

انواع روش‌های سنجش معلمان زبان انگلیسی مقطع متوسطه در تهران و عوامل مؤثر بر آن

اسمعیل علی سلیمی^۱

تاریخ دریافت: ۹۵/۰۱/۲۲

تاریخ پذیرش: ۹۵/۰۷/۱۵

چکیده

فرایند پیچیده سنجش معلمان را وادار به تصمیم‌گیری حرفه‌ای می‌کند تا واقعیت‌های کلاس درس خود را با تأثیر عوامل بیرونی همسو کنند. معلمان تصمیمات مربوط به سنجش و نمره دهی خود را بر دانش، عقیده، انتظارات و ارزش‌های تدریس و یادگیری استوار می‌کنند. پژوهش حاضر به بررسی تصمیم‌گیری برای نمره دهی معلمان دوره متوسطه در تهران می‌پردازد. در این راستا، عوامل مورد استفاده در نمره دهی مورد بررسی قرار می‌گیرد. بدین منظور پرسشنامه‌ای بین ۳۴۸ معلم زبان انگلیسی مقطع متوسطه در پایه‌های یازدهم و پیش-دانشگاهی توزیع شد. تحلیل توصیفی پرسشنامه‌ها نشان داد که معلمان زبان انگلیسی تهرانی عوامل پیشرفت تحصیلی و غیر از پیشرفت تحصیلی را در نمره دهی لحاظ می‌کنند و تأکید بیشتری بر روی عواملی غیر از پیشرفت تحصیلی نظیر تلاش، تکالیف، و عادت‌های مطالعه می‌گذارند. همچنین معلمان از انواع مختلف سنجش نظیر سنجش عملکرد محور، تکلیف محور، سنجش معلم ساخته و آزمون‌های سنتی کاغذ و مدادی استفاده نمودند. نتایج تحلیل واریانس چندگانه نشان داد که عوامل درونی و بیرونی مانند پایه تدریس معلم، آموزش‌های سنجش که معلمان داشته‌اند، و اندازه کلاس بر جنبه‌های مختلف تصمیم‌گیری آن‌ها در نمره دهی تأثیر گذاشته است. نتایج این پژوهش به‌طور کلی به درک تصمیم‌گیری برای نمره دهی معلمان زبان انگلیسی و علی‌الخصوص معلمان زبان انگلیسی تهرانی کمک می‌کند.

واژگان کلیدی: سنجش، عوامل بیرونی و درونی، معلمان زبان انگلیسی، دوره متوسطه

۱. استادیار گروه زبان و ادبیات انگلیسی، دانشگاه علامه طباطبائی (نویسنده مسئول) easalimi@atu.ac.ir

مقدمه

معلمان، مدیران آموزشی و محققان همواره در مورد کفایت و کارایی نظام‌های نمره دهی مختلف تردید داشته‌اند (بلک و ویلیام^۱، ۱۹۸۸؛ گاسکی و بیلی^۲، ۲۰۰۱، تئاف^۳، ۱۹۶۴). دست‌اندرکاران آموزشی همواره به پایایی نمره دهی معلمان توجه داشته و نظام‌های نمره دهی پایا در محیط‌های آموزشی را تجزیه و تحلیل می‌کرده‌اند. در سال‌های اخیر با رشد روزافزون مفهوم ارزشیابی برای یادگیری (هیوم و کول^۴، ۲۰۰۹؛ سیف، ۲۰۰۰؛ مارشال و دراموند^۵، ۲۰۰۶) و همچنین علایق فزاینده در خصوص پژوهش‌های کلاس‌های درس (اندراد^۶، ۲۰۰۹) راجع به تصمیم‌گیری برای نمره دهی معلمان این نکته را نشان می‌دهد که معلمان عوامل مختلف پیشرفت تحصیلی و عواملی غیر از پیشرفت تحصیلی را در تصمیم‌گیری‌های خود لحاظ می‌کنند (گاسکی، ۲۰۱۱؛ رندال و انگلهارد^۷، ۲۰۰۹؛ یسبک^۸، ۲۰۱۱). این عوامل می‌تواند با هم در تضاد باشند و فرایند نمره دهی معلمان را با مشکل مواجه کنند (بروکهارت^۹، ۱۹۹۱؛ کراس و فراری^{۱۰}، ۱۹۹۶).

پژوهش‌های به‌عمل آمده در مورد روش‌های سنجش معلمان زبان انگلیسی به‌عنوان زبان خارجی تقریباً جدید هستند (بیرجندی و سیری، ۲۰۱۰؛ بریندلی^{۱۱}، ۲۰۰۷؛ ریا-دیکین^{۱۲}، ۲۰۰۴). طبیعتاً مطالعات زیادی در مورد عوامل مؤثر بر نمره دهی معلمان انجام نشده است. چنگ و وانگ^{۱۳} (۲۰۰۷) و دیویدسون^{۱۴} (۲۰۰۴) نمره دهی معلمان در هنگ‌کنگ، استرالیا،

-
1. Black & William
 2. Guskey & Bailey
 3. Teaf
 4. Hume & Coll
 5. Marshall & Drummond
 6. Andrade
 7. Randall & Engelhard
 8. Yesbeck
 9. Brookhart
 10. Cross & Frary
 11. Brindley
 12. Rea-Dickins
 13. Cheng and Wang
 14. Davison

کانادا را مقایسه کردند. همه این مطالعات پیچیدگی شیوه‌های سنجش معلمان را نشان می‌دهد که عمدتاً تحت‌الشعاع عقاید آن‌ها و عوامل مربوط به بافت‌های بیرونی قرار می‌گیرد. با این حال، هنوز هم عوامل تعیین‌کننده نمره دهی معلمان و روش‌های سنجش آنان نامعلوم هستند.

پژوهش در مورد شیوه‌های تصمیم‌گیری برای نمره دهی می‌تواند تحولات آموزشی مهمی در ایران به دنبال داشته باشد که متأثر از نقش بااهمیت و دیرینه آزمون‌ها در کشور است (بیرجندی و سیری، ۲۰۱۰). اخیراً بحث‌های فراوانی در مورد ارزشیابی نظام باز و اصلاحات در نمره دهی در خصوص آزمون‌های مهم و سطح بالا در کشور رخ داده است (قربانی و نیساری، ۲۰۱۵؛ حسینی، ۲۰۰۷؛ حشمتی، ۲۰۰۴). همچنین مطالعاتی در مورد فرایندهای مختلف نمره دهی مثل نمره دهی صدکی و نمره دهی حروفی و توصیفی (صالح‌آبادی، ۲۰۰۵) و نمره دهی استاندارد (قربانی، ۲۰۱۲) انجام شده است. با وجود این، مطالعات مرتبط با تصمیم‌گیری برای نمره دهی معلمان در کشور بسیار محدود است.

شیوه‌های نمره دهی معلمان. نیوتون^۱ (۲۰۰۷) نمره دهی را صرفاً به‌عنوان فرایند تعیین نمره ارزش‌گذارانه می‌داند که در بستر یک فرایند فنی و استاندارد حاصل می‌شود. این موضوع نشان می‌دهد که نمره دهی معلمان می‌تواند از طریق آموزش حرفه‌ای بهبود یابد؛ اما پژوهش‌های بسیاری نشان داده است که معلمان از شیوه‌های ارائه‌شده در زمینه نمره دهی استفاده نمی‌کنند. مثلاً در دوره‌های ضمن خدمت و پیش از خدمت نمره دهی منحصرأ بر مبنای عامل پیشرفت تحصیلی است (دایرنس و دایرنس^۲، ۲۰۰۸؛ مک میلان^۳، ۲۰۰۸؛ اوکانر^۴، ۲۰۰۷)؛ اما معلمان در عمل عوامل متعددی نظیر تلاش، عادت‌های کاری و پیشرفت تحصیلی را در نمره دهی اعمال می‌کنند (گاسکی، ۲۰۱۱؛ رندال و انگلهدارد، ۲۰۰۹؛ یسبک، ۲۰۱۱). بروکهارت (۱۹۹۱) خاطر نشان می‌کند که حتی معلمانی که از علم سنجش و ارزشیابی آگاهی دارند نیز صرفاً بر اساس پیشرفت تحصیلی دانش‌آموزان نمره را تعیین می‌کنند. تفاوت بین شیوه‌های

1. Newton
2. Dyrness & Dyrness
3. McMillan
4. O'Connor

نمره دهی معلمان و پیشنهادهای ارائه‌شده نقش دوگانه معلمان در ارزشیابی کلاسی را نشان می‌دهد که همان نقش قاضی و مربی است (بیشاپ^۱، ۱۹۹۲؛ ویلسون^۲، ۱۹۹۶). معلم به‌عنوان یک قاضی باید نمره دانش‌آموز را بر اساس پیشرفت تحصیلی معین کرده و در این راستا بر عدالت، برابری و قطعیت تصمیم‌گیری‌های خود اتکا کند. به‌عنوان یک مربی او باید به مواردی که برای رشد دانش‌آموز مفید است توجه کند که شامل عواملی غیر از پیشرفت تحصیلی مانند تلاش، تشویق و بهبود از طریق نمره دهی می‌شود. بیشاپ (۱۹۹۲) معتقد است که معلمان نمی‌توانند همزمان قاضی و مربی باشند و آن‌ها باید قضاوت را رها و بر روابط مشاوره‌ای با دانش‌آموزان تمرکز کنند.

مک میلان و نش^۳ (۲۰۰۰) یک مدل ارزشیابی کلاسی و تصمیم‌گیری در مورد نمره دهی را پیشنهاد دادند. مدل آن‌ها تأثیر عوامل داخلی مثل نظرات معلمان و ارزش‌های آن‌ها و عوامل بیرونی مانند تأثیر والدین را نشان می‌دهد. این عوامل و دیگر واقعیت‌های کلاس درس بر استفاده از روش‌های مختلف ارزشیابی مانند آزمون‌ها، آزمونک و سنجش کارایی اثر می‌گذارد. در یک مطالعه موردی زوکلر^۴ (۲۰۰۵) به بررسی نحوه‌ی نمره دهی عادلانه معلمان آمریکایی پرداخت و تأثیر عوامل مربوط به پیشرفت و عواملی غیر از پیشرفت تحصیلی به همراه انتظارات معلم را مورد بررسی قرار داد. نتایج پژوهش او نشان داد که نمره دهی به‌طور کلی تحت تأثیر یک نظام نمره دهی، درک معلم از تلاش، و توجه آن‌ها بر رشد اخلاقی قرار گرفته است.

چنگ و وانگ (۲۰۰۷) نمره دهی معلمان را در سه بافت مؤسسه‌ای در کانادا، هنگ‌کنگ و چین مورد بررسی قرار دادند. آن‌ها چندین عامل بافتی که می‌توانست بر نظام نمره دهی در این کشورها مؤثر باشد را شناسایی کردند. مثلاً آن‌ها ترجیح معلمان برای روش‌های مختلف نمره دهی و عقاید آن‌ها در مورد جهت‌گیری ارزشیابی را تجزیه و تحلیل کردند. آن‌ها همچنین خاطرنشان کردند که واقعیت‌های کلاس درس به‌ویژه اندازه کلاس عامل دیگری

-
1. Bishop
 2. Wilson
 3. McMillan & Nash
 4. Zoeckler

بود که بر ارزشیابی معلم اثر گذاشت. مطالعات زیادی رابطه بین شیوه نمره دهی معلمان و ویژگی‌هایی چون جنسیت و سابقه تدریس آن‌ها (سیزک، فیتزجرالد و راجر^۱، ۱۹۹۵)، پایه مورد تدریس آن‌ها (رنдал و انگلهدارد، ۲۰۰۱) و موضوع مورد تدریس (مک میلان، ۱۹۹۸) را مورد بررسی قرار داده‌اند. نتایج مختلف این پژوهش‌ها نشان می‌دهد که شیوه‌های نمره دهی بسیار متغیر و غیرقابل پیش‌بینی هستند. یکی از عواملی که اخیراً مورد توجه قرار گرفته است میزان آموزش سنجش است که معلم‌ها دریافت می‌کنند. بروکهارت (۲۰۰۴) معتقد است تدریس و آموزش سنجش در نحوه تفکر معلمان برای نمره دهی تفاوت‌هایی ایجاد می‌کند اما تأثیری بر نوع تفکر آن‌ها در خصوص تحولات ارزشی و نتایج اجتماعی نمره دهی آن‌ها ندارد. چن، راجرز و وانگ^۲ (۲۰۰۸) خاطرنشان می‌کنند که اگرچه آموزش می‌تواند بر انتخاب نوع سنجش معلمان اثر بگذارد، بافت کلی تدریس نسبت به ویژگی‌های معلمان نقش مهم‌تری را ایفا می‌کند. در بستر آموزشی زبان انگلیسی به‌عنوان زبان خارجی در دوره متوسطه اندازه کلاس درس هم یکی از ویژگی‌های محیط آموزشی است؛ بنابراین پژوهش حاضر سه عامل مؤثر در تصمیم‌گیری برای نمره دهی شامل آموزش سنجش، پایه تدریس و اندازه کلاس را مورد بررسی قرار می‌دهد.

به‌طور خلاصه تحقیقات قبلی در مورد شیوه نمره دهی معلمان پیچیدگی این امر را آشکار ساخته است. در حقیقت فرایند پیچیده سنجش معلمان را وادار به تصمیم‌گیری حرفه‌ای می‌کند تا واقعیت‌های کلاس درس خود را با تأثیر عوامل بیرونی همسو کنند. معلمان تصمیمات مربوط به سنجش و نمره دهی خود را بر دانش، عقیده، انتظارات و ارزش‌های تدریس و یادگیری استوار می‌کنند. عقیده معلمان و ارزش‌های آنان اغلب با فشارهای بیرونی و مقتضیات آن در تضاد است که برای معلمان چالش‌برانگیز است چراکه از طرفی باید اطلاعات روشن و قابل‌فهمی را از طریق نمره به مسئولین گزارش دهند و از طرف دیگر، نقش نمره در افزایش انگیزه و یادگیری و همین‌طور عواقب سوء آن را نیز در نظر بگیرند.

1. Cizek, Fitzgerald, & Rachor
2. Cheng, Rogers, & Wang

از این رو، پژوهشگران (بروکهارت، ۲۰۰۳؛ مک میلان، ۲۰۰۳؛ ماس^۱، ۲۰۰۳) به بازنگری تئوری سنجش و اندازه‌گیری پرداخته‌اند تا آن را با بافت سنجشی در کلاس درس مرتبط‌تر کنند. به همین منظور، چگونگی تصمیم‌گیری معلمان برای نمره دهی در بسترهای خاص اجتماعی، فرهنگی و آموزشی حائز اهمیت است.

پژوهش حاضر به دنبال بررسی تأثیر سه متغیر معلم محور و آموزش محور پایه تدریس، آموزش سنجش و اندازه کلاس در تصمیم‌گیری برای نمره دهی معلمان زبان خارجی مقطع متوسطه شهر تهران است. در همین راستا، پژوه حاضر تلاش می‌کند به سؤالات ذیل پاسخ دهد.

۱. معلمان زبان انگلیسی مقطع متوسطه در ایران برای نمره دهی تا چه اندازه عوامل متعددی را لحاظ کرده و انواع مختلف سنجش را به کار می‌برند؟
۲. جنبه‌های مختلف تصمیم‌گیری معلمان زبان انگلیسی برای نمره دهی در مقطع متوسطه ایران تا چه میزان تحت تأثیر عواملی چون پایه تدریس، آموزش سنجش و اندازه کلاس قرار می‌گیرد؟
۳. چه رابطه‌ای بین عوامل مدنظر معلمان زبان انگلیسی در مقطع متوسطه و انواع مختلف سنجش برای نمره دهی آن‌ها وجود دارد؟

روش پژوهش

شرکت‌کنندگان این پژوهش ۳۴۸ نفر از معلمان زبان انگلیسی دوره متوسطه (پایه یازدهم و پیش‌دانشگاهی) از ۲۲ مدرسه در پنج منطقه آموزش و پرورش تهران است (۷۳،۳٪ خانم و ۲۶،۷٪ آقا). انتخاب شهر تهران به جهت دارا بودن جمعیت بسیار بیشتر آن در مقایسه با شهرهای دیگر برای شرکت در آزمون ورودی دانشگاه‌ها صورت گرفت که این امر هم از نگاه معلمان و دانش‌آموزان حائز اهمیت است و طبیعتاً باعث شده در نمره دهی معلمان تأثیر داشته باشد.

1. Moss

اکثر معلمان بین ۲۵ تا ۳۹ سال بودند که در پایه‌های یازدهم (۱۸۳ نفر) و پیش‌دانشگاهی (۱۶۵ نفر) مقطع متوسطه مشغول تدریس بودند. (۲۵-۲۹=۲۶/۳؛ ۳۰-۳۴=۲۹/۸؛ ۳۵-۳۹=۲۶/۳) سابقه تدریس معلمان از ۴ تا ۲۹ سال بود. حدود ۵۹٪ آن‌ها دارای مدرک کارشناسی آموزش زبان انگلیسی و ۴۱٪ دارای مدرک کارشناسی زبان انگلیسی (ادبیات و مترجمی) بودند. از میان همه معلمان حدود ۳۸٪ دوره کامل ضمن خدمت سنجش و ارزشیابی را گذرانده بودند و ۳۳٪ دوره ناتمام سنجش و ارزشیابی را گذرانده بودند و ۲۹٪ آموزش سنجش و ارزشیابی نداشتند. متوسط ساعت کاری آن‌ها در هفته حدود ۱۶ بود و متوسط اندازه کلاس ۴۰ دانش‌آموز بود.

ابزار پژوهش پرسشنامه‌ای به زبان انگلیسی بود که برای فهم بهتر معلمان برخی از واژگان و اصطلاحات آن به فارسی ارائه شد. پرسشنامه بر اساس مطالعات قبلی در زمینه سنجش و ارزشیابی معلمان و شیوه نمره دهی آن‌ها طراحی و ساخته شد (بروکهارت، ۱۹۹۳؛ فراری، کراس و وبر، ۱۹۹۳). پرسشنامه دارای چهار بخش است. بخش اول که شامل ۱۷ گویه است میزان عوامل مختلف لحاظ شده در نمره دهی معلمان را اندازه‌گیری می‌کند. این بخش بر اساس مقیاس لیکرت ۱ پنج‌تایی {هرگز (صفر) تا همیشه (پنج)} تنظیم شده است. گویه‌های این قسمت مربوط به عواملی مثل عملکرد دانش‌آموزان، تلاش و پیشرفت تحصیلی آن‌ها است که معلمان در نمره دهی لحاظ می‌کنند.

بخش دوم پرسشنامه مشتمل بر ۱۰ گویه با مقیاس لیکرت پنج‌تایی برای سنجش روش‌های مختلف سنجش و ارزشیابی طراحی شد. گویه‌های این بخش شامل انواع سنجش مثل آزمون کاغذ و مدادی (کتبی)، سنجش عملکرد و سنجش معلم ساخته است. بخش سوم شامل سه سناریو و سؤالات باز مربوطه است که مفهوم و ارزش نمرات معلمان را مورد تفحص قرار می‌دهد. به دلیل محدودیت فضا این بخش از پرسشنامه در این نوشتار گزارش نمی‌شود. بخش آخر پرسشنامه شامل هشت گویه است که اطلاعات جمعیت شناختی نظیر سن و جنسیت، سابقه تدریس، ساعات کاری، اندازه کلاس و آموزش‌های سنجش و

1. Likert Scale

ارزشیابی را جمع‌آوری کرد. پرسشنامه با تعداد ۱۳ نفر از معلمان مشابه نمونه آماری مورد آزمایش قرار گرفت و بر اساس بازخورد آن‌ها و تجزیه و تحلیل گویه‌ها اصلاحاتی انجام شد. مطالعه حاضر عوامل و روش‌های مختلف سنجش و ارزشیابی معلمان را گزارش می‌کند. ضریب پایایی آلفای کرونباخ^۱ در این پژوهش برای بخش اول و دوم پرسشنامه به ترتیب ۰/۸۷٪ و ۰/۷۹٪ بود.

با کمک مدیران گروه‌های آموزشی زبان انگلیسی مناطق مختلف آموزش و پرورش موفق شدیم تا با معلمان تماس گرفته و پرسشنامه را در اختیار آن‌ها قرار دهیم. معلمان از اهداف پژوهش و هدف شرکت آن‌ها و همچنین محرمانه ماندن اطلاعات آگاهی یافتند. مدت زمان لازم برای تکمیل پرسشنامه حدود ۲۰ تا ۳۰ دقیقه پیش‌بینی شده بود. مقرر شده بود شرکت‌کنندگان پس از تکمیل پرسشنامه‌ها را به مدیران گروه‌ها تحویل دهند. در نهایت ۹۳٪ از پرسشنامه‌ها پر شدند.

آمار توصیفی برای متغیرهای پرسشنامه محاسبه شد و داده‌های پرت^۲ مورد شناسایی قرار گرفت. مثلاً در مورد ساعات کاری متوسط کمتر از ۴/۷ داده پرت تلقی شد چراکه حداقل ساعت کاری ۶ جلسه بود. نتایج آزمون Little's MCAR برای مقادیر گمشده^۳ نشان داد که تعداد آن‌ها به صورت تصادفی کم بوده است.

بنابراین چهار فرایند تحلیلی مورد استفاده قرار گرفت. در ابتدا آمار توصیفی برای دو بخش اول پرسشنامه در خصوص عوامل متعدد مدنظر معلمان و انواع مختلف سنجش در نمره دهی آن‌ها محاسبه و طبقه‌بندی شد. در مرحله دوم، تحلیل مؤلفه‌های اصلی برای عوامل نمره دهی و سپس برای انواع مختلف سنجش اجرا شد. هدف تحلیل مؤلفه‌های اصلی یافتن الگوهایی است که در داده‌ها بروز می‌کند (براون^۴، ۲۰۰۹)؛ بنابراین تحلیل عامل اکتشافی در این مرحله اجرا نشد. آزمون کولموگروف-اسمیرنوف^۵ برای کفایت داده‌های عوامل

-
1. Cronbach's alpha
 2. outliers
 3. Missing data
 4. Brown
 5. Kolmogorov-Smirnov

مختلف نمره دهی ۰/۸۸ و برای روش‌های مختلف سنجش و ارزشیابی ۰/۸۶ بود. آزمون بارتلت^۱ نیز در سطح ($p < .001$) هم در خصوص عوامل و هم شیوه‌های سنجش معنادار بود. روش واریماکس^۲ برای چرخش داده‌ها در تحلیل مؤلفه‌های اصلی و به حداقل رسانی تعداد متغیرها و بار بالای عاملی یک مؤلفه مورد استفاده قرار گرفت. برای انتخاب مؤلفه‌ها دو معیار مقادیر ویژه (یک و بالاتر) و محل قرارگیری آن در بالای انحنای اسکری^۳ استفاده شد. نمره مؤلفه‌ها با استفاده از روش اندرسون-روبین^۴ محاسبه شد چراکه در این روش وزن هر گویه مؤلفه مورد لحاظ قرار می‌گیرد.

در مرحله سوم دو مجموعه آزمون تحلیل واریانس چند متغیره با استفاده از نمرات مؤلفه‌ها اجرا شد تا تأثیر متغیرهای پایه تدریس، آموزش سنجش و اندازه کلاس را اندازه‌گیری کند. مجموعه اول آزمون تحلیل واریانس چند متغیره برای عوامل لحاظ شده نمره دهی و مجموعه دوم برای انواع مختلف سنجش اجرا شد. در اینجا باید تحلیل واریانس چند متغیره سه راهه اجرا می‌شد ولی به دلیل حجم کم نمونه این آزمون به کار برده نشد؛ بنابراین سه آزمون جداگانه تحلیل واریانس دوبه‌دو چند متغیره اجرا شد.

در اینجا شرکت کنندگان در پایه تدریس (پایه یازدهم و پیش‌دانشگاهی)، آموزش سنجش (معلمان آموزش دیده و آموزش ندیده) و اندازه کلاس (۴۰ نفر و بیش از ۴۱ نفر) گروه‌بندی شدند. توزیع فراوانی اندازه کلاس نشان داد که توزیع دو حالتی بوده است و گروه بالایی بیش از ۴۶ و پایینی کمتر از ۴۰ نفر بوده است. برای بررسی یکپارچگی واریانس در بین سه عوامل گروهی، ۱۲ آزمون لوین^۵ اجرا شد و میزان واریانس‌ها مورد محاسبه قرار گرفت. آزمون لوین برای هیچ کدام از متغیرهای وابسته به لحاظ آماری معنادار نبود. فرض نرمال بودن تک متغیری از طریق چولگی و کشیدگی و همچنین بافت نگاشت^۶ مورد آزمون قرار گرفتند. آزمون‌های تحلیل چند متغیره با تحلیل عملکرد تفاوت همراه شدند. هدف این

1. Bartlett test
2. varimax
3. scree plot
4. Anderson-Rubin
5. Levene's tests
6. histograms and plots

آزمون توضیح تفاوت بین معلمان پایه‌های یازدهم، کلاس‌های کوچک و بزرگ، و معلمانی که آموزش سنجش دیده بودند و آن‌هایی که ندیده بودند است. در نهایت سه آزمون رگرسیون چندگانه برای پیش‌بینی نوع سنجش با استفاده از عواملی که معلمان لحاظ کردند اعمال شد. فرض خطی بودن با مقدار بحرانی ۰/۰۵ برای آزمون در نظر گرفته شد.

نتایج

این بخش در ابتدا به تجزیه و تحلیل عوامل دخیل در نمره دهی و سپس انواع مختلف سنجش و ارزیابی معلمان می‌پردازد. سپس یافته‌های مربوط به تأثیر متغیرهای آموزش محور و معلم محور (پایه تدریس، آموزش سنجش و اندازه کلاس) بر روی شیوه نمره دهی و رابطه آن‌ها با روش‌های ارزشیابی ارائه می‌کند.

عوامل مدنظر معلمان در نمره دهی

معدل نمره گویه‌های بخش اول پرسشنامه برای مشخص کردن تأثیر عوامل مختلف در نمره دهی معلمان لحاظ شد. جدول ۱ معدل و انحراف معیار این گویه‌ها را نشان می‌دهد. آماره‌های توصیفی نشان داد که عوامل مختلفی در نمره دهی تأثیر داشت. محدوده معدل در مقیاس پنج‌تایی بین ۲/۷۱ و ۴/۴۵ متغیر بود. شش گویه با معدل بیش از ۴ وجود داشت و "تلاش" دانش‌آموزان مهم‌ترین عامل از نظر معلمان بود. کمترین معدل مربوط به گوی‌های بود که به توزیع نمره دهی معلمان دیگر می‌پرداخت. معدل گویه‌های دیگر از ۳/۱۷ تا ۳/۸۹ متغیر بود. معدل دو گویه مربوط به پیشرفت تحصیلی و یادگیری اهداف مشخص به ترتیب ۳/۵۹ و ۳/۴۹ بودند. انحراف معیار هر ۱۷ گویه نسبتاً بالا بود و اختلاف قابل توجهی را در عوامل مختلف نمره دهی نمایان می‌کرد.

جدول ۱. آمار توصیفی عوامل لحاظ شده در نمره دهی معلمان

انحراف معیار	میانگین	تعداد	عوامل مورد استفاده در تعیین نمرات
.۹۲	۴.۲۷	۳۴۴	تلاش دانش‌آموز
.۹۱	۴.۲۷	۳۴۶	شاخص‌های غیر آزمونی برای موارد مرزی
.۹۹	۴.۲۵	۳۴۶	بهبود کارایی علمی
.۹۳	۴.۲۴	۳۴۶	کیفیت تکالیف انجام شده
۱.۰۵	۴.۱۲	۳۳۷	عادت‌های مطالعه
۱.۰۱	۴.۱۱	۳۴۶	انجام تکالیف
۱.۱۱	۳.۹۷	۳۴۸	رفتار مخرب دانش‌آموز
۱.۰۲	۳.۸۹	۳۴۶	میزان توجه و مشارکت دانش‌آموز در کلاس
۱.۰۰	۳.۸۹	۳۴۵	سطح توانایی تحصیلی
.۹۲	۳.۶۱	۳۴۵	عملکرد تحصیلی
۱.۱۲	۳.۵۲	۳۴۶	یادگیری اهداف مشخص
۱.۱۱	۳.۴۸	۳۴۶	مقایسه عملکرد با دانش‌آموزان دیگر
۱.۱۸	۳.۳۴	۳۴۴	سیاست رسمی و غیررسمی مدرسه
۱.۰۲	۳.۲۹	۳۳۹	لحاظ کردن نمره صفر برای تکالیف ناتمام
۱.۱۶	۳.۲۳	۳۳۷	نمره اضافه برای عملکرد غیر آموزشی
۱.۲۸	۳.۱۲	۳۴۳	مقایسه عملکرد با دانش‌آموزان سال‌های قبل
۱.۳۱	۲.۸۳	۳۴۳	نمرات معلمان دیگر

جدول ۲. بار عاملی برای عوامل مؤثر در نمره دهی

مؤلفه			گویه
۳	۲	۱	
.۱۱۹	.۱۷۸	.۷۱۳	نمره معلمان دیگر
.۱۳۹	.۱۵۵	.۶۹۷	یادگیری اهداف مشخص
.۳۷۰	.۰۹۹	.۶۷۰	لحاظ کردن نمره صفر برای تکالیف ناتمام
.۲۱۶	.۰۶۹	.۶۴۲	سیاست رسمی و غیررسمی مدرسه
.۲۴۱	.۰۶۷	.۶۰۲	مقایسه عملکرد دانش‌آموزان سال‌های ماقبل
.۱۷۲	.۳۲۷	.۵۷۲	میزان توجه و مشارکت دانش‌آموز در کلاس
.۳۲۲	.۷۲۲	.۲۱۸	انجام تکالیف
.۳۰۳	.۷۰۴	.۰۷۹	تلاش دانش‌آموز
.۴۳۵	.۶۸۹	.۱۵۶	کیفیت تکالیف انجام شده

۰.۴۴۸	۰.۶۰۴	۰.۲۱۶	عادت‌های مطالعه
-۰.۰۸۲	۰.۵۶۷	۰.۰۸۹	بهبود عملکرد تحصیلی
-۰.۰۵۵	۰.۵۴۵	۰.۴۳۸	رفتار مخرب دانش‌آموز
۰.۷۳۳	۰.۰۰۴	۰.۳۱۳	عملکرد تحصیلی
۰.۶۷۲	۰.۳۳۰	۰.۰۳۴	سطح توانایی تحصیلی
۰.۵۹۴	۰.۱۵۰	۰.۳۲۵	نمره اضافی برای عملکرد غیر تحصیلی
۰.۵۵۲	۰.۱۸۶	۰.۲۹۰	مقایسه عملکرد دانش‌آموزان دیگر
۰.۴۳۰	۰.۴۲۴	۰.۲۲۷	علائم غیر آزمون‌ی برای موارد مرزی

تحلیل مؤلفه‌های اصلی منجر به پیدایش سه مؤلفه شد (جدول ۲). این بارهای عاملی همبستگی میان مؤلفه‌ها و متغیرهای مشاهده‌شده را نمایان می‌کند. موضوعات مشترک میان بارهای بالا می‌تواند سازه‌ها را مشخص کند. شش گویه دارای بار عاملی بالا بود و این مؤلفه‌ها شامل نمره دهی معلمان دیگر، تسلط بر اهداف یادگیری، تکالیف ناتمام، سیاست مدرسه، مقایسه عملکرد دانش‌آموزان با سال‌های قبل و توجه و مشارکت آن‌ها بودند؛ بنابراین این مؤلفه عامل هدف‌دار یا هنجارمدار نام دارد. در مؤلفه دوم نیز شش گویه بارگذاری شد که ابتدائاً شامل تلاش، تکالیف، پیشرفت، عادات کاری و رفتار مخرب بودند. این مؤلفه به‌عنوان عامل تلاش تعریف شد. گویه‌هایی که بر روی این مؤلفه بارگذاری شدند معدل‌های نسبتاً بالا و انحراف معیارهای پایینی داشتند (جدول ۱). درنهایت چهار گوی‌های که در مؤلفه سوم بارگذاری شدند مربوط به عملکرد علمی و غیرعلمی، توانایی علمی و مقایسه آن با دانش‌آموزان دیگر بود؛ بنابراین این مؤلفه عامل عملکرد نامیده شد.

انواع مختلف ارزیابی در نمره دهی

جدول ۳ معدل و انحراف معیار انواع مختلف روش‌های سنجش و ارزشیابی را نشان می‌دهد. این آماره‌های توصیفی نشان می‌دهند که معلمان از انواع مختلف روش‌های ارزشیابی در نمره دهی خود استفاده کردند.

جدول ۳. آمار توصیفی برای روش‌های سنجش مورد استفاده معلمان

انحراف معیار	میانگین	تعداد	
.۹۴	۴.۳۶	۳۴۸	امتحانات اصلی
۱.۰۸	۳.۷۸	۳۴۶	آزمونک‌ها
۱.۱۶	۳.۷۱	۳۴۵	سنجش عینی
۱.۰۹	۳.۶۰	۳۴۵	سنجش خودساخته
۱.۱۰	۳.۵۹	۳۴۸	تکلیف‌های فردی
.۹۹	۳.۵۲	۳۴۷	سنجش عملکرد
۱.۲۰	۳.۵۱	۳۴۸	تکلیف گروهی
۱.۲۳	۳.۳۳	۳۴۷	ارائه شفاهی
۱.۱۳	۳.۰۸	۳۳۵	ارزشیابی ارائه شده توسط ناشران
۱.۳۰	۲.۸۹	۳۴۷	سؤالات باز

جدول ۴. بار عاملی انواع روش‌های سنجش برای نمره دهی

گویه	مؤلفه		
	۱	۲	۳
کارهای گروهی	.۸۲۷	.۰۹۳	.۰۴۲
سؤالات باز	.۷۸۶	.۱۷۸	-.۱۰۳
ارزشیابی ارائه شده توسط ناشران	.۶۷۱	.۱۴۵	.۰۲۳
کارهای فردی	.۶۴۷	-.۰۰۹	.۳۰۳
ارزشیابی عملکرد	.۵۳۳	.۴۲۰	.۱۲۸
ارزشیابی خودساخته	.۱۷۷	.۷۶۵	.۱۱۸
ارزشیابی عینی	.۰۱۶	.۷۴۱	.۲۶۴
ارائه شفاهی	.۱۸۷	.۷۱۲	-.۱۶۴
آزمون‌های اصلی	.۰۰۵	.۰۰۷	.۸۰۲
آزمونک‌ها	.۱۴۶	.۱۷۷	.۷۸۸

معدل ده روش سنجش و ارزیابی از ۲/۸۹ تا ۴/۳۶ متغیر بود. امتحانات اصلی با معدل ۴/۳۶ و انحراف معیار ۰.۹۴. بیشترین نوع سنجش مورد استفاده بود. سؤالات نوع مقاله با معدل

۲/۸۹ و انحراف معیار ۱/۳۰ کمترین نوع سنجش استفاده شده بود. معدل انواع دیگر ارزیابی از ۳/۰۶ تا ۳/۷۹ با انحراف معیار ۰/۹۸ تا ۱/۱۹ متغیر بود.

تحلیل مؤلفه‌های اصلی سه نوع مؤلفه سنجشی را ارائه کرد (جدول چهار). پنج گویه بر روی مؤلفه اول بارگذاری شد. این گویه‌ها عمدتاً مربوط به سنجش عملکرد/ تکلیف محور بودند. سه گویه دیگر بر روی مؤلفه سنجش معلم-ساخته بارگذاری شد. دو گویه‌های که بر روی مؤلفه سوم بارگذاری شد شامل آزمون‌ها و آزمونک‌ها بودند که بالاترین معدل را داشتند.

تأثیر پایه تدریس، آموزش سنجش و اندازه کلاس بر عوامل نمره دهی. سه آزمون تحلیل واریانس چند متغیره برای برآورد تأثیر پایه تدریس، آموزش سنجش و اندازه کلاس اجرا شد. نمرات عوامل (عامل هنجارمدار، عامل تلاش، و عامل عملکرد) در این سه مؤلفه به‌عنوان متغیر وابسته استفاده شد. مجذور اتای تفکیکی^۱ برای نشان دادن میزان واریانس این آزمون‌ها گزارش می‌شود.

پایه تدریس و آموزش سنجش تأثیر تعاملی معناداری تولید کرد. $F(3,166) = 7.59, p < .001, \text{partial } \eta^2 = .12$ و این مقدار برای تأثیر اصلی آموزش سنجش نیز صادق بود. $F(3,166) = 3.01, p = .03, \text{partial } \eta^2 = .05$. تحلیل واریانس‌های بعدی تعامل معنادار تأثیر عامل هدف مدار $F(1, 168) = 14.50, p < .001, \text{partial } \eta^2 = .08$ و تلاش $F(1, 168) = 4.50, p = .035, \text{partial } \eta^2 = .03$ را نشان داد.

تأثیر تعاملی آموزش سنجش و پایه تدریس نشان می‌دهد که آموزش برای معلمان پایه یازدهم و پیش‌دانشگاهی مهم تلقی می‌شود. با این حال جهت تأثیر برای این دو گروه از معلمان متفاوت است. اثر این دو عامل پس از دریافت آموزش سنجش برای معلمان پیش‌دانشگاهی کاهش یافت ولی برای معلمان پایه یازدهم افزایش یافت.

تحلیل عامل چندگانه برای متغیرهای اندازه کلاس و پایه تدریس تأثیر تعاملی و یا اصلی به همراه نداشت؛ اما تحلیل عامل چندگانه برای متغیرهای آموزش سنجش و اندازه کلاس

1. partial η^2

حاکمی از آن است که آموزش سنجش تأثیر اصلی معنادار بر عامل تلاش و عامل هدف/هنجارمداری و همین‌طور تأثیر اصلی معنادار اندازه کلاس بر عامل تلاش دارد. تحلیل عامل تشخیصی (افتراقی)^۱ متعاقب عامل افتراقی را نشان داد که دو گروه تحت آموزش سنجش را به‌طور معناداری متمایز می‌کرد. $(R^2 = .06, \eta^2 = .94, \chi^2(3) = 11.04, p = .012)$. همبستگی‌های بین عوامل پیش‌بین و عامل افتراقی حاکمی از آن بود که عامل عملکرد بار عاملی سنگینی بر روی متغیر افتراقی داشت و درحالی‌که بار عاملی عامل هدف/هنجارمدار و عامل تلاش بر روی عامل افتراقی نسبتاً متعادل بود. (به ترتیب $r = .53$ و $r = .40$).

تأثیر پایه تدریس، آموزش سنجش و اندازه کلاس بر انواع روش‌های سنجش برای نمره دهی. سه تحلیل واریانس چندگانه دوراهه برای بررسی تأثیر پایه تدریس معلمان، آموزش سنجش و اندازه کلاس بر انواع مختلف روش سنجش آن‌ها به‌عنوان متغیرهای مستقل برای نمره دهی اجرا شد. نمرات عامل (عامل هدف/هنجارمدار، عامل تلاش، و عامل عملکرد) به‌عنوان متغیر وابسته در نظر گرفته شد. تحلیل واریانس چندگانه پایه تدریس و آموزش سنجش هم تأثیر تعاملی معناداری نشان داد $(F(3, 188) = 25.56, p < .001, \text{partial } \eta^2 = .29)$ که در مورد پایه تدریس هم اثر اصلی معناداری داشت $(F(3, 188) = 13.49, p < .001, \text{partial } \eta^2 = .18)$. تحلیل واریانس‌های بعدی نیز تعامل معناداری برای سنجش عملکرد / تکلیف محور نشان داد $(F(1, 190) = 18.72, p < .001, \text{partial } \eta^2 = .09)$ که در مورد سنجش معلم ساخته نیز صادق بود $(F(1, 190) = 50.95, p < .001, \text{partial } \eta^2 = .21)$...

تحلیل‌های آماری نشان می‌دهد جهت تأثیر آموزش سنجش برای معلمان پایه یازدهم و پیش‌دانشگاهی متفاوت بوده است. استفاده معلمان پایه یازدهم از سنجش عملکرد/تکلیف-محور با آموزش سنجش کاهش یافت درحالی‌که عکس این حالت برای معلمان پیش‌دانشگاهی اتفاق افتاد.

1. Discriminant function analysis

آموزش سنجش بر ارزیابی معلم-ساخته معلمان در هر دو مقطع پایه یازدهم و پیش‌دانشگاهی مؤثر بود. معلمان پایه یازدهم که آموزش سنجش دیده بودند بیش از همکاران آموزش ندیده خود در همان مقطع از ارزیابی معلم-ساخته استفاده کردند در حالی که در مقطع پیش‌دانشگاهی معلمان آموزش دیده نسبت به همکاران آموزش ندیده از این روش سنجش بیشتر استفاده نکردند.

تحلیل واریانس چندگانه^۱ اندازه کلاس و پایه تدریس اثر تعاملی معناداری را نشان نداد. داشت $(F(3, 197) = 4.76, p = .003, \text{partial } \eta^2 = .07)$ اما اندازه کلاس اثر اصلی معناداری

تحلیل واریانس‌های بعدی حاکی از اثر اصلی معنادار اندازه کلاس هم بر ارزیابی معلم-ساخته و آزمون مداد-کاغذی (کتبی) بود. $(F(1, 116) = 7.24, p = .008, \text{partial } \eta^2 = .06)$ نشان دادند. $(F(1, 116) = 4.51, p = .005, \text{partial } \eta^2 = .11)$. معلمان که کلاس‌های کوچک‌تری داشتند بیشتر از ارزیابی خودساخته استفاده کردند ولی معلمان که کلاس‌های بزرگ‌تری داشتند به آزمون داد-کاغذی (کتبی) رغبت

جدول ۵. تحلیل عامل‌های افتراقی انواع روش‌های سنجش مورد استفاده معلمان برای نمره دهی

همبستگی متغیر کانونی (r)			عامل افتراقی			عضویت پیش‌بینی شده
PPT***	TSDA**	PPBA*	p	X ²	Λ	
-.۳۳	.۹۰	-.۲۴	.۰۱۱	۱۱.۱۰	.۹۶	.۰۵
.۹۵	.۱۹	.۰۲	.۰۲۰	۹.۸۱	.۹۶	.۰۴
.۸۷	-.۴۳	-.۲۴	.۰۱۷	۱۰.۱۶	.۹۴	.۰۷

*PPBA = سنجش عملکرد / تکلیف محور

**TSDA = سنجش خودساخته معلم

***PPT = آزمون کاغذ و مداد

1. MANOVA

جدول ۵ نتایج تحلیل عامل‌های تشخیصی (افتراقی) را نشان می‌دهد. برای هر یک از متغیرهای وابسته (پایه تدریس، آموزش سنجش، اندازه کلاس) یک عامل افتراقی یافت شد که هر دو گروه داخل خود را به‌طور چشم‌گیری متمایز می‌کرد؛ اما میزان اثر این عوامل افتراقی از ۰/۰۵ تا ۰/۰۷ و بسیار کم بود. $\chi^2(3) = 11.04, p = .012, \lambda = .94$. تحلیل عامل‌های تشخیصی (افتراقی) را نشان داد که متغیرهای پیش‌بین با تک‌تک عوامل افتراقی همبستگی بالایی داشت. به‌عنوان مثال، سنجش خودساخته معلم بار عاملی سنگینی بر عوامل افتراقی متمایزکننده دو گروه تحت آموزش و عدم آموزش سنجش ($r = 0.95$) و دو گروه مربوط به کلاس بزرگ و کوچک داشت ($r = 0.87$). آزمون کاغذ و مداد بر روی عامل افتراقی متمایزکننده دو گروه مربوط به دو پایه تدریس بار عاملی سنگینی داشت ($r = 0.95$) اما بر عامل افتراقی متمایزکننده دو گروه مربوط به اندازه کلاس ($r = -0.43$) و دو گروه مربوط به آموزش و عدم آموزش بار عاملی ملایمی داشت ($r = -0.32$).

روابط بین عوامل مورد استفاده و روش‌های سنجش. رگرسیون چندگانه^۱ برای بررسی رابطه بین عوامل مورد استفاده و انواع روش‌های سنجش معلمان اجرا شد. نمرات سه مؤلفه (عامل هدف/هنجارمدار، عامل تلاش، و عامل عملکرد) به‌عنوان متغیر مستقل و نمرات انواع روش‌های سنجش (سنجش عملکرد/تکلیف محور، سنجش معلم ساخته، آزمون کاغذ و مداد) به‌عنوان متغیر وابسته در نظر گرفته شد. جدول ۶ نشان می‌دهد که به چه میزان می‌توان سه روش سنجش معلمان برای نمره دهی را با سه عامل مؤثر در نمره دهی پیش‌بینی کرد. به‌طور کلی، این عوامل می‌تواند به‌خوبی استفاده از سنجش عملکرد/تکلیف محور ($R^2 = .26, p < .001$) و سنجش معلم ساخته را پیش‌بینی کند ($R^2 = .32, p < .001$). مقادیر بتای استاندارد (β) در واحد انحراف معیار اندازه‌گیری شد و بنابراین مستقیماً قابل مقایسه است و نشان داد که متغیر وابسته به ازای یک انحراف معیار تغییر در متغیر مستقل به تعداد چند انحراف معیار تغییر خواهد کرد. عامل عملکرد پیش‌بینی کننده قوی سنجش عملکرد/تکلیف محور بود ($\beta = .47, p < .001$) و عامل هدف/هنجار محور پیش‌بینی کننده قوی

1. Multiple Regression

سنجش معلم ساخته بود ($\beta = .51, p < .001$). با این حال، قدرت پیش‌بینی سه عامل مدنظر در پیش‌بینی آزمون‌های کاغذ و مدادی ضعیف بود و تنها ۸٪ از تغییر متغیر وابسته را تبیین کرد.

جدول ۶. نتایج رگرسیون چندگانه در استفاده از انواع سنجش و عوامل مورد استفاده در نمره دهی

متغیر	β	T	P	95% CI	R^2
(ثابت)		-.۵۹	.۵۵۷	[-.۱۴, .۰۸]	
عامل ارجاعی	.۱۱	۲.۰۴	.۰۴۲	[.۰۰, .۲۲]	.۲۶*
عامل تلاش	.۱۶	۳.۰۷	.۰۰۲	[.۰۶, .۲۷]	
عامل عملکرد	.۴۷	۸.۷۶	.۰۰۰	[.۳۷, .۵۹]	
متغیر وابسته: سنجش عملکرد / تکلیف محور - تعداد = ۲۶۹					
(ثابت)		.۶۳	.۵۳۰	[-.۰۷, .۱۳]	
عامل ارجاعی	.۵۱	۱۰.۰۷	.۰۰۰	[.۴۱, .۶۰]	.۳۲*
عامل تلاش	.۲۴	۴.۷۳	.۰۰۰	[.۱۴, .۳۴]	
عامل عملکرد	.۰۵	۱.۰۷	.۲۸۴	[-.۰۵, .۱۶]	
متغیر وابسته: سنجش معلم ساخته - تعداد = ۲۶۹					
(ثابت)		.۱۵	.۸۷۹	[-.۱۰, .۱۲]	
عامل ارجاعی	.۱۳	۲.۲۴	.۰۲۶	[.۰۲, .۲۴]	.۰۸*
عامل تلاش	.۱۸	۲.۹۷	.۰۰۳	[.۰۶, .۲۸]	
عامل عملکرد	.۱۶	۲.۷۴	.۰۰۷	[.۰۴, .۲۷]	
متغیر وابسته: آزمون کاغذ و مداد - تعداد = ۲۶۹					

CI = فاصله اطمینان

بحث

یافته‌های این پژوهش پیچیدگی‌های تصمیم‌گیری در نمره دهی معلمان زبان انگلیسی در دبیرستان‌های تهران را نشان می‌دهد. در وهله اول عواملی غیر از پیشرفت تحصیلی برای معلمان از اهمیت بالایی برخوردار هست. نتایج این مطالعه با پژوهش‌های انجام‌شده در آمریکای شمالی مشابه است (کراس و فراری، ۱۹۹۶). معلمان زبان انگلیسی در تهران اهمیت ویژه‌ای به تلاش دانش‌آموزان در نمره دهی خود دادند که در بافتار نظام آموزشی ایران

اعتقاد بر این است که یادگیری در نتیجه تلاش صورت می‌گیرد و نه توانایی (بیرجندی و سیری، ۲۰۱۰).

علاوه بر این، نتایج تحلیل واریانس چند متغیره نشان داد که تفاوت معناداری بین معلمانی که آموزش سنجش دیده‌اند و آن‌هایی که آموزشی دریافت نکرده‌اند وجود دارد. باین حال تأثیر آموزش سنجش بر ملاحظات معلمان باید با توجه به پایه تدریس آن‌ها در نظر گرفته شود. آموزش سنجش برای معلمان پایه یازدهم و پیش‌دانشگاهی در این پژوهش می‌تواند به روش‌های متفاوتی صورت پذیرد. مثلاً آموزش سنجش می‌تواند تأثیر متفاوتی بر سنجش معلم ساخته و سنجش عملکرد/ تکلیف محور معلمان پایه یازدهم و پیش‌دانشگاهی داشته باشد. یکی از توضیحات احتمالی این امر می‌تواند تأثیر چشمگیر آزمون بسیار مهم ورودی دانشگاه‌ها بر آموزش معلمان پایه یازدهم و پیش‌دانشگاهی باشد. عملکرد دانش‌آموزان در آزمون ورودی در سطح کشور با هم مقایسه می‌شود و به‌عنوان پاداش یا تنبیه در قالب عدم پذیرش مورد استفاده قرار می‌گیرند. یکی از اثرات مهم آموزش منطبق کردن سنجش کلاسی معلمان با آزمون‌های خارج از کلاس است که از اهمیت بالایی برای دانش‌آموزان، معلمان، والدین و مدیران آموزشی دارد (چنگک، ۲۰۰۸؛ شی، ۲۰۰۵). این نوع آموزش سنجش برای معلمان نمونه سؤالاتی را فراهم می‌کند که در سطح کلان کشوری طراحی می‌شود. معلم به استفاده از این نوع آزمون‌ها در کلاس خود تشویق می‌شود. در یک مطالعه مشابه برخی از دانش‌آموزان معیارهای معلم خوب را در کسی می‌بینند که در آزمون‌ها به آنان کمک می‌کنند (شی، ۲۰۰۶).

استفاده از آزمون‌های کاغذ و مدادی میان‌ترم و پایان‌ترم عامل دیگری بود که بین معلمان مقطع متوسطه تمایز ایجاد کرد. این تفاوت باز هم تأثیرات آزمون مقطع بالاتر را در تصمیم‌گیری معلمان در مورد شیوه‌های سنجش نشان می‌دهد. فارغ‌التحصیلان دوره پیش‌دانشگاهی در این پژوهش در امتحان ورودی دانشگاه شرکت می‌کنند تا به مؤسسات آموزش عالی راه یابند و زبان انگلیسی یک از دروس تعیین‌کننده در این گونه آزمون‌ها است. (چنگک و

قی^۱، ۲۰۰۶). به دست آوردن نمرات بالا هدف معلمان و دانش‌آموزان است چراکه ورود به دانشگاه‌های بهتر هدف اکثر دانش‌آموختگان دوره متوسطه است. معلمان دوره متوسطه نیز سنجش خود را با آزمون ورودی دانشگاه‌ها هماهنگ کرده و از آن‌ها بیشتر در امتحانات پایان‌ترم استفاده می‌کنند. البته این نوع آزمون تأثیرات مخربی بر آموزش و یادگیری دارد (تنگ و بیگز^۲، ۱۹۹۶).

نوع سنجش نیز می‌تواند معلمان پایه یازدهم و پیش‌دانشگاهی را از هم متمایز کند. به‌طور کلی تأثیر آموزش سنجش، پایه تدریس و اندازه کلاس در استفاده معلمان از انواع مختلف سنجش بیشتر از تأثیر عواملی است که به خود معلمان مربوط می‌شود. این یافته‌ها نشان می‌دهد که آموزش سنجش، پایه تدریس و اندازه کلاس در مقایسه با عوامل مربوط به نمره دهی تأثیر بیشتری بر انواع سنجش دارد.

تأثیر قابل توجه اندازه کلاس در استفاده از سنجش معلم ساخته و آزمون‌های کاغذ و مدادی نشان می‌دهد که اندازه کم کلاس این امکان را به معلمان می‌دهد تا ارزشیابی‌های مخصوص خود را طراحی کنند.

نتایج رگرسیون چندگانه نشان داد که نمره دهی معلمان می‌تواند یک فرایند تصمیم‌گیری باشد و روش‌های سنجش می‌تواند توسط عوامل مربوط به نمره دهی پیش‌بینی شود؛ بنابراین مطالعه این عوامل تحولات مهمی برای آموزش سنجش معلمان دارد. درحالی که افزودن بر توانایی معلمان برای استفاده از روش‌های مختلف سنجش مهم است مسئله مهم‌تر تحقیق در مورد قضاوت ارزشی آن‌ها در نمره دهی است. این نتایج همچنین پیچیدگی نمره دهی را نشان می‌دهد که خود نه تنها متأثر از عوامل معلم-محور است بلکه متأثر از واقعیت‌های کلاس مثل اندازه آن و عوامل محیطی و آموزشی است. استفاده کم از آزمون‌های کاغذ و مدادی می‌تواند تأثیر کم سه عامل مورداستفاده معلمان در نمره دهی و انواع سنجش را نشان دهد. باین حال، برای تجزیه و تحلیل آزمون‌های پایان‌ترم نیاز به درک کاملی از بستر آموزش داریم.

1. Cheng & Qi
2. Tang & Biggs

نتیجه‌گیری

نتایج این پژوهش با مطالعات مشابه قبلی همسو است (گل‌سکی، ۲۰۱۱؛ رندال و انگل‌هارد، ۲۰۰۹؛ یسبک، ۲۰۱۱). تمرکز بر عواملی غیر از پیشرفت تحصیلی در بستر این پژوهش بسی نمایان است. مطالعات قبلی نشان داده است که اعمال عواملی غیر از پیشرفت تحصیلی در نمره دهی روایی سازه‌ای را با خطر مواجهه می‌کند. پس بهتر است توجه خود را به ارزش‌های معلمان در مورد آموزش و یادگیری و نتیجه نمره دهی آن‌ها در یک محیط آموزشی و اجتماعی متمرکز کنیم (بروکهارت، ۱۹۹۳؛ چنگ و همکاران، ۲۰۰۸).

روش‌های سنجش مورد استفاده تا حدود زیادی جمع‌آوری داده‌های پیشرفت تحصیلی را نمایان می‌کند و نتایج متناقض و پیچیده‌ای را به همراه خواهد داشت. نتایج پژوهش حاضر می‌تواند به دلیل نمونه محدود روش‌های سنجش باشد. پژوهش‌های بیشتر با طرح‌های آزمایشی می‌تواند تبیین بهتری ارائه نماید. در حقیقت نمره دهی معلمان فرایند پیچیده‌ای است که متأثر از عوامل زمینه‌ای و محیطی است.

علاوه بر این، تأثیر آموزش سنجش برای معلمان متفاوت بود. برای درک بهتر تصمیم‌گیری معلمان در نمره دهی مطالعات بیشتری نیاز است تا اصالت و محتوای این امر را مورد تفحص قرار دهد. همچنین تحقیقات بیشتری مورد نیاز است تا بر تحلیل بسترهای آموزشی و اجتماعی تمرکز کند (چنگ و همکاران، ۲۰۰۸).

تحقیق حاضر بر اهمیت تأثیر عوامل بافتی در شیوه نمره دهی معلمان می‌افزاید. یادگیری زبان انگلیسی یک اولویت ملی در ایران محسوب می‌شود و نمرات آن برای ادامه تحصیل در مقاطع بالاتر بسیار مهم تلقی می‌شود. در اینجا آزمون مقطع بالاتر (کنکور) تأثیر چشمگیری بر سنجش کلاسی معلم دارد. علاوه بر این، آزمون‌های مقطع بالاتر بر نوع آموزش‌های سنجش که معلمان دریافت می‌کنند اثر می‌گذارد. پس برای درک بهتر تأثیر آموزش سنجش بر نمره دهی معلمان اهمیت این تأثیر باید در نظر گرفته شود. عواقب نمره دهی بر یادگیری دانش‌آموزان از اولویت‌های معلمان محسوب می‌شود. عوامل دخیل در نمره دهی معلمان عقاید و فرض‌های آنان در مورد یادگیری را منعکس می‌کند.

به لحاظ روش‌شناختی نیازی به جمع‌آوری مجدد داده‌های کیفی در مورد فرایند نمره دهی معلمان از دیدگاه دانش‌آموزان نیز احساس می‌شود تا نتایج محکم‌تری ارائه شود. تنها با جمع‌آوری داده‌ها از معلمان و دانش‌آموزان روایی نمره دهی بهتر درک می‌شود. تحلیل داده‌های این پژوهش در صورتی گسترش پیدا خواهد کرد که هر سه متغیر مستقل در تحلیل واریانس چندگانه به سه ماتریکس دوه‌دو در تحلیل واریانس تبدیل شود. بدون لحاظ کردن این محدودیت، یافته‌های این پژوهش بر درک ما از نمره دهی معلمان را می‌افزاید و دربرگیرنده تحولات مهمی در مورد این مسئله است.

منابع

- حشمتی، عبدالرضا و دیگران (۱۳۸۳). طراحی یک آزمون خوب. اداره‌ی کل سنجش و ارزشیابی تحصیلی. انتشارات عابد. تهران.
- سیف، علی‌اکبر (۱۳۷۹). روش‌های اندازه‌گیری و ارزشیابی آموزشی. نشر دوران.
- موسوی، فرشته، (۱۳۸۴). چهارگام اساسی در ارزشیابی پیشرفت تحصیلی اداره‌ی کل سنجش و ارزشیابی تحصیلی. انتشارات عابد. تهران.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership*, 49(6), 15-18.
- Birjandi, P. & Siyyari, M. (2010). Self-assessment and Peer-assessment: A Comparative Study of Their Effect on Writing Performance and Rating Accuracy, *Iranian Journal of Applied Linguistics*, 13(1), 23-45.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Brindley, G. (2007). Editorial. *Language Assessment Quarterly*, 4(1), 1-5.
- Brookhart, S. M. (1991). Letter: Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.

- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.
- Brown, J. D. (2009). Principal components analysis and exploratory factor analysis Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 13(1), 26–30.
- Cheng, L. (2010). The history of examinations: Why, how, what and whom to select? In L. Cheng, & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 13–26). New York, NY: Routledge.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159–179.
- Cross, L., & Frary, R. (1996, April). Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53(1), 1–21.
- Dyrness, R., & Dyrness, A. (2008). Making the grade in middle school. *Kappa Delta Pi Record*, 44(3), 114–118.
- Ghorbani, M. R. (2012). Washback effect of the university entrance examination: Iranian case. Germany: Lambert Academic Publishing (LAP)
- Ghorbani, M. R. & Neissari, M. (2015). Washback Effect of the Iranian Concours on Senior High School Students' EFL Learning Activities. *Iranian Journal of language Testing*, 5(1), 1-28.
- Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership*, 69(3), 17–21.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Hosseini, S. M. H. (2007). ELT in higher education in Iran and India – A critical view. *Language in India*, 7, 1-11. Retrieved October 9, 2008, from <http://www.languageinindia.com/dec2007/eltinindiaandiran.pdf>
- Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice*, 16(3), 269–290.
- Lam, A. S. L. (2005). *Language education in China: Policy and experience from 1949*. Hong Kong SAR, China: Hong Kong University Press.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd ed.). Thousand Oaks, CA: Sage Publications.

- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203–213.
- Merwin, J. C. (1989). Evaluation. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 185–192). Oxford, UK: Pergamon Press.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stake test. *Language Testing*, 22(2), 142–173.
- Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *Journal of Educational Research*, 102(3), 175–185.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shi, L. (2006). The successors to Confucianism or a new generation? A questionnaire study on Chinese students' culture of learning English. *Language, Culture and Curriculum*, 19(1), 122–147.
- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meanings and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343.
- Teaf, H. M. (1964). Grades: Their dominion is challenged. *The Journal of Higher Education*, 35(2), 87–88.
- Wang, F. (2008). Motivation and English achievement: An exploratory and confirmatory factor analysis of a new measure for Chinese students of English Learning. *North American Journal of Psychology*, 10(3), 633–646.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 34(3), 14–27.
- Yesbeck, D. M. (2011). Grading practices: Teachers' considerations of academic and non-academic factors (Unpublished doctoral dissertation). Virginia Commonwealth University, Richmond, Virginia.
- Zoeckler, L. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education*, 35(2), 83–102.