها که در یک عامل قرار دارند ارائه می‌شود. با توجه به نوع ترکیب گویه‌ها و وضعیت خطاهای آن‌ها (با خطای همبسته و بدون خطای همبسته) فرمول‌های چندگانه‌ای پیشنهاد شده است که قابل تبدیل به یکدیگر می‌باشند. در حالت کلی و مدل‌های متجانس و هم‌چنین گویه‌های وزن‌بندی نشده، اعتبار مرکب به‌وسیله این فرمول به‌دست می‌آید (آدامسون و همکاران، 2000، 973؛ رایکوف،2004، 345):

Archive of SID

www.SID.ir

$$\text{اعتبار مرکب (7)} = \frac{(\sum_{i=1}^{k} b_i)^2}{(\sum_{i=1}^{k} b_i)^2 + \sum_{i=1}^{k} \theta_{ii}}$$

دراین عبارت $\sum_{i=1}^{k} b_i$ مجموع ضرایب بین متغیرهای مشاهده شده و نهفته (بارعاملی) k گویه و $\sum_{i=1}^{k} \theta_{ii}$ مجموع واریانس‌های خطای k گویه را نشان می‌دهد. در صورتی‌که بین خطاهای اندازه‌گیری همبستگی وجود داشته باشد، از فرمول زیر استفاده می‌شود (رایکوف، 2004، 344):

$$\text{اعتبار مرکب (8)} = \frac{(\sum_{i=1}^{k} b_i)^2}{(\sum_{i=1}^{k} b_i)^2 + \sum_{i=1}^{k} \theta_{ii} + 2\sum_{1\le i<j\le k} \theta_{ij}}$$

در این عبارت $\theta_{jj(1\le i<j\le k)}$ کوواریانس خطای غیرصفر است.

این دو فرمول قابل تبدیل به وضعیت گویه‌های وزن‌بندی شده هستند. در این‌حالت، گویه‌ها با وزن‌های خاصی با همدیگر ترکیب شده و نمره مرکبی را به‌وجود می‌آورند. نمره مرکب از طریق عبارت زیر به‌دست می‌آید:

$$Y = w_1 y_1 + w_2 y_2 + \ldots + w_k y_k \quad (9)$$

دراین عبارت وزن‌های w_1، w_2، ...، w_k از پیش مشخص هستند (چگونگی به‌دست آوردن وزن‌ها در قسمت‌های بعدی توضیح داده خواهد شد). اعتبار مرکب یک آزمون وزن‌بندی شده از طریق فرمول زیر محاسبه می‌گردد (رایکوف، 2004، 344):

$$\text{اعتبار مرکب (10)} = \frac{(\sum_{i=1}^{k} w_i b_i)^2}{(\sum_{i=1}^{k} w_i b_i)^2 + \sum_{i=1}^{k} w_i^2 \theta_{ii}}$$

در حالت خطاهای همبسته (رایکوف، 2004، 345):

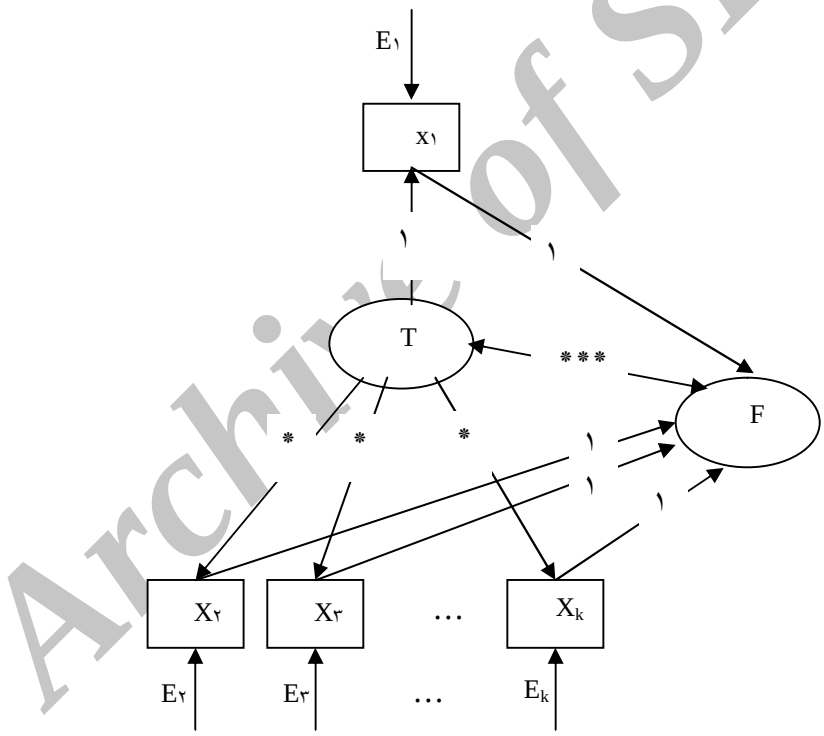
$$\text{اعتبار مرکب (11)} = \frac{(\sum_{i=1}^{k} w_i b_i)^2}{(\sum_{i=1}^{k} w_i b_i)^2 + \sum_{i=1}^{k} w_i^2 \theta_{ii} + 2\sum_{1\le i<j\le k} w_i w_j \theta_{ij}}$$

روش‌هایی که در بالا توضیح داده شد، برآورد نقطه‌ای اعتبار مرکب را نشان می‌دهند. می‌توان با استفاده از روش‌هایی برآورد فاصله‌ای اعتبار مرکب را نیز محاسبه کرد. برآورد فاصله‌ای در مطالعات اعتبار می‌تواند مفید باشد. بر پایه چنین روش‌هایی نتایج حاصل از برآورد اعتبار به فواصل مشخصی تعریف می‌شود (رایکوف و شراوت، 2002).

Archive of SID

www.SID.ir

علاوه بر روش بالا می‌توان اعتبار مرکب را از طریق تعریف پارامترهای جدید در یک مدل از قبل «برازش شده»[۱۷] در مدل‌ها محاسبه نمود. این‌کار از طریق تعریف «متغیر *خیالی*»[۱۸] امکان‌پذیر است. در این روش با محدود کردن ضرایب در نمرات مشاهده شده، نمره‌های مشاهده شده به مجموع نمرات خطا و نمره واقعی تجزیه می‌شوند. سپس این نمرات که X_i نامیده می‌شوند، با همدیگر ترکیب می‌گردند تا نمره متغیر خیالی (F) به‌وجود آید. مجذور همبستگی متغیر خیالی با متغیر نهفته اعتبار مرکب آزمون را نشان خواهد داد (رایکوف، 1997؛ گراهام، 2006). شکل نمادین ساخت متغیر خیالی در شکل 1 مشخص شده است.



شکل 1- چگونگی ساخت متغیر خیالی (منبع: رایکوف، 1997، 176).

می‌توان به‌جای محدود کردن پارامترها در عدد یک، از وزن‌های شناخته شده‌ای (همچون w_1، w_2، ...، w_k) برای هریک از مسیرها استفاده کرد. در انتخاب این وزن‌ها می‌توان یا از پژوهش‌های قبلی و یا از دانش نظری توسط صاحب‌نظران استفاده کرد.

Archive of SID

www.SID.ir

هم‌چنین با توجه به قابلیت عدم الزام برای تشخیص دقیق ضرایب مسیر، می‌توان از برآوردهای بار عاملی (λ_i) به عنوان وزن استفاده نمود (رایکوف، 1997). بنابراین با محدود کردن ضرایب نمرات مشاهده شده در $w_i=\lambda_i$ می‌توان متغیر خیالی را به‌وجود آورده و از طریق آن برآوردی از اعتبار مرکب وزن‌بندی شده را به‌دست آورد. به‌طور کلی کاربرد این روش با استفاده از نرم‌افزارهای آماری مدل معادلات ساختاری قابل انجام است. دراستفاده از این روش باید به‌خاطر داشت که از مدل‌هایی استفاده گردد که قبلاً ساختار آنان برازش یافته باشد.

- در نهایت، روش آخری که مورد بحث قرار می‌گیرد، ضریب «*اعتبار بیشینه‌ای*»[19] است که به عنوان شاخص اعتبار سازه تعریف می‌شود که می‌توان سازه نهفته را به‌صورت مقداری تعریف شده و یا «*لحاظ شده*» توسط شاخص‌های سازه درنظر گرفت (رایکوف و هنکوک، 2005). اعتبار بیشینه‌ای به عنوان آخرین شاخص برای برآورد اعتبار در روش مدل معادلات ساختاری پیشنهاد شده است. این شاخص از ترکیب مجموعه‌ی از پیش تعریف شده‌ای از مولفه‌های متجانس با عبارت‌های خطای ناهمبسته تشکیل شده است. این روش می‌تواند در ساخت، توسعه و تجدیدنظر مقیاس بهینه به‌کار آید. قابلیت مهم آن در سنجش تغییرات اعتبار بیشینه‌ای در نتیجه حذف و یا افزودن یک یا چند اندازه است. هم‌چنین با استفاده از ترکیب خطی وزن‌بندی شده می‌توان اعتبار مرکبی را به‌وجود آورد که دارای بالاترین درجه تفکیک‌پذیری در جنبه نهفته مورد نظر بوده و حداقل واریانس خطای نسبی ممکن را داشته باشد. با این ترکیب وزن‌بندی شده آماره رضایت‌بخشی از نمره توانایی نهفته به‌دست می‌آید. بررسی‌های تجربی نشان می‌دهند هنگامی‌که نمرات عامل با روش برآورد حداقل «*مجذورات عمومی شده*»[20] در یک مدل تک‌عاملی برآورد می‌شود، اعتبار بیشینه‌ای با روش متغیر خیالی (مجذور همبستگی بین عامل و ترکیب برآورد کننده‌ها) مساوی است (رایکوف و هنکوک، 2005).

- در روش «*اعتبار بیشینه‌ای*» تعیین وزن‌ها از اهمیت خاصی برخوردار است. روش‌های متفاوتی برای این‌کار پیشنهاد شده که در قبل به چند مورد از آن‌ها اشاره شد. در تکمیل این بحث دو روش دیگر نیز معرفی می‌شود که اولی به صورت نسبت بار

Archive of SID

www.SID.ir

عاملی به خطای آن است که می‌توان آن‌را به‌صورت زیر نوشت (کانجر، 1980، به نقل از رایکوف و هنکوک، 2005):

$$w_i^* = \frac{\beta_i}{\theta_i} (i = 1,2,...,m) \quad (12)$$

دومین روش که کاربرد زیادی در اعتبار بیشینه‌ای دارد به نسبت مجذور بارهای اندازه‌ها در جنبه مشترک (η_1) به خطای مرتبط با آن تعریف می‌گردد که به این صورت آن‌را می‌توان نشان داد:

$$w_i^* = \frac{b_i^{۲}}{q_i} (i = ۱,۲,...,m) \quad (13)$$

با ملاحظه چگونگی محاسبه وزن‌ها، ضریب اعتبار بیشینه‌ای جامعه (ρ_i^*) به این صورت محاسبه می‌شوند (رایکوف و هنکوک، 2005، 68):

$$\hat{\rho}^*_{(m)} = \frac{\sum_{i=1}^{m} (\hat{\beta}_i^{۲} / \hat{\theta}_i)}{۱ + \sum_{i=1}^{m} (\hat{\beta}_i^{۲} / \hat{\theta}_i)} \quad (14)$$

¨ آزمون تغییرات در اعتبار بیشینه‌ای به عنوان نتیجه‌ای از حذف و افزودن متغیرها یکی ازجنبه‌های مهم این روش است. بنابر این روش، اگر بخواهیم که ابزاری با بالاترین اعتبار بیشینه‌ای را از طریق مجموعه از پیش‌تعریف شده‌ای با m اندازه متجانس در اندازه معینی بسازیم (مثلاً k اندازه)، (m>k)، باید بالاترین نسبت‌های مجذور بار شاخص در واریانس خطای مرتبط $(\frac{\beta_i^{۲}}{\theta_i})$ را انتخاب کنیم. در نمونه‌های معین از جامعه مورد علاقه، ترتیب رتبه و انتخاب باید بر حسب نسبت‌های $\frac{\hat{b}_i}{\hat{q}_i}$ $(i = ۱,۲,...,m)$ از پارامترهای برآورد شده در هنگام برازش مدل با m مولفه انجام می‌شود. پس از انتخاب k مولفه، برآورد بیشینه‌ای مجموعه کوچک‌تر محاسبه می‌شود و سپس آزمون معنی‌داری تغییرات در اعتبار بیشینه‌ای انجام می‌شود. فرض صفر در این آزمون به این صورت است:

$$H_0 = r^*_{(m)} = r^*_{(k)} \quad (15)$$

آماره آزمون بر اساس $\Delta c^{۲}_{(m,k)} = c^{۲}_m - c^{۲}_k$ محاسبه شده که برای نمونه‌های بزرگ از توزیع خی‌دو مرکزی با m-k درجه آزادی پیروی می‌کند. اگر خی‌دو تفاوت معنی‌دار بود، فرض صفر رد می‌شود و نتیجه گرفته می‌شود که اعتبار بیشینه‌ای به عنوان نتیجه‌ای از حذف m-k مولفه از مجموعه اولیه m اندازه، کاهش خواهد یافت و یا با افزودن m-k

Archive of SID

www.SID.ir

اندازه در مجموعه اولیه k اندازه اعتبار افزایش خواهد یافت (رایکوف و هنکوک، 2005). ذکر این نکته لازم است چنان‌چه در اعتبار بیشینه‌ای فرض می‌شود هنگام حذف اندازه‌ها از مجموعه اولیه اندازه‌های متجانس، هیچ‌گاه اعتبار بیشینه‌ای بالاتری به‌دست نمی‌آید و تنها می‌توان بررسی کرد که آیا کاهش چندین اندازه می‌تواند از نظر آماری بر کاهش اعتبار بیشینه‌ای موثر باشد یا نه؟ برآورد نقطه‌ای میزان تغییرات اعتبار بیشینه‌ای نیز می‌تواند با استفاده از فرمول مربوطه و هم‌چنین متغیر خیالی به‌دست آید.

¨ در استفاده از «*اعتبار بیشینه‌ای*» توجه به ملاحظاتی لازم است؛ از جمله آن‌که تعداد اندازه‌ها در کوتاه‌ترین ابزار مورد بررسی باید به میزان k=2 و همراه با رعایت m>k باشد. هم‌چنین برای کاربرد این روش با نرم‌افزارهای موجود کنونی، باید مولفه‌های (تقریباً) پیوسته‌ای را مورد بررسی قرار داد، در غیر این‌صورت می‌توان از ماتریس همبستگی *پلی‌کوریک* و یا «*ساخت بسته*»[۲۱] استفاده نمود. علاوه‌براین، این روش بر اساس مولفه‌های متجانس و خطای اندازه‌گیری ناهمبسته استوار شده است. بنابراین، اگر حداقل یکی از این پیش‌فرض‌ها رعایت نشود، مدل توصیف شده ممکن است نتایج گمراه کننده‌ای را به‌بار آورده و مستلزم ترکیب ابزار «*زیربهینه‌ای*»[۲۲] باشد (رایکوف و هنکوک، 2005). بنابراین قبل از این‌که از این روش استفاده شود، باید مدل تک‌عاملی در همه مولفه‌ها برازش شده و برازش آن برای قابل دفاع بودن و مناسب بودن شاخص‌های اصلاح در کوواریانس خطا مشخص شود.

● روش

¡ مشارکت‌کنندگان 340 دانش آموز سال اول متوسطه شاغل به تحصیل در مدارس دولتی و غیرانتفاعی شهر تهران با روش نمونه گیری «احتمالات متناسب با حجم» (PPS)[23] انتخاب شدند. 160 نفر (47 درصد) از نمونه پسر و بقیه (180 نفر) دانش آموزان دختر بودند. محدوده سنی پاسخ‌گویان بین 13 تا 15 سال بود.

¡ **ابزار**«*مقیاس‌های تجدیدنظر شده‌ی نگرش ریاضی فنما و شرمن*»[۲۴] برای گردآوری نظرات دانش‌آموزان نسبت به ریاضیات مورد استفاده قرار گرفت. این مقیاس در سال 1976 ساخته شده و در سال 2001 مورد تجدیدنظر طراحان قرار گرفته است. این مقیاس یک ابزار سنجش مولفه عاطفی نگرش نسبت به ریاضی است. در مقیاس

Archive of SID

www.SID.ir

اصلی، نُه خرده آزمون وجود دارد که شامل: مقیاس نگرش در زمینه موفقیت ریاضی، مقیاس ریاضیات به عنوان یک حوزه مردانه، «*مقیاس مادر*»، «*مقیاس پدر*»، «*مقیاس اطمینان نسبت به توانایی خود در ریاضی*»، «*مقیاس اضطراب ریاضی*»، «*مقیاس انگیزش عاطفی در ریاضی*»، «*مقیاس سودمندی دریافت شده ریاضی*» و «*مقیاس ادراک از نگرش معلم*». از مجموع 9 مقیاس این ابزار، مقیاس «*ادراک از نگرش معلم*» برای بررسی در این مطالعه انتخاب شد. در این مقیاس 9 گویه وجود داشت. هر پاسخگو نظر خود را نسبت به هر یک از گویه‌ها بر اساس طیف پنج گزینه‌ای (از کاملاً مخالفم تا کاملاً موافقم) ابراز کرد. نمرات پاسخ‌گویان پس از هم‌جهت کردن گویه‌ها از طریق کدگذاری مجدد در محدوده 9 تا 45 قرار گرفت. نمره بالا هر پاسخگو، نگرش مثبت به ریاضی و نمره پائین نشان‌دهنده نگرش منفی نسبت به ریاضی بود.

● نتایج

قبل از تحلیل نتایج حاصل از اجرای مقیاس نگرش ریاضی، پیش‌فرض نرمال‌بودن مورد ارزیابی قرار گرفت. بدین منظور دو شاخص «چولگی» و «کشیدگی» به‌کار رفت. نتایج نشان داد که گویه‌ها دارای شاخص‌های قابل قبولی بوده و در نتیجه نرمال در نظر گرفته شدند. برای بررسی ساختار عاملی مقیاس از «*تحلیل عامل تائیدی*» استفاده شد. در جدول 1 آماره‌های توصیفی و ماتریس ضرایب همبستگی گویه‌ها و در جدول 2 پارامترهای مدل و شاخص‌های مربوط به محاسبات اعتبار ارائه شده است.

جدول 1- آماره‌های توصیفی و ماتریس ضرایب همبستگی گویه‌ها

گویه	میانگین	انحراف معیار	1	2	3	4	5	6	7	8
1	3/766	1/024	1							
2	2/988	1/399	0/186**	1						
3	3/22	1/277	0/271**	0/117*	1					
4	3/439	1/269	0/285**	0/137*	0/220**	1				
5	3/516	1/118	0/363**	0/232**	0/284**	0/311**	1			
6	3/277	1/287	0/403**	0/265**	0/261**	0/323**	0/418**	1		
7	3/405	1/157	0/232**	0/141**	0/077	0/126*	0/184**	0/272**	1	
8	3/383	1/202	0/333**	0/209**	0/297**	0/272**	0/392**	0/554**	0/188**	1
9	3/59	1/127	0/370**	0/223**	0/242**	0/227**	0/415**	0/312**	0/314**	0/366**

Archive of SID

www.SID.ir

جدول 2- پارامترهای مدل و شاخص‌های مربوط به محاسبات اعتبار

گویه	ضریب معیار شده مسیر (λ)	خطای اندازه‌گیری (θ)	ضرایب مجذور همبستگی‌های چندگانه (SMC)	λ^2/θ	θ^2
1	0/578	0/666	0/33	0/502	0/444
2	0/347	0/88	0/12	0/127	0/774
3	0/414	0/829	0/17	0/207	0/687
4	0/446	0/801	0/20	0/248	0/641
5	0/628	0/605	0/39	0/652	0/366
6	0/704	0/504	0/50	0/983	0/254
7	0/348	0/879	0/12	0/138	0/773
8	0/657	0/568	0/43	0/76	0/323
9	0/567	0/678	0/32	0/474	0/46

برازش مدل اندازه‌گیری با توجه به شاخص‌های $\chi2 = 43/15$ با 27 درجه آزادی، RMSEA = 0/042 با فواصل اطمینان (0/065، 0/015)، CFA =0/97، SRMR= 0/036، GFA = 0/97 و AGFA =0/95 مورد تائید قرار گرفت. برای بررسی انطباق این مدل با مدل‌های محدودتر (مدل معادل-تاو)، بارهای عاملی مدل در تساوی با همدیگر محدود شدند. با ایجاد این محدودیت $\chi2 = 266/51$، و df = 65 به‌دست آمد. با بررسی میزان تفاوت در $\chi2$ و درجه آزادی مدل، $\chi\Delta2 = 266/51 - 43/15 = 223/36$ و $\Delta df = 95 - 27 = 38$، نتیجه گرفته می‌شود که ایجاد این محدودیت تأثیر نامطلوبی بر روی برازش مدل داشته و در نتیجه مدل فوق به عنوان مدل متجانس درنظر گرفته می‌شود.

ضرایب به‌دست آمده از مدل همگی دارای مقادیر t معنی‌داری هستند. در ستون ضرایب مجذور همبستگی‌های چندگانه، شاخص‌های اعتبار گویه‌ها مورد مقایسه قرار گرفته‌اند. این مقایسه نشان می‌دهد که گویه‌های 6، 8 و5 بالاترین اعتبارها و گویه‌های 2، 7 و 3 اعتبارهای پائین‌تری نسبت به بقیه دارند. در ستون بعد نسبت مجذور بار عاملی به خطای اندازه‌گیری مشخص شده است. در این ستون نیز ملاحظه می‌شود که همین گویه‌ها بالاترین و پائین‌ترین نسبت‌ها را در میان گویه‌های دیگر در اختیار دارند. برای محاسبه اعتبار مرکب با استفاده از فرمول (14) اندازه‌های زیر به‌دست می‌آید:

$$\text{اعتبارمرکب}(14) = \frac{(۰/۵۷۸+۰/۳۴۷+...+۰/۶۵۷+۰/۵۶۷)^۲}{(۰/۵۷۸+۰/۳۴۷+...+۰/۶۵۷+۰/۵۶۷)^۲+(۰/۶۶۶^۲+۰/۸۸^۲+...+۰/۶۵۸^۲+۰/۶۷۸^۲)}$$

Archive of SID

www.SID.ir

$$= \frac{۲۱/۹۸۷}{۲۱/۹۸۷ + ۴/۷۲۲} = \frac{۲۱/۹۸۷}{۲۶/۷۰۹} = ۰/۸۲۱$$

بر اساس محاسبات، شاخص اعتبار مرکب این عامل به میزان 0/86 به‌دست آمد. علاوه‌براین، با ایجاد متغیر خیالی در نرم‌افزار LISREL، همبستگی متغیر خیالی با متغیر نهفته به میزان 0/877 به‌دست آمد که مجذور آن، 0/769 اعتبار مقیاس را نشان می‌دهد. علاوه بر محاسبات فوق، اعتبار بیشینه‌ای با استفاده از فرمول (14) محاسبه شد. شاخص‌های نسبت مجذور بار عاملی در خطای اندازه‌گیری از ستون مربوطه در جدول مورد استفاده قرار می‌گیرند:

$$\text{اعتبار بیشینه‌ای} = \frac{(۰/۵۰۲ + ۰/۱۳۷ + ... + ۰/۷۶ + ۰/۴۷۴)}{۱ + (۰/۵۰۲ + ۰/۱۳۷ + ... + ۰/۷۶ + ۰/۴۷۴)} = \frac{۴/۱۰۱}{۱ + ۴/۱۰۱} = ۰/۸۰۴$$

از عبارت بالا مشخص می‌گردد که اعتبار بیشینه‌ای این مقیاس 0/804 است. در نهایت ضریب آلفای کرونباخ این مقیاس به‌دست می آید که برابر با 0/763 است. مقایسه برآوردهای به‌دست آمده با استفاده مدل ساختاری (اعتبار مرکب = 0/821؛ اعتبار بیشینه‌ای = 0/804؛ اعتبار با استفاده از متغیر خیالی = 0/769) با ضریب آلفا (0/763) نشان می‌دهد که استفاده از روش‌های مورد اشاره افزایش قابل ملاحظه‌ای در ضرایب اعتبار به‌وجود می‌آورد.

هم‌چنین به عنوان عملکرد دیگری از «*اعتبار بیشینه‌ای*»، آزمون تغییرات بر اثر کاهش چندین گویه که دارای نسبت کم‌تری از مجذور بار عاملی به خطای اندازه‌گیری بودند، مورد بررسی قرار گرفت. در این تحلیل اثرات حذف سه گویه 2، 7 و3 که دارای کم‌ترین نسبت‌مورد بحث بودند، بررسی شد. فرم کوتاه شده مقیاس (k)، مورد برازش مجدد قرار گرفته و خی‌دو برابر با 25/03 با 9 درجه آزادی به‌دست آمد. این مدل آشیانه با مدل اصلی (m) مقایسه گردید. آماره برازش تغییر، 18/12= 25/03– 43/15= $\Delta c^2 = c^2_m - c^2_k$ و $\Delta df = df_m - df_k = 27 - 9 = 18$ نشان می‌دهد که فرم کوتاه شده در زمینه میزان اعتبار تفاوت معنی‌داری با فرم اصلی نمی‌کند. با وجود آن‌که اعتبار بیشینه‌ای فرم کوتاه شده به میزان 0/784 به‌دست می‌آید و این ضریب، کاهشی برابر با 0/020 در اعتبار مقیاس به‌وجود می‌آورد، با استناد به عدم معنی‌داری تفاوت بین دو فرم کوتاه شده و اصلی می‌توان نتیجه گرفت که کاهش در اعتبار فرم کوتاه شده، کاهش

Archive of SID

www.SID.ir

معنی‌داری به‌شمار نمی‌رود. تذکر مجدد این نکته لازم است که کوتاه کردن مقیاس و حذف کردن گویه‌ها به کاهش در اعتبار بیشینه‌ای منجر می‌شود و در آزمون تغییرات، این موضوع بررسی می‌شود که آیا کوتاه کردن مقیاس از نظر آماری به کاهش معنی‌داری در اعتبار مقیاس می‌انجامد و یا تفاوت حاصله قابل چشم‌پوشی است؟

● بحث و نتیجه‌گیری

i در این پژوهش با تمرکز بر موضوع اعتبار، قابلیت‌ها و محدودیت‌های روش آلفا و روش‌های جایگزین به‌ویژه روش‌های برگرفته از مدل معادلات ساختاری مورد بررسی قرار گرفت. استفاده از آلفا به علت پیروی از پیش‌فرض‌های محدودکننده، باعث برآورد شاخص‌های غیر دقیقی می‌شود. در مقابل روش‌های جایگزینِ معرفی شده، دارای قابلیت غلبه بر مشکلات روش آلفا را دارند. این روش‌ها در مورد مقیاس «ادراک از نگرش معلم» از مجموعه مقیاس‌های نگرش ریاضی فنما و شرمن مورد محاسبه قرار گرفته و با روش آلفا مقایسه گردید.

i روش سنتی برای برآورد اعتبار مجموعه گویه‌ها، روش آلفا است که توسط کرونباخ معرفی شده است. علی‌رغم سادگی و رواج استفاده از این روش، ملاحظاتی در کاربرد آن وجود دارد که بی‌توجهی به آن‌ها باعث بروز برآوردهای نادقیقی می‌گردد. از جمله این ملاحظات توجه به پیش‌فرض‌های اساسی این روش است که شامل برابری با مدل اساساً معادل-تاو و عدم همبستگی بین خطاهای اندازه‌گیری گویه‌ها می‌باشند. عدم رعایت پیش‌فرض برابری با «*مدل اساساً معادل-تاو*» باعث کاهش اعتبار و وجود همبستگی در بین خطاهای اندازه‌گیری باعث تورم در مقدار اعتبار می‌گردد. به‌طور کلی، بی‌توجهی به این پیش‌فرض‌ها به برآوردهای غیردقیقی منجر می‌گردد. روش‌های موجود در مدل معادلات ساختاری از چندین جنبه مزایایی نسبت به روش آلفا دارند که از آن جمله می‌توان به امکان آزمون کردن پیش‌فرض‌ها، عدم نیاز به پیش‌فرض برابری با مدل اساساً معادل-تاو و قابلیت محاسبه در مدل متجانس، امکان برآورد نااعتبار در صورت همبستگی خطای اندازه‌گیری و امکان استفاده از روش وزن‌بندی گویه‌ها در محاسبه اعتبار اشاره نمود. مجموعه روش‌های توصیف شده در مدل معادلات ساختاری می‌توانند به‌صورت بررسی تک‌تک گویه‌ها (از طریق مجذور همبستگی چندگانه) و یا

Archive of SID

www.SID.ir

روش‌های ترکیب چندین گویه انجام شوند. روش‌های ترکیبی شامل اعتبار مرکب (شامل موارد مربوط به گویه‌های وزن‌بندی شده و وزن‌بندی نشده و با محاسبه و بدون محاسبه همبستگی بین خطاهای اندازه‌گیری)، استفاده از متغیر خیالی و اعتبار بیشینه‌ای می‌گردد.

برای ملاحظه قابلیت عملی این روش‌ها و ارائه مثال کاربردی، مجموعه روش‌های توصیف شده بر روی داده‌های واقعی مورد آزمایش قرار گرفت. نتایج برآورد اعتبار مقیاس را با استفاده از روش‌های اعتبار مرکب به میزان 0/821، متغیر خیالی به میزان 0/769 و اعتبار بیشینه‌ای به میزان 0/804 نشان داد. مقایسه این مقادیر با مقدار محاسبه شده از روش آلفا (0/763) نشان داد که استفاده از روش‌های مذکور مقادیر بالاتری از اعتبار مقیاس را به‌دست می‌دهد.

وجود تفاوت بین مقادیر حاصل از روش آلفا و روش‌های برگرفته از مدل معادلات ساختاری به علت منطبق نبودن مقیاس مورد بررسی با مدل معادل-تاو بود. در بررسی انطباق با مدل‌های اندازه‌گیری مشخص شد که مقیاس با مدل متجانس منطبق بود ولی نمی‌توانست با مدل معادل-تاو انطباق داشته باشد.

در مباحث پیشینه بیان شد که عدم انطباق مقیاس با مدل معادل- تاو باعث کم‌برآوردی در مقدار اعتبار می‌گردد. این موضوع در مقایسه بین روش آلفا و روش‌های مدل معادلات ساختاری قابل بررسی است. با این‌حال، استفاده از مقیاس 9 گویه‌ای به بهتر شدن برآوردهای اعتبار با استفاده از روش آلفا منجر می‌گردد. برطبق توصیه‌های رایکوف (2001؛ 2004) هنگامی‌که تعداد زیادی گویه (بیشتر از 6 گویه) در مقیاس وجود داشته باشد و در این گویه‌ها بین خطای اندازه‌گیری همبستگی وجود نداشته باشد، برآورد اعتبار با استفاده از روش آلفا می‌تواند قابل توجیه باشد. بنابراین اندازه مناسب مقیاس باعث شده است که کم‌برآوردی ضریب آلفا به علت عدم انطباق با مدل معادل-تاو پوشش داده شده و بسیار زیاد نگردد.

با توجه به دشواری بررسی و برآورد پیش‌فرض‌های محدودکننده روش آلفا پیشنهاد می‌گردد که از روش‌های برگرفته از مدل معادلات ساختاری استفاده گردد. در این روش‌ها علاوه بر بررسی دقیق پیش‌فرض‌های لازم در سایر روش‌ها، پیش‌فرض‌های کم‌تری دربرگرفته می‌شود و به برآوردهای ناارریبی از اعتبار منجر

Archive of SID

www.SID.ir

می‌شود. هم‌چنین قابلیت‌های دیگری چون مقایسه بین فرم کوتاه شده و فرم اصلی، بررسی‌های بین گروهی و موارد مشابه نیز از جمله مزایای استفاده از این روش‌هاست.

i i i

یادداشت‌ها

1 .essentially τ-equivalence
2 . underestimated
3 . parallel model
4. congeneric model
5. additive
6 . nested models
7 . optimal
8 . over-estimated
9 . statistical independence
10. pure
11 . weighted least square
12 . overidentified
13. just identified
14. saturated model
15. squared multiple coefficient (SMC)
16. composite Reliability
17. fitted
18. phantom
19. maximal reliability
20. generalized least square
21. parceling
22. suboptimal
23. probability proportional to size
24. Fennema and Sherman's Math Attitude Scales

منابع

- آناستازی، ا. (1379). *روان‌آزمایی*. ترجمه محمد نقی براهنی. تهران: انتشارات دانشگاه.
- قاضی طباطبایی، م. (1377). ارزیابی سازه‌ای: نخستین گام ضرور در مطالعات بین فرهنگی (مورد پرسشنامه گیبسون و دمبو). *نامهٔ علوم اجتماعی*، 12، 135-107.

Abedi, J. (2002). A latent-variable modeling approach to assessing reliability of a creativity instrument. *Creativity Research Journal, 14* (2), 267-276.

Adamson, G., Shevlin, M., Lloyd, N. S. V., and Lewis, C. A. (2000). An integrated approach for assessing reliability and validity: An application of structural equation modeling to the measurement of religiosity. *Personality and Individual Difference, 29*, 971-979.

Bacon, D. R., Sauer P. L., & Young. M. (1995). Composite reliability in structural equations modeling. *Educatioal and Psychological Measurement*, 55(3), 394-406.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educatioal and Psychological Measurement*, 66(6), 930-944.

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, 7(2), 251-270.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago: Lawrence Erlbaum Associates Inc.

Archive of SID

www.SID.ir

Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated errors on coefficient alpha. *Applied Psychological Measurement, 21*(4), 337–348.

Liu, Y., & Zumbo, B. D. (2007). The impact of outliers on cronbach's coefficient alpha estimate of reliability: Visual analogue scales. *Educational and Psychological Measurement, 67*(4), 620-634.

Lucke, J. F. (2005). "Rassling the hog": The influence of correlated item error on internal consistency, classical reliability, and congeneric reliability. *Applied Psychological Measurement, 29*(2), 106-125.

Raines-Eudy, R. (2000). Using structural equation modeling to test for differential reliability and validity: An empirical demonstration. *Structural Equation Modeling*, 7(1), 124-141.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173-184.

Raykov, T. (1998a). A method for obtaining standard errors and confidence intervals of composite reliability with congeneric items. *Applied Psychological Measurement*, 22(4), 369–374.

Raykov, T. (1998b). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*(4), 375–385.

Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*(1), 69–76.

Raykov, T. (2004). Point and interval estimation of reliability for multiple-component measuring instruments via linear constraint covariance structure modeling. *Structural Equation Modeling, 11*(3), 342-356.

Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology, 58*, 65-82.

Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and internal estimation using a structural equation modeling approach. *Structural Equation Modeling, 9*(2), 195-212.

Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating ωh for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement, 31*(2), 135–157.

i i i

Archive of SID

www.SID.ir