

میزان اتکاپذیری نمره‌های ارزیابی دانشجویان کارشناسی از عملکرد آموزشی اعضای هیئت علمی

ابراهیم علیزاده*

چکیده

هدف پژوهش، برآورد میزان اتکاپذیری نمره‌های ارزیابی دانشجویان کارشناسی از عملکرد آموزشی استادان در چهارچوب نظریه تعمیم‌پذیری بود. برای این منظور، اطلاعات ارزیابی عملکرد استادان که در نیم‌سال اول سال تحصیلی ۹۴-۱۳۹۳ در پایگاه سامانه آموزشی دانشگاه علامه طباطبائی ثبت شده بود به عنوان جامعه آماری، تعریف شد. سپس به صورت تصادفی چند مرحله‌ای، نمره‌های ارزیابی عملکرد آموزشی دوازده استاد به عنوان نمونه آماری انتخاب شد. داده‌ها با استفاده از یک طرح دو وجهی آشیان شده، تحلیل شد تا ضریب اتکاپذیری داده‌ها برآورد شود. یافته‌ها نشان می‌دهد که اتکاپذیری نظر دانشجویان نسبت به عملکرد آموزشی استاد در همه گروه‌های آموزشی برابر نیست. همچنین میانگین نمره‌های ارزیابی، هشت استاد از بین دوازده استاد مورد مطالعه، به این علت که ضریب اتکاپذیری آنها کمتر از میزان مورد انتظار است نباید مبنای تصمیم‌گیری قرار بگیرد. بر اساس نتایج پیشنهاد می‌شود که پیش از به‌کارگیری نظرات دانشجویان در ارتقاء یا تمدید قرارداد استادان مدعو یا پیمانی، ابتدا ضریب اتکاپذیری نظرات دانشجویان برآورد شود و چنانچه ضریب اتکاپذیری، کمتر از میزان قابل قبول به‌دست آمد، نظر آنان مبنای تصمیم‌گیری قرار نگیرد.

واژگان کلیدی: نظریه تعمیم‌پذیری، ضریب اتکاپذیری، عملکرد آموزشی، ارزیابی استاد، دانشجویان کارشناسی.

تاریخ دریافت مقاله: ۹۴/۰۶/۰۲

تاریخ پذیرش مقاله: ۹۴/۱۲/۲۰

* استادیار گروه سنجش و اندازه‌گیری دانشکده علوم تربیتی و روان‌شناسی دانشگاه شهید بهشتی (نویسنده مسئول)

مقدمه

ارزیابی عملکرد آموزشی استاد با نظرخواهی از دانشجویان، جایگاه ویژه‌ای در مدیریت آموزش عالی دارد؛ زیرا، بر اساس آئین‌نامه ارتقای مرتبه اعضای هیئت علمی، امتیاز ارزیابی عملکرد در تمدید قرارداد و ارتقای علمی در نظر گرفته می‌شود. بر اساس این آئین‌نامه، دانشگاه‌ها و مؤسسات آموزش عالی کشور پس از پایان ترم به دانشجویان پرسشنامه‌ای ارائه می‌کنند تا عملکرد آموزشی استادان خود را ارزیابی کنند. پس از تکمیل پرسشنامه‌ها، میانگین نمره‌های ارزیابی دانشجویان در هر یک از درس‌ها محاسبه و کارنامه‌ای برای هر استاد صادر می‌شود تا او از وضعیت عملکرد آموزشی خود، آگاه شود.

نتایج ارزیابی عملکرد استاد توسط دانشجویان به این شرط می‌تواند مبنای تصمیم‌گیری قرار بگیرد که نظرات دانشجویان، پایایی مناسبی داشته باشد. اما نتایج پژوهش‌های انجام شده (عبداللهی، ۱۳۹۲؛ رئوفی و همکاران، ۱۳۸۹؛ امینی و هنردار، ۱۳۸۷) نشان می‌دهد که بسیاری از استادان بر این باورند که نمره‌های ارائه شده توسط دانشجویان، نمایانگر عملکرد آموزشی آنها نیست و بیشتر دانشجویان هنگام تکمیل پرسشنامه، عملکرد آموزشی استاد را منصفانه ارزیابی نمی‌کنند. ارزیابی این دسته از دانشجویان، تحت تأثیر خطای هاله‌ای است و پیش‌داوری‌های آنها در نمره ارزیابی دخالت داده می‌شود.

منابع بالقوه‌ای از خطای ارزیابی در بافت ارزیابی عملکرد وجود دارد که باعث ناپایا شدن میانگین نمره‌های به‌دست آمده می‌شود. برای مثال، ناهمگنی سؤال‌های پرسشنامه ارزیابی، نظر متفاوت دانشجویان درون هر کلاس نسبت به عملکرد آموزشی استاد، تعامل بین نظر دانشجویان درون هر کلاس با سؤال‌های پرسشنامه ارزیابی (کیلی، فور و بوسکیست^۱، ۲۰۱۰)، ویژگی‌های ظاهری استاد مانند چهره و لهجه، ویژگی‌های دانشجو مانند جنسیت، پایه تحصیلی و نمره مورد انتظار، ساعت تشکیل کلاس از آن جمله هستند (دهقانی و نخعی، ۱۳۹۱).

برای افزایش دقت نتایج ارزیابی عملکرد و پیش از برآورد ضریب پایایی، نیاز است که خطاهای منظم موجود در بافت ارزیابی، مورد توجه قرار گیرد تا مشخص شود که داده‌های به دست آمده در چه شرایطی پایایی دارد و با تغییر در شرایط و موقعیت

¹. Keeley, Furr & Buskist

اندازه‌گیری (مانند افزایش تعداد سؤال‌ها یا تعداد پاسخگویان به پرسشنامه)، پایایی داده‌ها تا چه اندازه می‌تواند کاهش یا افزایش یابد؛ به همین دلیل، پژوهش‌هایی در کشور انجام شده است تا میزان پایایی نظرات دانشجویان را برآورد کنند. تقریباً در همه پژوهش‌های انجام شده از فرمول آلفای کرونباخ برای برآورد پایایی استفاده شده است (باستانی، روح‌اللهی، طاهرنژاد، ۱۳۹۳؛ امدادی و همکاران، ۱۳۹۲؛ بیگدلی، ۱۳۷۹). فرمول آلفای کرونباخ به دو دلیل نمی‌تواند پایایی نظرات دانشجویان را با دقت زیادی برآورد کند: اول اینکه هدف فرمول آلفای کرونباخ، برآورد میزان همسانی درونی سؤال‌های پرسشنامه است؛ در حالی که هدف اندازه‌گیری در مطالعه ارزیابی عملکرد آموزشی استاد، اندازه‌گیری تفاوت‌های واقعی بین نظرات دانشجویان درون کلاس‌هایی است که استاد برای آنها تدریس کرده است. دوم اینکه این فرمول در چهارچوب نظریه کلاسیک آزمون^۱ قرار دارد و به همین دلیل هنگام برآورد پایایی، اثر خطاهای منظم (مانند نوع سؤال‌ها، تعامل سؤال‌ها با نظر دانشجویان و پراکندگی نظر دانشجویان درون هر کلاس) را در نظر نمی‌گیرد (سان، والیجا و ژائو^۲، ۱۹۹۷).

فرمول‌های برآورد پایایی مبتنی بر نظریه کلاسیک آزمون، بر این فرض بنا شده‌اند که برای موضوع مورد مطالعه نمره‌های واقعی وجود دارد که در زمان‌ها و موقعیت‌های مختلف ثابت است و هرگونه تفاوت در داده‌ها از موقعیتی به موقعیت دیگر از خطاهای تصادفی ناشی می‌شود (مورگسون و کمپین^۳، ۲۰۰۰). در حالی که در مطالعات ارزیابی عملکرد، همان‌طور که پیش از این اشاره شد، منابع بالقوه‌ای از خطای ارزیابی وجود دارد که باعث نقض مفروضه ثابت بودن نمره واقعی در نظریه کلاسیک آزمون می‌شوند.

برآورد دقیق پایایی نمره‌های دانشجویان درباره عملکرد آموزشی استادان نیاز به روشی پیشرفته دارد تا بتواند منابع خطای اندازه‌گیری منظم و تصادفی را هنگام برآورد پایایی در نظر بگیرد. برای این منظور به یک مدل برآورد پایایی نیاز است. نظریه تعمیم‌پذیری، مدل برآورد پایایی است که چنین امکانی را فراهم می‌سازد (برنان^۴، ۲۰۰۱ الف)؛ اما جستجو در پایگاه اطلاع‌رسانی سازمان اسناد و مدارک علمی ایران نشان داد که با استفاده از این روش، تاکنون پژوهشی در کشور منتشر نشده است؛

1. Classical test theory

2. Sun, Valiga, & Gao

3 - Morgeson & Campion

4. Brennan

بنابراین، در این باره سؤال‌هایی مطرح است. نخست آنکه تا چه اندازه می‌توان به نظرات دانشجویان درباره عملکرد آموزشی استادان اتکا کرد و آنها را مبنای تصمیم‌گیری درباره ترفیع و ارتقای استادان قرار داد. دوم اینکه هر یک از منابع بالقوه خطای اندازه‌گیری قابل مطالعه (شامل، ناهمگنی سؤال‌های پرسشنامه ارزیابی، نظر متفاوت دانشجویان درون هر کلاس نسبت به عملکرد آموزشی استاد، تعامل بین نظر دانشجویان کلاس‌های مختلف با سؤال‌های پرسشنامه ارزیابی) چند درصد در تولید خطای اندازه‌گیری نقش دارند. پژوهش حاضر تلاش کرد که با استفاده از نظریه تعمیم‌پذیری، پاسخ‌های مناسبی برای پرسش‌های طرح شده بیابد.

پیشینه پژوهش

نظریه تعمیم‌پذیری، بر خلاف نظریه کلاسیک آزمون بر این فرض استوار است که نمره واقعی با توجه به موقعیت و شرایط اندازه‌گیری قابل تغییر است؛ زیرا، هنگام جمع‌آوری داده‌ها، نمره مشاهده شده تحت تأثیر عوامل مختلفی قرار می‌گیرد. نظریه تعمیم‌پذیری جایگاه ارزشمندی در برآورد پایایی نمره‌های ارزیابی عملکرد آموزشی استادان دارد. نظریه تعمیم‌پذیری به منظور افزایش دقت برآورد پایایی داده‌هایی که تحت تأثیر خطاهای منظم اندازه‌گیری قرار می‌گیرند، از سوی کرونباخ و همکارانش، معرفی شد (کرونباخ و همکاران^۱، ۱۹۷۲؛ کرونباخ، راجاراتانام و گلسر^۲، ۱۹۶۳). در هر اندازه‌گیری منابع بالقوه‌ای از خطا وجود دارد. نظریه تعمیم‌پذیری، رویکردی برای برآورد دقت اندازه‌گیری در موقعیت‌هایی است که داده‌های جمع‌آوری شده در معرض منابع چندگانه خطا قرار دارد. برای این منظور با برآورد مؤلفه‌های واریانس^۳، زمینه لازم برای برآورد سهم نسبی عوامل تشکیل‌دهنده خطای منظم و خطای نامنظم در نمره مشاهده شده ایجاد می‌شود. پس از شناخت عوامل تولید خطای منظم، تحلیلگر می‌تواند طرح اندازه‌گیری مؤثری را برای جمع‌آوری داده‌ها طرح‌ریزی کند (برینان، ۲۰۱۱؛ برینان، ۲۰۰۱ الف؛ شولسون و وب^۴، ۱۹۹۱).

1. Cronbach et al.

2. Cronbach, Rajaratnam & Gleser

3. variance components

4. Shavelson & Webb

در بافت نظریه تعمیم‌پذیری به هر یک از منابع خطا، "وجه"^۱ و به سطوح تشکیل‌دهنده هر وجه، شرایط یا موقعیت اندازه‌گیری گفته می‌شود (شولسون، وب و رولی^۲، ۱۹۸۹). برای مثال، اگر دو دانشجو، یک استاد را از لحاظ عملکرد آموزشی، ارزیابی کنند، دانشجو یک وجه اندازه‌گیری با دو سطح خواهد بود، چون دو دانشجو وجود دارد که بنا به دلایل مختلف ممکن است که عملکرد آموزشی استاد را همانند هم ارزیابی نکنند. تفاوت زیاد بین نظرات دانشجویان می‌تواند منبع تولید خطای اندازه‌گیری منظم شود.

در نظریه تعمیم‌پذیری، وجه‌های اندازه‌گیری به دو دسته ثابت و تصادفی تقسیم می‌شوند. وجه ثابت و تصادفی در نظریه اندازه‌گیری معادل با عامل ثابت و عامل تصادفی در طرح‌های تحلیل واریانس است. در وجه ثابت، سطوح تشکیل‌دهنده وجه مورد نظر محدود است و از یک موقعیت اندازه‌گیری به موقعیت دیگر، تغییری در آنها ایجاد نمی‌شود؛ اما در وجه تصادفی، سطوح مورد بررسی به صورت تصادفی از میان مجموعه مرجع سطوح ممکن آن وجه انتخاب شده‌اند و بنابراین در هر بار اجرای اندازه‌گیری، سطوح مورد نظر می‌تواند تغییر کند (برینان، ۲۰۱۱). برای نمونه وجه دانشجویان در ارزیابی عملکرد استادان، هم می‌تواند به عنوان یک وجه تصادفی و هم یک وجه ثابت در نظر گرفته شود. چنانچه در یک شغل، جامعه آماری ارزیابان بزرگ باشد و ارزیابان به صورت تصادفی از میان آنها انتخاب شوند، وجه ارزیاب یک وجه تصادفی خواهد بود و اگر تعداد ارزیابان کم باشد و شغل مورد نظر را همه ارزیابان، ارزیابی کنند، وجه ارزیاب، ثابت خواهد بود.

در نظریه تعمیم‌پذیری با توجه به تفسیری که از نمره مشاهده شده به عمل می‌آید، دو نوع ضریب پایایی با عنوان ضریب تعمیم‌پذیری^۳ و ضریب اتکاپذیری^۴ برآورد می‌شود؛ بر این اساس، اگر یک ارزشیاب آموزشی بخواهد نمره‌های ارزیابی چندین استاد را با یکدیگر مقایسه کند، نیاز دارد که ضریب تعمیم‌پذیری نظرات دانشجویان را برآورد کند. در مقابل، ضریب اتکاپذیری زمانی مورد استفاده قرار می‌گیرد که ارزشیاب آموزشی می‌خواهد، تفسیر مطلق از نمره مشاهده شده به عمل آورد؛ یعنی، میانگین نمره مشاهده شده هر استاد را با یک ملاک از پیش تعیین شده مقایسه کند.

1. facet

2. Shavelson, Webb & Rowley

3. generalizability coefficient

4. dependability coefficient

برای نمونه، هنگامی که ارزشیاب آموزشی، نمره ارزیابی هر استاد را با نمره چهار از پنج (یعنی خوب از عالی) مقایسه می‌کند، از نمره مشاهده شده، تفسیر مطلق به عمل آورده است.

برای برآورد پایایی نظرات دانشجویان، پژوهشگران (ایبراهیم^۱، ۲۰۱۱؛ کیلی و همکاران، ۲۰۱۰؛ ونلیوون، دورمودی و سیورس^۲، ۱۹۹۹؛ بارنز و بارنز^۳، ۱۹۹۳؛ کانه، گلیمور و کروکس^۴، ۱۹۷۶) در چهارچوب نظریه تعمیم‌پذیری از مدل‌های مختلفی برای برآورد پایایی داده‌های ارزیابی عملکرد آموزشی استادان، استفاده کرده‌اند. برای نمونه، کیلی و همکاران (۲۰۱۰) با یک طرح دو وجهی متقاطع، این مسئله را بررسی کردند که آیا دانشجویان می‌توانند بین عملکرد استادان خود تمایز بگذارند و آنها را با دقت بالایی ارزیابی کنند. برای این منظور دانشجویان، عملکرد آموزشی استادان خود را در سه حالت ارزیابی کردند. در حالت اول عملکرد بدترین استاد و در حالت دوم، عملکرد بهترین استادی را که با وی واحد درسی را گذرانده بودند. در حالت سوم عملکرد استادی را که با وی درس روان‌شناسی عمومی را گذرانده بودند. نتایج نشان داد که دانشجویان قادر بوده‌اند که بین عملکردهای مختلف آموزشی استادان خود تفکیک قائل شوند.

سان و همکاران (۱۹۹۷) از طرح آشیان شده‌ای که دارای سه وجه دانشکده، استاد و دانشجو بود، استفاده کردند. در طرح مورد استفاده سؤال‌های پرسشنامه به‌عنوان هدف اندازه‌گیری تعیین شده بود و وجه دانشجویان، درون وجه استاد و وجه استاد، درون وجه دانشکده آشیان کرده بود. نتایج نشان داد که ۲۹ درصد پراکندگی نمره مشاهده شده ناشی از پراکندگی در نظرات دانشجویان است و برای افزایش اتکاپذیری نمره به‌دست آمده، نیاز است که تعداد بیشتری از دانشجویان عملکرد یک استاد را ارزیابی کنند.

ونلیوون و همکاران (۱۹۹۹) واریانس نظرات دانشجویان را با طرح دو وجهی آشیان شده تحلیل کردند. آنان به این نتیجه رسیدند که پراکندگی در نظرات دانشجویان، بزرگ‌ترین منبع تغییرپذیری نمره ارزیابی عملکرد استادان محسوب می‌شود، به طوری که یک سوم خطای منظم موجود در نمره مشاهده شده، ناشی از پراکندگی

1. Ibrahim

2. VanLeeuwen, Dormody & Seevers

3. Barnes & Barnes

4. Kane, Gillmore & Crooks

نظرات دانشجویان هر کلاس است؛ اما سؤال‌های پرسشنامه مورد استفاده تأثیر بسیار ناچیزی در کاهش اتکاپذیری نمره دارد. در مقابل، بارنز و بارنز (۱۹۹۳) این مسئله را بررسی کردند که آیا استفاده از پرسشنامه می‌تواند برای ارزیابی عملکرد آموزشی استادان در همه واحدهای درسی استفاده شود. تحلیل داده‌های جمع‌آوری شده از دانشکده‌های مختلف با استفاده از دو طرح دو وجهی آشپان شده، نشان داد که سؤال‌های پرسشنامه، واریانس خطای منظم تولید می‌کند و بنابراین برای افزایش اتکاپذیری نظرات دانشجویان نیاز است که برای درس‌های مختلف، سؤال‌های اختصاصی طراحی شود.

روش تحقیق

این پژوهش از نوع پژوهش‌های توصیفی است که هدف آن برآورد میزان اتکاپذیری نظرات دانشجویان نسبت به عملکرد آموزشی استادان بود. جامعه آماری پژوهش، اطلاعات موجود در پایگاه سامانه آموزشی دانشگاه شامل نظرات همه دانشجویان کارشناسی دانشگاه علامه طباطبائی بود که در نیم‌سال اول سال تحصیلی ۹۴-۱۳۹۳ عملکرد آموزشی استادان خود را ارزیابی کرده‌اند. در این بانک اطلاعاتی، نظرات هر دانشجو نسبت به عملکرد آموزشی هر استاد به تفکیک سؤال‌های پرسشنامه وجود دارد و در این پژوهش نمره‌های خام نظرات دانشجویان، مبنای تحلیل تعمیم‌پذیری قرار گرفته است. ذکر این نکته ضروری است که پیش از نمونه‌گیری، استادانی که کمتر از سه دانشجو آنها را ارزیابی کرده بودند و یا فقط یک کلاس در دوره کارشناسی داشتند از فایل داده‌ها، کنار گذاشته شدند. همچنین، نظر آن دانشجویانی که با استاد مورد نظر واحد درسی خود را گذرانده و پرسشنامه ارزیابی را به طور کامل تکمیل کرده بودند، مبنای تحلیل قرار گرفت.

برای انتخاب نمونه آماری، ابتدا سه دانشکده (شامل روان‌شناسی و علوم تربیتی، مدیریت و حسابداری، ادبیات) به صورت تصادفی ساده از بین دانشکده‌های موجود دانشگاه انتخاب شدند. سپس در هر دانشکده نیز به صورت تصادفی ساده از بین استادانی که در مقطع کارشناسی تدریس می‌کردند، چهار استاد انتخاب شد. بر این اساس، دوازده استاد از گروه‌های آموزشی زبان فرانسه، ادبیات عرب، ادبیات فارسی، زبان‌شناسی همگانی، روان‌شناسی کودکان استثنائی، روان‌شناسی بالینی، مشاوره، تکنولوژی آموزشی، حسابداری، جهانگردی، مدیریت صنعتی و مدیریت بازرگانی

انتخاب شدند. به‌طور متوسط هر استاد برای دو تا سه کلاس تدریس داشت و تعداد دانشجویان هر کلاس هم از هفت نفر تا ۴۷ نفر متغیر بود. پس از انتخاب استادان، نظرات دانشجویانی که عملکرد آموزشی را ارزیابی کرده بودند از بانک اطلاعاتی، استخراج شد.

ابزار اندازه‌گیری

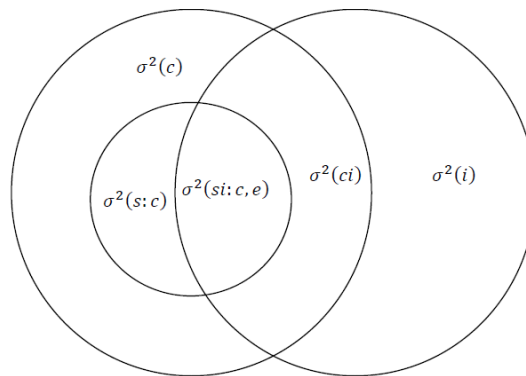
پرسشنامه استفاده شده برای ارزیابی عملکرد آموزشی استاد از دیدگاه دانشجو، ۱۸ سؤال دارد که در دانشگاه طراحی شده است و در سامانه آموزشی دانشگاه قرار دارد تا دانشجویان پس از پایان کلاس‌ها و پیش از آغاز امتحانات پایان ترم به صورت اینترنتی، کیفیت عملکرد آموزشی استاد را در مقیاس پنج درجه‌ای (عالی تا ضعیف) ارزیابی کنند. این پرسشنامه عملکرد آموزشی استاد را در شش حوزه، شامل توانایی علمی، مهارت تدریس، ایجاد انگیزه در دانشجو، ساماندهی و اداره کلاس، رعایت مقررات آموزشی و نحوه برخورد با دانشجو ارزیابی می‌کند. در این پژوهش برای بررسی همسانی درونی این پرسشنامه از ضریب آلفای کرونباخ استفاده شد. ضریب به‌دست آمده (۰/۹۱) نشان داد که سؤال‌های پرسشنامه با یکدیگر همگن هستند. علاوه بر این ساختار عاملی پرسشنامه با استفاده از تحلیل عاملی اکتشافی برآورد گردید. نتایج نشان داد که اولین عامل استخراج شده از ماتریس داده‌ها ۵۵/۴ درصد واریانس را تبیین می‌کند و این در حالی است که دومین عامل، فقط ۷/۰۲ درصد واریانس را تبیین نموده است و بر این اساس می‌توان بیان کرد که ساختار عاملی پرسشنامه از یک عامل اشباع شده است.

تحلیل داده‌ها

شناسایی طرح مناسب اندازه‌گیری، نقش تعیین کننده‌ای در برآورد پایایی هدف اندازه‌گیری دارد. برای این منظور، ابتدا مشخص شد که ارتباط وجه یا وجه‌های اندازه‌گیری با یکدیگر و همچنین با هدف اندازه‌گیری به صورت متقاطع است یا به صورت آشیانه‌ای و یا ترکیبی از هر دو (کاردینت، جانسون و پینی^۱، ۲۰۱۰). در پژوهش حاضر دانشجویان و سؤال‌های پرسشنامه ارزیابی به عنوان دو وجه

^۱ . Cardinet, Johnson & Pini

اندازه‌گیری تعیین شدند. ارتباط بین وجه‌های اندازه‌گیری به این صورت بود که دانشجویان درون وجه کلاس (هدف اندازه‌گیری) آشیان کرده‌اند. همچنین وجه اندازه‌گیری سؤال‌های پرسشنامه ارزیابی با هدف اندازه‌گیری و وجه دانشجویان متقاطع شده است. هدف اندازه‌گیری طرح حاضر، میزان تفاوت واقعی بین نظرات دانشجویان کلاس‌هایی بود که هر استاد در آن تدریس کرده بود. بر این اساس طرح دو وجهی آشیان شده برای برآورد ضریب اتکاپذیری داده‌ها به کار گرفته شد. طرح اندازه‌گیری مطالعه حاضر را می‌توان به صورت شکل شماره (۱) نشان داد:



شکل (۱) مدل برآورد پایایی نظرات دانشجویان نسبت به عملکرد آموزشی هر استاد در مطالعه حاضر

این طرح به صورت هم‌زمان، پنج مؤلفه واریانس را برآورد می‌کند که همه مؤلفه‌های واریانس در جدول (۱) توصیف شده است.

جدول (۱) توصیف مؤلفه‌های واریانس مربوط به طرح اندازه‌گیری

مؤلفه‌های واریانس	نماد	توصیف مؤلفه‌های واریانس
هدف اندازه‌گیری	C	به تفاوت‌های واقعی بین نظرات کلاس‌هایی اشاره دارد که هر استاد در آنها تدریس کرده است.
وجه دانشجویان آشیان شده درون کلاس‌ها	S:C	دانشجویان هر کلاس، عملکرد استاد را متفاوت از یکدیگر ارزیابی می‌کنند و نظر بسیار متفاوتی نسبت به یکدیگر دارند.
وجه سؤال‌های پرسشنامه ارزیابی	I	سؤال‌های پرسشنامه با یکدیگر تفاوت دارند و سهم یکسانی در ارزیابی عملکرد استاد ندارند.

مؤلفه‌های واریانس	نماد	توصیف مؤلفه‌های واریانس
وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی	Ci	معنای سؤال‌های پرسشنامه در همه کلاس‌ها به صورت یکسان درک نشده است؛ یعنی دانشجویان نظر ویژه‌ای نسبت به برخی از سؤال‌ها داشته‌اند.
خطای اندازه‌گیری تصادفی	s:i,e	خطای اندازه‌گیری تصادفی و تعامل‌های مرتبه بالاتر که از یکدیگر، قابل تفکیک نیستند.

مؤلفه‌های واریانس با استفاده از نرم‌افزار یورجنوا^۱ (برنان، ۲۰۰۱ ب) و در قالب تحلیل اثرات تصادفی برآورد شدند. در قالب تحلیل اثرات تصادفی، فرض بر این بود که سؤال‌های پرسشنامه از جامعه بزرگی از سؤال‌های ارزیابی استاد، انتخاب شده‌اند و همچنین دانشجویان نیز به صورت تصادفی از بین جامعه دانشجویانی که می‌توانستند عملکرد آموزشی استاد را ارزیابی کنند، انتخاب شده‌اند.

نتایج

در ادامه سهم هر یک از مؤلفه‌های واریانس در نمره ارزیابی مشاهده شده برای هر یک از استادان به تفکیک دانشکده‌های آنها تشریح می‌شود. تفسیر مؤلفه‌های واریانس برآورد شده، اهمیت دارد که از این نظر می‌توان سهم هر مؤلفه را در واریانس کل نمره‌های مشاهده شده، بررسی کرد. بزرگ بودن مؤلفه‌های مربوط به وجه‌های اندازه‌گیری نشان می‌دهد که آن مؤلفه تا چه مقدار در خطای اندازه‌گیری نقش دارد.

¹.urGENOVA

جدول (۲) مؤلفه‌های واریانس برآورد شده برای استادان دانشکده ادبیات

مؤلفه‌های واریانس برآورد شده (سهم مؤلفه در واریانس نمره کل به درصد)				منابع واریانس
استاد زبان‌شناسی همگانی	استاد ادبیات فارسی	استاد ادبیات عرب	استاد زبان فرانسه	
۰/۱۲۱ (/۶)	۰/۲۳۰ (/۱۸)	۰/۱۳۰ (/۱۷)	۰/۰۴۳ (/۳)	هدف اندازه‌گیری
۱/۵۲۳ (/۷۱)	۰/۷۰۶ (/۵۷)	۰/۴۲۶ (/۵۵)	۱/۱۹ (/۸۲)	وجه دانشجویان آشیان شده درون کلاس‌ها
۰/۰۴۱ (/۲)	۰/۰۱۹ (/۲)	۰/۰۰۶ (/۱)	۰/۰۱۲ (/۱)	وجه سؤال‌های پرسشنامه ارزیابی
۰/۰۱۵ (/۱)	۰ (/۰)	۰ (/۰)	۰/۰۱۲ (/۱)	وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی
۰/۴۵۰ (/۲۱)	۰/۲۸۸ (/۲۳)	۰/۲۱۱ (/۲۷)	۲۰ (/۱۰)	خطای اندازه‌گیری تصادفی

همان‌طور که در جدول (۲) مشاهده می‌شود، مؤلفه واریانس وجه دانشجویان هر یک از استادان گروه‌های آموزشی زبان فرانسه، ادبیات عرب، ادبیات فارسی و زبان‌شناسی همگانی بیش از نیمی از واریانس نمره مشاهده شده را به خود اختصاص داده‌اند؛ به بیان دیگر، دانشجویان هر کلاس، عملکرد استاد خود را متفاوت از دانشجویان هم‌کلاسی خود ارزیابی کرده‌اند. در مقابل، واریانس وجه سؤال‌های پرسشنامه ارزیابی و همچنین وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی، سهم بسیار ناچیزی در واریانس نمره مشاهده شده داشته‌اند که این وضعیت نشان می‌دهد این دو عامل نقشی در تولید خطای اندازه‌گیری منظم نداشته‌اند.

مؤلفه واریانس هدف اندازه‌گیری در استاد زبان فرانسه و زبان‌شناسی همگانی، سهم بسیار اندکی در واریانس نمره مشاهده شده دارد. علت این است که مؤلفه واریانس وجه دانشجویان، بیشتر سهم واریانس را از آن خود کرده است و برای افزایش پایایی نمره‌های ارزیابی این استادان، باید سهم واریانس وجه دانشجویان به حداقل ممکن کاهش یابد.

خطاهای اندازه‌گیری تصادفی در همه نمره‌های ارزیابی استادان مطالعه شده، تقریباً کمتر از یک پنجم نمره مشاهده شده را به خود اختصاص داده‌اند و این وضعیت نشان

می‌دهد که وجه‌های اندازه‌گیری پنهان که بتوانند تولید خطای اندازه‌گیری منظم کنند، وجود ندارد. معمولاً زمانی که سهم خطای تصادفی در واریانس نمره‌های مشاهده شده بزرگ باشد، احتمال دارد که یک یا چند وجه پنهان بر هدف اندازه‌گیری تأثیر داشته باشد (برنان، ۲۰۰۱ الف).

جدول (۳) مؤلفه‌های واریانس برآورد شده برای استادان دانشکده روان‌شناسی و علوم تربیتی

مؤلفه‌های واریانس برآورد شده (سهم مؤلفه در واریانس نمره کل به درصد)				منابع پراکندگی
استاد مشاوره	استاد تکنولوژی آموزشی	استاد روان‌شناسی بالینی	استاد روان‌شناسی کودکان استثنائی	
۰/۱۵۱ (٪۱۵)	۰/۰۷۴ (٪۷)	۰/۰۳۷ (٪۴)	۰/۱۶۹ (٪۱۳)	هدف اندازه‌گیری
۰/۳۸۶ (٪۳۹)	۰/۷۲۲ (٪۶۹)	۰/۷۴۸ (٪۷۰)	۰/۸۳۲ (٪۶۳)	وجه دانشجویان آشیان شده درون کلاس‌ها
۰/۱۰۶ (٪۱۱)	۰/۰۱۵ (٪۱)	۰/۰۲۹ (٪۳)	۰/۰۱۱ (٪۱)	وجه سؤال‌های پرسشنامه ارزیابی
۰/۰۰۷ (٪۱)	۰/۰۰۳ (٪۱)	۰ (٪۰)	۰/۰۰۹ (٪۱)	وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی
۰/۳۲۷ (٪۳۳)	۰/۲۳۶ (٪۲۲)	۰/۲۵۱ (٪۲۴)	۰/۲۹۵ (٪۲۲)	خطای اندازه‌گیری تصادفی

مؤلفه واریانس وجه دانشجویان هر یک از استادان گروه‌های آموزشی روان‌شناسی کودکان استثنائی، روان‌شناسی بالینی، تکنولوژی آموزشی بیش از سه پنجم واریانس نمره مشاهده شده را به خود اختصاص داده است؛ به بیان دیگر، دانشجویان درون هر کلاس عملکرد استاد خود را بسیار متفاوت از دانشجویان هم کلاسی خود ارزیابی کرده‌اند؛ اما مؤلفه واریانس وجه دانشجویان استاد گروه آموزشی مشاوره کمتر از دو پنجم است و این وضعیت نشان می‌دهد که دانشجویان درون هر یک از کلاس‌هایی که این استاد برایشان تدریس داشته است، دیدگاه تقریباً مشابهی به عملکرد آموزشی استاد خود داشته‌اند.

واریانس وجه سؤال‌های پرسشنامه ارزیابی، و همچنین وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی سهم بسیار ناچیزی در واریانس نمره مشاهده شده داشته‌اند و این وضعیت نشان می‌دهد که این دو عامل، خطای اندازه‌گیری منظم ایجاد نکرده‌اند. در واریانس نمره مشاهده شده، نمره ارزیابی استاد روان‌شناسی بالینی و تکنولوژی آموزشی، سهم مؤلفه واریانس هدف اندازه‌گیری بسیار کم است. علت این است که مؤلفه واریانس وجه دانشجویان، بیشتر سهم واریانس را از آن خود کرده است و برای افزایش پایایی نمره‌های ارزیابی این استادان، باید سهم واریانس وجه دانشجویان به حداقل ممکن کاهش یابد.

در همه نمره‌های ارزیابی استادان روان‌شناسی کودکان استثنائی، روان‌شناسی بالینی و تکنولوژی آموزشی، خطاهای اندازه‌گیری تصادفی حدود یک پنجم نمره مشاهده شده را به خود اختصاص داده‌اند و می‌توان استنباط کرد که وجه‌های اندازه‌گیری پنهان که بتوانند تولید خطای اندازه‌گیری منظم کنند، وجود ندارد. در مقابل، خطاهای اندازه‌گیری تصادفی در نمره ارزیابی استاد مشاوره، بیش از یک سوم واریانس نمره مشاهده شده را از آن خود کرده است و این نشان می‌دهد که وجه‌های اندازه‌گیری دیگری به صورت پنهان، وجود دارند که خطای منظم تولید می‌کنند.

جدول (۴) مؤلفه‌های واریانس برآورد شده برای استادان دانشکده مدیریت و حسابداری

مؤلفه‌های واریانس برآورد شده (سهم مؤلفه در واریانس نمره کل به درصد)				منابع پراکندگی
استاد حسابداری	استاد جهانگردی	استاد مدیریت صنعتی	استاد مدیریت بازرگانی	
۰	۰	۰	۰/۰۸۷ (٪۱۱)	هدف اندازه‌گیری
۰/۸۰۶ (٪۷۳)	۱/۰۸ (٪۶۷)	۰/۴۶ (٪۶۴)	۰/۴۰ (٪۵۳)	وجه دانشجویان آشپز شده درون کلاس‌ها
۰	۰/۰۹۳ (٪۶)	۰/۰۱۳ (٪۲)	۰/۰۰۹ (٪۱)	وجه سؤال‌های پرسشنامه ارزیابی
۰/۰۲۶ (٪۲)	۰/۰۱۸ (٪۱)	۰/۰۱ (٪۱)	۰/۰۰۹ (٪۱)	وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزیابی
۰/۲۷۰ (٪۲۵)	۰/۴۲ (٪۲۶)	۰/۲۳ (٪۳۳)	۰/۲۶ (٪۳۴)	خطای اندازه‌گیری تصادفی

مؤلفه واریانس هدف اندازه‌گیری در نمره ارزشیابی استاد حسابداری، مدیریت جهانگردی و مدیریت صنعتی، صفر بود. فقدان واریانس هدف اندازه‌گیری نشان می‌دهد که در نمره‌های ارزشیابی دانشجویان، واریانس نمره واقعی مربوط به عملکرد آموزشی استاد وجود ندارد. فقط مؤلفه واریانس هدف اندازه‌گیری مربوط به استاد مدیریت بازرگانی در حدود یک دهم واریانس نمره مشاهده شده است که واریانس خیلی زیادی به شمار نمی‌آید.

مؤلفه واریانس وجه دانشجویان هر یک از استادان گروه‌های آموزشی حسابداری، مدیریت جهانگردی، مدیریت صنعتی تقریباً دو سوم واریانس نمره مشاهده شده را به خود اختصاص داده‌اند؛ به بیان دیگر، دانشجویان درون هر کلاس عملکرد استاد خود را متفاوت از دانشجویان هم‌کلاسی خود ارزشیابی کرده‌اند و نظرشان خیلی با یکدیگر اختلاف داشته است؛ در مقابل، واریانس وجه سؤال‌های پرسشنامه ارزشیابی (به غیر از استاد گروه مدیریت جهانگردی)، و همچنین وجه تعامل کلاس‌ها با سؤال‌های پرسشنامه ارزشیابی، سهم بسیار ناچیزی در واریانس نمره مشاهده شده داشته‌اند و بنابراین، این دو عامل خطای اندازه‌گیری منظم، ایجاد نکرده‌اند.

پس از برآورد مؤلفه‌های واریانس پایایی داده‌های به دست آمده برآورد شد. در چهارچوب نظریه تعمیم‌پذیری، پایایی به عنوان نسبت واریانس نمره مرجع^۱ (واریانس هدف اندازه‌گیری) به واریانس نمره مشاهده مورد انتظار^۲ تعریف می‌شود (برنان، ۲۰۰۱ الف؛ برنان، ۲۰۱۱). بر این اساس، ضریب اتکاپذیری مطالعه حاضر مطابق با مدل تدوین شده در شکل (۱) با تقسیم نمره مرجع به نمره مشاهده شده مورد انتظار (ترکیبی از نمره مرجع به علاوه واریانس‌های خطای مطلق) به صورت زیر فرمول‌بندی شد:

$$E\rho^2 = \frac{\sigma^2_c}{[\sigma^2_c] + \left[\frac{\sigma^2(s:c)}{n_s} + \frac{\sigma^2(cI)}{n_i} + \frac{\sigma^2(I)}{n_i} + \frac{\sigma^2(s:c)I}{n_s \cdot n_i} \right]}$$

^۱. universe score

^۲. expected observed score

با جای‌گذاری مقادیر مربوط به مؤلفه‌های واریانس، ضریب آتکاپذیری داده‌های ارزیابی عملکرد آموزشی استادان برآورد شد. بر این اساس ضریب آتکاپذیری میانگین نمره‌های ارزیابی استادان مورد مطالعه در جدول (۵) ارائه شده است.

جدول (۵) ضریب آتکاپذیری میانگین نمره‌های ارزیابی استادان مورد مطالعه

گروه آموزشی استادان	ضریب آتکاپذیری
زبان فرانسه	۰/۳۰
ادبیات عرب	۰/۸۰
ادبیات فارسی	۰/۶۸
زبان شناسی همگانی	۰/۳۵
حسابداری	۰
مدیریت جهانگردی	۰/۰۴
مدیریت صنعتی	۰
مدیریت بازرگانی	۰/۸۱
روان‌شناسی کودکان استثنائی	۰/۸۱
روان‌شناسی بالینی	۰/۳۳
تکنولوژی آموزشی	۰/۴۰
مشاوره	۰/۷۵

نتایج نشان می‌دهد که میانگین نمره‌های ارزیابی پنج استاد از آتکاپذیری (پایایی) لازم را دارند. بر این اساس از بین استادان دانشکده زبان و ادبیات فارسی فقط نمره ارزیابی یک استاد، از بین استادان دانشکده مدیریت و حسابداری فقط نمره ارزیابی یک استاد و از بین استادان دانشکده روان‌شناسی و علوم تربیتی نیز فقط نمره ارزیابی دو استاد، با ضریبی بیش از ۷۰ درصد آتکاپذیری لازم را داشتند. همچنین ضریب آتکاپذیری نمره‌های ارزیابی دو استاد حسابداری و مدیریت صنعتی صفر بود. به‌طور کلی بر اساس نتایج به دست آمده می‌توان بیان کرد که میانگین نمره‌های ارزیابی هشت استاد (زبان فرانسه، زبان شناسی همگانی، حسابداری، مدیریت جهانگردی، مدیریت صنعتی، روان‌شناسی بالینی و تکنولوژی آموزشی) به این علت که ضریب آتکاپذیری آنها کمتر از میزان مورد انتظار است نباید مبنای تصمیم‌گیری قرار بگیرد.

بحث و نتیجه‌گیری

هدف ارزیابی عملکرد آموزشی هر استاد از دیدگاه دانشجویان، فراهم آوردن اطلاعاتی پایا و معتبر درباره عملکرد آموزشی ایشان است. بنابراین، هدف مقاله حاضر برآورد سهم هر یک از منابع بالقوه خطای اندازه‌گیری قابل مطالعه با استفاده از نظریه تعمیم‌پذیری بود تا آشکار شود که نظرات متفاوت دانشجویان در کلاس‌های مختلف تا چه اندازه بیانگر تفاوت‌های واقعی است. سپس با مؤلفه‌های واریانس برآورد شده، میزان اتکاپذیری میانگین نظرات دانشجویان، برآورد شود تا مشخص شود که آیا نظرات آنان می‌تواند مبنای تصمیم‌گیری درباره ترفیع و ارتقای استادان قرار گیرد.

یافته‌های به دست آمده نشان داد که نظر دانشجویان نسبت به عملکرد آموزشی استاد در همه گروه‌های آموزشی تحلیل شده، اتکاپذیری (پایایی) یکسانی نداشت و هدف اندازه‌گیری به نسبت قابل قبولی اندازه‌گیری نشده بود. این نتیجه، مشابه پژوهش‌های انجام شده پیشین در سطح دانشگاه‌های کشور (عبداللهی، ۱۳۹۲؛ رئوفی و همکاران، ۱۳۸۹؛ امینی و هنردار، ۱۳۸۷) شواهدی مبنی بر ناپایا بودن نظر دانشجویان نسبت به عملکرد آموزشی استادان فراهم کرده است. اما این یافته‌ها از این جنبه که نشان می‌دهد نظر دانشجویان همیشه ناپایا نیست و در برخی از موارد می‌تواند اتکاپذیری قابل قبولی داشته باشد، تا اندازه‌ای با پژوهش‌های پیشین متفاوت است.

یافته‌های این پژوهش شرایطی را فراهم آورد تا منابعی از خطای اندازه‌گیری را شناسایی کنیم که تأثیر منفی بر پایایی نتایج ارزیابی عملکرد آموزشی استاد دارند. نتایج نشان داد که نظر متفاوت دانشجویان درون هر کلاس نسبت به عملکرد آموزشی استاد به طور متوسط، نیمی از واریانس نمره مشاهده شده را تشکیل می‌دهد. این وجه اندازه‌گیری بزرگ‌ترین مؤلفه‌های واریانس خطای منظم است و باعث کاهش پایایی نمره‌های به دست آمده شده بود. در حالی که وجه سؤال‌های پرسشنامه و همچنین تعامل بین نظر دانشجویان درون هر کلاس با سؤال‌های پرسشنامه ارزیابی سهم بسیار ناچیزی در خطای اندازه‌گیری داشت. همچنین، مشخص شد که سؤال‌های پرسشنامه مورد استفاده با یکدیگر همگن بودند و یک خصیصه واحد را اندازه می‌گرفتند. علاوه بر آن، دانشجویان کلاس‌های مختلف درک یکسانی از سؤال‌های پرسشنامه داشتند.

این نتیجه با پژوهش انجام شده توسط کیلی و همکاران (۲۰۱۰) نیز همسو است. این مطالعه شواهدی فراهم کرد، مبنی بر اینکه واریانس نمره مشاهده شده در همه گروه‌های آموزشی مورد مطالعه در بردارنده واریانس هدف اندازه‌گیری نیست. به این

معنی که در دو سوّم از گروه‌های آموزشی، نظرات دانشجویان در بین کلاس‌های مختلفی که یک استاد در آن تدریس کرده بود، نمایانگر تفاوت‌های واقعی نبوده است. پیامد این مسئله باعث کاهش اتکاپذیری (پایایی) و در برخی از موارد موجب از بین رفتن اتکاپذیری میانگین نمره‌های ارزیابی شده است (در نمره ارزیابی استادان حسابداری، مدیریت صنعتی). این یافته با پژوهش‌های انجام شده توسط کیلی و همکاران (۲۰۱۰)، ونلیوون و همکاران (۱۹۹۹)، بارنز و بارنز (۱۹۹۳)، کانه و همکاران (۱۹۷۶) تفاوت دارد؛ زیرا، این پژوهشگران حداقل اتکاپذیری میانگین نمره‌های ارزیابی دانشجویان را ۰/۶۵ برآورد کرده‌اند.

ضریب اتکاپذیری برآورد شده در چهارچوب نظریه تعمیم‌پذیری همانند ضریب پایایی برآورد شده در چهارچوب نظریه کلاسیک آزمون تفسیر می‌شود؛ بنابراین حداقل این ضریب صفر و حداکثر آن یک خواهد بود (شولسون و وب، ۱۹۹۱: ۱۴) و بهتر است تا حدّ ممکن مقدار این ضریب به عدد یک نزدیک‌تر باشد. بر اساس نتایج به‌دست آمده می‌توان بیان کرد که میانگین نمره‌های ارزیابی هشت استاد (زبان فرانسه، زبان شناسی همگانی، حسابداری، مدیریت جهانگردی، مدیریت صنعتی، روان‌شناسی بالینی و تکنولوژی آموزشی) به علّت ضریب اتکاپذیری کمتر از میزان انتظار نباید مبنای تصمیم‌گیری قرار بگیرد (کمتر از ۰/۷۰).

نظرات دانشجویان، همواره تحت تأثیر خطاهای بالقوه اندازه‌گیری قرار دارد که می‌توانند واریانس خطای منظم تولید کنند (کیلی و همکاران، ۲۰۱۰؛ دهقانی و نخعی، ۱۳۹۱). نادیده گرفتن چنین خطاهایی هنگام محاسبه پایایی، باعث عدم اندازه‌گیری هدف مورد مطالعه می‌شود. در این باره، با تفسیر مؤلفه‌های واریانس در هر یک از نمره‌های ارزیابی استادان نشان دادیم که هر وجه اندازه‌گیری، چند درصد در واریانس نمره مشاهده شده اثر داشته و چند درصد واریانس نمره مشاهده شده مربوط به هدف اندازه‌گیری است. بنابراین، اگر برای برآورد پایایی داده‌های مطالعه حاضر از نظریه تعمیم‌پذیری استفاده نمی‌شد، فرض‌های موجود درباره میزان تغییرپذیری نتایج بر حسب سؤال‌های پرسشنامه ارزیابی عملکرد آموزشی، در حدّ یک فرضیه باقی می‌ماند. اما برآورد مؤلفه‌های واریانس، این امکان را فراهم کرد تا مشخص شود که هر یک از وجه‌های اندازه‌گیری تا چه اندازه باعث کاهش پایایی داده‌ها شده‌اند.

نظریه تعمیم‌پذیری برای برآورد میزان اتکاپذیری نمره‌های ارزیابی عملکرد آموزشی استاد بسیار توانمند است و مطالعه حاضر هماهنگ با پژوهش‌های پیشین (ابراهیم،

۲۰۱۱؛ کیلی و همکاران، ۲۰۱۰؛ ونلیوون، دورمودی و سیورس، ۱۹۹۹؛ بارنز و بارنز، ۱۹۹۳؛ کانه، گلیمور و کروکس، ۱۹۷۶) این مسئله را تأیید می‌کند. بنابراین به ارزشیابان آموزشی شاغل در دانشگاه‌ها و مؤسسات آموزشی عالی کشور پیشنهاد می‌شود که برای برآورد میزان اتکاپذیری نمره‌های جمع‌آوری شده مربوط به نظر دانشجویان درباره عملکرد آموزشی استادان، از این نظریه استفاده کنند. طرح‌های موجود در نظریه تعمیم‌پذیری، ارزشیاب آموزشی را قادر می‌سازد که علاوه بر وجه‌های اندازه‌گیری مطالعه شده در این پژوهش، سایر وجه‌های اندازه‌گیری احتمالی اثرگذار بر میزان اتکاپذیری نمره‌ها (مانند ویژگی‌های دانشجو مانند جنسیت، پایه تحصیلی و نمره مورد انتظار، ساعت تشکیل کلاس، نوع درس) را شناسایی کند تا برای افزایش اتکاپذیری داده‌ها، تغییرات مورد نیاز در شرایط یا موقعیت اندازه‌گیری ایجاد شود.

نظریه تعمیم‌پذیری علاوه بر برآورد میزان اتکاپذیری به ارزشیاب آموزشی کمک می‌کند تا برای بهبود آن، طرح اندازه‌گیری مناسبی را طراحی کند. برای این منظور ارزشیاب آموزشی با استفاده از مؤلفه‌های واریانس برآورد شده برای وجه‌های اندازه‌گیری (برای نمونه تعداد ارزیابان یا تعداد سؤال‌ها) و تعامل آنها با هدف مورد مطالعه، می‌تواند سهم هر یک از منابع خطای منظم را در موقعیت اندازه‌گیری شناسایی کند. سپس برای بهبود سیستم ارزیابی عملکرد استادان مشخص کند که تغییر در کدام یک از وجه‌های اندازه‌گیری، باعث کاهش خطای منظم و افزایش پایایی خواهد شد.

پرسش‌هایی در خلال پژوهش حاضر برای پژوهشگر ایجاد شده است که به علاقه‌مندان حوزه ارزیابی عملکرد آموزشی پیشنهاد می‌شود که آنها را در آینده، مورد پژوهش قرار دهند. نخست آنکه در این مطالعه در گروه‌های آموزشی که نظرات دانشجویان آنها پایایی لازم را داشت، میزان خطای منظم تفکیک نشده که با خطای تصادفی آمیخته شده، بیش از چهل درصد بود. بنابراین پیشنهاد می‌شود که در پژوهش‌های آینده، علاوه بر وجه دانشجویان درون کلاس از وجه‌های اندازه‌گیری دیگری مانند جنسیت و معدل تحصیلی دانشجویان و نوع درس تدریس شده توسط استاد نیز استفاده شود تا سهم آنها در خطای اندازه‌گیری منظم شناسایی شود. علاوه بر این، در پژوهش‌های آینده، داده‌های حاصل از نظرات دانشجویان مقطع کارشناسی

ارشد و مقطع دکترا نیز تحلیل شود و نتایج به دست آمده با آنچه در این پژوهش به دست آمد، مقایسه شود.

به منظور افزایش اتکاپذیری نظرات دانشجویان دانشگاه علامه طباطبائی تهران درباره عملکرد آموزشی استادان خود، چند راهکار بر مبنای نتایج به دست آمده پیشنهاد می‌شود. نخست آنکه میزان اتکاپذیری نمره ارزیابی دانشجویان نسبت به هر یک از استادان برآورد شود و چنانچه ضریب اتکاپذیری، کمتر از میزان قابل قبول (دست کم ۰/۷۰) به دست آید، نظر دانشجویان به هیچ‌وجه مبنای ارتقا یا تمدید قرارداد استادان مدعو یا پیمانی قرار نگیرد. علاوه بر این، آشکار شد که میزان اتکاپذیری نظرات دانشجویان در برخی از دانشکده‌ها بیشتر است؛ بنابراین انجام تحقیق کیفی لازم است تا آشکار شود که چرا دانشجویان برخی از دانشکده‌ها نظرات پایاتری در مقایسه با سایر دانشجویان دارند. برای یافتن پاسخ دقیق، نیاز است که پژوهش‌های دقیقی درباره نحوه اجرا و نگرش دانشجویان نسبت به نحوه کاربست نتایج ارزیابی و همچنین میزان اعتماد دانشجویان به محرمانه بودن نظرات‌شان انجام شود تا راهکارهای موثرتری برای بهبود نظام ارزیابی عملکرد استادان شناسایی شود.

سرانجام لازم است به این نکته اشاره شود که نتایج به دست آمده در این پژوهش، باید با احتیاط به موقعیت‌های دیگر (نظرات سایر دانشجویان در رشته‌های دیگر و مقاطع کارشناسی ارشد و دکترا) تعمیم داده شود؛ زیرا نتایج حاصل از نمونه مورد مطالعه، فقط شامل نظرات دانشجویان کارشناسی درباره عملکرد آموزشی ۱۲ استاد بود و علاوه بر این، نمونه آماری آن هم از سه دانشکده مربوط به یک دانشگاه انتخاب شده است.

منابع

- بیگدلی، ایمان‌الله (۱۳۷۹). بررسی اعتبار پرسشنامه ارزیابی دانشجویان از اساتید. *فصلنامه روانشناسی*، ۱۶ (۴)، ۴۰۵-۴۱۶.
- رئوفی، شهین؛ شیخیان، علی؛ ابراهیم‌زاده، فرزاد؛ طراحي، محمد جواد و احمدی، پروانه (۱۳۸۹). طراحی فرم جدید ارزیابی کیفیت تدریس نظری اساتید بر اساس دیدگاه‌های ذینفعان و اصول شش گانه دانش پژوهی کلاسیک. *فصلنامه پزشکی هرمزگان*، ۱۴ (۳)، ۱۶۷-۱۷۶.
- امینی، میترا و هنردار، مرضیه (۱۳۸۷). ارزیابی استاد از دیدگاه اساتید و دانشجویان دانشگاه علوم پزشکی جهرم. *فصلنامه دانشگاه علوم پزشکی سمنان*، ۹ (۳)، ۱۷۱-۱۷۷.
- امدادی، شهره؛ امانی، فیروز؛ سلطانیان، علیرضا؛ ایمنی، بهزاد؛ مقصود، امیرحسین؛ شجاعی، ستاره و دیگران (۱۳۹۲). مطالعه روایی و پایایی فرم ارزیابی استاد توسط دانشجو در درس تئوری و برخی عوامل مؤثر بر ارزیابی دانشجو. *فصلنامه گام‌های توسعه در آموزش پزشکی*، ۱۰ (۱)، ۸۷-۹۴.
- باستانی، پیوند؛ روح‌اللهی، نیره‌سادات و طاهرنژاد، علی (۱۳۹۳). بررسی روایی و پایایی فرم‌های ارزیابی اساتید از دیدگاه دانشجویان دانشگاه علوم پزشکی تهران. *دو فصلنامه آموزش پزشکی*، ۳ (۱)، ۷-۱۴.
- عبداللهی، حسین (۱۳۹۲). تحلیل سازوکار ارزیابی عملکرد آموزشی اعضای هیئت علمی: مطالعه موردی دانشگاه علامه طباطبائی. *فصلنامه مطالعات و ارزیابی آموزشی*، ۳ (۳)، ۱۰۱-۱۲۶.
- دهقانی، مسعود و نخعی، نوذر (۱۳۹۱). ارزیابی استاد توسط دانشجو: مروری بر نقدها. *فصلنامه گام‌های توسعه در آموزش پزشکی*، ۹ (۲)، ۱۰۲-۱۰۹.

Barnes, L. L. B., & Barnes, M. W. (1993). Academic discipline and generalizability of student evaluations of instruction. *Research in Higher Education*, 34 (2), 135-149.

- Brennan, R. L. (2011). Generalizability Theory and classical test theory. *Applied Measurement in Education*, 24, 1–21.
- Brennan, R. L. (2001a). *Generalizability Theory*. New York: Springer-Verlag.
- Brennan, R. L. (2001b). Manual for urGENOVA. Iowa City, IA: Iowa Testing Programs, University of Iowa.
- Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG*. New York, NY: Routledge.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cronbach, L. J., Raj aratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.
- Ibrahim, A. M. (2011). Using generalizability theory to estimate the relative effect of class size and number of items on the dependability of student ratings of instruction. *Psychological reports*, 109 (1), 252-258.
- Kane, M. T., Gillmore, G. M., & Crooks, T. J. (1976). Student evaluations of teaching: The generalizability of class means. *Journal of Educational Measurement*, 13, 171-183.
- Keeley, J., Furr, R. M., & Buskist, W. (2010). Differentiating psychology student's perceptions of teachers using the teacher behavior checklist. *Journal of Teaching of Psychology*, 37, 16–20.
- Morgeson, F. P., Campion, M. A. (2000). Accuracy in job analysis: toward an inference-based model. *Journal of Organizational Behavior*, 21, 819-827.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer: A Primer (Vol. 1)*. Sage Publications.
- Shavelson, R.J., Webb, N.M., Rowley, G.L. (1989). Generalizability Theory. *American Psychological Association*, 44(6), 922-932.
- Sun, A., Valiga, M. J., & Gao, X. (1997). Using generalizability theory to assess the reliability of student ratings of academic advising. *The Journal of experimental education*, 65 (4), 367-379.
- VanLeeuwen, D.M., Dormody, T.J., Seevers, B.S. (1999). Assessing the Reliability of Student Evaluations of Teaching (SETS) with Generalizability Theory. *Journal of Agricultural Education*, 40 (4), 1-9.

