مجله علوم رفتاری
دوره ۳ ، شماره ۲ ، تابستان ۱۳۸۸
صفحات: ۱۳٤-۱۲۷

Journal of Behavioral Sciences
Vol. 3, No. 2, Summer 2009
Pages: 127-134

رواسازی نسخه ایرانی پرسش‌نامه ارزیابی دانشجویان از کیفیت تدریس

Validation of the Iranian version of student's evaluation of educational quality questionnaire

تاریخ دریافت: ۸۷/۱۰/٤ تاریخ پذیرش: ۸۸/۲/۲۰

سیاوش طالع‌پسند✉، مرتضی نظیفی[۱]، ایمان‌اله بیگدلی[۲]

Tale'pasand S. *PhD*✉, Nazifi M. *MSc*, Bigdeli I. *PhD*

چکیده

مقدمه: هدف این مطالعه رواسازی نسخه ایرانی پرسش‌نامه ارزیابی دانشجویان از کیفیت تدریس و آزمون مقدماتی الگوی *مارش* در ارزشیابی کیفیت تدریس بود.

روش: در کل ۳۳۹ دانشجوی دانشگاه سمنان در این مطالعه شرکت داشتند. شرکت‌کنندگان پرسش‌نامه ارزیابی دانشجویان از کیفیت تدریس را تکمیل نمودند و نمرات امتحان پایان‌ترم آنها جمع‌آوری شد. تحلیل عاملی اکتشافی، همبستگی سئوال- نمره‌ی کل و تحلیل اعتبار برای بررسی ویژگی‌های روان‌سنجی ابزار اجرا شد. تحلیل عاملی تاییدی، به‌منظور فراهم آوردن شواهد روایی اجرا شد.

یافته‌ها: تحلیل عاملی اکتشافی، مدلی ۶عاملی را آشکار کرد. درحالی‌که مدل ۶عاملی با روش تحلیل عاملی تاییدی برازنده نبود. هماهنگی درونی پرسش‌نامه ارزیابی دانشجویان از کیفیت تدریس بالا بود (α=۰/۹۳۶). نمرات ارزیابی از کیفیت تدریس با پیشرفت تحصیلی دانشجویان به‌طور معنی‌دار مرتبط بود (r=۰/۵۰).

نتیجه‌گیری: نتایج تحلیل عاملی اکتشافی و تاییدی، ساختار ۹عاملی *مارش* را تایید نمی‌کند، اما با این دیدگاه *مارش* که کیفیت تدریس سازه‌ای چندبُعدی است، همسو است. بر پایه نتایج، استفاده از این ابزار برای ارزیابی کیفیت تدریس اساتید و تفسیر آن باید با احتیاط همراه باشد.

کلیدواژه‌ها: کیفیت تدریس، ارزیابی، آموزش عالی، روایی سازه، روایی پیش‌بینی

Abstract

Introduction: The objective of this study was to validate the Iranian version of students' evaluations of educational quality questionnaire (PSEEQ) and to initially tested *Marsh* model in the context of Iranian students.

Method: A total of 339 Iranian students of Semnan University participated in the study. Participants completed an Iranian version of PSEEQ and to be collected the scores of their final exam. To assess the psychometric properties of PSEEQ, exploratory factor analysis, item-total correlation and reliability analysis were undertaken. Confirmatory factor analysis used to provide further validation for the questionnaire.

Results: Exploratory factor analyses showed a six-factor model of the Iranian version of PSEEQ. The fit of the proposed models examined. The internal consistency of PSEEQ (31 items) was high (Cronbach's α=0.936). PSEEQ correlated with the scores of final exam (r=0.50).

Conclusion: The results of confirmatory factor analysis do not support the nine-factor framework of the SEEQ with Iranian students on the basis of *Marsh* model, but it supports that educational quality is a multidimensional construct, as *Marsh* stated. Taken together, the data suggest that PSEEQ should be interpreted with caution when used to assess educational quality of instructors.

Keywords Educational Quality, Evaluations, Higher Education, Construct Validity, Prediction Validity

✉گروه روان‌شناسی تربیتی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه سمنان، سمنان، ایران
[۱] گروه روان‌شناسی، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران
[۲] گروه روان‌شناسی بالینی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه سمنان، سمنان، ایران

✉**Corresponding Author:** Department of Educational Psychology, Faculty of Psychology & Educational Sciences, Semnan University, Semnan, Iran
Email: stalepasand@semnan.ac.ir

مقدمه

ارزیابی دانشجویان از کارآیی تدریس، روش بسیار رایجی در ارزیابی تدریس است. درباره ارزیابی‌های دانشجویان از کارآیی تدریس تاکنون هزاران پژوهش انجام شده است که اوج این پژوهش‌ها در دهه‌های هفتاد و هشتاد میلادی بوده است [۱]. *هوگوتو برنارد* از دهه هفتاد به بعد، سه دوره را در پژوهش‌های مرتبط با ارزیابی‌های دانشجویان شناسایی کرده است: الف) در سال‌های ۸۰-۱۹۷۰ کانون اصلی پژوهش، ارزیابی تکوینی بود و بررسی‌ها بر نشانگرهای کارآیی تدریس تمرکز داشتند؛ ب) در سال‌های ۱۹۹۰-۱۹۸۰ کانون اصلی پژوهش، بیشتر استفاده‌های اجرایی از ارزیابی بود و پژوهش‌ها، روایی درجه‌بندی‌های دانشجویان را بررسی می‌نمودند و نیز به دنبال روش‌های متنوع و چندگانه برای به عمل‌آوردن ارزیابی جامع بودند؛ و ج) در سال‌های ۲۰۰۰-۱۹۹۰ اندیشه سوابق تدریس غلبه یافته و پژوهش‌ها بر تاثیر و کارآیی ارزیابی‌های دانشجویان تمرکز داشته است [۱].

علیرغم کاربرد گسترده درجه‌بندی‌های دانشجویان از کیفیت تدریس، به نظر برخی از استادان این درجه‌بندی نوعی کمّی‌سازی بی‌معنی است. آنها نگرانند که مبادا دانشجویان بیش از حد از نیروی قلم خود در جهت برکنار کردن استادان استفاده کنند و هشدار می‌دهند که ممکن است موضوع ارزیابی کارآیی تدریس، تبدیل به نوعی طغیان عوام گردد [۲]. در واقع، اساتید عموما معتقدند که برخی از عوامل که به عقیده‌ی آنها ربطی به کارآیی تدریس ندارد، ارزیابی‌های دانشجویان را جهت‌دار می‌سازد. در پیمایشی که از اعضای هیات‌علمی به‌عمل آمد، فهرستی از خصوصیات ارایه شد که باعث جهت‌داری درجه‌بندی‌های دانشجویان می‌گردید: ۷۲٪ دشواری درس را عنوان کردند، ۶۸٪ آسانی نمره‌دهی، ۶۳٪ محبوبیت استاد، ۶۲٪ علاقه قبلی دانشجو به مطلب پیش از شروع درس، ۶۰٪ فشار کاری درس، ۶۰٪ اندازه کلاس و ۵۵٪ دلایل انتخاب آن درس را مطرح نمودند [۳].

برخی دیگر از متغیرهای بالقوه تورش‌زا عبارتند از: سطح درس (کارشناسی یا کارشناسی‌ارشد)، نوع درس (اختیاری یا الزامی)، رشته تحصیلی دانشجو (مثلا، علوم انسانی و علوم اجتماعی در برابر علوم فیزیک و مهندسی)، خصوصیات جمعیت‌شناختی (مانند، جنسیت استاد و دانشجو)، سال‌های تدریس به طور کلی و سال‌های تدریس دروس خاص، بیان هیجانی استاد، هدف دانشجویان از ارزیابی (برای مثال، زمانی که دانشجویان می‌دانند که ارزیابی‌های آنها به منظور تصمیم‌گیری در زمینه حفظ یا ارتقای اساتید به کار می‌رود آنها آشکارا درجه‌بندی‌های بالاتری به اساتید می‌دهند) و وجود مسایلی در اجرای پرسش‌نامه‌های ارزیابی دانشجویان [۴].

آنچه تا به حال گفته شد به گستره اعتبار و روایی پرسش‌نامه‌های ارزیابی تدریس مرتبط بود. پس از نشان‌دادن اعتبار درجه‌بندی‌های دانشجویان، میزان روایی این درجه‌بندی‌ها سنجیده شد. برای نشان‌دادن روایی درجه‌بندی‌های دانشجویان بایستی آنها را از یک سو با نمرات کارآیی تدریس و از سوی دیگر با عوامل تورش‌زای احتمالی همبسته ساخت. چنانچه این درجه‌بندی‌ها روا باشند، دانشجویان درجه‌بندی‌های خوب را به استادان کارآمد و درجه‌بندی‌های ضعیف را به استادان ناکارآمد خواهند داد و درجه‌بندی‌های آنها همبستگی معنی‌داری با متغیرهای تورش‌زای احتمالی نشان نخواهد داد. به این ترتیب، اندازه همبستگی بین درجه‌بندی‌ها و کارآیی تدریس، شاخص دقیقی از روایی درجه‌بندی‌های دانشجویان در اختیار خواهد گذاشت.

در خصوص متغیرهای تورش‌زا، مرور اساسی بر متون گذشته نشان داده است که متغیرهای تورش‌زای احتمالی در مجموع فقط ۵ تا ۱۵٪ از واریانس درجه‌بندی‌های دانشجویان را تبیین می‌کنند [۴]. پس اگر قرار است در پژوهش‌های رواسازی از نمرات کارآیی تدریس استفاده شود، کدام اندازه‌ها می‌توانند به عنوان ملاک کارآیی تدریس عمل کنند؟ پژوهشگران مدت مدیدی است در مورد ملاک‌های کاملی از این سازه کاوش می‌کنند. دسته‌ای از ملاک‌ها که آنها مورد بررسی قرار داده‌اند عبارتند از: میزان یادگیری دانشجویان، درجه‌بندی‌های دانشجویان قبلی، مشاهده‌های کلاسی از سوی متخصصان، خودارزیابی‌های اساتید، ارزیابی همتایان یا همکاران، اثرات دستکاری‌های آزمایشی همچون مداخله‌های بازخوردی و تغییرات در رفتارهای دانشجویان [۲، ۵، ۶، ۷، ۸، ۹، ۱۰]. پژوهش‌های مرتبط با این ملاک‌ها نشان داده‌اند که درجه‌بندی‌های دانشجویان از دروس به‌طور معنی‌دار و قابل‌توجهی با پیشرفت/یادگیری دانشجویان و با درجه‌بندی‌های دانشجویان قبلی رابطه دارد [۱]. همچنین، در پژوهشی که به‌صورت دستکاری آزمایشی انجام شد، نشان داده شد که شیوایی گفتار استاد از روایی درجه‌بندی‌های دانشجویان پشتیبانی می‌کند [۱]. در این مطالعه بین رفتارهای روشن‌ساز خاص و درجه‌بندی‌های دانشجویان، همبستگی بسیار بالایی به‌دست آمد. به‌علاوه، پژوهش‌ها نشان دادند زمانی که استادان از درجه‌بندی‌های دانشجویان بازخورد دریافت کرده و در بحثی عاری از تهدید درباره ضعف‌های تدریس خود با مشاور صحبت می‌کردند، درجه‌بندی‌های دانشجویان و افزایش یادگیری آنها بهبود می‌یافت [۳، ۱۱]. در مداخله‌ای بازخوردی با هدف بررسی بُعد خاصی از درجه‌بندی‌های دانشجویان، نشان داده شد که کارآیی تدریس به‌طور کلی افزایش می‌یابد ولی بیشترین تاثیر روی همان ابعاد خاص از درجه‌بندی‌ها است که هدف مداخله قرار گرفته بودند [۱۱]. شماری از مطالعات حاکی از آن است که کیفیت تدریس، سازه‌ای چندبعدی است. بنابراین طراحی ابزارهایی که قرار است کارآیی تدریس را بسنجند بایستی به گونه‌ای باشد که این ابعاد چندگانه را نشان دهد. زمانی که چنین ابزاری طراحی و به کار گرفته شود، پژوهش‌ها نشان می‌دهند که دانشجویان بین ابعاد مختلف تدریس تمایز قایل می‌شوند. تحلیل عاملی این‌گونه پرسش‌نامه‌ها چندین بُعد

را آشکار می‌سازد [۱]. مطالعات نشان دادند که درجه‌بندی‌های دانشجویان به‌طور نامساعد تحت تاثیر ابعاد همپوش قرار نمی‌گیرند. به‌عبارت دیگر، خصوصیات استاد (مانند صمیمیت، رفتار دوستانه، پذیرا بودن) بر فقدان سازماندهی یا استفاده از راهبردهای انگیزشی ضعیف توسط همان استاد تاثیری ندارد [۱۲، ۱۳].

پرسش‌نامه ارزیابی دانشجویان از کیفیت تدریس، یکی از پرسش‌نامه‌هایی است که به خوبی طراحی شده و دارای ویژگی‌های روان‌سنجی مطلوبی است. این پرسش‌نامه توسط *مارش* ساخته شده و در طول سال‌ها به کوشش خود وی مورد بازنگری و اصلاح قرار گرفته [۱۴] و ضریب اعتبار بالایی (دامنه آلفا ۰/۸۸ تا ۰/۹۷) نشان داده است [۱۵]. به گفته *مارش* در چندین پژوهش بزرگ، تاثیر هم‌نهاده بسیاری از تورش‌های احتمالی نتوانست بیش از ۱۵٪ واریانس درجه‌بندی‌های پرسش‌نامه ارزیابی‌های دانشجویان از کیفیت تدریس را تبیین کند. هر چند که روابط متوسطی بین پاسخ‌های ابزار مذکور و تعداد اندکی از متغیرهای زمینه‌ای (مانند، علاقه قبلی به موضوع، نمره مورد انتظار، فشارکاری/سختگیری، اندازه کلاس، جنسیت استاد و دانشجو) وجود داشت ولی بررسی دقیق ماهیت این اثرها و معادل‌های همین روابط که بر اساس خود ارزیابی‌های استاد از تدریس خویش به دست آمدند، نشان داد که آنها نمی‌توانند تورش ایجاد کنند [۱۶].

پاسخ‌های SEEQ در پژوهش‌های چندبخشی رواسازی با موفقیت در رابطه با یادگیری [۱۷، ۱۸] درجه‌بندی‌های دانشجویان قبلی [۱۸، ۱۹]، خود ارزیابی‌های استاد از کارآیی تدریس خویش [۲۰، ۲۱]، مداخله بازخوردی که مقیاس‌های خاصی از SEEQ را هدف قرار می‌دادند [۱۱] و در رابطه با انواعی از ملاک‌ها [۱۳] رواسازی شده‌اند. بنابراین، درجه‌بندی‌های SEEQ اصولاً تابعی است از استادی که درسی را تدریس می‌کند و نه درسی که مورد ارزیابی قرار می‌گیرد [۲۲، ۲۳].

در خصوص روایی سازه، *مارش* اظهار می‌کند که بیش از ۳۰ پژوهش تحلیل عاملی بر روی پاسخ‌های SEEQ انجام شده است که عوامل زیربنایی آن را شناسایی کرده‌اند [۱۱، ۱۳، ۲۰، ۲۴، ۲۵، ۲۶]. پژوهش‌های تحلیل عاملی روی پاسخ‌های به‌دست‌آمده از ۵۰/۰۰۰ کلاس (که معرف نزدیک به یک میلیون پرسش‌نامه SEEQ بودند) پشتیبانی روشنی برای ساختار عاملی SEEQ فراهم نمود [۲۶]. در تحلیل‌های جداگانه بر روی پاسخ‌های به‌دست‌آمده از ۲۱ گروه متفاوت که معرف سطوح مختلف آموزش (مانند، دروس مربوط به سطح کارشناسی و کارشناسی‌ارشد) و انواع رشته‌های تحصیلی بودند، مجموعه یکسانی از عامل‌های SEEQ شناسایی گردید. همچنین، زمانی که اساتید با همان پرسش‌نامه‌ای که دانشجویان تکمیل کرده بودند، کارآیی تدریس خویش را ارزیابی می‌کردند، تحلیل عامل بر روی درجه‌بندی‌های دانشجویان و بر روی ارزیابی‌های خود اساتید، هر دو عامل‌های یکسانی را در SEEQ شناسایی نمودند [۲۰، ۲۱]. در مطالعه‌ای برای یک نسل از اساتید متشکل از ۲۲۱ استاد که در طول یک دوره ۱۳ ساله به طور منظم با SEEQ مورد ارزیابی قرار گرفته بودند (به طور متوسط ۲۵ مجموعه درجه‌بندی برای هر استاد) نیمرخ پاسخ‌های SEEQ بررسی شدند [۲۷]. نتایج نشان داد که نه تنها درجه‌بندی‌های مربوط به مقیاس‌های مجزای SEEQ در طول زمان ثابت است [۲۸] بلکه نیمرخ‌های چند بعدی این درجه‌بندی‌ها نیز در طول زمان تغییر نمی‌کنند. نیمرخ مربوط به هر استاد (مثلاً در علاقه‌مندی تدریس بالا ولی در سازماندهی پایین) با نیمرخ‌های مربوط به اساتید دیگر تفاوت داشت و همان نیمرخ بر زمان و سطح درس قابل تعمیم بود. این پژوهش‌ها تعمیم‌پذیری گسترده عامل‌های SEEQ در طی زمان، رشته‌های تحصیلی و پاسخ‌های مربوط به دانشجویان و اساتید را نشان می‌دادند.

هدف پژوهش حاضر، بررسی ویژگی‌های روان‌سنجی نسخه ترجمه‌شده SEEQ است. از این جهت که آیا این ابزار با توجه به تفاوت‌های زبانی و فرهنگی دانشجویان ایرانی می‌تواند برای ارزیابی استادان دانشگاه‌های ایران مناسب باشد؟ هدف دیگر این مطالعه بررسی روایی پیش‌بین آن از طریق همبسته ساختن پاسخ دانشجویان به SEEQ با نمرات امتحان پایان ترم آنها است.

روش

کلیه دانشجویان دانشگاه سمنان در سال تحصیلی ۸۸-۸۷ جامعه آماری (N=۶۱۵۰) این پژوهش را تشکیل می‌دهند. نمونه‌ای به حجم ۳۳۹ دانشجو از ۵ دانشکده دانشگاه سمنان، SEEQ را تکمیل نمودند. دانشجویان به طور داوطلبانه و بی‌نام، سه هفته قبل از شروع امتحان پایان ترم ابزار را تکمیل نمودند. پرسش‌نامه ارزیابی، ابزار اندازه‌گیری دانشجویان از کیفیت تدریس بود. این پرسش‌نامه ۴۱ سوال دارد که به منظور ارزیابی کیفیت تدریس طراحی شده است. ۳۱ سوال این ابزار ۹ بُعد تدریس کارآمد و ۱۰ سوال ویژگی‌های دانشجویان و درس را به عنوان متغیرهای بالقوه تورش‌دار می‌سنجد. این ابزار ویژگی‌های روان‌سنجی بسیار مطلوبی در مطالعات قبلی نشان داده است. گویه‌های این ابزار در یک طیف لیکرت پنج درجه‌ای (خیلی خوب=۵ تا خیلی ضعیف=۱) درجه‌بندی می‌شوند. نمره هر بُعد با جمع نمرات گویه‌های آن و حذف اثر تعداد گویه به دست می‌آید. دامنه نمره هر بُعد بین ۱ تا ۵ است.

از دانشجویان خواسته شد ابزار را در مورد یکی از بهترین استادان خود تکمیل نمایند. همچنین، از آنجا که برای بررسی روایی پیش‌بین، نمرات دانشجویان در پایان ترم باید جمع‌آوری می‌شد، تعداد ۹۲ دانشجو اعلام همکاری نمودند. از آنها خواسته شد که نام معلم اول دبستان و نام دبستان سال اول ابتدایی خود را بنویسند. در انتهای ترم، دو اطلاع مذکور بر روی تابلوی اعلانات نصب شد و طی اعلامیه از آنها خواسته شد تا نمرات خود را با کد مورد نظر یادداشت

نموده و در جعبه‌ای که در سالن نصب شده بود بیاندازند. به این ترتیب دانشجویان بدون اینکه شناسایی شوند توانستند اطلاعات مورد نظر پژوهشگران را تکمیل نمایند. از این تعداد در نهایت، ۶۰ نفر نمرات خود را با کد مربوطه اعلام کردند.

یافته‌ها

نتایج توصیفی در جدول ۱ گزارش شده است. در همه ابعاد پرسش‌نامه، میانگین‌های به دست آمده بالاتر از میانه مقیاس است و دامنه انحراف معیارها بین ۰/۶۵ تا ۰/۹۹ است و دامنه همبستگی‌های خرده‌مقیاس‌های ابزار نیز ۰/۲۵ تا ۰/۵۹ است.

جدول ۱) میانگین، انحراف استاندارد، ضرایب آلفای کرونباخ و همبستگی پیرسون بین ابعاد ۹گانه پرسش‌نامه (تعداد ۳۳۹)

	M	SD	۱	۲	۳	۴	۵	۶	۷	۸	۹
۱	۴.۱	.۷۰	.۷۹۰	.۵۹۱	.۵۷۴	.۳۵۸	.۳۷۳	.۳۹۱	.۴۱۸	.۳۳۷	.۴۶۵
۲	۴.۲	.۶۵		.۷۵۵	.۵۲۹	.۴۷۱	.۵۲۵	.۴۲۵	.۴۱۵	.۳۰۹	.۵۰۸
۳	۴.۲	.۷۱			.۷۵۹	.۴۵۶	.۴۳۳	.۴۹۸	.۵۲۰	.۳۸۹	.۵۰۹
٤	۳.۷	.۹۳				.۸۶۳	.۳۹۷	.۴۷۹	.۴۶۶	.۳۷۸	.۳۵۸
۵	۳.۹	.۷۷					.۷۵۰	.۴۰۳	.۴۵۱	.۲۵۰	.۳۷۳
٦	۳.۷	.۹۴						.۷۵۵	.۵۵۳	.۳۷۹	.۳۳۲
۷	۳.۷	.۹۹							.۸۰۹	.۵۱۰	.۴۲۹
۸	۳.۸	.۹۳								.۷۳۰	.۳۶۲
۹	۴.۴	.۷۰									.۶۳۷

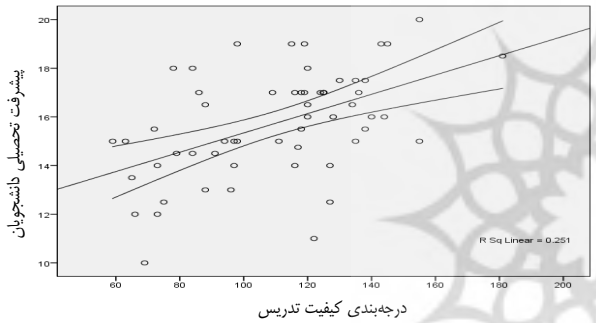
همه همبستگی‌ها در سطح p<۰/۰۰۱ معنی‌دار هستند.

ضرایب اعتبار بر روی قطر اصلی ماتریس بالا مثلثی قرار دارند.

(۱. یادگیری؛ ۲. علاقمندی؛ ۳. سازمان‌دهی؛ ۴. تعامل گروه؛ ۵. رابطه استاد/دانشجو؛ ۶. جامعیت؛ ۷. امتحانات؛ ۸. تکالیف؛ ۹. ارزیابی کلی)

اعتبار: برای بررسی هماهنگی درونی هر یک از ابعاد ابزار، ضریب آلفای کرنباخ محاسبه شد. همه ابعاد (به استثنای ارزیابی کلی) دارای ضرایب اعتبار قابل قبول هستند. دامنه ضرایب اعتبار ابعاد از ۰/۶۳ (ارزیابی کلی) تا ۰/۸۶ (تعامل گروهی) است. در بُعد یادگیری دامنه همبستگی‌های اصلاح شده سوال- نمره کل از r=۰/۵۶۲ (ایجاد تحرک‌فکری و برانگیزندگی) تا r=۰/۶۲۷ (یادگیری و درک مطلب) به دست آمد. در بُعد علاقه‌مندی استاد، دامنه همبستگی‌های اصلاح شده سوال- نمره کل از r=۰/۴۹۸ (علاقه استاد به تدریس) تا r=۰/۶۱۱ (حفظ علاقه دانشجویان در طول مدت کلاس با سبک تدریس خود) است. در بُعد سازماندهی مطالب، دامنه همبستگی‌های اصلاح شده سوال- نمره کل از r=۰/۴۷۵ (ارایه درس به نحوی که یادداشت‌برداری ممکن باشد) تا r=۰/۶۳۰ (آماده‌سازی مطالب و تشریح دقیق آن) است. در بُعد تعامل گروهی، دامنه همبستگی‌های اصلاح‌شده سوال- نمره کل از r=۰/۶۸ (تشویق دانشجویان به بیان اندیشه‌های خود) تا r=۰/۷۴ (تشویق دانشجویان به پرسش و پاسخ‌دادن به سوال‌های آنها) است. در بُعد رابطه استاد- دانشجو، دامنه همبستگی‌های اصلاح شده سوال- نمره کل از r=۰/۳۹۲ (دسترسی کافی به استاد در طول ساعات اداری یا پس از کلاس) تا r=۰/۶۵ (ابراز علاقه واقعی به فرد فرد دانشجویان) است. در بُعد جامعیت مطالب، دامنه همبستگی‌های اصلاح‌شده سوال- نمره کل از r=۰/۵۱ (بحث در مورد پیشرفت‌های کنونی مرتبط با هر زمینه درسی) تا r=۰/۶۱ (توضیح پیش‌زمینه‌های مفاهیم و مطالب درسی) است. در بُعد امتحانات، دامنه همبستگی‌های اصلاح‌شده سوال- نمره کل از r=۰/۶۳ (ارائه آگاهی‌های ارزشمند از نتایج امتحانات مستمر) تا r=۰/۷۱ (روش‌های مناسب و عادلانه ارزشیابی) است. در بُعد تکالیف، همبستگی‌های اصلاح شده سوال-نمره کل برای هر ۲ سوال r=۰/۵۷ (سوال ۲۸ و ۲۹) است. در بُعد ارزیابی کلی، همبستگی‌های اصلاح‌شده سوال- نمره کل برای هر ۲ سوال r=۰/۴۶ (سوال ۳۰ و ۳۱) است (جدول ۱).

روایی پیش‌بین: روایی پیش‌بین از طریق همبسته‌کردن نمره کل پرسش‌نامه ارزیابی کیفیت تدریس با نمرات پیشرفت تحصیلی دانشجویان در درس مربوطه و در انتهای ترم محاسبه شد. وارسی نمودار پراکنش دو متغیر حاکی از وجود ۲ مقدار پرت بود که با حذف آنها ضریب همبستگی پیرسون r=۰/۵ محاسبه شد (نمودار ۱).



نمودار ۱) پراکنش متغیرهای پیشرفت تحصیلی و درجه‌بندی کیفیت تدریس

تحلیل عاملی اکتشافی: به منظور تعیین تعداد عامل‌های زیربنایی و بررسی ویژگی‌های ۳۱ سوال پرسش‌نامه ارزیابی کیفیت تدریس، تحلیل عاملی اکتشافی با روش مولفه‌های اصلی بر روی کل داده‌های نمونه اجرا شد. آزمون *کرویت بارتلت* معنی‌دار (p<۰/۰۰۱، df=۴۶۵ و X^2=۳۸۰۵/۱۷) و اندازه کفایت نمونه‌برداری KMO=۰/۹۰۹ است. در این تحلیل، ۶ عامل با ارزش ویژه بیش از یک استخراج شد. به هر حال، وارسی نمودار *اسکری* یک عامل اولیه بزرگ را نشان داد (λ=۱۰/۸۹۱ که ۳۵/۱٪ واریانس کل را تبیین می‌کرد) و به دنبال آن تعدادی از مولفه‌های کوچکتر هستند (مولفه ۲، λ=۲/۱ با ۷٪؛ مولفه ۳، λ=۱/۶۵ با ۵/۳٪؛ مولفه ۴، λ=۱/۶ با ۵/۲٪ و ... واریانس کل را تبیین می‌کردند). به این ترتیب، سعی شد با اجرای تحلیل‌های متعدد، راه‌حل‌های عاملی مختلف بررسی شود. برای ساده‌سازی ساختار عاملی از چرخش *واریماکس* استفاده شد.

برای راه‌حل ۲عاملی در حدود ۴۸٪ سوال‌ها با بار عاملی بیش از ۰/۳ بر روی هر ۲ عامل بار گرفته بودند. اساسا شواهد اندکی برای پذیرش یک راه‌حل ۲عاملی وجود داشت. برای مثال، سوال‌هایی از هر بُعد بر روی عامل ۱ بار گرفته بودند که اساس نظری منطقی نداشتند.

جدول ۲) ماتریس چرخش یافته عاملی فرم کوتاه ۲۱سئوالی پرسش‌نامه ارزیابی کیفیت تدریس

	مولفه‌ها					
سؤال	۱	۲	۳	۴	۵	۶
۱۶	۰/۸۲۰					
۱۵	۰/۷۹۶					
۱۴	۰/۷۹۶					
۱۳	۰/۷۵۷					
۴		۰/۷۵۱				
۲		۰/۷۱۴				
۱		۰/۷۱۰				
۳		۰/۷۰۹				
۲۹			۰/۷۸۳			
۲۸			۰/۷۶۴			
۲۷			۰/۶۹۳			
۲۶			۰/۵۶۵			
۱۷				۰/۸۳۵		
۱۹				۰/۸۰۲		
۱۸				۰/۷۴۸		
۲۱					۰/۷۹۰	
۲۳					۰/۷۴۲	
۲۲					۰/۵۸۴	
۲۴					۰/۵۲۷	
۵						۰/۸۱۳
۶						۰/۷۳۵

در راه‌حل ۳عاملی سوال‌های ۵ و ۶ از بُعد علاقه‌مندی استاد، سوال‌های ۳۰ و ۳۱ بُعد ارزیابی کلی، سوال‌های ۹، ۱۰ و ۱۲ از بُعد سازماندهی و سوال‌های ۱ تا ۴ از بُعد یادگیری عامل ۱ را تشکیل می‌دادند. در این راه حل، سوال‌های ۱۳ تا ۱۶ بُعد تعامل گروهی و سوال ۲۸ از بُعد تکلیف عامل ۲ را تشکیل می‌دادند. عامل ۳، سوال‌های ۱۷ تا ۱۹ از بُعد رابطه استاد- دانشجو و سوال ۷ از بُعد علاقه‌مندی را شامل می‌شدند. سایر سوال‌ها در این راه حل ساختار روشنی نشان ندادند. بعضی از آنها بر روی ۳ عامل همزمان (۲۱، ۲۳، ۲۴ و ۲۷) بار شدند و برخی دیگر بر روی ۲ عامل همزمان بار شدند (۸، ۱۱، ۲۰ ، ۲۲، ۲۵، ۲۶ و ۲۹).

جدول ۳) شاخص‌های برازندگی مدل‌های فرضی مختلف (تعداد ۳۳۹)

مدل	χ^2	RMSEA	CFI	RMSEA بازه اطمینان ٪۹۰	SRMR	GFI	AGFI
A	۲۴۵۹/۶۹	۰/۱۱۸	۰/۶۶	۰/۱۱ -۰/۱۲	۰/۰۸۲	۰/۶۸	۰/۶۳
B	۱۰۶۰/۸۶	۰/۰۷	۰/۸۶	۰/۰۶۵ -۰/۰۷۵	۰/۰۵۸	۰/۸۳	۰/۷۹
C	۹۶۹/۷۳	۰/۰۷۳	۰/۸۷	۰/۰۶۷ -۰/۰۷۸	۰/۰۶۰	۰/۸۳	۰/۷۹
D	۱۰۴۲/۴۰	۰/۰۷۶	۰/۸۶	۰/۰۷۰ -۰/۰۸۱	۰/۰۶۲	۰/۸۲	۰/۷۹
E	۷۰۱/۸۰	۰/۰۷۱	۰/۸۹	۰/۰۶۵ -۰/۰۷۷	۰/۰۵۲	۰/۸۶	۰/۸۲

(**A**. تک‌عاملی؛ **B**. نه‌عاملی؛ **C**. هشت‌عاملی؛ **D**. هفت‌عاملی؛ **E**. شش‌عاملی)

در راه‌حل ۴عاملی سوال‌های ۱ تا ۴ از بُعد یادگیری، سوال‌های ۹، ۱۰ و ۱۲ از بُعد سازماندهی، سوال‌های ۳۰ و ۳۱ از بُعد ارزیابی کلی و سوال‌های ۵ و ۶ از بُعد علاقه‌مندی عامل ۱ را تشکیل می‌دادند. در این راه حل، سوال‌های ۲۸ و ۲۹ از بُعد تکلیف، سوال‌های ۲۵ و ۲۷ از بُعد امتحـانات و سوال‌های ۲۱ و ۲۴ از بُعد جامعیت عامل ۲ را تشکـیل مـی‌دادند. عامل ۳، سوال‌های ۱۷ تا ۱۹ از بُعد رابطه استاد- دانشجو را شامل می‌شد و عامل ۴ شامل سوال‌های ۱۳ تا ۱۶ بُعد تعامل گروهی بود. سایر سوال‌ها در این راه حل ساختار روشنی نشان نمی‌دادند، بعضی از آنها بر روی ۲ یا ۳ عامل همزمان بار شدند (این سوال‌ها گزارش نشده‌اند).

در راه‌حل ۵عاملی سوال‌های ۱ تا ۴ از بُعد یادگیری، سوال ۹ از بُعد سازماندهی، سوال‌های ۵ و ۶ از بُعد علاقه‌مندی عامل ۱ را تشکیل می‌دادند. عامل ۲ شامل سوال‌های ۱۳ تا ۱۶ بعد تعامل گروهی بود، عامل ۳ سوال‌های ۲۱ و ۲۴ از بعد جامعیت و سوال ۲۰ از بعد رابطه استاد- دانشجو، عامل ۴ شامل سوال‌های ۱۷ و ۱۸ از بعد رابطه استاد- دانشجو و عامل ۵ سوال ۱۷ از بعد امتحانات و سوال‌های ۲۸ و ۲۹ از بعد تکلیف را شامل می‌شد. سایر سوال‌ها در این راه حل ساختار روشنی نشان نمی‌دادند. برخی از آنها بر روی ۲ یا ۳ عامل همزمان بار شدند (این سوال‌ها گزارش نشده‌اند).

در راه‌حل ۶عاملی سوال‌های ۱ تا ۴ بعد یادگیری عامل ۱ را تشکیل می‌دادند، عامل ۲ شامل سوال‌های ۱۳ تا ۱۶ بعد تعامل گروهی بود، عامل ۳ سوال ۲۰ از بعد رابطه استاد- دانشجو و سوال‌های ۲۱، ۲۳ و ۲۴ از بعد جامعیت را شامل می‌شد، عامل ۴ شامل سوال‌های ۱۷ و ۱۸ از بعد رابطه استاد- دانشجو، عامل ۵ سوال ۳۱ از بعد ارزیابی کلی و سوال‌های ۵ و ۶ از بعد علاقه‌مندی، عامل ۶ سوال‌های ۲۸ و ۲۹ از بعد تکلیف و سوال ۲۷ از بعد امتحانات را شامل می‌شد.

از آنجا که راه‌حل ۶عاملی بیشترین تفسیرپذیری را به لحاظ نظری دارا بود، سوال‌هایی که ساختار عاملی روشنی نشان نمی‌دادند (سوال‌هایی که همزمان بر روی ۲، ۳ یا ۴ عامل بار می‌شدند) به طور متوالی حذف شدند. این فرآیند نه تنها سبب شد که سوال‌های مربوط به هر عامل نظری به گونه‌ای روشن‌تر بر روی عامل‌های مرتبط بار شوند بلکه منجر به افزایش بار عاملی هر سوال بر عامل مربوطه شد. ماتریس چرخش‌یافته عامل در راه‌حل ۶عاملی در جدول ۲ گزارش شده است. همانطور که ملاحظه می‌شود نخست اکثر سوال‌ها بر روی بُعد نظری پیش‌بینی‌شده بار شده‌اند و دوم آنکه ۵ عامل از ۶ عامل استخراج‌شده کاملا منطبق بر ابعاد نظری پیش‌بینی‌شده بود و صرفا عامل ۳ به صورت ترکیبی از بعد امتحانات و تکالیف نمایان شد (جدول ۲).

تحلیل سوال مدل ۶عاملی: در ادامه هماهنگی درونی (ضریب آلفای کرنباخ) راه حل ۶ عاملی و همبستگی اصلاح‌شده سوال- نمره‌کل بررسی شد. برای این مدل، اعتبار عامل‌های یادگیری، تعامل گروهی و جامعیت، قبلا گزارش شده است (جدول ۱). اعتبار عامل علاقه‌مندی استاد (α=۰/۶۹)، عامل رابطه استاد- دانشجو (α=۰/۷۸) و عامل امتحانات/تکالیف (α=۰/۷۹) محاسبه شد. برای عامل علاقه‌مندی استـاد، همبستگی‌های اصلاح‌شده سوال- نمره‌کل برای هـر ۲ سوال r=۰/۵۲ بود. دامنه همبستگـی‌های اصلاح‌شده سوال- نمره‌کل برای عامل رابطه استاد- دانشجو (۰/۶۰ تا ۰/۶۶) و برای عامل امتحانات/ تکالیف (۰/۵۸ تا ۰/۶۳) بود.

تحلیل عاملی تاییدی: همه تحلیل‌ها با استفاده از نرم افزار LISREL 8.51 انجام شده است [۲۹]. کشیدگی تک‌متغیری و چندمتغیری و برآورد بیشینه درست‌نمایی با تعدیل نرمال استفاده شد. شاخص‌های برازندگی متعدد برای سنجش برازندگی داده- مدل مورد استفاده قرار گرفت: ریشه دوم واریانس خطای تقریب، ریشه دوم واریانس پس‌ماند استاندارد شده، شاخص برازندگی مقایسه‌ای، شاخص نیکویی برازش و شاخص نیکویی برازش تعدیل شده (جدول ۳).

برای شاخص‌های برازندگی، برش‌های متعددی توسط متخصصان مطرح شده است. برای مثال، مقادیر مساوی یا کمتر از ۰/۰۵ برای RMSEA، مساوی یا بالاتر از ۰/۹۶ برای CFI، مساوی یا کمتر از ۰/۰۷ برای SRMR پیشنهاد شده است [۳۰]. دامنه مقادیر برای GFI و AGFI بین صفر تـا ۱ است و مقـادیر بالاتر نشان‌دهنده برازندگی بیشتر داده- مدل است [۳۱]. برای دو شاخص مذکور، نقطه برشی برابر یا مساوی با ۰/۰۵ در نظر گرفته شد.

سنجش برازندگی داده- مدل: جدول ۳، شاخص‌های برازندگی را برای مدل‌های مفروض نشان می‌دهد. ابتدا فرض شد که سازه کیفیت تدریس یک مدل تک عاملی است. برازندگی داده- مدل برای مدل تک عاملی در نمونه مورد بررسی بسیار ضعیف بود ($p=۰/۰۰۱$، $X^2=۲۴۵۹/۶۹$، CFI=۰/۶۶، SRMR=۰/۸۲، RMSEA=۰/۱۱۸، AGFI=۰/۸۰، GFI=۰/۸۳). سپس بر اساس مدل نظری *مارش* یک مدل ۹عاملی طراحی و آزمون شد و نتایج حاکی از نابرازندگی مدل ۹عاملی *مارش* است. تنها شاخص حاکی از برازندگی مدل، مقدار SRMR است که کمتر از ۰/۰۷ است. سپس فرض شد که بُعد ارزیابی کلی احتمالا، به عنوان بُعدی است که در سایر ابعاد منعکس خواهد شد. به این ترتیب، سوال‌های این بُعد حذف شده و یک مدل ۸عاملی فرض و آزمون شد. نتایج حاکی از کاهش معنی‌دار در مقدار آماره X^2 در مقایسه با مدل ۹عاملی است. به هر حال، به استثنای SRMR سایر شاخص‌ها حاکی از نابرازندگی مدل ۸عاملی است. مدل فرضی دیگر، ادغام بُعد تکالیف و امتحانات در یک عامل و آزمون یک مدل ۷عاملی بود. فرض شد امتحانات و تکالیف، احتمالا می‌توانند به عنوان یک عامل در نظر گرفته شوند (به جدول ۲ نگاه کنید). نتایج برازندگی این مدل چندان مطلوب نیست (به جدول ۳ نگاه کنید). در نهایت، مدل ۶عاملی حاصل از تحلیل عاملی اکتشافی (جدول ۲) آزمون شد. این مدل به وضوح منجر به کاهش معنی‌دار در آماره X^2 نسبت به مدل ۹عاملی شد. شاخص SRMR کمتر از ۰/۰۷ است و سایر شاخص‌ها از جمله GFI،CFI و AGFI نسبت به مدل ۹عاملی بهبود یافت اما به هر حال برای این مدل، شاخص‌ها به حد نصاب کافی نرسید. به ویژه شاخص RMSEA=۰/۰۷ حاکی از نابرازندگی مدل مذکور است.

بحث

در کل، نتایج این مطالعه هماهنگی درونی و روایی پیش‌بین پرسش‌نامه ارزیابی کیفیت تدریس را تایید کرد. اعتبار این پرسش‌نامه بسیار بالا است و از این لحاظ نتایج این مطالعه با مطالعات پیشین همسو است [۱۴، ۱۵]. نتایج این مطالعه نشان داد که نمرات پرسش‌نامه ارزیابی کیفیت تدریس با نمرات پیشرفت دانشجویان در انتهای ترم همبسته بوده و از لحاظ روایی پیش‌بین این نتیجه با نتایج مطالعات پیشین همخوان است [۱۱، ۱۲، ۱۳]. پژوهش‌ها نشان داده‌اند که درجه‌بندی‌های دانشجویان از دروس به طور معنی‌دار و قابل توجه با پیشرفت/یادگیری دانشجویان و درجه‌بندی‌های دانشجویان قبلی رابطه دارند [۱].

فرم کوتاه‌شده ۲۱ سوالی پرسش‌نامه مذکور یک ابزار چندبعدی با ۶ عامل است که از اعتبار کافی برخوردار است. ساختار عاملی این پرسش‌نامه شامل عامل‌های یادگیری، تعامل گروهی، رابطه استاد دانشجو، جامعیت، علاقه‌مندی و امتحانات/تکالیف است. گرچه این ساختار با ساختار ۹عاملی مورد ادعای *مارش* [۱۱، ۱۳، ۲۰، ۲۱، ۲۴، ۲۵ و ۲۶] صرفا در ۶ عامل همخوان بود اما با نظر *مارش* که کیفیت تدریس را سازه‌ای چندبعدی می‌داند همسو است.

مطالعات طولی حاکی از ثبات عوامل ۹گانه کیفیت تدریس و نیمرخ‌های چند بعدی است [۲۷، ۲۸] که چند بُعدی بودن سازه کیفیت تدریس در این مطالعه نیز تایید شده است. وجه تفاوت این مطالعه با ادبیات پژوهش در مورد تعداد عوامل است. یک تفسیر احتمالی این اختلاف، مربوط به داوطلب بودن نمونه است. شاید افراد داوطلب در مقایسه با افراد عادی برخی از عوامل را در درجه‌بندی اساتید مهمتر تلقی می‌کنند.

مطالعات بیشتری برای روشن‌ساختن سودمندی این ابزار لازم است و سوالاتی از این دست هنوز باقی است. آیا ترکیب خطی جزیی جمع‌پذیر از نمرات سوال‌های این پرسش‌نامه (که به یک نمره کل از آن در هر بُعد منتهی می‌شود) شاخص موثرتری از سازه کیفیت تدریس است؟ به عبارت دیگر آیا روش نمره‌گذاری با وزن‌های مساوی، شاخص رواتری از کیفیت تدریس اساتید به دست می‌دهد؟ تا چه اندازه نمرات هر بُعد معرف سازه کیفیت تدریس در هر بُعد هستند؟ تفسیرهای درست بر پایه نمرات حاصل از هر بُعد چیست؟ چه نوع تصمیم‌هایی بر پایه نمرات حاصل از این ابزار می‌توان گرفت؟ از لحاظ روایی همزمان باید مطالعاتی طراحی شوند که نمرات این ابزار را با سایر ابزارهای مناسب و ملاک‌های لازم برای تفسیرپذیری نمرات همبسته مقایسه نمایند. آیا نمرات حاصل از این ابزار با رفتار واقعی اساتید در کلاس درس همبسته هستند؟ عامل فرهنگ بر نمرات حاصل از پرسش‌نامه چه تاثیری دارد؟

نخستین محدودیت این مطالعه آن است که نمونه مورد بررسی، دانشجویان داوطلب دانشگاه سمنان بودند، لذا نتایج این مطالعه صرفا قابل تعمیم به جامعه دانشجویان داوطلب این دانشگاه است.

محدودیت دوم مربوط به ملاک مورد نظر است. در این مطالعه از نمرات درسی پایان ترم همان درس استفاده شده است. این شاخص به دلایلی، چند سوگیرانه است. نخست آن که آزمون پایان ترم درس مورد نظر، استاندارد نیست و لذا ممکن است از روایی و اعتبار لازم برخوردار نباشد. دوم آنکه این آزمون به شدت وابسته به نظر ارزیاب است و نمره حاصل از آن ممکن است از جهتی نشانگر جنبه‌های هیجانی نمره‌دهی استاد باشد. محدودیت دیگر این مطالعه حجم نمونه است. هر چند نمونه مورد نظر حداقل‌های لازم را برای آزمون مدل دارا بود ولی شاید نمونه‌ای با حجم بالاتر، نتایج دیگری ارائه دهد. محدودیت دیگر این مطالعه مربوط به روش تحلیل ابزار است. بسیاری از اندازه‌های روان‌شناختی که ساختار عاملی آنها با روش تحلیل عامل اکتشافی به خوبی برازنده شده‌اند با روش تحلیل عامل تاییدی به طور مناسبی برازنده نمی‌شوند. این امر به دلیل ساختار عاملی آنها است که با مدل خوشه‌های مستقل مقیدکننده که به طور خاص در مطالعات تحلیل عاملی تاییدی استفاده می‌شود، سازگار نیست. در این مدل‌ها به هر سوال اجازه داده می‌شود بر روی یک عامل بار شود و بار عاملی آن بر روی عامل‌های دیگر مقید به صفر می‌شود [۳۲]. برای رفع این مشکل پیشنهاد می‌شود پژوهشگران از روش مدل‌سازی معادلات ساختاری اکتشافی استفاده کنند. این روش به طور ساده تحلیل عاملی تاییدی و اکتشافی را وحدت می‌بخشد و بسیاری از مشکلات روش تحلیل عاملی تاییدی را رفع می‌کند [۳۳].

نتیجه‌گیری

نتایج تحلیل عاملی اکتشافی و تاییدی، ساختار ۹عاملی مارش را تایید نکرد ولی با دیدگاه *مارش*، مبنی بر چندبُعدی بودن سازه کیفیت تدریس، همسو است. بنابراین، استفاده از این ابزار برای ارزیابی کیفیت تدریس اساتید و تفسیر آن بایستی با احتیاط صورت گیرد.

منابع

1- Germain-Rutherford A. Student evaluation questionnaires: Dispelling misconceptions and looking at the literature. 2003. Available from: http://www.spiral.univ-lyon.fr/files_m/M152/files196630-2014.doc.

2- Kulik JA. Student ratings: Validity, utility and controversy. New Directions for Institutional Research. 2001;109:9-25.

3- Marsh HW, Overall JU. Validity of students' evaluations of teaching: A comparison with instructor self evaluations by teaching assistants, undergraduate faculty and graduate faculty. Los Angeles: University of Southern California Office of Institutional Studies; 1979.

4- Palmunen Lewis A, Knox K, Portner B, Zipser N. Course Evaluation Literature Review. 2006. Available from: http://www.fas.harvard.edu/~secfas/CUE_Literature_Review.pdf.

5- Braskamp LA,Ory JC.Assessing faculty work: Enhancing individual and institutional performance. San Francisco: Jossey-Bass; 1994.

6- Cashin WE. Student ratings of teaching: The research revisited (IDEA Paper No. 32). Manhattan: Kansas State University. 1995.

7- Centra JA. Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass; 1993.

8- Feldman KA. The association between student ratings of specific instructional dimensions and student achievement. Research in Higher Education. 1989;30(6):583-645.

9- Feldman KA. Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (Neutral) observers. Research in Higher Education. 1989;30(2):113-35.

10- Marsh HW, Dunkin MJ. Students' evaluations of university teaching: A multidimensional perspective. In: Smart J, editor. Effective teaching in higher education: Research and practice. New York: Agathon Press; 1997.

11- Marsh HW, Roche L. The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal. 1993;30:217-51.

12- Aleamoni LM. On the invalidity of student ratings for administrative personnel decisions. Journal of Higher Education. 1976;47(5):607-10.

13- Marsh HW. Students' evaluations of university teaching: Research findings, methodological issues and directions for future research. International Journal of Educational Research. 1987;11(3):253-388.

14- Devlin M, editor. Communicating outcomes of students' evaluations of teaching and learning: One-size-fits-all? Refereed Proceedings of the 2004 Evaluation Forum: Communicating Evaluation Outcomes: Issues and Approaches; Monash University, Melbourne; 2004

15- Coffey M, Gibbs G. The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education. Assessment & Evaluation in Higher Education. 2001;26(1):89-93.

16- Marsh HW, Hau KT, Chung CM, Siu TLP. Students' evaluations of university teaching: Chinese version of the Students' Evaluations of Educational Quality (SEEQ) instrument. Journal of Educational Psychology. 1997;89:568-72.

17- Marsh HW, Fleiner H, Thomas CS. Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology. 1975;67:833-9.

18- Marsh HW, Overall JU. Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology. 1980;72:468-75.

19- Marsh HW. The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal. 1977;14:441-7.

20- Marsh HW. Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. Journal of Educational Psychology. 1982;74:264-7.

21- Marsh HW, Overall JU, Kesler SP. Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. Journal of Educational Psychology. 1979;71:149-60.

22- Marsh HW. Students' evaluations of tertiary instruction: Testing the applicability of American surveys in an Australian setting. Australian Journal of Education. 1981;25:177-92.

23- Marsh HW, Overall JU. The relative influence of course level, course type and instructor on students' evaluations of college teaching. American Educational Research Journal. 1981;18:103-12.

24- Marsh HW. Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. Journal of Educational Psychology. 1983;75:150-66.
25- Marsh HW. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. Journal of Educational Psychology. 1984;76:707-54.
26- Marsh HW, Hocevar D. Multidimensional perspective on students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level and course level. Teaching & Teacher Education. 1991;7:9-18.
27- Marsh HW, Bailey M. Multidimensionality of students' evaluations of teaching effectiveness: A profile analysis. Journal of Higher Education. 1993;64:1-18.
28- Marsh HW, Hocevar D. Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. Teaching & Teacher Education. 1991;7:303-14.
29- Jöreskog KG, Sörbom D. LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software; 1993.
30- Yu C, Muthén B. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Paper presented at the annual meeting of the American Educational Research Association; 2002; New Orleans.
31- Jöreskog K, Sörbom D. Analysis of linear structural relationships by maximum likelihood and least squares methods. Sweden: University of Uppsala; 1981.
32- Asparouhov T. Muthén B. Exploratory structural equation modeling. Structural Equation Modeling. 2009. Available from: http://statmodel.com/download/EFACFA810.pdf.
33- Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A. Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. 2009 [In Print].

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی