

تعیین ضرایب خرده آزمون‌های ارزشیابی صلاحیت‌های حرفه‌ای (اصلاح) دانش‌آموختگان دانشگاه فرهنگیان

حسن مشکاتی*
علی دلاور**
نورعلی فرخی***
اصغر مینایی****
غلامرضا کیانی*****

چکیده

این پژوهش با هدف تعیین ضرایب خرده آزمون‌های ارزشیابی صلاحیت‌های حرفه‌ای (اصلاح) دانش‌آموختگان دانشگاه فرهنگیان اجرا شد. پژوهش از نظر هدف، کاربردی و از حیث روش، از نوع همبستگی است. جامعه پژوهش، همه مهارت‌آموزان ماده ۲۸ سال تحصیلی ۹۵-۹۴ دانشگاه فرهنگیان به تعداد ۲۸۲۳ نفر بود. با توجه به پراکندگی جامعه در تمام استان‌های کشور، از نمونه در دسترس استفاده شد و مهارت‌آموزان استان خراسان رضوی به تعداد ۱۹۰ نفر، به‌عنوان نمونه انتخاب شدند. قسمت اصلی داده‌های این پژوهش، نتایج حاصل از آزمون اصلاح دانشگاه فرهنگیان بود که در شهریور ۱۳۹۵ برگزار شده است. برای بررسی موفقیت معلمان از منظر همکاران، از آزمون محقق‌ساخته استفاده شد که پایایی آن با استفاده از روش آلفای کرونباخ ۰/۸۹ به‌دست آمد و روایی آن با استفاده از تحلیل عاملی بررسی شد. بر اساس شیوه‌نامه آزمون اصلاح دانشگاه فرهنگیان، سهم خرده‌آزمون «فرهنگی»، «تدریس خرد»، «علمی» و «عملکردی» به ترتیب ۲۰، ۳۵، ۲۰ و ۲۵ درصد است که بر اساس نتایج این پژوهش برای افزایش روایی ملاکی پیش‌بین، پیشنهاد می‌شود که سهم خرده‌آزمون‌ها به ترتیب ۱۵، ۱۱، ۳۸ و ۳۶ درصد باشد. همچنین نتایج نشان داد که طبقه‌بندی افراد به گروه‌های دارای صلاحیت دائمی، صلاحیت موقت و عدم احراز صلاحیت، با روش به‌کار رفته در آزمون و روش پیشنهادی این پژوهش، به نتایج نهایی متفاوتی خواهد انجامید.

واژگان کلیدی: آزمون اصلاح دانشگاه فرهنگیان، روایی، روایی ملاکی پیش‌بین، نمره ترکیبی، سنجش صلاحیت

* دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی

** استاد گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی

***دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی (نویسنده مسئول: farrokhinoorali@yahoo.com)

**** استادیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی

***** دانشیار گروه زبان انگلیسی، دانشگاه تربیت مدرس

مقدمه

در هر کشوری نظام آموزشی عاملی اصلی برای کامل کردن استراتژی توسعه پایدار است. تربیت نیروی انسانی متخصص و کارآموده، اساس توسعه همه‌جانبه کشور است و این وظیفه نظام آموزشی است (کریمی، ۱۳۹۴). بر این اساس، مفهوم ارتقای مهارت‌های حرفه‌ای معلمان، به‌عنوان عنصر اصلی نظام‌های آموزشی، از اواسط قرن بیستم پدیدار شده است و صاحب‌نظران امر تعلیم و تربیت نظریات مختلفی را در این باره بیان کرده‌اند؛ از جمله هابرمن و بومهان^۱ (۲۰۰۵) که به مؤلفه‌های روش‌های تدریس و ارزشیابی اشاره کرده‌اند. ویلگاس-ریمرس^۲ (۲۰۰۷) نیز نقش مهارت‌های حرفه‌ای معلم را جهت تغییر در نظام آموزشی بسیار ضروری دانسته و به مؤلفه‌های روش‌های جدید تدریس، فناوری آموزشی و ارزشیابی اشاره کرده‌اند. مهرمحمدی (۱۳۷۸) به توانایی‌های مربوط به روش‌های تدریس و ارزشیابی و استفاده از تکنولوژی آموزشی معلم اعتقاد دارد. این مهارت‌ها شامل: آگاهی معلمان از روش‌های تدریس، سازمان‌دهی و اداره کلاس، مشاوره و راهنمایی دانش‌آموزان و ارزشیابی است (نقل در نادری و همکاران، ۱۳۸۹).

بررسی عمیق عملکرد حرفه‌ای معلمان، به‌طور فزاینده‌ای به‌عنوان یک ابزار و سیاست برای بهبود آموزش و یادگیری با استفاده از تمایز بین معلمان از سطوح مختلف عملکرد استفاده می‌شود (تیلر و تیلر^۳، ۲۰۱۲؛ گلزرمن و همکاران^۴، ۲۰۱۱). بر اساس گزارش اینگوارسون و همکاران^۵ (۲۰۱۳) تأیید صلاحیت معلمان برای ورود به عرصه شغلی شاید از مهم‌ترین تصمیمات است و می‌تواند تأثیر زیادی بر بهبود آموزش و یادگیری در نظام‌های آموزشی داشته باشد. زمانی که معلم به‌طور رسمی به حوزه آموزش راه می‌یابد شاید بخواهد به مدت ۳۰ سال یا حتی بیشتر تدریس کند و بنابراین بر رشد صدها دانش‌آموز تأثیرگذار باشد. در برخی نظام‌های آموزشی، حذف معلمان نالایق بدون درگیر شدن در فرایندهای حقوقی طولانی، دشوار است. طبق نظر هماهنگ‌کنندگان تحقیقات ملی^۶ (NRCs) در طرح تربیت معلم و توسعه ریاضیات^۱

1. Huberman & Boomhan

2. Villegas_reimers

3. Taylor & Taylor

4. Glazerman et al

5. Ingvarson et al

6. National Research Coordinators (NRCs)

(TEDS-M)، قبلاً بسیاری از کشورها تصور می‌کردند، دانش‌آموخته شدن از دوره تربیت‌معلم به‌خودی‌خود، ضامن کیفیت تدریس معلم است ولی در سال‌های اخیر، این اعتماد تضعیف شده است. از گزارش‌های هماهنگ‌کنندگان تحقیقات ملی به‌خوبی مشخص بود که سیاست‌ها و اقدامات مربوط به اعطای گواهینامه در زمان اجرای طرح تربیت‌معلم و توسعه ریاضیات، تفاوت‌های گسترده‌ای دارد. پس از بررسی گزارش کشورها مشخص شد که شرکت‌کنندگان در طرح تربیت‌معلم و توسعه ریاضیات به سه دسته تقسیم می‌شوند که در جدول (۱) نشان داده شده است. درعین‌حال، همچنان در میان کشورهای هر گروه، تفاوت‌هایی وجود دارد.

جدول (۱) پیش‌نیازهای حاکم بر ورود به عرصه تدریس در کشورهای مشارکت‌کننده در (TEDS-M)

کشور	مقررات تربیت معلم
بوتسوانا، شیلی، گرجستان، مالزی، نروژ، لهستان، فدراسیون روسیه، سنگاپور، اسپانیا (مدرسه‌های خصوصی)، سوئیس و تایلند	گروه ۱. کشورهایی که دانش‌آموختگی از دانشگاه به‌طور خودکار به ورود رسمی به عرصه تدریس می‌انجامد.
کانادا، عمان، فیلیپین و اسپانیا (مدارس دولتی)	گروه ۲. کشورهایی که در آنها ورود به حرفه آموزش، منوط به قبولی در آزمون‌های دیگری است که نهادهای بیرونی اجرا می‌کنند (همچون آزمون‌های گواهی دانش حرفه‌ای).
چین تایپه، آلمان، امریکا	گروه ۳. کشورهایی که در آنها ورود به این حرفه و استخدام در آن، منوط به موفقیت در امتحانات دانش شغلی و ارزیابی عملکرد است.

در ایران تا سال ۱۳۹۴، دانشجویان تربیت‌معلم و دانشگاه فرهنگیان، پس از دانش‌آموختگی در آموزش و پرورش مشغول به خدمت می‌شدند تا اینکه معاونت نظارت، ارزیابی و تضمین کیفیت دانشگاه فرهنگیان، طرحی برای ارزشیابی

1. Teacher Education and Development Study in Mathematics (TEDS-M)

صلاحیت‌های حرفه‌ای دانش‌آموختگان دانشگاه فرهنگیان مصوب کرد که به‌اختصار «اصح» نامیده شد. این آزمون، در شهریور سال ۱۳۹۵ برای دانش‌آموختگان ماده ۲۸ دانشگاه فرهنگیان اجرا شد. دانش‌آموختگان ماده ۲۸، داوطلبان شغل معلمی هستند که دارای مدرک حداقل کارشناسی از دانشگاه‌های معتبرند و در آزمون استخدام آموزش و پرورش پذیرفته شده، دو ترم در دانشگاه فرهنگیان، مهارت‌های اساسی معلمی را فرا می‌گیرند. شرط اشتغال این دانش‌آموختگان در آموزش و پرورش، موفقیت در آزمون اصح است. در راستای صرفه‌جویی در بودجه‌های دولت و جایگزینی سریع نیرو به جای حجم زیاد نیروهای در حال بازنشست شدن، سیاست کنونی مدیران آموزش و پرورش و دانشگاه فرهنگیان بر به‌کارگیری دانش‌آموختگان سایر دانشگاه‌ها با ارائه دوره یک‌ساله مهارت‌های معلمی، در کنار تربیت شبانه‌روزی دانشجویان چهار ساله کارشناسی است؛ بنابراین قرار است هر سال، آزمون استخدام آموزش و پرورش برگزار شود و روند جذب و به‌کارگیری دانشجویان ماده ۲۸ ادامه یابد. در این آزمون، با توجه به طراحی کلان برنامه درسی تربیت معلم (برنامه درسی تربیت معلم جمهوری اسلامی ایران)، از شایستگی‌های تصریح شده در این برنامه به‌عنوان شایستگی‌های پایه استفاده می‌شود. بر این اساس، شایستگی‌های حرفه‌ای معلمی، ذیل چهار بُعد کلیدی قرار می‌گیرد که در جدول (۲) همراه با سهم هر کدام در آزمون ارائه شده است (به نقل از طرح تفصیلی اصح دانش‌آموختگان دانشگاه فرهنگیان، ۱۳۹۴).

جدول (۲) شایستگی‌ها و مؤلفه‌های آزمون اصح

شایستگی	مؤلفه	سهم (درصد)
دانش عمومی	آزمون کتبی	۲۰
دانش موضوعی	آزمون کتبی (چندگزینه‌ای و عملکردی)	۲۵
دانش تربیتی موضوعی	عمل حرفه‌ای تدریس	۳۵
فرهنگی	عملکرد فرهنگی	۲۰
	جمع	۱۰۰

در این آزمون، شایستگی‌های چهارگانه جدول (۲) به‌عنوان شاخص‌های شایستگی، با هم ترکیب می‌شوند و نمره نهایی آزمون محاسبه می‌شود. اگرچه ایده چندین شاخص، روایی صوری قابل‌توجهی دارد، پیاده‌سازی آن، پرسش‌های مهمی را درباره مفهوم‌سازی و ویژگی فنی این شاخص‌ها و استفاده مناسب از آنها برای ارزیابی تکوینی و پایانی عملکرد معلم ایجاد می‌کند (بروکهارت^۱، ۲۰۰۹؛ گو و کرفت^۲، ۲۰۰۹) و با وجود استفاده گسترده از شاخص‌های چندگانه، مفروضه‌های نظری و ویژگی‌های تجربی نظام‌های ارزیابی معلم به‌خوبی شناخته شده نیستند. این مفهوم رایج که شاخص‌های چندگانه پایایی را بهبود می‌بخشد در تمامی زمینه‌ها تأیید نشده است. می‌توان انتظار داشت که پایایی مرکب وقتی که شاخص‌های یک سازه یک‌بُعدی با هم ترکیب می‌شوند، افزایش یابد اما در نظام‌های ارزیابی معلم از شاخص‌های چندگانه برای اندازه‌گیری ساختارهای متفاوت یا اجزای جداگانه یک ساختار گسترده استفاده می‌شود. در این موارد شاخص‌های چندگانه ممکن است روایی را بهبود بخشد اما تأثیر آنها بر پایایی قابل‌پیش‌بینی نیست و حتی ممکن است منفی باشد (کان و کیس^۳، ۲۰۰۴؛ ماس^۴، ۱۹۹۴). در واقع همان‌طور که مهرنز^۵ (۱۹۸۹) پیشنهاد می‌کند، با یک تعریف خوب از ساختار، شاخص‌های چندگانه باعث مبهم شدن آن می‌شود و فهم ما را از عملکرد معلم بهبود نمی‌بخشد. ادبیات سنجش معلم از منظر همکاران و دانش‌آموزان نشان می‌دهد که ویژگی‌های شاخص‌های ترکیبی در مدل‌های مختلف، به‌ویژه هنگام استفاده از قوانین تصمیم‌گیری پیچیده یا ترکیبی به‌طور قابل‌ملاحظه‌ای تغییر می‌کنند (کرونباخ، لین، برینان و هرتل^۶، ۱۹۹۷؛ داگلاس و میسلوی^۷، ۲۰۱۰) اجزای ناپایا می‌توانند حد بالایی را در پایایی نمره‌های مرکب ایجاد کنند (چستر^۸، ۲۰۰۳)؛ و اینکه تعداد و جایگاه نقاط برش، همسانی و دقت

1. Brookhart

2. Goe & Croft

3. Kane & Case

4. Moss

5. Mehrens

6. Cronbach Linn, Brennan & Haertel

7. Douglas & Mislevy

8. Chester

آزمون را تحت تأثیر قرار می‌دهد (کروکر و آلیجینا^۱، ۲۰۰۶). تحقیقات درباره مدل‌های وزن‌دهی در ارزیابی عملکرد معلم به نتایج یکسانی نرسیده است. برای مثال میهالی، مک‌کافری، استیگر و لک‌وود^۲ (۲۰۱۳) برای افزایش پایایی وزن‌دهی بیشتر اجزایی که موضوع ارزیابی را با دقت بیشتری پیش‌بینی می‌کند، پیشنهاد می‌کنند. به‌طور کلی، فهم ما از مفروضه‌های زیربنایی طراحی نظام‌های ارزیابی معلم مبتنی بر شاخص‌های چندگانه، ویژگی‌های روان‌سنجی نتایج، شاخص‌های ترکیبی در مدل‌های مختلف، پیامدهایی که برای تک‌تک معلمان دارد و دلالت‌هایی که این مفروضه‌ها و انتخاب‌ها برای ارزیابی عملکرد معلم دارد و موارد دیگر، ناقص است (بروکهارت و لودمن^۳، ۱۹۹۲).

هدف از آزمون اصلاح، تشخیص یا تعیین میزان توفیق مهارت‌آموزان در شغل معلمی در آینده است. بنابراین هرچه روایی پیش‌بین آزمون بالاتر باشد در تعیین موفقیت معلمان کارایی بهتری خواهد داشت. در محاسبه روایی پیش‌بین، هدف این است که با استفاده از آزمون، جایگاه افراد را در توزیع بخصوصی که فقط در آینده در دسترس خواهد بود، پیش‌بینی کنیم. یعنی آزمون نتایجی را پیش‌بینی می‌کند که پس از فاصله زمانی معینی به‌دست خواهد آمد. نتایج مورد نظر با متغیر ملاک، اندازه‌گیری می‌شود. به‌عنوان مثال نمره‌های درسی در پایان دوره کارآموزی با درجه‌بندی افراد بعد از مدتی اشتغال به کار متناسب باشد. در آزمون‌هایی که برای راهنمایی حرفه‌ای، یا گزینش یا طبقه‌بندی افراد جهت کارآموزی یا استخدام ساخته شده‌اند باید روایی پیش‌بین بررسی و به‌صورت ضریب روایی ارائه شود (مگنوسون^۴، ۱۹۶۷؛ ترجمه براهنی، ۱۳۵۲).

ونگ و استنلی^۵ (۱۹۷۰) پژوهش‌هایی را درباره استفاده از وزن‌ها برای حداکثر کردن همبستگی‌های چندگانه با معیار خارجی، جمع‌بندی کرده‌اند. به‌عنوان مثال، برای یک مجموعه آزمون ورودی، یک ترکیب می‌توان یافت که همبستگی چندگانه بین نمره‌های دانشگاهی و نمره در این مجموعه آزمون را به حداکثر برساند. کان و کیس

1. Crocker & Algina

2. Mihaly, McCaffrey, Staiger & Lockwood

3. Brookhart & Loadman

4. Magnusson

5. Wang & Stanley

(۲۰۰۴)، نشان دادند که وزن بیشتر دادن به بخش پایاتر^۱ آزمون، پایایی نمره‌های ترکیبی را افزایش می‌دهد و از یک نقطه به بعد، روایی را نیز افزایش می‌دهد و وزن بخش پایاتر اگر خیلی زیاد شود، روایی را کاهش می‌دهد؛ یعنی در شرایط معین، ترکیب‌هایی با پایایی بیشتر، دارای روایی ملاکی کمتری هستند. برینان^۲ (۲۰۰۱)، مباحثی درباره بهینه کردن هم‌زمان پایایی و روایی مطرح کرده است. راندر^۳ (۲۰۰۱) نشان داده است که در شرایط معین، حداکثر کردن روایی، به پایایی کمتر منجر می‌شود.

مارتینز، اسپویگ و گلداسمیت^۴ (۲۰۱۶)، به این نتیجه رسیدند که دقت در مدل‌های مختلف و نقاط برش مختلف متفاوت است و مدل‌هایی با دقت مشابه ممکن است به طبقه‌بندی‌های متفاوتی از معلمان بیانجامد. دقت و ثبات، هنگامی بیشترین مقدار است که نمره‌های ترکیبی برای حداکثر کردن پایایی ساخته شوند و هنگامی که به دنبال پیش‌بینی بهینه نمره دانش‌آموزان باشیم، دقت و ثبات، کمترین مقدار خواهد بود. بالدوین^۵ (۲۰۱۵) با داده‌های شبیه‌سازی به این نتیجه رسید که پیش‌بینی کننده‌های مطرح شده، در عمل می‌تواند به نمره‌های ترکیبی با روایی بالاتر بینجامد.

ترکیب خطی نمره‌های آزمون، بیشترین کاربرد را در تصمیم‌گیری برای پیش‌بینی موفقیت و گزینش افراد دارد و مطابق با روش وزن‌دهی خرده‌آزمون‌ها برای افزایش روایی ملاکی است. در این روش، مجموعه‌ای از پیش‌بینی کننده‌های وابسته به عملکرد در ملاک مورد نظر تعیین می‌شود. مثلاً نمره‌های کلامی و غیرکلامی (پیش‌بینی کننده‌ها) با موفقیت تحصیلی (ملاک) مربوط می‌شوند. می‌توان این دو پیش‌بینی کننده را با ضرایب دلخواه برای هرکدام، با هم ترکیب و موفقیت تحصیلی را پیش‌بینی کرد. به این ضرایب، ضرایب رگرسیون می‌گویند. رگرسیون چندمتغیری، یک روش آماری است که شیوه تعیین و محاسبه ضرایب را نشان می‌دهد. از این ضرایب برای به‌دست آوردن دقیق‌ترین پیش‌بینی خطی یک ملاک با استفاده از چند متغیر پیش‌بینی کننده استفاده می‌شود (دلاور، ۱۳۹۵).

1. more reliable

2. Brennan

3. Runder

4. Martínez, Schweig & Goldschmidt

5. Baldwin

سنجش صلاحیت معلمان در بیشتر کشورهای پیشرفته دنیا مرسوم است و در ایران هم برای اولین بار در شهریور ۱۳۹۵ برای مهارت‌آموزان ماده ۲۸ اجرا شد. این آزمون طبق جدول (۲) از چهار خرده‌آزمون تشکیل شده است. ضریب این خرده‌آزمون‌ها بر اساس نظر متخصصان تعیین شده است و پژوهشی علمی برای بررسی ابعاد مختلف آن مانند روایی و پایایی صورت نگرفته است. هدف از اجرای این آزمون، تشخیص عملکرد آینده معلمان است. از سوی دیگر، تحقیقات نشان داده است که ترکیب‌ها و ضرایب مختلف خرده‌آزمون‌ها می‌تواند بر ویژگی‌های روان‌سنجی آزمون و نتایج نهایی تأثیر بگذارد (کرونباخ و همکاران، ۱۹۹۷؛ داگلاس و میسلوی، ۲۰۱۰؛ کان و کیس، ۲۰۰۴؛ ماس، ۱۹۹۴). تخصیص ضرایب به این خرده‌آزمون‌ها بر اساس ضرایب رگرسیون پیش‌بینی موفقیت در عملکرد معلمی، می‌تواند به افزایش روایی ملاکی پیش‌بین کمک کند (میهایلی و همکاران، ۲۰۱۳؛ ونگ و استنلی، ۱۹۷۰؛ گالیکسن، ۱۹۵۰). بنابراین دو پرسش در این پژوهش، مطرح است: یکی اینکه ضرایب خرده‌آزمون‌ها بر اساس روایی ملاکی چند است؟ و دوم اینکه آیا نتایج آزمون بر اساس این ضرایب در مقایسه با شرایط اجرای کنونی متفاوت خواهد بود؟

روش اجرای پژوهش

پژوهش حاضر، از نظر هدف، کاربردی و از حیث روش، از نوع همبستگی است. جامعه پژوهشی، همه مهارت‌آموزان ماده ۲۸ سال تحصیلی ۹۵-۹۴ دانشگاه فرهنگیان به تعداد ۲۸۲۳ تن بود. این مهارت‌آموزان بعد از فراغت از دوره مهارت‌آموزی، در مدرسه‌های استان‌های مختلف کشور مشغول به خدمت می‌شوند و دسترسی به آنها دشوار است؛ به همین دلیل از نمونه در دسترس استفاده شد و مهارت‌آموزان استان خراسان رضوی به تعداد ۱۹۰ نفر به‌عنوان نمونه انتخاب شدند. پس از گردآوری اطلاعات، مشخص شد هشت پرسشنامه ناقص است که کنار گذاشته شد و سرانجام اطلاعات ۱۸۲ پرسشنامه تحلیل شد.

قسمت اصلی داده‌های این پژوهش، نتایج به دست آمده از آزمون اصلح دانشگاه فرهنگیان بود که در شهریور ۱۳۹۵ برگزار شده است. این آزمون، معمولاً در پایان دوره مهارت‌آموزی دانشجویان ماده ۲۸ دانشگاه فرهنگیان، برای صدور گواهی

صلاحیت تدریس اجرا می‌شود و برای نخستین بار در شهریور ۱۳۹۵ اجرا شد که شامل چهار بخش زیر است:

الف- آزمون تدریس خرد: در این بخش مهارت آموز باید یک تدریس خرد در حد ۱۵ دقیقه ارائه دهد. این تدریس از منظر دو داور بر اساس ملاک‌های از پیش تعیین شده، ارزیابی می‌شود: یک متخصص در حوزه تدریس و یک داور تربیتی که از نظر فنون تدریس آن را بررسی می‌کند و در مجموع نمره‌ای بین صفر تا ۳۵۰ به مهارت آموزان می‌دهند.

ب- آزمون علمی کتبی: شامل ۴۵ پرسش چهارگزینه‌ای دارای نمره منفی است. مهارت آموزان در این بخش، می‌توانند حداکثر نمره ۲۰۰ را کسب کنند.

ج- آزمون عملکردی کتبی: در این آزمون، مهارت آموز می‌بایست به‌عنوان یک طراح آموزشی، مبحثی مرتبط با رشته مهارت‌آموزی را در قالب یک واحد یادگیری، بر اساس محورهای ذکر شده طراحی کند. این طراحی توسط دو داور ارزیابی شده و در نهایت نمره‌ای بین صفر تا ۲۵۰ کسب می‌کند.

د- نمره پوشه‌کار فرهنگی: مهارت آموز باید گزارشی درباره فعالیت‌های فرهنگی در طول دوره مهارت‌آموزی ارائه دهد و این گزارش توسط دو داور در سطح کشور بر اساس ملاک‌های از پیش تعیین شده، ارزیابی و نمره‌گذاری می‌شود و می‌تواند نمره‌ای بین صفر تا ۲۰۰ بگیرد.

برای بررسی موفقیت معلمان از منظر همکاران، از پرسشنامه محقق‌ساخته استفاده شد. برای ساخت این پرسشنامه، با توجه به فهرست‌ها، طبقه‌بندی‌ها و ابزارهای اندازه‌گیری موجود، ابعاد مختلف عملکرد معلمی، شامل: توان علمی، مهارت‌های ارتباطی، بیان مطلب، تسلط بر روش‌های تدریس، کلاس‌داری، اخلاق‌مداری و نظم و آراستگی مشخص شد. پس از آن، بر اساس تعاریف هریک از ابعاد، عبارت‌هایی تدوین شد تا آن بعد را اندازه‌گیری کند. سپس به‌منظور بررسی روایی محتوا، این پرسشنامه ۴۷ سؤالی در اختیار چند نفر از مدیران گروه‌های آموزشی استان و مدیران مدرسه‌ها و استادان دانشگاه فرهنگیان قرار گرفت تا اشکالات آن را برطرف کنند؛ پرسش‌های نارسا را تغییر دادند و بعضی پرسش‌ها با پرسش‌های بهتر جایگزین شد یا برخی پرسش‌های غیر قابل اصلاح، حذف شد؛ فرم اولیه در اختیار ۱۶ نفر از مدیران، معاونان، مربیان پرورشی و نمایندگان معلمان مدرسه‌ها قرار گرفت تا ضمن

پاسخگویی، نظر خود را پیرامون سؤالات بازگو کنند. بعد از تحلیل جواب‌های این ۱۶ نفر، تغییرات لازم در فرم نهایی داده شد تا پرسشنامه آماده اجرا شود. سپس، فرم نهایی آزمون با ۴۳ پرسش تهیه شد. پایایی این فرم در اجرای نهایی با استفاده از روش آلفای کرونباخ ۰/۸۹ به دست آمد که پایایی خوب و قابل قبولی است. پایایی و شاخص‌های توصیفی نمره خرده‌مقیاس‌ها و کل آزمون در جدول (۳) ارائه شده است.

جدول (۳) پایایی و شاخص‌های توصیفی نمره خرده‌مقیاس‌ها

انحراف استاندارد	میانگین	میانه	حداکثر نمره	حداقل نمره	تعداد پرسش‌ها	پایایی (آلفای کرونباخ)	خرده‌مقیاس
۳/۸۲	۱۶/۵۳	۱۷	۲۵	۵	۵	۰/۶۸۱	توان علمی
۳/۲۹	۱۰/۹۵	۱۱	۲۰	۴	۴	۰/۶۳۳	بیان مطالب
۴/۴۴	۱۸/۱۲	۱۸	۳۰	۶	۶	۰/۶۷۹	مهارت‌های آموزشی
۵/۰۵	۱۷/۲۷	۱۷	۳۰	۶	۶	۰/۷۹۷	کلاس داری
۳/۷۷	۱۴/۰۴	۱۴	۲۵	۵	۵	۰/۶۳۵	اخلاق اسلامی
۴/۰۹	۱۴/۷۴	۱۵	۲۵	۵	۵	۰/۷۴۸	نظم و آراستگی
۲/۸۶	۱۶/۸۹	۱۷	۲۵	۵	۵	۰/۵۳۷	مهارت‌های ارتباطی
۳/۲۲	۲۲/۰۱	۲۲	۲۸	۷	۷	۰/۷۰۱	سنجش
۲۱/۱۱	۱۳۰/۵۴	۱۳۱	۲۱۵	۴۳	۴۳	۰/۸۹۲	کل آزمون

به‌منظور تعیین روایی سازه پرسشنامه، داده‌های به دست آمده از پرسشنامه محقق ساخته، با استفاده از روش تحلیل عاملی اکتشافی تجزیه و تحلیل شد. در پژوهش حاضر برای انتخاب سؤال‌های هریک از خرده‌مقیاس‌های پرسشنامه برای انجام تحلیل عاملی، از ملاک ضریب همبستگی هر پرسش با نمره کل خرده‌مقیاس مربوط به آن، استفاده شده است. تحلیل عاملی دربرگیرنده مراحل نسبتاً متنوع و پیچیده‌ای است که در آن، مراحل سه‌گانه زیر به کار برده می‌شود: ۱- آماده کردن ماتریس همبستگی^۱؛

^۱. Correlation matrix

۲- استخراج عوامل اولیه^۱؛ ۳- چرخش^۲ برای دستیابی به راه حل نهایی، جستجو برای کشف عوامل ساده و قابل تفسیر (دلاور، ۱۳۷۷). پس از انجام مراحل تحلیل عاملی، نتایج نشان داد که مقدار ارزش ویژه برای عامل اول ۸/۲۱۴، برای عامل دوم ۳/۱۴۶، عامل سوم ۲/۹۲۸، عامل چهارم ۲/۸۸۷، عامل پنجم ۲/۶۳۶، عامل ششم ۲/۴۷۵ عامل هفتم ۱/۹۸۰ و برای عامل هشتم ۱/۸۳۸ است.

یافته‌های پژوهش

پرسش اول: بهترین ضرایب خرده‌مقیاس‌های آزمون اصلح دانشگاه فرهنگیان بر اساس روایی ملاکی چیست؟

برای محاسبه بهترین ضرایب خرده‌آزمون‌ها بر اساس روایی ملاکی، ضرایب رگرسیون خرده‌آزمون‌ها برای پیش‌بینی موفقیت در شغل معلمی محاسبه شد. برای ورود متغیرها به معادله رگرسیون و به دست آوردن بهترین زیرمجموعه از متغیرهای پیش‌بین، به دلیل اینکه روش رگرسیون گام‌به‌گام^۳ احتمالاً زیرمجموعه‌های رضایت‌بخش کوچک‌تری از متغیرها را انتخاب می‌کند، (هومن، ۱۳۸۵) از روش رگرسیون گام‌به‌گام، استفاده شد.

بررسی مفروضات رگرسیون: یکی از مفروضات رگرسیون این است که متغیرهای وابسته و پیش‌بین دارای توزیع نرمال باشد. نرمال بودن داده‌ها با استفاده از آزمون کولموگروف-اسمیرنف^۴ بررسی شد. نتایج نشان داد که به‌جز داده‌های آزمون کتبی و عملکردی که با توزیع نرمال تفاوت معنی‌داری ندارد، نمره‌های سه خرده‌آزمون دیگر با توزیع نرمال تفاوت معنی‌داری دارد. به‌منظور رفع این مشکل از لگاریتم داده‌ها به‌جای داده‌ها استفاده می‌شود (مؤمنی و فعال قیومی، ۱۳۸۹).

یکی دیگر از مفروضات تحلیل رگرسیون، استقلال خطاهاست که با استفاده از آزمون دوربین-واتسون بررسی شد و نشان داد که خطاها مستقل هستند.

از دیگر مفروضات، عدم هم‌خطی متغیرهای مستقل یا به‌عبارت دیگر همبسته نبودن متغیرهای مستقل است. نتایج نشان‌دهنده امکان وجود هم‌خطی بین متغیرهای

1. Initial factors extraction

2. Rotation

3. Stepwise regression

4. Kolmogorov-Smirnov Test

مستقل است. یکی از روش‌های حل این مشکل نیز استفاده از لگاریتم اعداد به‌جای خود آنهاست (همان منبع). ماتریس همبستگی بین متغیرها در ادامه آمده است.

جدول (۴) میانگین، انحراف استاندارد و ماتریس همبستگی متغیرها

متغیر	میانگین	انحراف استاندارد	۱	۲	۳	۴
۱- فرهنگی	۱۵۲/۵۹۳	۱۵/۴۵				
۲- تدریس خرد	۳۲۹/۲۷۱	۱۶/۸۲	۰/۱۲۴			
۳- علمی	۱۰۲/۹۱۶	۴۵/۲۲	۰/۱۳۵	۰/۰۲۴		
۴- عملکردی	۱۳۸/۴۴۳	۵۰/۹۲	۰/۱۴۰	۰/۱۳۷	۰/۲۹۶*	
۵- موفقیت معلمان	۱۸۲/۳۷۰	۱۶/۶۷	۰/۲۹۹*	۰/۱۸۰*	۰/۶۸۷*	۰/۵۶۵*
+P<0/05 *P<0/01 N=۱۸۲						

همان‌طور که مشاهده می‌شود، همبستگی بین خرده‌آزمون علمی با آزمون عملکردی و موفقیت معلمان با تمام متغیرها معنی‌دار است.

اجرای رگرسیون: همه خرده‌آزمون‌ها سهم معنی‌داری در معادله داشتند. R برابر $۰/۸۶۵$ و R^2 برابر $۰/۷۴۸$ است. یعنی تمام خرده‌آزمون‌ها در پیش‌بینی ملاک، سهم معنی‌داری دارند و نزدیک به ۷۵ درصد از متغیر ملاک را تبیین می‌کنند. همان‌طور که در جدول (۵)، مشخص است، F محاسبه شده برای این درصد از واریانس، برابر $۱۳۱/۱۲۵$ و به لحاظ آماری در سطح کمتر از $۰/۰۱$ معنی‌دار است.

جدول (۵) ضرایب متغیرها در معادله رگرسیون

متغیر پیش‌بین	β	T	p	R^2	F	df	p
نمره‌های فرهنگی	۰/۱۴۶	۳/۷۷۴	P<0/01	۰/۷۴۸	۱۳۱/۱۲۵	۱۷۷،۴	P<0/01
تدریس خرد	۰/۱۱۱	۳/۸۸۰	P<0/01				
آزمون علمی	۰/۵۴۳	۱۳/۷۴۵	P<0/01				
آزمون عملکردی	۰/۴۷۴	۱۱/۹۱۲	P<0/01				

همان‌طور که ملاحظه می‌شود، ضریب استاندارد رگرسیون خرده‌آزمون فرهنگی برابر ۰/۱۴۶، تدریس خرد برابر ۰/۱۱۱، علمی برابر ۰/۵۴۳ و برای خرده‌آزمون عملکردی برابر ۰/۴۷۴ است که همگی در سطح ۰/۰۱ معنی‌دار است. بر اساس شیوه‌نامه آزمون اصلاح دانشگاه فرهنگیان، سهم خرده‌آزمون «فرهنگی»، «تدریس خرد»، «علمی» و «عملکردی»، به ترتیب ۲۰، ۳۵، ۲۰ و ۲۵ درصد است که بر اساس نتایج این پژوهش، برای افزایش روایی ملاکی پیشنهاد می‌شود که به ترتیب ۱۵، ۱۱، ۳۸ و ۳۶ درصد باشد.

برای محاسبه سهم هر خرده‌مقیاس، همبستگی تفکیکی نمره‌های هرکدام از ۴ خرده‌مقیاس با نمره‌های موفقیت در شغل معلمی با حذف تأثیر نمره‌های سه خرده‌مقیاس دیگر بر این رابطه، محاسبه شده است. سپس این همبستگی‌های تفکیکی برای ۴ خرده‌مقیاس با هم جمع و سهم هر خرده‌مقیاس در این مجموع با تقسیم هر ضریب تفکیکی بر جمع ضرایب، محاسبه شده است.

پرسش دوم: آیا طبقه‌بندی افراد به گروه‌های دارای صلاحیت دائمی، صلاحیت موقت و عدم احراز صلاحیت، با روش آزمون و روش پیشنهادی این پژوهش، متفاوت است؟

در این آزمون، نمره کلی ۱۰۰۰ بوده است: نمره خرده‌آزمون کتبی ۲۰۰، آزمون عملکردی ۲۵۰، آزمون تدریس خرد ۳۵۰ و فرهنگی ۲۰۰ بوده است. داوطلبان بر اساس ملاک‌های زیر به سه گروه زیر تقسیم شدند:

۱- گروهی که در کل آزمون نمره حداقل ۶۵۰ از ۱۰۰۰ می‌گیرند و حداقل ۲۵ درصد نمره آزمون‌های کتبی یعنی آزمون علمی و کتبی عملکردی را نیز گرفته‌اند، گواهی‌نامه صلاحیت دائمی می‌گیرند.

۲- گروهی که در کل آزمون، نمره بین ۵۰۰ تا ۶۵۰ کسب می‌کنند و حداقل ۲۵ درصد نمره آزمون‌های کتبی یعنی آزمون علمی و کتبی عملکردی را نیز گرفته‌اند، گواهی‌نامه صلاحیت موقت می‌گیرند و باید برای جبران نمره خود در آزمونی که بعداً اعلام می‌شود، شرکت کنند و حداقل شرایط گواهی‌نامه دائم را کسب کنند.

۳- گروهی که در کل آزمون، نمره کمتر از ۵۰۰ می‌گیرند یا حداقل ۲۵ درصد از نمره آزمون‌های کتبی، یعنی آزمون علمی و کتبی عملکردی را نمی‌گیرند، احراز

صلاحیت نشده‌اند و باید دوره مهارت‌آموزی را دوباره طی و دوباره با مهارت‌آموزان سال بعد در آزمون شرکت کنند.

با استفاده از ضرایب به دست آمده در این پژوهش، نمره‌های کل داوطلبان بر مبنای ۱۰۰۰ دوباره حساب شد و با توجه به این نمره‌های جدید، افراد دوباره مطابق شرایط سه‌گانه فوق به سه گروه تقسیم شدند. نتایج در جدول (۶) ارائه شده است.

جدول (۶) نتایج آزمون اصلح در دو روش وزن‌دهی

جمع	تصمیم‌گیری بر اساس نتایج این پژوهش			تصمیم‌گیری بر اساس دستورالعمل
	رد صلاحیت	صلاحیت مشروط	احراز صلاحیت	
۱۵۱	۸	۷۵	۶۸	احراز صلاحیت
۲۰	۱۸	۲	۰	صلاحیت مشروط
۱۱	۱۱	۰	۰	رد صلاحیت
۱۸۲	۳۷	۷۷	۶۸	جمع

برای بررسی معنی‌داری تفاوت بین دو طبقه‌بندی، از آزمون خی‌دو استفاده شد. نتایج آزمون در جدول (۷) درج شده است.

جدول (۷) نتایج آزمون فریدمن

تعداد	آماره خی‌دو	درجه آزادی	سطح معنی‌داری
۱۸۲	۱۲۴/۳۶۲	۴	$P < 0/01$

آماره خی‌دو برابر ۱۲۴/۳۶۲ و سطح معنی‌داری از ۰/۰۱ کمتر است؛ یعنی تفاوت معنی‌داری بین نتایج دو روش طبقه‌بندی وجود دارد و ضریب دادن به خرده‌مقیاس‌ها برای افزایش روایی، به نتایج نهایی متفاوتی می‌انجامد.

بحث و نتیجه‌گیری

سنجش صلاحیت‌های حرفه‌ای معلمان پیش از ورود به حرفه معلمی، موضوعی کاملاً جدید در ایران است که برای نخستین بار در شهریور ۱۳۹۵ برای مهارت‌آموزان یک‌ساله دانشگاه فرهنگیان اجرا شد. با توجه به جدید بودن این آزمون و اهمیت نتایج آن در سرنوشت مهارت‌آموزان و نظام آموزشی ایران، پژوهش در این زمینه ضروری به نظر می‌رسد. با توجه به اینکه این آزمون از چهار خرده‌آزمون تشکیل شده است، یکی از مباحث مهم، نحوه ترکیب نمره‌های این خرده‌آزمون‌ها و سهم هر کدام در نمره کلی است. از بهترین روش‌های به دست آوردن ضرایب بهینه، محاسبه ضرایب رگرسیون این خرده‌آزمون‌ها در پیش‌بینی میزان موفقیت مهارت‌آموزان پس از اشتغال در شغل معلمی است که به آن روایی ملاکی پیش‌بین می‌گویند.

در روایی ملاکی پیش، این پرسش وجود دارد که آیا ابزار اندازه‌گیری می‌تواند آنچه را که ادعا می‌کند، پیش‌بینی کند؟ به‌عنوان مثال، سازمان‌هایی که برای استخدام کارکنان کارآمد، اقدام به گزینش و آزمون می‌کنند، از یک سری آزمون‌ها و مصاحبه‌ها بهره می‌برند و هدف آنها جذب بهترین کارکنان برای افزایش کارایی سازمان است. بنابراین، در صورت عدم روایی آزمون‌ها و مصاحبه‌ها، نمی‌تواند بهترین کارکنان را جذب کنند. آزمون اصلح دانشگاه فرهنگیان هم باید مهارت‌آموزانی را که در آینده، معلمان دارای عملکرد خوب را از معلمان با عملکرد ضعیف، تشخیص دهد. به این منظور، ضریب‌دهی بهینه به خرده‌آزمون‌ها بر اساس ملاک موفقیت آنها باعث خواهد شد که آزمون در تشخیص، بهتر عمل کند. تاکنون پژوهشی پیرامون ویژگی‌های روان‌سنجی و سهم هر خرده‌آزمون در نمره کل انجام نگرفته است. بنابراین پژوهش حاضر، با هدف تعیین ضرایب خرده‌آزمون‌های ارزشیابی صلاحیت‌های حرفه‌ای (اصلح) دانش‌آموختگان دانشگاه فرهنگیان، اجرا شد.

بخش عمده داده‌های این پژوهش، نتایج حاصل از آزمون اصلح دانشگاه فرهنگیان است که برای نخستین بار در شهریور ۱۳۹۵ برگزار شده و دارای چهار خرده‌مقیاس است. برای بررسی موفقیت معلمان از منظر همکاران، از پرسشنامه محقق‌ساخته استفاده شد. این پرسشنامه ۸ بُعد از عملکرد معلم را ارزیابی می‌کرد. فرم نهایی آزمون ۴۳ پرسش داشت. پایایی این فرم در اجرای نهایی با استفاده از روش آلفای کرونباخ

۰/۸۹ به دست آمد که پایایی خوب و قابل قبولی است. روایی پرسشنامه با استفاده از تحلیل عاملی بررسی شد و نتایج نشان‌دهنده روایی بالای پرسشنامه بود. یکی از پرسش‌های پژوهش این بود که بهترین ضرایب خرده‌مقیاس‌ها برای به دست آوردن بهترین پیش‌بینی موفقیت مهارت‌آموزان در شغل معلمی چیست؟ برای پیدا کردن این ضرایب، از رگرسیون استفاده شد. بدین منظور، ابتدا مفروضات رگرسیون در داده‌ها بررسی و سپس ضرایب رگرسیون محاسبه شد. نتایج نشان داد که R برابر ۰/۸۶۵ و R^2 برابر ۰/۷۴۸ است. تمام خرده‌آزمون‌ها در پیش‌بینی ملاک سهم معنی‌داری دارند و نزدیک به ۷۵ درصد از متغیر ملاک را تبیین می‌کنند. ضریب استاندارد رگرسیون برای خرده‌آزمون‌های فرهنگی برابر ۰/۱۴۶، تدریس خرد برابر ۰/۱۱۱، علمی برابر ۰/۵۴۳ و عملکردی برابر ۰/۴۷۴ است. بر اساس شیوه‌نامه آزمون اصلح دانشگاه فرهنگیان، سهم خرده‌آزمون «فرهنگی»، «تدریس خرد»، «علمی» و «عملکردی» به ترتیب ۲۰، ۳۵، ۲۰ و ۲۵ درصد است که بر اساس نتایج این پژوهش برای افزایش روایی ملاکی پیش‌بین، پیشنهاد می‌شود که به ترتیب، ۱۵، ۱۱، ۳۸ و ۳۶ درصد باشد.

پرسش دیگر پژوهش، این بود که آیا طبقه‌بندی افراد به گروه‌های دارای صلاحیت دائمی، صلاحیت موقت و عدم احراز صلاحیت با روش به‌کار رفته در آزمون و روش پیشنهادی این پژوهش، متفاوت است؟ برای پاسخ به این پرسش، طبقه‌بندی افراد به مقوله‌های دارای صلاحیت دائمی، صلاحیت موقت و عدم احراز صلاحیت با روش به‌کار رفته در آزمون و روش پیشنهادی این پژوهش، با استفاده از آزمون خی‌دو بررسی شد. نتایج نشان داد، تفاوت معنی‌داری بین دو روش طبقه‌بندی وجود دارد و ضریب دادن به خرده‌مقیاس‌ها بر اساس روایی ملاکی پیش‌بین، به نتایج نهایی متفاوتی می‌انجامد. این نتایج با پژوهش مارتینز، اسچویگ و گلداسمیت (۲۰۱۶) همخوانی دارد. آنها نشان دادند که دقت، در مدل‌های مختلف و نقاط برش مختلف متفاوت است و اینکه مدل‌هایی با دقت مشابه ممکن است به طبقه‌بندی‌های متفاوتی از معلمان بیانجامد. طبق این پژوهش هم روش‌های مختلف وزن‌دهی به نتایج متفاوتی می‌انجامد. نتایج این پژوهش همچنین با پژوهش ونگ و استنلی (۱۹۷۰) همخوانی دارد. آنها به این نتیجه رسیدند که برای یک مجموعه آزمون ورودی، یک ترکیب می‌توان یافت که همبستگی چندگانه بین نمره دانشگاهی و نمره در این مجموعه آزمون را به حداکثر برساند. همچنین نتایج پژوهش بالدوین (۲۰۱۵) نشان داد که

پیش‌بینی کننده‌های مطرح شده در عمل می‌توانند به نمره‌های ترکیبی با روایی بالاتر بیانجامند که با یافته‌های این پژوهش، همخوانی دارد. با نتایج کرونباخ و همکاران (۱۹۹۷) و داگلاس و میسلوی (۲۰۱۰) نیز هماهنگ است زیرا آنها نشان دادند که ویژگی‌های شاخص‌های ترکیبی در مدل‌های مختلف، به‌ویژه هنگام استفاده از قوانین تصمیم‌گیری پیچیده یا ترکیبی به‌طور قابل ملاحظه‌ای تغییر می‌کنند.

برای افزایش روایی ملاکی نتایج نشان می‌دهد که سهم خرده‌آزمون‌ها باید متفاوت باشد. سهم خرده‌آزمون تدریس خرد در شیوه‌نامه، بیشتر از سایر خرده‌آزمون‌هاست. اما بر اساس نتایج این پژوهش، کمترین سهم را داراست. سهم آزمون کتبی علمی از دو خرده‌آزمون در شیوه‌نامه کمتر است، اما بر اساس نتایج این پژوهش، بیشترین سهم را به خود اختصاص می‌دهد. این نتیجه احتمالاً ناشی از رعایت شرایط استاندارد خرده‌آزمون علمی کتبی است؛ زیرا پرسش‌ها در سطح ملی و با رعایت اصول علمی سنجش طراحی شده و هم شرایط اجرای استاندارد تا حد ممکن رعایت شده، اما شرایط اجرای تدریس خرد، آن‌طور که باید، نبوده است. یعنی؛ در فرصتی بسیار محدود، باید تمامی مهارت‌آموزان، آزموده می‌شدند. از این‌رو، به دست‌اندرکاران این آزمون در دانشگاه فرهنگیان پیشنهاد می‌شود، برای آشنایی استادانی که تدریس خرد را بررسی می‌کنند، جلسات توجیهی برگزار کنند و دقت و حساسیت کارشان مورد تأکید قرار گیرد تا از دخالت احساسات در نمره‌گذاری پرهیز شود. همچنین به‌گونه‌ای برنامه‌ریزی کنند که فرصت مناسب برای اجرای تدریس خرد در نظر گرفته شود و از شتاب‌زدگی در اجرا جلوگیری شود. در مجموع، به تصمیم‌گیران دانشگاه فرهنگیان پیشنهاد می‌شود با توجه به ذهنی بودن نمره‌گذاری تدریس خرد، سهم آن را در نمره کل کمتر کرده و بر سهم آزمون علمی و عملکردی بیفزایند. همچنین نتایج این پژوهش نشان می‌دهد که شرایط خرده‌آزمون فرهنگی نیز مناسب نبوده است، چون سهم آن ۲۰ درصد بوده اما بر اساس نتایج پژوهش باید ۱۵ درصد باشد. بنابراین پیشنهاد می‌شود سهم این خرده‌آزمون کمتر شود یا اینکه معیارها و شرایط استاندارد آن بهبود یابد تا نتایج معتبرتری از آزمون به‌دست آید. به پژوهشگران پیشنهاد می‌شود که این پژوهش را با داده‌های سال‌های آینده که تدریس خرد با دقت بیشتری اجرا می‌شود، اجرا و نتایج را مقایسه کنند. همچنین پیشنهاد می‌شود چنین پژوهشی را با نمونه‌هایی از سایر استان‌ها اجرا کنند.

منابع

- دلاور، علی (۱۳۹۵). مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی). تهران: انتشارات سمت.
- دلاور، علی (۱۳۷۷). مقدمه‌ای بر تحلیل عاملی. تهران: انتشارات دانشگاه آزاد اسلامی، واحد تهران مرکزی.
- سرمد، زهره؛ بازرگان، عباس و حجازی، الهه (۱۳۹۵). روش‌های تحقیق در علوم رفتاری. تهران: انتشارات آگاه.
- طرح تفصیلی ارزشیابی صلاحیت‌های حرفه‌ای (اصلاح) دانش‌آموختگان دانشگاه فرهنگیان (۱۳۹۴). معاونت نظارت، ارزیابی و تضمین کیفیت، مرکز سنجش شایستگی‌های حرفه‌ای، ویراست بیست و یکم، منتشر نشده.
- کریمی، مسلم (۱۳۹۴). آموزش و پرورش محور شبکه‌ی تضمین کیفیت در نظام آموزشی. مجموعه مقالات دومین کنفرانس ملی و نهمین همایش ارزیابی کیفیت نظام‌های دانشگاهی. تهران، دانشگاه فرهنگیان، صص ۷۴-۶۳.
- مگنوسون، داوید (۱۹۶۷). مبانی نظری آزمون‌های روانی؛ ترجمه محمدتقی براهنی. تهران: انتشارات دانشگاه تهران.
- مؤمنی، منصور و فعال قیومی، علی (۱۳۸۹). تحلیل آماری با استفاده از SPSS، ویرایش سوم. تهران: انتشارات مؤلف.
- نادری، عزت‌الله؛ حاجی‌زاده، محمد؛ شریعتمداری، علی؛ و سیف نراقی، مریم (۱۳۸۹). بررسی و مقایسه مهارت‌های حرفه‌ای معلمان دروس علوم پایه و علوم انسانی دوره راهنمایی شهرستان بهشهر به‌منظور ارائه روش‌های ارتقای کیفی این مهارت‌ها. تحقیقات مدیریت آموزشی، ۲ (۲)، ۷۵-۹۶.
- هومن، حیدرعلی (۱۳۸۵). تحلیل داده‌های چندمتغیری در پژوهش رفتاری. تهران: نشر پیک فرهنگ.

- Baldwin, P. (2015). Weighting Components of a Composite Score Using Naïve Expert Judgments about Their Relative Importance. *Applied Psychological Measurement*, 39 (7) 539-550.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Brookhart, S. M.; & Loadman, W. E. (1992). Schooluniversity collaboration: Across cultures. *Teaching Education*, 4 (2), 53-68.

- Chester, M. D. (2003). Multiple measures and highstakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice*, 22, 32–41.
- Cronbach, L. J.; Linn, R. L.; Brennan, R. L. & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational & Psychological Measurement*, 57, 373–399.
- Douglas, K. M.; & Mislevy, R. J. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Educational and Behavioral Statistics*, 35, 1–27.
- Glazer, S.; Goldhaber, D.; Loeb, S.; Raudenbush, S.; Staiger, D. O.; & Whitehurst, G. J. (2011). *Passing muster: Evaluating evaluation systems*. Washington, DC: Brown Center on Education Policy at Brookings.
- Goe, L.; & Croft, A. (2009). *Methods of evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Huberman, F.; & Boomhan, T. (2005). *Applying TQM, philosophy to the teaching and learning process*. Monash University, Malaysia.
- Ingvarson, L. J.; Schulle, M. T. T.; Glenn R.; Ray P.; & Sharon L. S. (2013). *An Analysis of Teacher Education Context, Structure, and Quality Assurance Arrangements in TEDS-M Countries, Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Australian Council for Educational Research Michigan State University, Amsterdam, the Netherlands, www.iea.nl
- Kane, M.; & Case, S. M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education*, 17, 221–240.
- Martínez, J. F.; Schweig, J.; & Goldschmidt, P. (2016). Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy. *Educational Evaluation & Policy Analysis*, 38 (4), 738–756.
- Mehrens, W. (1989). Combining evaluation data from multiple sources. In J. Millman & L. DarlingHammond (Eds.), *the new handbook of teacher evaluation: Assessment of elementary and secondary school teachers* (pp. 322–336). Newbury Park, CA: SAGE.
- Mihaly, K.; McCaffrey, D.; Staiger, D.; & Lockwood, J. R. (2013). *A composite estimator of effective teaching (MET Project)*. The RAND Corporation. Retrieved from http://www.rand.org/pubs/external_publications/EP50155.html
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23 (2), 5–12.

- Runder, L. (2001). Informed test component weighting, *Educational Measurement: Issues and Practice*, 0, pp. 16-19.
- Taylor, E. S.; & Tyler, J. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102, 3628–3651.
- villegas-Reimers, E. (2007). *Teacher professional development: an international review of the literature*. UNESCO; International Institute for Educational Planning
- Wang, M. W.; & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 4, 663–704.