

Validity and Reliability of Student's Evaluation of Educational Quality Islamic Azad University

Vaziri, Shahram

The Major goal of the present research is to study the validity and reliability of students Evaluation of Educational Quality. The Samples consist of all the students of different fields of Islamic Azad University Roodchen branch. 319 (180 Female & 239 male) students were randomly selected from different fields. The tool of the research was SEEQ. The obtained results of data analysis showed that this questionnaire has no validity and reliability for evaluating.

اعتبار و پایایی پرسشنامه ارزیابی اعضای هیات علمی

شهرام وزیری^۱

چکیده: هدف پژوهش حاضر بررسی اعتبار و پایایی فرم از پیش آماده‌شده‌ی نظرسنجی دانشجویان نسبت به اساتید دانشگاه است. از آنجایی که هدف این پژوهش بررسی اعتبار و پایایی فرم ارزشیابی اعضای هیات علمی دانشگاه آزاد اسلامی رودهن بوده است، دانشجویان دانشگاه آزاد اسلامی رودهن در رشته‌های مختلف به عنوان نمونه‌ی پژوهش، در نظر گرفته شدند. پاسخ‌های ۳۱۹ نفر (۱۸۰ نفر دختر و ۱۳۹ نفر پسر) مورد تجزیه و تحلیل قرار گرفت. بررسی‌های آماری حکایت از آن دارند، که فرم ارزیابی نمونه‌های معرف از حوزه‌ی رفتار مورد اندازه‌گیری را دربر نمی‌گیرد.

مقدمه

دانشگاه محل رسمی آموزش، کانون پرورش نخبگان، برگزیدگان و کارگزاران آتیه‌ی کشور است. دانشگاه به عنوان مرکز تحقیق و پژوهش علمی و فرهنگی، مهد نوسازی اجتماع و محلی است که در آن افکار و اندیشه‌های جدید، دانش و معرفت نو، تکنولوژی متحول و فرهنگی خلاق و پویا رشد و گسترش می‌یابد. یک موقعیت دانشگاهی حداقل سه عنصر لازم دارد: معنا و شرایط فیزیکی (آموزشگاه) کسی که علمی را منتقل می‌کند

^۱ - Ph.D ، عضو هیات علمی دانشگاه آزاد اسلامی ایران واحد رودهن

(آموزگار)، کسی که به دنبال یادگیری علم است (دانشجو یا طلبه). هر سه جنبه یاد شده باید دایما مورد بررسی و ارزشیابی قرار گیرند تا شرایط تولید علم مهیا شود. ارزشیابی فضا و شرایط فیزیکی آموزشگاه فهم مطالبات نوعی آن، درک تفاوت‌های بنیادی آن نسبت به دبستان، راهنمایی و دبیرستان و اسباب و ادوات الزامی در آنها و در مجموع ارزشیابی سه عنصر موجود در موقعیت دانشگاهی به‌خصوص ارزیابی عملکرد آموزگاران و تولیدکنندگان علم می‌تواند در به‌روز نگهداشتن صنعت «تولید علم»، موثر باشد.

زمانی که ما از توانایی ارزشیابی دقیق حرف می‌زنیم حداقل دو عنصر را باید مد نظر داشته باشیم. اولی پایایی^۱ وسیله است که به ثبات و پایداری اندازه‌های گرفته شده مربوط می‌شود و دومی به اعتبار^۲ وسیله باز می‌گردد و نشان می‌دهد که آیا وسیله‌ی اندازه‌گیری مورد نظر، آنچه را که ادعای اندازه‌گرفتن آن را دارد، به واقع اندازه‌گیری می‌کند یا خیر؟ مقیاس‌ها یا ابزارهای اندازه‌گیری علاوه بر پایایی و اعتبار از نظر سودمندی^۳ یا ارزش عملی نیز متفاوتند و باید از این نظر نیز مورد قضاوت قرار گیرند تا معلوم شود که آیا ابزار مورد نظر اطلاعات ارزشمندی را فراهم می‌سازند یا خیر؟ زیرا می‌خواهیم با تکیه بر این موارد بدانیم یک شخص یا یک شی‌چه مقدار از یک ویژگی را دارد (گی،^۴ ۱۹۹۱).

هر اندازه‌گیری فرایند یا فعالیتی نظام‌دار و دربرگیرنده‌ی قواعدی است (نانالی،^۵ ۱۹۷۸) که طی نوعی فعالیت و از طریق وسیله‌ای معین که به آن وسیله‌ی اندازه‌گیری می‌گویند، ما را به اهداف تعیین‌شده‌ای رهنمون می‌سازد. از آنجایی که متخصص اندازه‌گیری معمولاً به ویژگی مورد نظر دسترسی مستقیم ندارد ناچار است، به رفتارها یا نمودهایی که به گمان وی معرف ویژگی مورد نظر اوست، مراجعه کند. سنجش^۶ نیز، نوعی اندازه‌گیری است که به تحلیل جامع و چندوجهی از عملکرد می‌پردازد (ویگینز،^۷ ۱۹۹۳). ارزشیابی یا

1. Reliability

2. Validity

3. Unility

4. Gay

5. Nunnally

6. Assessment

7. Wiggins

ارزیابی^۱ به طور ساده، به تعیین ارزش برای هر چیزی یا دآوری ارزشی گفته می‌شود. ارزشیابی فرآیندی نظام‌دار برای جمع‌آوری، تحلیل و تفسیر اطلاعات است و تعیین می‌کند که آیا هدف‌های مورد نظر تحقق یافته‌اند یا در حال تحقق یافتن هستند (گی، ۱۹۹۱).

از نظر ورتن و سندرز^۲ (۱۹۸۷)، ارزشیابی فعالیتی رسمی است که برای تعیین کیفیت اثربخشی، یا ارزش یک برنامه، فرآورده، پروژه، فرایند، هدف، یا برنامه‌ی درسی به اجرا در می‌آید. علیرغم اینکه ارزشیابی اساتید از نظر اسمی در این مقوله قرار می‌گیرد ولی واقعیت آن است که با آن متفاوت است، چون فرآیند متفاوتی را پی گرفته است. از طرف دیگر، این موضوع ممکن است در حیطه‌ی سنجش بوده باشد که در ظاهر امر نیز چنین است، زیرا پرسشنامه‌ی ارزیابی عملکرد اساتید، براساس دستورالعمل‌های مطرح آن، تلاش دارد تا به تحلیل جامع و چندوجهی از عملکرد اساتید بپردازد. اما، آیا ساختار خاص سوال‌ها به گونه‌ای هست که بتواند، از چنین چارچوبی دفاع کند؟

اگر با کمی اغماض به موضوع بنگریم، فرم ارزیابی، نوعی اندازه‌گیری ویژگی‌های خاص است که از طرف طراحان آن، معرف ویژگی‌های یک مدرس ایده‌آل به حساب آمده، اما متأسفانه این چارچوب هیچ‌گونه منبع و مرجعی ندارد، طراحان آن هیچ‌گونه نشانه‌ای از مولفه‌های اساسی و روان‌سنجی آن ارائه نکرده‌اند، در مورد پایایی آن، هیچ یافته‌ای در دست نداریم و در بعد اعتبار، چه از نظر ملاک مرجع و یا نرم مرجع، نشانه یا توضیحی ارائه نشده است. در صورتی که چنین پرسشنامه‌هایی صرفاً ابزاری آزمایشی یا پیمایشی باشند مشکلی نخواهیم داشت. ولی مساله این است که، این پرسشنامه براساس بخشنامه‌ی سازمان مرکزی به عنوان عنصری قوی در وضعیت و ارتقا مدرسین در نظر گرفته شده است و گاهی اوقات مبنایی برای معرفی اساتید موفق و برجسته تلقی می‌شود و درست همین امر، نیاز جدی به بررسی علمی آن را مطرح می‌سازد.

براین اساس و با توجه به آنچه گذشت ما در این پژوهش با این سوال اصلی مواجه هستیم که: آیا فرم نظرسنجی عملکرد اساتید از پایایی، اعتبار و سودمندی کافی برخوردار است؟

1. Evaluation

2. Sanders & Worthen

همان‌گونه که سیف (۱۳۷۶) نوشته است، مهم‌ترین ارزش‌گذاری به معلم را می‌توان بهبود روش‌های آموزشی و کمک به مدیران سازمان‌های آموزشی، برای تصمیم‌گیری درست‌تر در استخدام و ترفیع و مانند اینها دانست. از آنجایی که معلم همواره از عملکرد بخش‌های آموزشی تاثیر می‌پذیرد، ارزشیابی باید به صورت سیستماتیک و با مشخص کردن نقاط قوت و ضعف فعالیت‌های کارکنان یک موسسه آموزشی اجرا شود. در غیر این صورت راه برای اعتراض اعضای هیات‌های علمی و آموزشی باز خواهد بود.

استفاده از نظرهای دانشجویان برای ارزشیابی از درس و فعالیت‌های آموزشی معلمان روشی بسیار معمول و شناخته شده است. اما، به رغم کاربرد فراوان آن، شواهد پژوهشی نشان می‌دهند که اظهارنظر شاگردان درباره‌ی کار معلمان، متاثر از عواملی است که با موضوع ارزشیابی، ارتباط چندانی ندارند. از نظر دونکین (۱۹۸۶)، پژوهش‌های انجام‌شده به وسیله‌ی روان‌شناسان در چند دهه‌ی اخیر به روشنی نشان داده‌اند که صفات شخصی و ویژگی‌های کلی محیط، ادارک‌ها و قضاوت‌های افراد را تحت تاثیر قرار می‌دهند و دلیلی وجود ندارد که انتظار داشته باشیم دانشجویان و دانش‌آموزان، در ارزشیابی خود از معلمان، از این نوع تاثیرها مصون باشند.

پژوهش‌های زیادی، رابطه‌ی بین اظهارنظر شاگردان و فرآیندهای آموزشی معلمان را بررسی کرده‌اند. مشهورترین این پژوهش‌ها، مطالعات معروف به «اثر دکتر فاکس»^۲ یا «اغواگری آموزشی»^۳ است. پژوهشگران پی برده‌اند که شنوندگان یک سخنرانی اغواکننده، ممکن است آن را خوب و مفید بدانند و فکر کنند که از آن چیزی آموخته‌اند، حتی اگر از نظر محتوا بسیار ضعیف باشد (نفقولین، ویر و دانلی،^۴ ۱۹۷۳). ویر و یلیامز^۵ (۱۹۸۰) در پژوهشی، علاوه بر سنجش نظر شنوندگان در مورد مطلوب بودن یا نامطلوب بودن سخنرانی، با استفاده از یک آزمون عینی چند گزینه‌ای، میزان یادگیری شنوندگان از سخنرانی را نیز، اندازه‌گیری کردند. آنها دریافتند استفاده از نظرخواهی‌های

1. Dunkin

2. Dr. Fox Effect

3. Educational Seduction

4. Naftulin; Ware & Donnelly

5. Ware & Williams

دانشجویان برای تصمیم‌گیری‌های مربوط به استخدام، تبدیل وضع استخدام آزمایشی به قطعی و ارتقای معلمان، ممکن است عملی ناروا باشد. معلمانی که با فوت‌وفن روش دکتر فاکس یا اغواگری آموزشی آشنا هستند ممکن است در نظرسنجی‌های شاگردان خود نمره‌های بالایی بگیرند، بدون اینکه از موضوع‌های درسی اطلاع کافی داشته باشند یا به شاگردان خود چیز زیادی بیاموزند.

ابرامی، لونتال و پری^۱ (۱۹۸۲)، یافته‌های این پژوهش‌ها را با روش فرا تحلیل مورد بازبینی قرار داده و به این نتیجه رسیدند که تاثیر روش آموزش یا طرز رایجی در درس بر اظهارنظر شاگردان نسبت به درس و روش معلم، معنادار است در حالی که تاثیر محتوای درس یا مطالب آموزش داده شده بر اظهارنظر شاگردان، این-گونه نیست. به طور کلی، می‌توان این نتایج را چنین خلاصه کرد: «ارزشیابی‌های دانشجویی، بیش از آنکه نشان‌دهنده تغییرات محتوایی باشند، نشان‌دهنده تغییرات روش آموزشی معلم یا طرز رایجی درس به وسیله‌ی او هستند» (دونکن، ۱۹۸۶).

ویژگی دیگر اظهارنظرهای شاگردان درباره‌ی معلم که موجب پایین آمدن اعتبار این روش ارزشیابی می‌شود، این است که در آن جزئیات دقیق فرایند آموزشی منعکس نمی‌شود. این اظهارنظرها غالباً تحت تاثیر عوامل کلی، از جمله شهرت معلم و ویژگی‌های دیگر چون سخت و آسان‌گیری و طرز برخورد او، قرار می‌گیرند. دانشجو، پس از کسب اطلاعات کلی درباره‌ی معلم یا بعد از برخورد اولیه با او، درباره‌ی او نگرشی کسب می‌کند و از آن پس بر این اساس نسبت به او اظهارنظر می‌نماید و صرف نظر از نوع سوالی که درباره‌ی معلم به آن جواب می‌دهد، همان نگرش کلی خود را ابراز می‌دارد.

1. Abrami, Leventhal & Perry

سیف (۱۳۵۶)، پس از جمع‌آوری نظر دانشجویان درباره‌ی کار معلمان، آنها را در اختیار معلمان مورد ارزشیابی قرار داد و از آنان خواست تا واکنش خود را نسبت به ارزشیابی دانشجویان خود ابراز دارند. گروهی از معلمان گفتند، با اینکه در راهنمای پرسشنامه‌های نظرخواهی دانشجویان، از آنان خواسته شده بود تا هنگام ارزشیابی از کار معلمان، سوال‌هایی را که در مورد درس و معلم به خصوص، مناسب‌ند ندارند بی‌جواب بگذارند، دانشجویان در بسیاری از موارد این موضوع را رعایت نکردند. برای نمونه، بعضی از معلمان می‌گفتند گرچه آنان در درس‌های مورد ارزشیابی هیچ‌گونه امتحانی از دانشجویان به عمل نیاورده بودند و فرم ارزشیابی پیش از امتحان پایان نیمسال اجرا شده بود با این حال، دانشجویان در سوال‌هایی که درباره‌ی امتحان طرح شده بود، معلم را مورد ارزشیابی قرار داده و به او نمره داده بودند.

روش

جامعه و نمونه‌ی پژوهش

همان‌گونه که می‌دانیم، جامعه جمعیتی کلی است که، تعمیم یافته‌ها متوجه اوست و جمعیت، گروه مشخصی است که شرکت‌کنندگان در بررسی از آن انتخاب می‌شوند و سرانجام، نمونه زیرمجموعه‌ای از جمعیت مورد نظر و شامل شرکت‌کنندگانی است که، در بررسی شرکت می‌کنند.

از آنجایی که هدف این پژوهش، بررسی اعتبار و پایایی فرم ارزشیابی اعضای هیات علمی و راههای بهینه کردن آن است، دانشجویان دانشگاه آزاد اسلامی رودهن، در رشته‌های مختلف به عنوان نمونه‌ی پژوهش در نظر گرفته شده‌اند.

براساس آمار مدیریت آموزشی دانشگاه، واحد رودهن در سال تحصیلی ۸۴-۸۳ تعداد ۱۷۱۷۴ دانشجو داشته است. برای نمونه‌گیری از این تعداد، ابتدا حداقل حجم نمونه، مورد محاسبه قرار گرفت (۲۹۳)، سپس نظر به اینکه کلاس‌های دانشگاه از

۱۵ تا بیش از ۶۰ نفر ظرفیت دارند، ۱۰ کلاس به طور تصادفی مورد انتخاب و در مجموع ۵۲۶ نفر انتخاب شدند. در آخر کار، بعد از حذف پاس‌نامه‌های مخدوش یا بدون رمز و یا غیر کامل، پاسخ‌های ۳۱۹ نفر (۱۸۰ نفر دختر و ۳۹ نفر پسر)، مورد تجزیه و تحلیل قرار گرفت.

ابزار پژوهش

فرم ارزشیابی اعضای هیات علمی، دارای ۱۰ ماده است که دانشجویان به هر یک از آنها از صفر تا ۹، نمره می‌دهند. مواد این پرسشنامه عبارتند از: دعوت به شیونات اسلامی؛ نظم استاد از نظر حضور و ترک به موقع کلاس؛ نحوه‌ی اداره‌ی کلاس؛ قدرت بیان و تفهیم مطالب؛ میزان تسلط بر مطالب گفته شده در سر کلاس؛ میزان نظم و ترتیب مطالب؛ مقایسه با محتوای درسی طبق سرفصل؛ میزان علاقه و توجه استاد به فراگیری دانشجو و پاسخ به سوالات دانشجو؛ میزان توجه استاد به ارزیابی دانشجویان؛ نحوه‌ی برخورد با دانشجو و اداره‌ی کلاس. برای انجام ارزیابی، دانشجویان با گرفتن شماره‌ی تلفن، وارد برنامه‌ی کامپیوتری می‌شوند که نام هر استاد را خوانده و از دانشجو می‌خواهد با فشار دادن دکمه‌های تلفن، امتیاز خاصی را در مورد وی انتخاب نماید. میانگین این میزان‌ها در مورد هر استاد، محاسبه و سرانجام نمره‌های صفر تا ۹ در کارنامه‌ی استاد درج می‌شود.

روش اجرا

برای اجرای پژوهش حاضر، ابتدا ۱۰ کلاس به صورت تصادفی انتخاب شدند. با حضور در هر کلاس ورقه‌ای در اختیار دانشجویان قرار گرفت و به آنها گفته شد که پژوهشی در مورد سیستم یونی کام دانشگاه انجام می‌دهیم و از آنها خواسته شد

تا در صورتی که مایل به شرکت در این پژوهش هستند، با شماره تلفنی که برای این کار در نظر گرفته شده، تماس بگیرند.

بعد از تماس، از آنها درخواست می‌شد تا نظر خود را نسبت به استاد خود (در همان کلاسی که انتخاب شده بودند) در پرسشنامه وارد کنند. برای این کار، آنها ابتدا رمزی را در محل مربوط در صفحه اول می‌نوشتند و سپس با گوش دادن به نوار ارزیابی شبیه روش اصلی، با نوشتن شماره ی ۰ تا ۹، نظر خود را در مورد همان استاد اعلام می‌کردند. پرسشنامه‌ای که در اختیار دانشجویان قرار می‌گرفت، شامل فرم ارزیابی و پنج سوال در رابطه با نحوه ی پاسخ‌گویی به این ارزیابی بود. بعد از پاسخ‌گویی، با گذاشتن یک قرار برای یک ماه دیگر از آنها می‌خواستیم تا در تاریخ تعیین شده، در محل حضور یابند. بعد از یک ماه، در صورت حضور از آنها خواسته می‌شد تا همان رمزی را که بار اول در صفحه نوشته بودند در برگه پرسشنامه وارد نمایند و سپس مانند مرحله ی نخست همان استادی را که یک ماه قبل ارزیابی کرده بودند، دوباره ارزیابی کنند. در تمام مراحل به دانشجویان توضیح داده می‌شد که هر چند پاسخنامه‌ها بدون اسم هستند با این حال، اطلاعات دریافتی و اسامی گروه نمونه، کاملاً محرمانه تلقی خواهند شد.

یافته‌ها

مقصود از پایایی، درجه ی ثبات نمره‌هاست (اناستازی، ۱۳۶۱). نظر به اینکه، تمام انواع پایایی مربوط به درجه ی ثبات یا توافق بین دو سلسله نمره است که به صورت مستقل به دست می‌آیند، همه ی آنها را می‌توان با ضرایب همبستگی نشان داد.

ضریب همبستگی، نشان‌دهنده‌ی درجه‌ی همخوانی یا رابطه‌ی بین دو سلسله‌نمره هستند. ضریب همبستگی را با توجه به نوع داده‌ها، می‌توان با روش‌های مختلف حساب کرد که رایج‌ترین آنها ضریب همبستگی گشتاوری پیرسون است. در محاسبه‌ی این ضریب همبستگی، نه تنها جایگاه فرد در گروه بلکه، انحراف نمره‌ی او از میانگین گروه نیز منظور می‌شود.

ساده‌ترین روش برای پیدا کردن پایایی نمره‌های یک آزمون این است که آن را یکبار دیگر اجرا کنیم. در این مورد ضریب پایایی، همبستگی بین نمره‌های حاصل از دو اجرای آزمون واحد، در گروهی واحد خواهد بود. این پایایی، نشان می‌دهد که تا چه اندازه می‌توان نمره‌های آزمون را به موقعیت‌های مختلف تعمیم داد.

در جدول شماره ۱، میانگین و انحراف معیار هر یک از ماده‌های ارزیابی اساتید، در دو بار اجرای این فرم با فاصله‌ی یک ماه، ارایه شده است. همان‌گونه که در جدول شماره ۲ مشاهده می‌شود ماده‌های این فرم در دو بار اجرا همبستگی‌های بسیار پایینی با یکدیگر نشان می‌دهند. به طوری که ملاحظه می‌شود میانگین این همبستگی‌ها ۰/۱۲۲ است. معمولاً در پاره‌ای از آزمون‌ها، مانند مقیاس‌های نگرش‌سنج یا همین فرم ارزیابی اساتید، برای سوال‌های آزمون، پاسخ درست یا غلط، وجود ندارد و از آزمودنی خواسته می‌شود تا درجه‌ی مورد نظر خود را برای هر ماده‌ای، ثبت یا انتخاب نماید. در این حالت هیچ یک از این پاسخ‌ها به خودی خود درست یا غلط نیستند، بلکه پاسخ آزمودنی به این گزاره یا اظهار نظر، موضع او را روی پیوستاری که از موافقت کامل تا مخالفت کامل را دربرمی‌گیرد، نشان می‌دهد. در این شرایط می‌توان از فرمول آلفای کرونباخ استفاده کرد.

جدول شماره ۱: میانگین و انحراف معیار نمره‌های آزمودنی‌ها به مواد فرم ارزیابی در دوبار اجرا

میانگین	انحراف معیار	ماده‌ها
۴/۴۵	۲/۰۱	میزان دعوت به شیونات اسلامی اجرای اول
۴/۱۳	۱/۷۸	میزان دعوت به شیونات اسلامی اجرای دوم
۴/۴۶	۲/۳۰	نظم استاد از نظر حضور و غیاب و ترک به موقع کلاس اجرای اول
۴/۶۸	۱/۸۵	نظم استاد از نظر حضور و غیاب و ترک به موقع کلاس اجرای دوم
۴/۳۰	۱/۷۸	نحوه ی اداره ی کلاس اجرای اول
۴/۵۶	۲/۰۷	نحوه ی اداره ی کلاس اجرای دوم
۴/۳۰	۱/۹۲	قدرت بیان و تفهیم مطالب اجرای اول
۴/۵۶	۱/۷۶	قدرت بیان و تفهیم مطالب اجرای دوم
۴/۲۸	۲/۲۷	میزان تسلط بر مطالب گفته‌شده در کلاس اجرای اول
۴/۷۱	۱/۸۲	میزان تسلط بر مطالب گفته‌شده در کلاس اجرای دوم
۴/۵۸	۱/۹۴	میزان نظم و ترتیب مطالب اجرای اول
۴/۳۲	۱/۶۴	میزان نظم و ترتیب مطالب اجرای دوم
۴/۱۹	۱/۸۶	تناسب منابع با محتوای درسی طبق سر فصل اجرای اول
۴/۴۳	۲/۱۴	تناسب منابع با محتوای درسی طبق سر فصل اجرای دوم
۴/۰۹	۱/۷۲	میزان علاقه و توجه استاد به فراگیری دانشجو و پاسخ به سوالات دانشجو اجرای اول
۴/۴۰	۱,۷۵۳۷	میزان علاقه و توجه استاد به فراگیری دانشجو و پاسخ به سوالات دانشجو اجرای دوم
۴/۸۵	۲/۴۸	میزان توجه استاد به ارزیابی دانشجویان اجرای اول
۴/۶۳	۱/۹۲	میزان توجه استاد به ارزیابی دانشجویان اجرای دوم
۴/۱۳	۱/۷۷	نحوه ی برخورد با دانشجو و اداره ی کلاس اجرای اول
۴/۱۱	۱/۷۷	نحوه ی برخورد با دانشجو و اداره ی کلاس اجرای دوم

جدول شماره ۲: همبستگی نمره‌های مواد فرم ارزیابی در دوبار اجرا

۲	۲(۱)	۲(۲)	۲(۳)	۲(۴)	۲(۵)	۲(۶)	۲(۷)	۲(۸)	۲(۹)	۲(۱۰)
۱(۱)	°-/۱۳۹	-/۰۲۴	-/۰۹۲	-/۰۲۱	°°-/۳۷۷	-/۰۷۹	-/۰۸۱	/۰۰۹	-/۰۷۷	°/۱۱۷
۱(۲)	-/۰۰۹	°/۰۴۵	/۰۴۱	/۰۱۶	/۱۰۲	-/۰۵۴	/۰۰۲	-/۰۳۷	/۰۸۶	-/۰۱۹
۱(۳)	°°/۵۶۱	°/۰۷۷	°/۱۲۰	/۰۵۰	°°/۱۵۰	/۰۳۳	/۰۰۵	/۰۲۲	°-/۱۲۸	/۰۵۶
۱(۴)	/۰۳۲	°°-/۱۸۱	/۰۷۸	°°/۳۰۰	-/۰۷۸	-/۰۱۹	-/۰۰۴	-/۰۲۲	°°/۱۸۳	-/۰۵۹
۱(۵)	/۰۸۵	°°/۱۹۸	-/۰۲۶	/۰۱۵	/۰۵۹	-/۰۲۳	-/۱۴۴	-/۰۴۳	°°-/۲۵۷	-/۰۸۲
۱(۶)	-/۰۱۵	/۰۱۹	°°-/۱۶۳	/۰۳۴	-/۰۰۴	°°-/۱۷۳	/۰۰۷	-/۰۸۴	/۰۳۸	/۰۴۷
۱(۷)	-/۰۹۳	°°-/۱۶۹	/۰۰۲	/۰۲۱	-/۰۶۱	-/۰۶۸	/۱۲۰	/۱۰۰	°°/۱۷۴	-/۰۵۱
۱(۸)	-/۰۱۵	/۰۴۴	/۰۰۷	°-/۱۴۰	-/۰۲۳	/۰۲۵	/۰۵۲	°/۱۱۹	°-/۱۱۰	/۰۳۹
۱(۹)	/۰۱۱	-/۰۸۲	-/۰۴۹	-/۰۰۷	°/۱۱۶	/۰۱۰	/۰۶۵	/۰۷۰	°/۱۱۳	-/۰۸۸
۱(۱۰)	-/۰۱۴	/۰۴۶	/۰۱۳	-/۱۱۵	°-/۱۳۷	-/۰۲۵	-/۰۴۹	-/۰۵۵	/۰۳۲	/۰۳۵

* Correlation is significant at the 0/05 level (2-tailed).

** Correlation is significant at the 0/0 level (2-tailed).

آلفای کرونباخ، برای برآورد همسانی درونی این نوع آزمون‌ها یا مقیاس‌هایی که پاسخ درست یا پاسخ غلط ندارند به کار می‌رود و نشان می‌دهد که سوال‌های آزمون تا چه میزان خصیصه ی واحدی را اندازه می‌گیرند (شریفی، ۱۳۷۲). به منظور بررسی همسانی درونی، نتایج بار اول اجرای فرم ارزیابی را با روش آلفای کرونباخ مورد مطالعه قرار داده‌ایم ($r_{tt} = 0.7528$).

یکی از راه‌های مناسب به منظور نشان دادن ارتباط سوال‌ها یا ماده‌ها با یکدیگر نشان دادن همبستگی سوال‌ها با یکدیگر است. با نگاهی به ماتریس‌های همبستگی سوال‌ها با یکدیگر در بار اول اجرای فرم (جدول شماره ی ۳) می‌بینیم که، از ۹۰ همبستگی ممکن بین سوال‌ها، فقط ۲۶ همبستگی معنادار هستند و جالب توجه اینکه از ۲۶ همبستگی معنادار بین سوال‌ها، ۱۹ مورد همبستگی منفی است. یعنی در ۱۹ رابطه از ۲۶ رابطه، سوال‌ها موضوعاتی متفاوت و عکس یکدیگر را اندازه گرفته‌اند. این همبستگی‌ها، در بار دوم اجرا به ۲۲ مورد کاهش یافته که از این تعداد نیز ۱۲ مورد، همبستگی معکوس است (جدول شماره ی ۴).

با نگاهی به ماتریس همبستگی سوال‌ها، می‌بینیم که از ۹۰ مورد ممکن، فقط ۷ مورد در بار اول و ۱۰ مورد در بار دوم، با یکدیگر همبستگی معنادار نشان می‌دهند. یعنی، این ماده‌ها همبستگی قابل قبولی با یکدیگر ندارند و نمی‌توان از همبستگی ماده‌های فرم ارزیابی، به عنوان یک فرم منسجم دفاع کرد.

مقصود از اعتبار آزمون، پاسخ دادن به این سوال است که آزمون چه چیزی را اندازه‌گیری می‌کند و تا چه اندازه از این لحاظ کارآیی دارد. به همین خاطر است که گفته می‌شود فقط با بررسی منابع عینی اطلاعات و عملیات تجربی که برای تشخیص اعتبار آزمون به عمل آمده است می‌توان صفتی که آزمون به خصوصی را می‌سنجد تعریف کرد (آناستازی، ۱۳۶۱).

جدول شماره ۳: ماتریس همبستگی ماده‌های فرم ارزیابی در اجرای اول

۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	
/۰۴۳	-/۱۲۳*	/۱۴۷**	-/۰۰۴	/۰۲۵	-/۰۳۱	/۰۷۳	-/۱۸۰**	-/۱۶۵**	۱	۱
-/۰۷۰	/۰۲۵	-/۰۷۴	/۰۶۴	/۰۰۲	-/۰۳۰	-/۱۰۷	-/۰۱۱	۱	-/۱۶۵**	۲
-/۰۵۴	-/۰۰۸	-/۰۰۲	-/۱۲۳*	/۰۲۴	/۱۳۴*	-/۰۹۰	۱	-/۰۱۱	-/۱۸۰**	۳
/۰۰۲	/۱۰۰	-/۰۸۳	/۰۲۸	-/۰۲۵	-/۱۴۳*	۱	-/۰۹۰	-/۱۰۷	/۰۷۳	۴
-/۱۴۹*	-/۰۲۰	-/۰۶۷	-/۱۵۹**	/۰۷۳	۱	-/۱۴۳*	/۱۳۴*	-/۰۳۰	-/۰۳۱	۵
/۰۴۷	/۰۸۸	-/۱۴۴**	/۰۷۹	۱	/۰۷۳	-/۰۲۵	/۰۲۴	/۰۰۲	/۰۲۵	۶
/۰۱۵	/۲۲۵**	-/۱۹۹**	۱	/۰۷۹	-/۱۵۹**	/۰۲۸	-/۱۲۳*	/۰۶۴	-/۰۰۴	۷
-/۱۱۰*	-/۰۰۸	۱	-/۱۹۹**	-/۱۴۴**	-/۰۶۷	-/۰۸۳	-/۰۰۲	-/۰۷۴	/۱۴۷**	۸
-/۰۶۰	۱	-/۰۰۸	/۲۲۵**	/۰۸۸	-/۰۲۰	/۱۰۰	-/۰۰۸	/۰۲۵	-/۱۲۳*	۹
۱	-/۰۶۰	-/۱۱۰*	/۰۱۵	/۰۴۷	-/۱۴۹**	/۰۰۲	-/۰۵۴	-/۰۷۰	/۰۴۳	۱۰

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

جدول شماره ۴: ماتریس همبستگی ماده‌های فرم ارزیابی در اجرای دوم

۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	
-/۰۲۳	-/۰۲۶	-/۰۱۶	-/۰۶۱	-/۰۰۱	/۰۹۹	۰/۱۷۵**	/۱۵۵**	/۰۳۴	۱	۱
-/۰۶۳	-/۰۶۹	-/۰۸۷	-/۰۹۲	-/۰۲۳	/۱۹۲**	-/۱۵۳**	-/۰۴۹	۱	/۰۳۴	۲
/۰۶۰	-/۱۱۴*	/۰۰۷	-/۱۲*	/۰۰۳	/۰۹۵	/۰۳۲	۱	-/۰۴۹	/۱۵۵**	۳
-/۰۰۸	-/۱۲۹*	-/۰۸۲	-/۰۰۵	/۰۸۸	/۰۹۳	۱	/۰۳۲	-/۱۵۳**	/۱۷۵**	۴
-/۱۲۲*	-/۰۳۳	-/۰۱۰۷	-/۰۸۲	/۱۸۲**	۱	/۰۹۳	/۰۹۵	/۱۹۲**	/۰۹۹	۵
-/۱۶۰**	/۰۷۳	/۰۶۱	-/۰۱۳	۱	/۱۸۲**	/۰۸۸	/۰۰۳	-/۰۲۳	-/۰۰۱	۶
-/۰۲۲	/۰۸۵	-/۰۸۵	۱	-/۰۱۳	-/۰۸۲	-/۰۰۵	-/۱۲۴*	-/۰۹۲	-/۰۶۱	۷
-/۰۳۱	/۰۷۵	۱	-/۰۸۵	۰۶۱	-/۰۱۷	-/۰۸۲	/۰۰۷	-/۰۸۷	-/۰۱۶	۸
-/۱۳۰*	۱	/۰۷۵	/۰۸۵	/۰۷۳	-/۰۳۳	/۱۲۹*	-/۱۱۴*	-/۰۶۹	-/۰۲۶	۹
۰۰۰/۱	-/۱۳۰*	-/۰۳۱	-/۰۲۲	-/۱۶۰	-/۱۲۲*	-/۰۰۸	/۰۶۰	-/۰۶۳	-/۰۲۳	۱۰

* Correlation is significant at the 0/05level (2-tailed).

** Correlation is significant at the 0/01 level (2-tailed).

می‌دانیم که اعتبار هر آزمون، باید در ارتباط با کاربردهای خاصی که برای آن پیش‌بینی شده است تعیین شود. به عقیده آناستازی (۱۳۶۱)، اصولاً تمام روش‌های تعیین اعتبار با این مساله سروکار دارند که بین کارکرد در آزمون و واقعیت‌های قابل مشاهده و مستقل درباره‌ی ویژگی‌های رفتار مورد نظر چه روابطی وجود دارد.

یکی از روش‌های برآورد اعتبار، اعتبار محتوا است. اعتبار محتوا به طور کلی متضمن بررسی منظم محتوای آزمون به منظور پاسخ‌دادن به این سوال است که آیا آزمون، نمونه‌های معرف از حوزه رفتار مورد اندازه‌گیری را دربر می‌گیرد یا نه؟ مشاهده‌ی همبستگی سوال‌ها با یکدیگر (جدول شماره ۳ و ۴)، نشان می‌دهد که فرم ارزیابی، نمونه‌های معرف از حوزه‌ی رفتار مورد اندازه‌گیری را دربر نمی‌گیرد.

از طرف دیگر، پیش‌بینی ۵ سوالی که از آزمودنی‌ها پرسیده شد، می‌تواند اعتبار فرم ارزیابی را در شکل موجود آن، زیر سوال ببرد. همان‌گونه که در جدول شماره ۵ نشان داده‌ایم، ۵۷/۱ درصد آزمودنی‌ها اعلام کرده‌اند که معمولاً ارزیابی را با دقت کافی انجام نمی‌دهند، در ۶۴/۶ درصد موارد به تجربه‌ی واقعی خود برای ارزیابی اساتید توجه ندارند، در ۸۴ درصد موارد خوش‌نمره‌بودن استاد و در ۸۲/۱ درصد موارد سهل‌گیر بودن او را مبنای ارزیابی خود قرار می‌دهند و سرانجام اینکه در ۶۱/۸ درصد موارد در صورت سخت‌گیر بودن استاد، نمره‌های او را پایین‌تر از حد طبیعی ارزیابی می‌کنند.

جدول شماره ۵: پاسخ‌های دانشجویان در مورد نحوه‌ی تکمیل فرم ارزیابی

درصد فراوانی	فراوانی		
۴۲/۹۹	۱۳۷	بله	معمولاً ارزیابی را با دقت کافی انجام می‌دهم
۵۷/۱	۱۸۲	خیر	
۳۵/۴	۱۱۳	بله	به تجربه‌ی واقعی خود برای ارزیابی اساتید توجه دارم
۶۴/۶	۲۰۶	خیر	
۸۴	۲۶۸	بله	در صورت خوش‌نمره‌بودن استاد نمره‌های او را بالاتر از حد طبیعی ارزیابی می‌کنم
۱۶	۵۱	خیر	
۸۲/۱	۲۶۲	بله	در صورت سهل‌گیر بودن استاد، نمره‌های او را بالاتر از حد طبیعی ارزیابی می‌کنم
۱۷/۹	۵۷	خیر	
۶۱/۸	۱۲۲	بله	در صورت سخت‌گیر بودن استاد، نمره‌های او را پایین‌تر از حد طبیعی ارزیابی می‌کنم
۳۸/۲	۱۹۷	خیر	

بحث و نتیجه‌گیری

در مجموع می‌بینیم که فرم ارزیابی در شکل موجود خود، فاقد پایایی قابل قبول یا همسانی درونی است. سوال‌های آن همبستگی قابل دفاعی با یکدیگر نشان نمی‌دهند، برخی از سوال‌ها موضوعاتی را می‌سنجد که ارتباط معکوس با بقیه‌ی ماده‌ها دارد. از طرف دیگر

اعتبار فرم ارزیابی به علت مخدوش بودن پایایی، همسانی درونی و همبستگی غیر قابل قبول ماده‌ها با هم، کاملاً زیر سوال می‌رود و مهم‌تر از همه دانشجویان اظهار می‌کنند که در تکمیل و پاسخگویی به فرم ارزیابی، دقت و از تجربه‌ی واقعی خود استفاده نمی‌کنند، به خوش‌نمره‌بودن و سهل‌گیری استاد متمایل‌اند و در صورت سخت‌گیر بودن استاد، نمره‌های او را پایین‌تر از حد طبیعی برآورد می‌کنند.

عقل سلیم حکم می‌کند که هر یک از موارد فوق به تنهایی برای نشان دادن ناکارآمدی این فرم، کافی است. هرچند در این پژوهش نشانه‌ای دال بر حمایت از فرم ارزیابی نیافتیم، با این حال، اگر سیستم بر کاربرد این فرم اصرار دارد، در جهت کارآمد ساختن فرم حاضر به- نظر می‌رسد که باید مجدداً فرم جدیدی تصویب یا ترجمه شود که از نظر اعتبار و پایایی قابل دفاع باشد. هرچند که هیچ فرمی نخواهد توانست بیانگر چهره‌ی علمی و واقعی اساتید باشد. چنین فرمی باید چارچوب زیر را دارا باشد:

ارزش علمی - یادگیری درس: شامل سؤالی در مورد ارزش و اعتبار محتوای درس و اهمیت علمی آن، مثال: «آیا شما از این درس چیز با ارزشی یاد گرفتید؟».

اشتیاق مدرس به موضوع درس: شامل سؤالی در مورد میزان علاقمندی استاد به موضوع تدریس، مثال: «آیا استاد شما نسبت به این درس علاقمند است؟».

روشنی و سازمان‌بندی مطالب: شامل سؤالی در مورد روشن بودن و سازمان‌بندی مطالب تدریس شده، مثال: «آیا توضیحات استاد واضح بود؟».

تعامل گروهی: شامل سؤالی در مورد قابل دسترس بودن معلم برای دانشجویان، مثال: «ایا دانشجویان برای رفع مشکلات خود به اندازه‌ی کافی به استاد دسترسی داشتند؟».

جامع بودن مطالب آموزش داده‌شده: شامل سؤالی در مورد گستردگی و جامع بودن مطالب تدریس شده مثال: «آیا در این درس استاد نظریه‌های متفاوت را با هم مقایسه می‌کرد؟».

امتحان و نمره‌گذاری: شامل سوالهایی درمورد کیفیت امتحان و نحوه‌ی نمره‌گذاری استاد. مثال: «آیا تاکید مواد و سوال‌های امتحانی با تاکید استاد بر موضوع‌های درس همخوانی داشت؟».

مطالب خواندنی و تکالیف: شامل سوالهایی درمورد کیفیت کتاب‌ها، جزوات و تکالیف درس، مثال: «آیا کتاب درسی شامل مطالب با ارزش بود؟».

نظر کلی: شامل سوالهایی درمورد کل درس و کیفیت آموزش استاد، مثال: «به طور کلی این استاد را در مقایسه با استادان دیگر چگونه ارزیابی می‌کنید؟»

منابع

- سیف، علی اکبر. (۱۳۷۶). روش تهیه‌ی پرسشنامه. دانشگاه آزاد اسلامی. رودهن
- شریفی، حسن پاشا. (۱۳۷۲). اصول روانسنجی و روان‌آزمایی. تهران: انتشارات رشد.
- سیف، علی اکبر. (۱۳۷۵). روش‌های اندازه‌گیری و ارزشیابی آموزشی. تهران: نشر دوران.
- اناستازی، ا. (بی تا). روان‌آزمایی. ترجمه محمدنقی براهنی (۱۳۶۱). تهران: دانشگاه تهران.
- سیف، علی اکبر. (۱۳۵۶). ارزشیابی اثربخشی فعالیت‌های آموزشی معلمان به وسیله‌ی دانشجویان. پژوهنده: علوم انسانی و اجتماعی (گزارش نهایی طرح‌های پژوهشی). ۱۸، ۹-۴۹.
- Abrami, P.C., Leventhal, V., & Perry, R.D (1982). Educational Seduction. (*Review of Educational Research*, 52,446–464.)
- Dunkin, J. (1986), Research on teaching in higher education. In M.C. Wittrock (Ed.), *Hand book of research on teaching* (3d ed.), New York: Macmillan, PP. 754-777.
- Gay, L.R. (1991). *Educational evaluation and measurement* (2nd ed.) Maxwell Macmillan. International.
- Naftulin, D.H., Ware, J.E., & Donnelly, F.A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48,630-635.
- Nunnally, J.C. (1978). *Psychometric Theory*. New York: Mc Graw – Hill.
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (PP. 244-271). Beverly Hills, CA: Sage.
- Scriven, M. (1981). summative teacher evaluation. In J. Millman (Ed.), *handbook of teacher evaluation* (pp 244-271). Beverly Hills, CA: Sage.
- Ware, J., Williams, R.G. (1980). A reanalysis of the Doctor Fox experiments. *Instructional Evaluations*. 4, 15-18.

Wiggins, G P (1993). *Assessing student performance: San Francisco: Jossey - Bass.*
Worthen, B.R., Sanders, J.K. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* (New York: Longman) .

