

## Examining Teachers' Assessment Literacy in EFL Classrooms: The Effect of Educational Background

Masoomeh Estaji\*

Associate Professor, Applied Linguistics Dept.,  
Allameh Tabataba'i University, Tehran, Iran

### Abstract

Classroom assessment, as the complementary part of the language learning process, is a powerful decision-making instrument. Nonetheless, more research is required regarding the ways teachers cope with these requirements, and how they affect their pedagogical practices. This study is an attempt to examine EFL teachers' perceptions of assessment literacy and the criteria they consider to assess their students. Moreover, it examines whether graduate and undergraduate teachers differ in terms of their assessment literacy. To this end, by using a survey and Ex-Post Facto research design and through a two-part questionnaire on assessment literacy (adapted from Plake, 1993; Plake, Impara, & Fager, 1993), a comparison was made between undergraduate (N=22) and graduate teachers (N=10) of English as a Foreign Language (EFL), English Literature, Translation studies, and Linguistics selected through purposive sampling on their perceptions of assessment literacy. The research results revealed a statistically significant difference between undergraduate and graduate teachers' perceptions of assessment literacy. Results also showed that graduate teachers had higher perceptions of assessment literacy than their undergraduate counterparts, representing the effect of their level of education and educational background. Thus, this study highlights the significance of giving sufficient and proper training to all soon-to-be language teachers on language assessment, argues for the need, suggests ways by which teachers can become more literate in the domain of language assessment, and presents ways teacher educators and language testing experts can assist in this path.

**Keywords:** language assessment literacy, classroom assessment, educational background, teachers' perceptions, EFL context

\* Corresponding Author: mestaji74@gmail.com

**How to Cite:** Estaji, M. (2021). Examining Teachers' Assessment Literacy in EFL Classrooms: The Effect of Educational Background. *Quarterly of Educational Measurement*, 12(45), 7-38. doi: 10.22054/jem.2022.59105.2153



## بررسی سواد ارزشیابی معلمان در کلاس‌های آموزش زبان انگلیسی به‌عنوان یک زبان خارجی: تأثیر پیشینه تحصیلی

دانشیار زبان‌شناسی کاربردی، گروه زبان و ادبیات انگلیسی، دانشگاه علامه طباطبائی، تهران، ایران

معصومه استاجی \*

### چکیده

ارزشیابی‌های کلاسی بخش جدایی‌ناپذیر فرایند یادگیری زبان و ابزاری قدرتمند برای تصمیم‌گیری آگاهانه و صحیح است. با این حال، شناخت نسبتاً کمی در مورد چگونگی برخورد معلمان با این نوع مطالبات وجود دارد و مهم‌تر اینکه، چگونه چنین شیوه‌های ارزشیابی در تدریس روزانه آن‌ها تأثیر می‌گذارد. این مقاله در نظر دارد تا سواد ارزشیابی معلمان زبان انگلیسی به‌عنوان یک زبان خارجی و معیار ارزشیابی آن‌ها از روش‌های ارزشیابی را بررسی کند. علاوه بر این، تأثیر پیشینه تحصیلی آنان در سطح سواد ارزشیابی‌شان مورد تجزیه و تحلیل قرار گرفته است. بدین منظور با استفاده از روش تحقیق پس‌رویدادی و با استفاده از یک پرسشنامه‌ی دویبخشی (Plake, 1993, Plake et al., 1993) در مورد سواد ارزشیابی، مقایسه‌ای بین معلمان با مدرک تحصیلی زبان انگلیسی و گرایش‌های مرتبط (آموزش زبان انگلیسی، ادبیات انگلیسی، مطالعات ترجمه و زبان‌شناسی) در مقطع کارشناسی (تعداد، ۲۲ نفر) و تحصیلات تکمیلی (کارشناسی ارشد و دکتری، تعداد ۱۰ نفر) که از طریق نمونه‌گیری هدفمند انتخاب شده بودند انجام شد. یافته‌ها نشان داد که تفاوت معناداری در سطح سواد ارزشیابی زبان معلمان با توجه به پیشینه تحصیلی آن‌ها وجود دارد. به‌خصوص، نتایج نشان می‌دهند که معلمان سطوح تحصیلات تکمیلی از سواد ارزشیابی زبان بالاتری نسبت به معلمان مقطع کارشناسی برخوردار هستند که نشان‌دهنده تأثیر سطح و پیشینه تحصیلات معلمان است. در نتیجه، این تحقیق بر اهمیت ارائه‌ی آموزش مناسب در ارتقاء سواد ارزشیابی برای تمامی معلمان زبان انگلیسی آینده تأکید می‌کند، نیاز به چنین دوره‌هایی را بررسی می‌نماید، راه‌های باسواد شدن در حوزه ارزشیابی زبان را به معلمان نشان می‌دهد و به ارائه‌ی روش‌هایی می‌پردازد که از طریق آن کارشناسان آزمون‌سازی و مربیان معلم می‌توانند در رسیدن به این مهم کارگشا باشند.

**کلیدواژه‌ها:** ارزشیابی کلاسی، ادراک معلمان، بافت انگلیسی به‌عنوان یک زبان خارجی، پیشینه تحصیلی، سواد ارزشیابی زبان

## مقدمه

ارزشیابی<sup>۱</sup> از دیرباز بخش مکمل فرایند تدریس و یادگیری بوده است (James et al., 2002) و آن را موتور محرک یادگیری می‌نامند (Cowan, 1998). یکی از روش‌های مؤثر تقویت یادگیری در دوران تحصیلات تکمیلی بهبود رویه‌های ارزشیابی است و ارزشیابی عملکرد زبان‌آموزان یکی از ضروری‌ترین وظایف هر معلم است. بخش اعظم زمان فعالیت حرفه‌ای معلمان به فعالیت‌های مرتبط با ارزشیابی شامل تعیین روش‌های ارزشیابی، اجرا، نمره‌دهی، تعبیر نتایج ارزشیابی، روش نمره‌دهی، اعلان و انتقال نتایج ارزشیابی و استفاده از نتایج ارزشیابی در تصمیم‌گیری‌های آموزشی اختصاص دارد. اقدامات اصلاحی گوناگونی معلمان را ترغیب می‌کنند تا خود را مسئول ارزشیابی یادگیری زبان‌آموزانشان بدانند. همان‌طور که Linn and Grolund (2000) بیان می‌کنند، مسئولیت‌پذیری آموزشی به معنای ایجاد الزامات بالاتر در ارزشیابی کلاسی و افزایش تعداد ارزشیابی‌های بایسته است. ارزشیابی و سنجش بر معلمان، زبان‌آموزان، والدین، مدارس، اصلاحات آموزشی و برنامه‌های آماده‌سازی معلم بسیار تأثیر می‌گذارند و در عرصه آموزش مسائل بسیار مطرح و بحث‌برانگیزی هستند (Phye, 1997). قانون هیچ کودک‌کی جا نماند<sup>۲</sup> در سال ۲۰۰۱ که به صورت یک قانون مصوب در سال 2002 در آمد، مدارس دولتی را ملزم به اجرای نظام‌های مسئولیت‌پذیری می‌نمود. با چنین روندی، معلمان مجبور هستند خود را برای ارزشیابی و سنجش یادگیری زبان‌آموزانشان آماده کنند، آموزش کلاسی خود را تقویت بخشند و بتوانند نتایج ارزشیابی دارای استانداردهای خارجی را تعبیر و تفسیر کنند. اساساً مدرسان باید ابرانگاره خود را تغییر دهند تا چگونگی تأثیر ارزشیابی بر آموزش و نیز تأثیر مثبت آن بر عملکرد و یادگیری زبان‌آموزان را درک کنند. از این رو، آن‌ها باید اهداف مختلف ارزشیابی را بشناسند و طبق آن‌ها عمل کنند.

به کارگیری مناسب این فعالیت‌ها نیازمند دانش و توانایی قوی معلم در سنجش ارزشیابی است که به این توانایی و دانش «سواد ارزشیابی»<sup>۳</sup> اطلاق می‌شود و شامل دانش در مورد چیستی، چرایی و چگونگی ارزشیابی، چالش‌های ممکن در روند ارزشیابی و چگونگی مرتفع کردن این موانع است (Stiggins, 1995). از آنجا که توانش‌های ارزشیابی زبان‌آموز

---

1. assessment  
2. No Child Left Behind (NCLB)  
3. assessment literacy

منجر به تصمیم گیری های دقیق و آگاهانه می شود، سواد ارزشیابی از این حیث بسیار در کیفیت آموزش تأثیرگذار است. معلمان باید تسلط مناسب بر اصول و تکنیک های ارزشیابی داشته باشند تا بتوانند قضاوت های پیچیده ای از اعتبار و روایی سیاست گذاری ها و روش های ارزشیابی در بافت های آموزشی متفاوت داشته باشند (Kane, 2006). بالعکس، برخورداری از سواد ارزشیابی ناکافی منجر به تصمیم گیری های آموزشی نادرست و ناآگاهانه می شود؛ بنابراین، سواد ارزشیابی معلم در موفقیت ارزشیابی آموزشی و کیفیت آموزش نقش بسیار تعیین کننده ای دارد.

پژوهشگران بر این باورند که ارزشیابی کلاسی باید آموزش و یادگیری زبان آموزان را پشتیبانی و تقویت کند (Shepard, 2001)؛ اما برآیند مطالعات صورت گرفته چنین است که معلمان پیوسته از عوامل مختلفی در روش های ارزشیابی خود استفاده می کنند و متعاقباً تصمیمات اشتباهی می گیرند. حتی بدتر آن که اکثر معلمان فاقد مهارت ها و دانش کارآمد ارزشیابی هستند؛ به عبارتی، آن ها هنگام سنجش موفقیت تحصیلی زبان آموز انگاشت های اشتباهی را در مورد روش های ارزشیابی نشان می دهند (Cizek et al., 1991)؛ (McMillan, 2001). به اجمال، گرچه بسیاری از معلمان ظاهراً این فرایند را درک می کنند، به نظر می رسد بیشتر آن ها درک درستی از آن ندارند و به اندازه کافی آمادگی حل این مسئله را ندارند. اگر سطح سواد ارزشیابی مدرسان بالاتر نرود، نمی توانیم به زبان آموزان کمک کنیم که به موفقیت های تحصیلی بیشتر دست یابند. در این راستا، معلمان به دلیل آگاهی یافتن از عدم آمادگی خود برای ارزشیابی، خواهان آموزش بیشتری هستند و پیوسته نقاط ضعف خود را در برنامه های دوره کارشناسی شان ابراز می کنند (Rogers, 1991).

مطالعات گسترده ای در حوزه سواد ارزشیابی معلمان انگلیسی در نتیجه توسعه حرفه ای معلمان یا دوره های تربیت مدرس (Wang et al., 2008) انجام شده است. متأسفانه، مطالعات انجام شده نشان می دهند سواد ارزشیابی معلمان ناکافی است (Xu & Brown, 2017) و بسیاری از مطالعات عدم تطبیق روش های ارزشیابی معلم با اهداف آموزشی را نشان می دهند (Berry et al., 2017؛ O'z & Atay, 2017). بعضی مطالعات نیز سواد ارزشیابی را در ارتباط با موفقیت زبان آموزان (Braney, 2010) بررسی کرده اند، اما تا امروز و تا آنجا که محققان مطلع هستند، مطالعه ای درباره سواد ارزشیابی معلمان انگلیسی بر اساس سطح تحصیلی شان صورت نگرفته است. از آنجا که سواد ارزشیابی معلمان نمایانگر

بخش مهمی از توسعه حرفه‌ای آنان است، لازم است مطالعات در حوزه ارزشیابی کلاسی را با در نظر گرفتن پیشینه تحصیلی معلمان گسترش دهیم. علاوه بر این، از آنجا که مسائل و مشکلات مطرح شده در این تحقیق برای ارتقاء آموزش و موفقیت زبان‌آموزان بسیار حائز اهمیت هستند، نتایج این مطالعه می‌تواند مرجع معنادار و خردمندانه‌ای برای مطالعات آتی باشد. به همین منظور، این مطالعه به بررسی سواد ارزشیابی مدرسان زبان انگلیسی به‌عنوان زبان خارجی با در نظر گرفتن پیشینه تحصیلی‌شان می‌پردازد و با استفاده از یک پرسشنامه دویبخشی که مبتنی بر معیارهای توانش معلم در ارزشیابی آموزشی زبان‌آموزان تدوین شده است، سواد ارزشیابی معلمان را مقایسه می‌کند.

### پیشینه پژوهش

نظر به مؤلفه اصلی پژوهش، آنچه در این بخش واکاوی می‌شود، پیشینه پژوهش‌های صورت گرفته در خصوص توانش معلم در ارزشیابی آموزشی یادگیرندگان و سواد ارزشیابی معلمان است.

توانش معلم در ارزشیابی آموزشی زبان‌آموزان: آنچه معلمان باید در مورد ارزشیابی بدانند، به عوامل مختلفی از قبیل نمره، آموزش در حوزه محتوای مختلف، نظام‌های ارزشیابی مورد استفاده در منطقه و الزامات منطقه‌ای و دولتی بستگی دارد. فهرست معیارهای توانش معلم در ارزشیابی آموزشی زبان‌آموزان<sup>۱</sup> (1990) اقدام مشترک انجمن آمریکایی معلمان، شورای ملی سنجش آموزش و انجمن ملی آموزش بود. این اقدام در سال ۱۹۸۷ شروع شد و هدف آن طراحی و گسترش معیارهایی برای توانش معلم در ارزشیابی زبان‌آموزان بود؛ به این امید که مزایای آموزشی بالقوه ارزشیابی زبان‌آموزان را کاملاً محقق سازند. هدف اولیه آن‌ها بررسی مشکل آموزش ناکافی ارزشیابی برای معلمان بود. فهرست معیارهای مذکور که به صورت چارچوبی مفهومی برای معلمان درآمد، شامل هفت مورد ذیل است:

۱. انتخاب روش‌های ارزشیابی مناسب با تصمیمات آموزشی
۲. طراحی روش‌های ارزشیابی مناسب با تصمیمات آموزشی
۳. اجرا، نمره‌دهی و تفسیر نتایج روش‌های ارزشیابی تدوین‌شده معلم و بیرونی

1. The Standards for Teacher Competence in the Educational Assessment of Students (American Federation of Teachers, AFT; The National Council on Measurement in Education, NCME; and The National Education Association, NEA, 1990)

۴. بهره‌گیری از نتایج ارزشیابی هنگام تصمیم‌گیری در مورد زبان‌آموزان، برنامه‌ریزی آموزش، طراحی برنامه‌ی درسی و ارتقاء مدرسه
۵. طراحی رویه‌های نمره‌دهی معتبر که از ارزشیابی‌های زبان‌آموزان بهره می‌گیرند
۶. انتقال نتایج ارزشیابی به زبان‌آموزان، والدین و سایرین
۷. تشخیص روش‌های ارزشیابی غیراخلاقی، غیرقانونی و هر نوع ارزشیابی ناصحیح دیگر و استفاده از اطلاعات ارزشیابی (۳۱-۳۲).

چنین معیارهایی به‌صراحت حوزه‌هایی را معرفی می‌کنند که معلمان در آن‌ها باید تسلط خیلی خوبی داشته باشند تا بتوانند یادگیری زبان‌آموزانشان را ارزشیابی کنند. این معیارها اهمیت آموزش معلم و رشد حرفه‌ای وی در عرصه‌ی ارزشیابی کلاسی را تأیید و تعیین می‌کنند (Brookhart, 2001). این هفت معیار همگی برای طراحی و استفاده از ارزشیابی‌های کلاسی، اهداف و مقاصد آموزشی توسط معلم به کار می‌روند. معیارهای شماره ۳، ۴، ۶ و ۷ برای ارزشیابی کلان-مقیاس از قبیل اجرا، تفسیر و بیان نتایج ارزشیابی، استفاده از اطلاعات برای تصمیم‌گیری و تشخیص شیوه‌های غیراخلاقی نیز به کار می‌روند (Brookhart, 2001). سواد ارزشیابی معلمان: در عصر حاضر، آزمون‌سازی و ارزشیابی اهمیت وافر پیدا کرده‌اند و طبق برآوردها، معلمان حدود ۵۰ درصد از وقت خود را صرف فعالیت‌های ارزشیابی می‌کنند (Plake, 1993). جدای از میزان زمان صرف‌شده، ارزشیابی کلاسی کارکردهای بسیار مهمی دارد (Brookhart, 1999, 1998). ارزشیابی با اهداف مختلفی انجام می‌شود: تشخیص نیازهای زبان‌آموزان، گروه‌بندی و نمره‌دهی آن‌ها، سنجش آموزش و انگیزه‌بخشی و دیگر موارد (Stiggins, 1999). ارزشیابی و نمره‌دهی صحیح به معلمان کمک می‌کند سبک آموزشی خود را تقویت کنند؛ همچنین انگیزه و سطح موفقیت زبان‌آموزان را بالا می‌برد (Brookhart, 1999). به باور Stiggins (1999)، «کیفیت آموزش در هر کلاسی به کیفیت ارزشیابی‌های به‌کاررفته بستگی دارد» (ص. ۲۰). از این‌رو، اطلاعات برگرفته از ارزشیابی‌ها باید معنادار و صحیح باشد؛ یعنی اطلاعات باید معتبر و موثق باشند. (Brookhart, 1999).

اخیراً، پژوهش‌های فراوانی معلمان را به کسب سواد ارزشیابی فراخوانده‌اند. Popham (2004) نبود چنین سوادی در معلمان را «خودکشی حرفه‌ای<sup>۱</sup>» (ص. ۸۲) نامیده و عنوان نموده است که معلمان موظف‌اند زمان خود را صرف کسب دانش در این حوزه نمایند. سواد

ارزشیابی عبارت است از دانش معلم در مورد اصول اساسی ارزشیابی از قبیل اصطلاحات، طراحی و استفاده از روش‌ها و تکنیک‌های ارزشیابی و آشنایی با معیارهای کیفیت ارزشیابی. این دانش هر روز بیشتر و بیشتر جایگزین شیوه‌های سنتی سنجش یادگیری می‌شود (Hearne, 2001).

برخی دیگر از پژوهشگران گرچه رسماً سواد ارزشیابی را تعریف نکرده‌اند، اما ویژگی‌های دارندگان آن را توصیف نموده‌اند. موارد زیر یکی از این دست توصیف‌هاست:

- معلمان دارای سواد ارزشیابی شیوه‌های درست ارزشیابی، سنجش و برقراری ارتباط را تشخیص می‌دهند.
- آن‌ها می‌دانند که برای جمع‌آوری اطلاعات صحیح و موفقیت زبان‌آموزان باید از کدام روش‌های ارزشیابی استفاده کنند.
- نتایج ارزشیابی را به روش‌های مؤثری بیان می‌کنند؛ چه از طریق کارنامه، نمرات آزمون، کارپوشه، یا کنفرانس.
- می‌توانند انگیزه و یادگیری زبان‌آموزان را با ارزشیابی به حداکثر برسانند و بدین منظور آن‌ها را کاملاً در ارزشیابی، ثبت پیشینه و برقراری ارتباط مشارکت دهند (Center for School Improvement and Policy Studies, 2007).

در این پژوهش، از سه روش برای بررسی شیوه‌های ارزشیابی معلمان و نیز میزان آمادگی آن‌ها برای ارزشیابی بهره گرفته شده است: پیمایش نگرش‌ها، باورها و شیوه‌ها؛ آزمون دانش ارزشیابی محور معلمان و مرور شیوه‌های واقعی ارزشیابی معلمان (Brookhart, 2001). پژوهش‌ها نشان می‌دهد که مهارت‌های ارزشیابی معلمان، صرف‌نظر از روش به کار گرفته شده، عموماً ضعیف است (Campbell et al., 2002). اگرچه بسیاری از منابع شیوه‌های تقویت سواد ارزشیابی را شناسایی کرده‌اند، تحقیقات بسیاری نقاط ضعف معلمان در این حوزه را نشان داده‌اند. شاید گویاترین شواهد نبود سواد ارزشیابی در میان معلمان مطالعاتی است که به ادراک آن‌ها از ارزشیابی کلاسی پرداختند. Brown (2004) پیمایشی شامل ۵۰ سؤال در میان ۵۲۵ معلم و مدیر مدرسه نیوزیلندی انجام داد. چهار عامل پیمایش عبارت بودند از: تقویت تدریس و یادگیری، مسئولیت‌پذیری مدرسه، مسئولیت‌پذیری زبان‌آموزان و نامرتب‌دانستن ارزشیابی. شرکت‌کنندگان با مفاهیم تقویت و مسئولیت‌پذیری مدرسه موافقت کردند، درحالی‌که با عامل نامرتب‌دانستن ارزشیابی مخالفت ورزیدند؛ اما این

شرکت کنندگان با این نگرش که ارزشیابی به نفع مسئولیت پذیری یادگیرندگان است، مخالفت ورزیدند. جالب این است که هیچ تفاوت آماری معناداری، صرف نظر از اطلاعات جمعیت‌شناسی معلم همچون سن، جنسیت، نقش یا اطلاعات مدرسه‌ای مانند مکان، اندازه، یا متغیرهای اجتماعی اقتصادی، در نمرات میانگین هر کدام از مفاهیم چهارگانه یافت نشد. در مطالعه دیگری که در مورد ادراک معلمان از ارزشیابی انجام شد، عقاید آن‌ها در مورد ارزشیابی آشکار گردید. یافته‌های این مطالعات می‌تواند جدل‌ها و کشمکش‌های مربوط به تغییر شیوه‌های ارزشیابی را تبیین کند. Hargreaves (2005) نظر ۸۳ معلم را در مورد تعریف «ارزشیابی» و «یادگیری» پرسید. با شیوه کدگذاری، دو مقوله‌ای که از تعاریف مختلف ارزشیابی برآمد، عبارت بود از: «ارزشیابی» به عنوان سنجش، یعنی ارزشیابی به مثابه قضاوت نهایی کار یادگیرندگان و ارزشیابی به معنای پرس و جو، یعنی کمک به یادگیرندگان برای کشف دانسته‌هایشان از محتوا. اولین مقوله یا ارزشیابی به معنای سنجش، تعریف متداول معلمان شرکت کننده در این مطالعه بود. به همین ترتیب، دو مفهوم مربوط به یادگیری، عبارت بود از یادگیری به عنوان تحقق اهداف و یادگیری به عنوان تولید دانش. مفهوم نخست متداول‌تر بود که دوباره بیانگر عقاید معلمان در این مورد است که باید ارزشیابی را به عنوان محصول نهایی استفاده نمود.

همچنین اخیراً، به منظور تعیین و درک هر چه بهتر مؤلفه‌های سواد ارزشیابی، پژوهشی توسط Kremmel and Harding (2020) صورت گرفته است که در آن با استفاده از پیمایشی مقیاس بالا و برخط از گروه‌های متفاوت از جمله معلمان، طراحان آزمون و محققان حوزه آزمون‌سازی و از طریق تحلیل عاملی اکتشافی نه‌سازه مجزای سواد ارزشیابی را مشخص کرده‌اند که با مدل پیشنهادی و سازه‌های Taylor (2013) بسیار منطبق است. این مؤلفه‌ها شامل طراحی و اجرای ارزشیابی‌های زبان، ارزشیابی در آموزش زبان، روش‌های محلی و سیاستی ارزشیابی، نگرش‌ها و اعتقادات شخصی، روش‌های تحقیق و آماری، تعبیرات و اصول ارزشیابی، طراحی، کاربرد و ساختار زبان، آمادگی و تأثیر آزمون<sup>۱</sup> و نمره‌دهی و امتیازدهی است. با استفاده از این ساختار می‌توان به نیازسنجی در خصوص سواد ارزشیابی زبان گروه‌های مختلف پرداخت.

در مطالعه دیگری، Lam (2019)، دانش، ادراک و روش‌های ارزشیابی کلاس‌محور نوشتار را در خصوص ۶۶ معلم مدرسه متوسطه در هنگ‌کنگ، بررسی کرد و دریافت که



اکثر معلمان از دانش ارزشیابی مرتبط و نظرات مثبت در مورد ارزشیابی‌های جایگزین نوشتار برخوردار بودند. برخی از معلمان درک ناقصی از ارزشیابی یادگیری و ارزشیابی به‌منظور یادگیری داشتند، اما از ارزشیابی به‌عنوان یادگیری، به دلیل استفاده از روش‌های ارزشیابی بدون درونی کردن آن‌ها، درک درستی نداشتند. همین‌طور Koh و همکاران (2018) موضوع طراحی تکلیف در سواد ارزشیابی را در ۱۲ معلم زبان چینی بررسی کردند و دریافتند اگرچه معلمان سریعاً ابعاد متفاوت طراحی تکلیف را متوجه شدند، به کارگیری معیار استفاده از دانش خاص در ارزشیابی را بسیار مشکل و بغرنج دیدند.

مطالعه‌ای که Razavipour و همکاران (2011) در ایران در مورد تعامل بین تأثیر آزمون (واش‌بک) و سواد ارزشیابی معلمان رشته آموزش زبان انگلیسی انجام دادند نیز نشان داد که سطح سواد ارزشیابی شرکت‌کنندگان پایین‌تر از حد انتظار بود. علاوه بر این، نتایج نشان می‌دهد که معلمان پایه دانش ضعیفی در ارزشیابی دارند؛ حال مهم نیست که سواد ارزشیابی‌شان چقدر است؛ آن‌ها تدریس و آزمون‌سازی زبان انگلیسی خود را با الزامات آزمون‌های خارجی تطبیق می‌دهند؛ اما ظاهراً اکثر معلمان رشته آموزش زبان انگلیسی به‌عنوان زبان خارجی با سواد ارزشیابی بیشتر احتمالاً از شیوه‌های فاقد تأثیر آزمون (غیر واش‌بک) در آموزش انگلیسی خود استفاده می‌کنند. همین‌طور، نتایج مطالعه مهمان‌دوست و همکاران (۱۳۹۷) تفاوت معناداری در سواد ارزشیابی اساتید دانشگاه بر اساس سوابق تدریس و پیشینه تحصیلی آنان نشان داد. این مطالعه، تجربه تدریس و داشتن آموزش‌های مرتبط در حوزه سنجش و ارزشیابی را عواملی تأثیرگذار بر شمرد و بر ارائه دوره‌های آموزشی منظم و هدفمند به‌منظور توانمندسازی معلمان تأکید کرد. در پژوهشی تازه‌تر نیز، Salimi and Farsi (2020) با مقایسه سواد ارزشیابی مدرسان بومی و غیربومی، به رابطه سواد ارزشیابی و تجربه تدریس در معلمان بومی پی بردند. این مطالعه نتیجه گرفت دو مؤلفه مشاهدات کلاسی و ثبات در سنجش و ارزشیابی از جمله عوامل مهم در سواد ارزشیابی مدرسان هستند. Hoyt (2005) بر اساس پژوهش خود در مورد معلمان زبان خارجی در ایالت ایندیانا استدلال کرد مدرسان زبان خارجی برای بهبود سواد ارزشیابی خود نیازمند جعبه‌ابزاری مملو از مهارت‌ها و راهکارهایی هستند که به آن‌ها یاری کند تا نتایج ارزشیابی را رمزگشایی کنند، معانی‌شان را تحلیل کنند، الزامات آنان را تأمین کنند و آن‌ها را در برنامه درسی و سنجشی خود به کارگیرند. درحالی‌که پیشینه پژوهشی فراوانی در مورد شیوه‌های ارزشیابی

معلمان آموزش زبان انگلیسی به عنوان زبان خارجی و زبان دوم (Cheng et al., 2008) وجود دارد، در حوزه سواد ارزشیابی و دانش معلمان خلأ وجود دارد. صرف گردآوری داده‌های ارزشیابی و ثبت و ضبط آن‌ها در گزارش‌های انفرادی زبان‌آموزان کافی نیست. با در نظر گرفتن این مشکلات و در راستای اهداف این پژوهش، سؤالات ذیل مطرح شدند.

۱. میزان سواد ارزشیابی معلمان مؤسسات انگلیسی به عنوان زبان خارجی دارای مدرک کارشناسی چقدر است؟

۲. میزان سواد ارزشیابی معلمان مؤسسات انگلیسی به عنوان زبان خارجی در مقطع تحصیلات تکمیلی و با مدرک کارشناسی ارشد و دکتری چقدر است؟

۳. تا چه میزان سواد ارزشیابی دو گروه معلمان مؤسسات انگلیسی به عنوان زبان خارجی دارای مدرک کارشناسی و تحصیلات تکمیلی تفاوت دارند؟

به منظور بررسی معناداری تفاوت گروه‌ها، در این پژوهش فرضیه جایگزین زیر برای سؤال سوم مطرح شد.

$H_1$  = سواد ارزشیابی دو گروه از معلمان مؤسسات انگلیسی به عنوان زبان خارجی دارای مدرک کارشناسی و تحصیلات تکمیلی تفاوت دارند.

## روش

در این مطالعه، ۳۲ معلم زبان انگلیسی به عنوان زبان خارجی شرکت نمودند. در میان آن‌ها، ۲۲ نفر دارای مدرک کارشناسی در رشته زبان انگلیسی و گرایش‌های مرتبط (آموزش زبان انگلیسی، ادبیات انگلیسی، مطالعات ترجمه و زبانشناسی) بودند و ۱۰ نفر دارای مدرک تحصیلات تکمیلی (مقطع کارشناسی ارشد و دکتری) در همین رشته‌ها بودند. پژوهشگر برای انتخاب شرکت‌کنندگان از روش نمونه‌گیری هدفمند برای گزینش شرکت‌کنندگان استفاده نمود. بر اساس نظر Patton (1990)، «هدف نمونه‌گیری هدفمند انتخاب موردهای با اطلاعات غنی است که مطالعه آن‌ها سؤالات پژوهش را روشن می‌سازد» (ص. ۱۶۹). پژوهشگر برای انتخاب شرکت‌کنندگان معیارهای زیر را لحاظ نمود: (۱) مدرسانی که در حال حاضر در یکی از آموزشگاه‌های زبان انگلیسی و در سطوح بالا (فوق میانی و پیشرفته) مشغول تدریس زبان انگلیسی باشند و (۲) مدرسانی که برای همکاری و مشارکت در این مطالعه مشتاق هستند. معلمان شرکت‌کننده در این پژوهش هم خانم و آقا و با تجربه تدریس

بین ۳ تا ۱۷ سال بودند. از میان شرکت‌کنندگان، ۶۹ درصد دارای مدرک کارشناسی (تعداد=۲۲) و ۳۱ درصد مدرک کارشناسی ارشد و دکتری (تعداد=۱۰) داشتند. پژوهش حاضر از یک پرسشنامه دویبخشی بهره گرفت. بخش اول (شامل ۷ گویه و برگرفته از Impara et al., 1993)، شامل پرسش‌هایی درباره پیشینه تحصیلی و تجارب ارزشیابی معلمان و علایق آن‌ها نسبت به یادگیری اصول ارزشیابی و نگرش‌هایشان نسبت به آزمون‌سازی بود (پیوست الف). در بخش دوم، از پرسشنامه سواد ارزشیابی کلاسی<sup>۱</sup> یا به اختصار CALI استفاده شد. این ابزار از پرسشنامه سواد ارزشیابی معلمان (Plake, 1993؛ Plake et al., 1993) برگرفته شد. این فهرست مبتنی بر معیارهای توانش معلم در ارزشیابی آموزشی زبان آموزان است (AFT، NCME و NEA، 1990). فهرست واری CALI شامل ۳۵ گویه محتوایی بود و برای هر معیار ۵ گویه طرح شده است (پیوست ب و ج). این پرسشنامه دارای پایایی قابل قبولی (ضریب آلفا=۰.۷۴) (Campbell et al., 2002) بود. افزون بر این، در این مطالعه، پرسشنامه سواد ارزشیابی معلمان مورد روایی سنجی محتوایی قرار گرفت. بدین منظور، ۳ متخصص زبانشناسی کاربردی بخش‌های مختلف این پرسشنامه را از نظر زبان و محتوا بازبینی و نظراتشان در این خصوص را بیان کردند. ضریب محتوایی ۰.۹۷ برای این پرسشنامه به دست آمد.

با استفاده از روش تحقیق پس‌رویدادی (Ary et al., 2010)، پژوهشگر در ابتدا ابزار پژوهش را با انجام یک مطالعه آزمایشی (پایلوت) بررسی کرد. این مرحله با هدف ارزیابی مشکلات انجام پژوهش و بررسی روایی و پایایی پرسشنامه انجام شد. در مرحله بعد، پس از ارسال درخواست مشارکت در این پژوهش و دریافت تأییدیه و نامه موافقت معلمان، فهرستی از کسانی که مشتاق به شرکت در مطالعه بودند، تهیه کرد. پس از شناسایی شرکت‌کنندگان، با آن‌ها از طریق شبکه‌های اجتماعی و پست الکترونیکی ارتباط برقرار نمود. شرکت‌کنندگان پیشاپیش از اهداف مطالعه و حقوق خود مطلع شدند و به آن‌ها اطمینان داده شد که پاسخ‌هایشان محرمانه خواهد ماند. همچنین، از معلمان شرکت‌کننده خواسته شد بخش اول پرسشنامه را که حاوی سؤالاتی در خصوص سطح تحصیلی آن‌ها و تجارب آموزشی و ارزشیابی‌شان بود، با دقت تکمیل نمایند. پس از کسب رضایت کتبی، معلمان پرسشنامه سواد ارزشیابی را تکمیل و به پست الکترونیکی پژوهشگر ارسال کردند.

تحلیل‌های توصیفی در سطح گویه‌های جداگانه و بر اساس فهرست معیارهای هفت‌گانه فهرست سواد ارزشیابی انجام شد. در تحلیل استنباطی، با استفاده از تحلیل واریانس چندمتغیره، نمرات میانگین گروه معلمان کارشناسی و تحصیلات تکمیلی برای هر یک از هفت معیار سواد ارزشیابی مقایسه شدند. همه تحلیل‌ها با SPSS 21 انجام شدند.

### یافته‌ها

در این بخش، نتایج مربوط به هر سؤال به صورت جداگانه ارائه می‌گردند. هدف از ارائه نتایج در این بخش ترکیب تمامی اطلاعات است که به واسطه آن بتوان یافته‌های پژوهش را بهتر مورد بحث قرار داد. یافته‌های سؤال اول سطح نسبتاً خوبی از پایایی و انسجام داخلی را برای گروه کارشناسی نشان داد (ضریب آلفا = ۰٫۷۶ و تعداد کل = ۲۲). گروه مدرسان کارشناسی تقریباً به کمتر از ۲۰ مورد از میان ۳۵ مورد پاسخ صحیح دادند. از میان هفت حوزه توانش مشخص شده در فهرست معیارها، بالاترین نمره اکتسابی گروه مدرسان کارشناسی مربوط به معیار اول یا انتخاب روش‌های مناسب ارزشیابی بود (میانگین ۳٫۲۳ و بالاترین نمره = ۵). حال آنکه کمترین نمره این گروه مربوط به معیار پنجم یا طراحی رویه‌های نمره‌دهی معتبر بود (میانگین ۲٫۰۵). نتایج نمرات گروه معلمان کارشناسی در هر کدام از هفت معیار در جدول ۱ نشان داده شده است. همچنین بررسی نسبت چولگی نشان می‌دهد که داده‌های مربوط به تمامی معیارها دارای توزیع نرمال است، زیرا تمامی نسبت‌های چولگی در بازه مجاز  $\pm ۱٫۹۶$  برای توزیع نرمال قرار داشتند.

جدول ۱. میانگین و انحراف معیار گروه معلمان کارشناسی در فهرست سواد ارزشیابی کلاسی

معیار	کمترین	بیشترین	میانگین	انحراف معیار	نسبت چولگی
معیار اول انتخاب روش‌های ارزشیابی مناسب	۲	۵	۳٫۲۳	۰٫۶۹	۱٫۳۳
معیار دوم طراحی روش‌های ارزشیابی مناسب	۲	۴	۲٫۷۷	۰٫۶۹	۰٫۶۶
معیار سوم اجرا، نمره دهی و تفسیر نتایج ارزشیابی	۲	۴	۳٫۱۸	۰٫۵۹	-۰٫۰۵
معیار چهارم استفاده از نتایج ارزشیابی برای تصمیم‌گیری	۱	۴	۲٫۶۸	۰٫۶۵	-۱٫۵۵

معیار	کمترین	بیشترین	میانگین	انحراف معیار	نسبت چولگی
معیار پنجم طراحی رویه‌های نمره‌دهی معتبر	۱	۳	۲,۰۵	۰,۵۸	۰,۰۳
معیار ششم بیان نتایج ارزشیابی	۲	۳	۲,۲۷	۰,۴۶	۱,۸۳
معیار هفتم تشخیص روش‌های غیراخلاقی یا غیرقانونی	۲	۴	۲,۶۸	۰,۵۷	۰,۱۰
نمره کل	۱۳	۲۶	۱۸,۸۶	۳,۳۳	۰,۵۱

در این گروه، بالاترین نمره اکتسابی برای حوزه توانشی، معیار سوم یا همان اجرا، نمره‌دهی و تفسیر نتایج ارزشیابی (میانگین ۳,۹۰ و بالاترین نمره ۵) و کمترین نمره به دست آمده برای معیار پنجم یا طراحی رویه‌های نمره‌دهی معتبر بود (میانگین ۲,۱۰). نتایج هر کدام از این هفت معیار در جدول ۲ نشان داده شده است. همچنین بررسی نسبت چولگی نشان می‌دهد داده‌های مربوط به تمامی معیارها دارای توزیع نرمال هستند، زیرا تمامی نسبت چولگی در بازه مجاز  $\pm ۱,۹۶$  برای توزیع نرمال قرار داشتند.

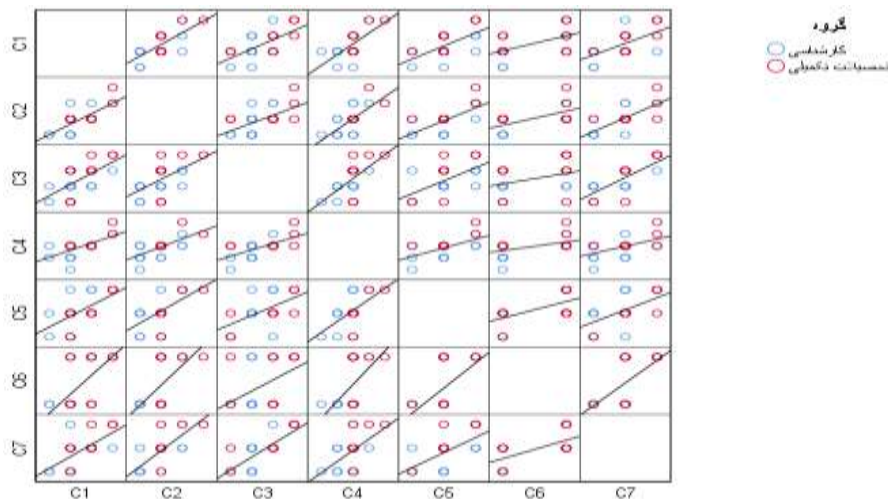
جدول ۲. میانگین و انحراف معیار گروه معلمان تحصیلات تکمیلی برحسب معیارهای سواد ارزشیابی کلاسی

معیار	کمترین	بیشترین	میانگین	انحراف معیار	نسبت چولگی
معیار اول انتخاب روش‌های ارزشیابی مناسب	۳	۵	۳,۸۰	۰,۷۹	۰,۵۹
معیار دوم طراحی روش‌های ارزشیابی مناسب	۳	۵	۳,۳۰	۰,۶۸	۱,۸۱
معیار سوم اجرا، نمره دهی و تفسیر نتایج ارزشیابی	۲	۵	۳,۹۰	۱,۱۰	-۱,۴۷
معیار چهارم استفاده از نتایج ارزشیابی برای تصمیم‌گیری	۳	۵	۳,۳۰	۰,۶۸	۱,۸۱
معیار پنجم طراحی رویه‌های نمره‌دهی معتبر	۱	۳	۲,۱۰	۰,۵۸	۰,۱۳
معیار ششم بیان نتایج ارزشیابی	۲	۳	۲,۶۰	۰,۵۲	-۰,۷۱

معیار	کمترین	بیشترین	میانگین	انحراف معیار	نسبت چولگی
معیار هفتم	۲	۴	۳,۲۰	۰,۶۳	-۰,۱۹
تشخیص روش‌های غیراخلاقی یا غیرقانونی	۱۶	۲۹	۲۲,۲۰	۴,۱۴	۰,۹۹
نمره کل					

برای پاسخ به سؤال سوم، از تحلیل واریانس چندمتغیره استفاده شد. پیش از انجام این آزمون، پیش‌فرض‌های آن مورد بررسی قرار گرفت. اولین پیش‌فرض نرمالیته بودن نمرات بود که پیش‌تر در جدول‌های ۱ و ۲ بررسی شده بود و نتایج حاکی از نرمالیته بودن تمام نمرات بود. پیش‌فرض بعدی عدم وجود داده‌های پرت<sup>۱</sup> بود که برای بررسی آن مقادیر ماهالانوبیس<sup>۲</sup> از طریق آزمون رگرسیون خطی به دست آمد. بنا به گفته Pallant (2013)، مقدار بیشینه بحرانی عدد ماهالانوبیس برای ۷ متغیر عدد ۲۴/۳۲ است و در صورتی که مقادیر مشاهده شده ماهالانوبیس از این میزان بیشتر باشد، نشان‌دهنده وجود داده پرت است. بررسی داده‌ها نشان داد مقادیر ماهالانوبیس در داده‌های تحقیق بین ۲/۸۵ و ۱۲/۵۳ است؛ بنابراین این پیش‌فرض نیز رعایت شده بود. پیش‌فرض بعدی خطی بودن روابط است که برای بررسی آن شمایل ماتریس طرح پراکنندگی داده‌ها مورد بررسی قرار گرفت (شکل ۱).

شکل ۱. ماتریس طرح پراکنندگی داده‌ها در دو گروه تحقیق



1. outlier
2. Mahalanobis

همان‌طور که از تصویر ۱ مشخص است، بین هیچ جفتی از متغیرها رابطه غیرخطی وجود نداشت؛ بنابراین این پیش‌فرض نیز تأیید شد. پیش‌فرض بعدی متجانس بودن ماتریس واریانس و کوواریانس است که از طریق آزمون باکس<sup>۱</sup> بررسی شد. نتایج آزمون ( $F = 0/830 > 0/05$ )، نشان‌دهنده عدم تفاوت معنادار بود. بنابراین، این پیش‌فرض هم رعایت شده بود. در نهایت، آخرین پیش‌فرض متجانس بودن واریانس‌ها بود که از طریق آزمون لون<sup>۲</sup> بررسی گردید. نتیجه این آزمون نیز نشان داد که در تمام معیارها تفاوت بین واریانس دو گروه معنی‌دار نبود؛ بنابراین، این پیش‌فرض نیز رعایت شده بود. با حصول اطمینان از برقراری تمام پیش‌فرض‌های موردنیاز، آزمون تحلیل واریانس چندگانه اجرا گردید (جدول ۳).

جدول ۳. تحلیل واریانس چندگانه: آزمون‌های مابین گروهی

منبع	متغیر وابسته	جمع مجذورات نوع سوم	درجه آزادی	مربع میانگین	F	معناداری	اندازه اثر
معیار اول		۲.۲۵	۱	۲.۲۵	۴.۳۷	۰.۰۴	۰.۱۲
معیار دوم		۱.۹۱	۱	۱.۹۱	۴.۲۰	۰.۰۴	۰.۱۲
معیار سوم		۳.۵۴	۱	۳.۵۴	۵.۸۵	۰.۰۲	۰.۱۶
گروه	معیار چهارم	۲.۶۲	۱	۲.۶۲	۶.۱۲	۰.۰۱	۰.۱۷
	معیار پنجم	۰.۰۲	۱	۰.۰۲	۰.۰۶	۰.۸۰	۰.۰۰
	معیار ششم	۰.۷۳	۱	۰.۷۳	۳.۲۶	۰.۰۸	۰.۰۹
	معیار هفتم	۱.۸۴	۱	۱.۸۴	۵.۳۳	۰.۰۲	۰.۱۵

همان‌طور که از جدول ۳ مشخص است، از میان هفت معیار، دو گروه تنها در معیارهای پنجم و ششم دارای تفاوت معنادار نبودند و در سایر معیارهای تفاوت بین دو گروه معنادار بوده است. بررسی اندازه اثرها که بر مبنای مربع اتای جزئی است، نشان می‌دهد که این تفاوت‌ها دارای اندازه اثر متوسط هستند. با توجه به میانگین‌های ارائه شده در جدول‌های ۱ و ۲ می‌توان نتیجه گرفت نمراتی که گروه معلمان تحصیلات تکمیلی در این پنج معیار کسب نموده‌اند، از نظر آماری نسبت به نمرات گروه معلمان کارشناسی به‌طور معناداری بالاتر است؛ بنابراین فرضیه صفر مربوط به این سؤال تحقیق رد می‌شود.

1. Box  
2. Levene

## بحث و نتیجه گیری

مطالعه حاضر به بررسی سواد ارزشیابی معلمان دارای مدرک کارشناسی و تحصیلات تکمیلی (کارشناسی ارشد و دکتری) پرداخته و سواد ارزشیابی معلمان را از این حیث مقایسه کرده است. اهداف این مطالعه در قالب سه سؤال مطرح شدند که در سؤال اول، سواد ارزشیابی معلمان دارای مدرک کارشناسی مطرح شد؛ سؤال دوم به بررسی سواد ارزشیابی مدرسان در مقطع تحصیلات تکمیلی پرداخت و در سؤال سوم، سواد ارزشیابی معلمان کارشناسی و تحصیلات تکمیلی مقایسه شد. داده‌های مربوطه از طریق یک پرسشنامه دوبخشی در مورد سواد ارزشیابی جمع آوری شد و یافته‌های مطالعه نشان داد مدرسان گروه تحصیلات تکمیلی نسبت به گروه کارشناسی نمرات بالاتری در اکثر معیارهای ارزشیابی کسب نمودند که نمایانگر تأثیر پیشینه تحصیلی مدرسان زبان انگلیسی و سطح آگاهی آنان از معیارهای سواد ارزشیابی است.

یافته‌های تحلیل‌ها نشان داد مدرسان تحصیلات تکمیلی از سواد ارزشیابی قابل ملاحظه تری نسبت به معلمان کارشناسی برخوردار بودند. این نتایج تا بخشی همانند یافته‌های مطالعه Mertler (2005) است که نشان داد مدرسان ضمن خدمت نسبت به مدرسان پیش از خدمت نمرات بالاتری در سواد ارزشیابی کسب نمودند. همچنین، نتایج این مطالعه همسو با یافته‌های برخی مطالعات پیشین بود (Plake, 1993؛ Plake et al., 1993) که از نسخه اولیه این ابزار استفاده کردند و بر سواد ارزشیابی معلمان ضمن خدمت تمرکز کردند. بر اساس پژوهش Plake (1993)، بالاترین عملکرد میانگین در معیار سوم یعنی اجرا، نمره‌دهی و تفسیر نتایج ارزشیابی دیده شد و پایین‌ترین عملکرد در معیار ششم یا همان بیان و انتقال نتایج ارزشیابی مشاهده گردید.

البته در این مطالعه بالاترین عملکرد در معیار سوم و کمترین عملکرد در معیار پنجم یا طراحی رویه‌های نمره‌دهی معتبر دانش آموز نیز مشاهده شد. نمره‌دهی به یادگیرندگان اهمیت بالایی در حرفه معلمی دارد. مدرسانی که این معیار را عملی نمی‌کنند، قادر نیستند رویه طراحی نمره‌های معتبر و متشکل از نمرات تکالیف، پروژه‌ها، فعالیت‌های درون کلاسی، آزمونک‌ها، آزمون‌ها یا دیگر ارزشیابی‌های مورد استفاده را تدبیر و اجرا کنند و توضیح دهند. هدف عمده سواد ارزشیابی کمک به تصمیم‌گیری‌های معلمان و بهبود یادگیری و آموزش است. ارزشیابی‌های به‌خوبی برنامه‌ریزی شده انتظارات را شفاف



می‌کنند، ایجاد اعتماد می‌کنند و درک یادگیرندگان از خودشان را شکل می‌دهند. معلمانی که پیشینه خوبی در ارزشیابی دارند، به خوبی می‌توانند آن را با فرایند آموزش ترکیب نمایند تا از شیوه‌های درست تدریس استفاده کنند.

اهمیت مطالعه حاضر در نشان دادن این موضوع بود که سواد ارزشیابی معلمان از اهمیت شایانی برخوردار است. از این رو، به نظر می‌رسد دوره آموزش ارزشیابی تأثیر شگرفی بر سواد ارزشیابی معلمان داشته باشد. Brindley (2001) «سه نوع شرط را برای دوره‌های آموزش سواد ارزشیابی مدرسان پیش رو نهاد که عبارت‌اند از: تمرکز بر ارزشیابی مبتنی بر برنامه درسی، انعطاف‌پذیری در تأمین نیازهای گوناگون و اتکا به دانش کنونی مدرسان» (ص. ۱۲۹). برگزارکنندگان این دوره‌ها با شناسایی نقاط ضعف و قوت معلمان در مهارت‌ها و دانش ارزشیابی‌شان، می‌توانند شیوه آموزش را اصلاح و سواد ارزشیابی مدرسان را تقویت کنند.

طبق پژوهش‌ها، دوره‌های سنتی آماده‌سازی معلمان در ارزشیابی کلاسی به خوبی با نیازهای آن‌ها در روش‌های کلاسی همخوانی ندارند (Schafer, 1993). در روش سنتی، عمدتاً بر آزمون‌های کلان‌مقیاس (استانده) تأکید شده است (Schafer, 1993). ولیکن اکنون این رویه در حال دگرگونی است. صرف ارائه تنها یک دوره آموزش ارزشیابی و سنجش واقعاً برای معلمان کافی نیست و تمام نیازهای آن‌ها را پوشش نمی‌دهد. این مسئله وقتی مشکل‌سازتر می‌شود که می‌بینیم بسیاری از نهادهای تربیت مدرس اصلاً به چنین دوره‌هایی نیاز ندارند (Campbell et al., 2002; Schafer, 1993). اکثر نهادها و کشورهای صرفاً مواد درسی ارزشیابی را در دیگر واحدهای آموزشی معلم می‌گنجانند. در این حالت، خود یادگیرندگان از معلمان و مدرسانی فنون و اصول ارزشیابی فرامی‌گیرند که تخصص و دانش کافی در این حوزه ندارند (Quilter, 1999).

مقایسه توانش‌های هفت‌گانه تفاوت‌های معناداری بین مدرسان دارای مدرک کارشناسی و تحصیلات تکمیلی در ۵ حوزه و نمرات کلی آنان آشکار نمود. در همه مواردی که تفاوت‌های معناداری دیده شد، گروه مدرسان تحصیلات تکمیلی نمره بالاتری نسبت به مدرسان با مدرک کارشناسی دریافت نمودند. در هر دو گروه، ضعیف‌ترین عملکرد در معیار پنجم یا طراحی رویه‌های نمره‌دهی معتبر و پس از آن معیار ششم یا بیان و انتقال نتایج ارزشیابی مشاهده شد. معلمان برای برقراری ارتباط مناسب با دیگران در مسائل ارزشیابی یادگیرندگان، باید بتوانند به درستی از اصطلاحات تخصصی ارزشیابی استفاده و معنی، محدودیت‌ها و

کاربردهای نتایج آن‌ها را بیان کنند؛ بنابراین، بر اساس نتایج به دست آمده، همچنان می‌شود ادعا کرد که هیچ‌یک از دو گروه مدرسان کارشناسی و تحصیلات تکمیلی نمی‌توانند از رویه‌های ارزشیابی خود دفاع کنند و نتایج آن را به درستی تفسیر نمایند. از این روست که مطالعه حاضر، حوزه‌های توانشی خاصی همچون طراحی رویه‌های نمره‌دهی معتبر و بیان نتایج ارزشیابی را مورد تأکید قرار می‌دهد که هر دو گروه مدرسان در آن نیازمند آموزش و حمایت بیشتر هستند.

حصول چنین نتیجه‌ای شاید به این دلیل است که مدرسان مقطع تحصیلات تکمیلی در مقایسه با مدرسان با مدرک کارشناسی در دوره‌های ارزشیابی بیشتری را تجربه می‌کنند و به واسطه انجام کارهای پژوهشی متفاوت، مقالات و منابع متنوعی را مطالعه می‌کنند؛ به علاوه، این گروه از مدرسان در کنفرانس‌ها و کارگاه‌های متفاوتی شرکت می‌کنند که این تجربیات دیدگاهی انتقادی نسبت به ارزشیابی و معیارهای دخیل در ارزشیابی را برایشان به ارمغان می‌آورد. پرواضح است که مدرسان با مدرک کارشناسی ارشد و دکتری ارتباط بیشتری با ارزشیابی آموزشی دارند و بیشتر در معرض آن‌ها قرار دارند. این دست فرصت‌های یادگیری، آموزشی و پژوهشی برای این مدرسان منافی ایجاد می‌کند که مدرسان مقطع کارشناسی از آن محروم هستند.

پژوهش حاضر، دانش و مهارت‌های ارزشیابی دو گروه مدرسان کارشناسی و تحصیلات تکمیلی را بررسی و به ارتقاء و حفظ چنین دانش و مهارت‌هایی در شیوه‌های ارزشیابی کلاسی تأکید کرده است. این مطالعه به مدرسان تأکید می‌کند پیوند بین شیوه‌های آموزش و ارزشیابی خود را آشکار سازند. همان‌طور که این مطالعه و سایر مطالعات پیشین نشان دادند، یکی از حوزه‌های مورد نیاز بیان و انتقال نتایج ارزشیابی و اتخاذ تصمیمات آموزشی بر اساس آن است؛ بنابراین، در پژوهش‌های آتی باید بر تقویت دانش و آگاهی مدرسان دارای مدرک کارشناسی از شیوه صحیح انتقال نتایج ارزشیابی یادگیرندگان تأکید نمود. بدین منظور، برنامه‌های آموزشی باید زمان بیشتری صرف تفسیر نتایج ارزشیابی در سطح غیررسمی کلاس و نیز سطوح بالاتر (مثلاً آزمون‌های استاندارد) کنند. یکی دیگر از موضوعات پژوهشی باید بررسی و بازبینی معیارهایی باشد که انجمن آمریکایی معلمان، شورای ملی سنجش آموزش و انجمن ملی آموزش در سال ۱۹۹۰ ارائه دادند. با توجه به سیاست‌های آموزشی و الزامات ارزشیابی، سنگینی و ارزش باید هر کدام از این هفت معیار

را طبق خواسته‌ها و تمرکز ارزشیابی دوباره اولویت‌بندی و برای اطمینان از سواد ارزشیابی معلمان اصلاح کرد. افزون بر این، مربی معلمان و متخصصان ارزشیابی باید مسئولیت ارتقاء سواد ارزشیابی معلمان را بر عهده گیرند.

از نتایج و یافته‌های این مطالعه می‌توان برای اصلاح و طراحی دوره‌های آموزش معلمان و انگیزه‌بخشی به آن‌ها برای ارتقاء سواد ارزشیابی خود در محیط‌های مبتنی بر مسئولیت بهره گرفت. در پایان، ارائه دوره‌های آموزش ارزشیابی قوی برای معلمان می‌تواند به یادگیرندگان آینده کمک کند یادگیری آکادمیک خود را تقویت بخشند. همچنین، این دوره‌ها اعتماد معلمان به مراحل ارزشیابی کلاسی را افزایش می‌دهد. این مراحل شامل هدف‌گذاری برای یادگیری، طراحی و استفاده از ارزشیابی‌های گوناگون، مشارکت یادگیرندگان در ارزشیابی و داشتن شیوه بیان درست موفقیت زبان‌آموزان هستند. همان‌طور که Stiggins (2002) اذعان می‌کند، شیوه‌های ارزشیابی کلاسی باید به گونه‌ای اصلاح شوند که بتوان فرایندهای ارزشیابی را در فرایند آموزش تلفیق نمود و از این رهگذر یادگیری زبان‌آموزان را ارتقاء بخشید، از تصمیم‌گیری آموزشی حمایت نمود و نسبت به کارآمدی آموزشی معلمان اظهار نظر کرد. معلمان دارای سواد ارزشیابی به هر نوع ارزشیابی‌ای که اقدام کنند، «می‌دانند چه چیزی را و به چه دلیل ارزشیابی کنند، چگونه به بهترین وجه ممکن ارزشیابی نمایند، چگونه نمونه‌های صحیحی از عملکرد را تولید کنند، چه مشکلاتی ممکن است پیش بیاید و چگونه از وقوعشان پیشگیری نمایند» (Stiggins, 1995).

به‌اجمال، پیامدهای مطالعاتی همچون مطالعه فعلی، می‌توانند مسئولان اجرایی، مدرسان کارشناسی و تحصیلات تکمیلی و مجموع متخصصان آموزشی را ترغیب نمایند تا پیشرفت حرفه‌ای خود را حفظ کنند و شیوه‌های ارزشیابی خود را بهبود بخشند تا کیفیت آموزش زبان و انگیزه یادگیری را ارتقا دهند. شایان ذکر است که ارزشیابی صرفاً سنجش یا نمره‌دهی نیست. اگر مدرسان هدف ارزشیابی را یادگیری بدانند و یادگیری را ارزشیابی نکنند، زبان‌آموزان انگیزه بیشتر و ترس کمتری خواهند داشت. در پایان، مسئله حرفه‌ای بودن در عرصه ارزشیابی زبان در ایران مطرح می‌شود. این بدان معنی است که هم معلمان کنونی و هم معلمان آتی نیازمند آموزش بیشتری در ارزشیابی زبان هستند؛ اما مسئولیت آموزش معلمان زبان در طراحی، استفاده، نمره‌دهی و تفسیر ارزشیابی‌های زبان به دوش نهادهای آموزشی‌ای است که دارای برنامه‌های آموزشی برای معلمان هستند. بر تمام معلمان آتی

بایسته است که دست‌کم یک دوره ارزشیابی و آزمون‌سازی زبان را پیش از شروع تدریس بگذرانند. آن‌ها باید بکوشند از طریق آموزش ضمن خدمت، همایش، کارگاه و دیگر روش‌ها سواد خود را تقویت نمایند تا برای فرهنگ ارزشیابی زبان بسترسازی کنند.

همچنین برنامه‌های پیشرفت حرفه‌ای باید به معلمان بیاموزند از داده‌های برگرفته از ارزشیابی‌ها استفاده و شیوه‌های تدریس خود را تنظیم کنند و شواهد استواری از عملکرد زبان‌آموزان برای اهداف بازمینی برنامه درسی فراهم آورند. دوره‌های ارزشیابی به‌عنوان بخشی از پیشرفت مدرس گامی در مسیر درست هستند؛ اما برنامه درسی دوره‌های ارزشیابی باید مطابق با پژوهش‌های این عرصه باشند. افزون بر این، مدرسان باید قادر باشند شیوه‌های ارزشیابی گوناگونی همچون پاسخ انتخابی، پاسخ مکتوب، ارزشیابی عملکرد و ارتباط شخصی مستقیم را انجام دهند. چنانچه دوره آموزشی بر پایه پژوهش‌های کنونی دایر باشد، مدرسان باید طی دوره کارشناسی یا سال‌های پیش از خدمتشان در خصوص شیوه‌های ارزشیابی در کلاس آموزش ببینند.

مدرسان زبان خارجی با ارتقاء سواد ارزشیابی خود نه تنها می‌توانند شیوه‌های ارزشیابی مناسب با اهداف خاصی همچون تعیین سطح زبان‌آموزان یا سنجش برنامه را شناسایی کنند، بلکه می‌توانند داده‌های تجربی را برای تقویت شیوه آموزشی خود بدون عواقب منفی تجزیه و تحلیل کنند. در چنین فرصتی، اولویت نوظهور این است که برای مدرسان زبان خارجی پیشرفت حرفه‌ای فراهم کنند تا سواد ارزشیابی به دست آورند. به‌علاوه، در روند آموزش باید بر ارائه بازخوردهای مناسب به یادگیرندگان، ایجاد انطباق میان شیوه آموزش و نیازها و کمک به یادگیرندگان از طریق استفاده از راهبردهای خودارزشیابی و هم‌ارزشیابی تمرکز شود. این نوع آموزش اعتماد معلمان به مراحل و گام‌های ارزشیابی را افزایش می‌دهد. این گام‌ها عبارت‌اند از: هدف‌گذاری برای یادگیری، طراحی و استفاده از ارزشیابی‌های گوناگون، مشارکت یادگیرندگان در ارزشیابی و شیوه بیان صحیح موفقیت یادگیرندگان. Bracey (2000) به بهترین وجه این موضوع را بیان نموده است که ممکن است زمانی برسد که آزمون و نمراتش از رونق بیافتند، اما آن زمان حالا نیست. با توجه به اهمیت ارزشیابی در نهادهای آموزشی امروزی در جهان، «سواد ارزشیابی» نیاز اساسی تمام مدرسان زبان است و در این زمینه، پژوهش‌های بیشتری در بافت‌های آموزشی مختلف و با

شرکت تعداد بیشتری از معلمان که یکی از محدودیت‌های اصلی این پژوهش بوده است، نیاز است.

## منابع

مهماندوست، زهره، علوی، سید محمد و کیوان پناه، شیوا. (۲۰۱۸). بررسی سواد سنجش و ارزشیابی مدرسان دانشگاهی زبان انگلیسی در ایران. پژوهش‌های زبان‌شناختی در زبان‌های خارجی، ۱(۲)، ۶۴۵-۶۶۳. <https://dx.doi.org/10.22059/jflr.2018.245602.422>

## References

- American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, and NEA) (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32. <https://doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education* (8th ed.). Belmont, CA: Wadsworth.
- Berry, V., Sheehan, S., & Munro, S. (2017). Exploring teachers' language assessment literacy: A social constructivist approach to understanding effective practices. In *Proceedings of the ALTE 6th International Conference on Learning and Assessment: Making the Connections* (pp. 201-207). Bologna: Italy.
- Bracey, G. (2000, August). *Thinking about tests and testing: A short primer in assessment literacy*. Paper presented at the American Youth Policy Forum, Washington, DC.
- Braney, B. T. (2010). *An examination of fourth grade teachers' assessment literacy and its relationship to students' reading achievement* (Master's thesis). East Carolina University, USA.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393-407. <https://doi.org/10.1191/026553201682430102>
- Brookhart, S. M. (1998, April). *Teaching about grading and communicating assessment results*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Brookhart, S. M. (1999a). *The art and science of classroom assessment: The missing part of pedagogy*. Washington, D.C.: ERIC Clearinghouse on Higher Education and Office of Educational Research and Improvement.
- Brookhart, S. M. (1999b). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13. <https://doi.org/10.1111/j.1745-3992.1999.tb00002.x>
- Brookhart, S. M. (2001, March). *The standards and classroom assessment research*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education*, 11(3), 301-318. <https://doi.org/10.1080/0969594042000304609>
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Paper

- presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Center for School Improvement and Policy Studies (2007). *What is assessment literacy?* Columbus, OH: Boise State University.
- Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classroom. *Assessment and Evaluation in Higher Education*, 33(1), 9-32.  
<https://doi.org/10.1080/02602930601122555>
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159-179.  
[https://doi.org/10.1207/s15326977ea0302\\_3](https://doi.org/10.1207/s15326977ea0302_3)
- Cowan, J. (1998). *On becoming an innovative university teacher: Reflection in action*. Buckingham, Philadelphia, PA: Society for Research into Higher education & Open University Press.
- Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213-224.  
<https://doi.org/10.1080/03057640500146880>
- Hearne, J. (2001). *Assessment as a tool for learning*. *New horizons for learning*. Retrieved from [www.newhorizons.org/strategies/assess/hearne.htm](http://www.newhorizons.org/strategies/assess/hearne.htm).
- Hoyt, K. (2005, April). *Assessment: Impact on instruction*. New Visions in Action, National Assessment Summit, Meeting conducted in Alexandria, VA.
- Impara, B. S., & Fager, J. J. (1993). Teachers' assessment background and attitudes towards testing. *Theory into Practice*, 32(2), 113-117.  
<https://doi.org/10.1080/00405849309543584>
- James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning at Australian Universities*. Melbourne, Australia: Center for the Study of Higher Education.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 17-64). Westport, CT: Praeger.
- Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264-288.  
<https://doi.org/10.1177/1362168816684366>
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120. <https://doi.org/10.1080/15434303.2019.1674855>
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78-89.  
<https://doi.org/10.1016/j.system.2019.01.006>
- Linn, R. L., & Grolund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Merrill Prentice Hall.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.  
<https://doi.org/10.1111/j.1745-3992.2001.tb00055.x>
- Mertler, C. A. (2005). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(2), 76-92.
- No Child Left Behind (2001). Retrieved November 22, 2005 from <http://www.ed.gov/nclb/landing.jhtml>.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.

- Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (4th ed). Crows Nest, NSW: Allen & Unwin.
- Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Phye, G. D. (1997). Classroom assessment: A multidimensional perspective, In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, adjustment, and achievement* (pp. 33-51). San Diego: Academic Press.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12.  
<https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Popham, W. J. (2004). All about accountability: Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Quilter, S. M. (1999). Assessment literacy for teachers: Making a case for the study of test validity. *Teacher Educator*, 34(4), 235-243.  
<https://doi.org/10.1080/08878739909555204>
- Razavipour, K., Riazzi, A., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *Journal of English Language Teaching*, 4(1), 156-161.  
<https://doi.org/10.5539/elt.v4n1p156>
- Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, 37(2), 179-192.
- Salimi and Farsi, M. (2020). Probing into EFL teachers' assessment literacy and teaching experience: The case of native ESL and non-native EFL teachers. *Journal of English language Teaching and Learning*, 12(25), 269-288.
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32(2), 118-126.
- Shepard, L.A. (2001). The role of classroom assessment in teaching and learning, In V. Richardson (Ed.), *Handbook of research on teaching* (pp.1066-1101). Washington DC: AERA.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (1999). Are you assessment literate? *The High School Journal*, 6(5), 20-23.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.  
<https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765. <https://doi.org/10.1177/003172170208301010>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.  
<https://doi.org/10.1177/0265532213480338>
- Wang, T., Wang, K., & Huang, S. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51(1), 448-462. <https://doi.org/10.1016/j.compedu.2007.06.010>
- Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.

پیوست «الف»: پرسشنامه‌ی تجربیات ارزشیابی معلمان

1. What is your gender? O female O male
2. Which of the following is the **most appropriate** description of the level at which you teach?  
O elementary O upper-intermediate  
O pre-intermediate O advanced  
O intermediate O other
3. Which **best** describes the educational level you have attained? What is your major?  
O B.A. or B.S. O Ed. D.  
O M.A. or M.S. O Ph. D.  
O Specialist Majoring in: ...
4. **Including the current year**, how many years of experience do you have as a classroom teacher?  
O 1-5 years O 16-20 years O more than 30 years  
O 6-10 years O 21-25 years  
O 11-15 years O 26-30 years
5. To the best of your knowledge, did you take a **standalone** course in classroom assessment as part of your undergraduate teacher preparation? O yes O no
6. Which of the following best describes your perception of the level of preparation **for the overall job of being a classroom teacher** that resulted from your undergraduate teacher preparation program?  
O very unprepared O somewhat prepared  
O somewhat unprepared O very prepared
7. Which of the following best describes your perception of the level of preparation **for assessing student performance** that resulted from your undergraduate teacher preparation program?  
O very unprepared O somewhat prepared  
O somewhat unprepared O very prepared



پیوست «ب»: بخش دوم پرسشنامه‌ی سواد ارزشیابی کلاسی

1. What is the **most** important consideration in choosing a method for assessing student achievement?
  - The ease of scoring the assessment.
  - The ease of preparing the assessment.
  - The accuracy of assessing whether or not instructional objectives were attained.
  - The acceptance by the school administration.
2. When scores from a standardized test are said to be “reliable,” what does it imply?
  - Student scores from the test can be used for a large number of educational decisions.
  - If a student retook the same test, he or she would get a similar score on each retake.
  - The test score is a more valid measure than teacher judgments.
  - The test score accurately reflects the content of what was taught.
3. Mrs. Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be **most** valid?
  - Select a textbook with a "teacher's guide" with a test developed by the authors.
  - Develop an assessment consistent with an outline of what she has actually taught in the class.
  - Select a standardized test providing a score on problem solving skills.
  - Select an instrument measuring students' attitudes about problem solving strategies.
4. What is the **most** effective use a teacher can make of an assessment requiring students to show their work (e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion)?
  - Assigning grades for a unit of instruction on problem solving.
  - Providing instructional feedback to individual students.
  - Motivating students to attempt innovative ways to solve problems.
  - None of the above.
5. Ms. Green, the principal, was evaluating the teaching performance of Mr. Williams, the fourth-grade teacher. One of the things Ms. Green wanted to learn was if the students were being encouraged to use higher order thinking skills in the class. What documentation would be the **most** valid to help Ms. Green to make this decision?
  - Mr. Williams' lesson plans.
  - The state curriculum guides for fourth grade.
  - Copies of Mr. Williams' unit tests or assessment strategies used to assign grades.
  - Worksheets completed by Mr. Williams' students, but not used for grading.
6. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the **best** evidence for this purpose?
  - Have other teachers judge whether the assessment strategy covers what was taught.
  - Match an outline of the instructional content to the content of the actual assessment.
  - Let students in the class indicate if they thought the assessment was valid.
  - Ask parents if the assessment reflects important learning outcomes.
7. Which of the following would most likely **increase** the reliability of Mrs. Lockwood's multiple-choice end-of-unit examination in physical science?

- Use a blueprint to develop the test questions.
  - Change the test format to true-false questions.
  - Add more items like those already on the test.
  - Add an essay component.
8. Ms. Gregory wants to assess her students' skills in organizing ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?
- compare, contrast, criticize
  - identify, specify, list
  - order, match, select
  - define, recall, restate
9. Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will **best** measure his instructional goal?
- "Spoke the raven, nevermore." comes from which of Poe's works?
  - True or False: Poe was an orphan and never knew his biological parents.
  - Edgar Allen Poe wrote: 1. Novels 2. Short stories 3. Poems 4. All of the above.
  - Discuss briefly your view of Poe's contribution to American literature.
10. Several students in Ms. Atwell's class received low scores on her end-of-unit test covering multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be **best** for her to use for grouping students?
- Use the test provided in the "teacher's guide."
  - Have the students take a test with separate items for each step of the process.
  - Look at the student's records and standardized test scores to see which topics the students had not performed well on previously.
  - Give students story problems to complete and have them show their work.
11. Many teachers score classroom tests using a 100-point percent correct scale. In general, what does a student's score of 90 on such a scale mean?
- The student answered 90% of the items on this test correctly.
  - The student knows 90% of the instructional content of the unit covered by this test.
  - The student scored higher than 90% of all the students who took the test.
  - The student scored 90% higher than the average student in the class.
12. Students in Mr. Jakman's science class are required to develop a model of the solar system as part of their end-of-unit grade. Which scoring procedure below will **maximize** the objectivity of assessing these student projects?
- When the models are turned in, Mr. Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
  - Mr. Jakman asks other teachers in the building to rate each project on a 5-point scale based on their quality.
  - Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
  - Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.
13. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed two

of her students—who had been performing well in their class assignments—scored much lower than other students. Which of the following types of additional information which would be most helpful in interpreting the results of this test?

- The gender of the students.
  - The age of the students.
  - Reliability data for the standardized social studies test she used as the model.
  - Reading comprehension scores for the students.
14. Frank, a beginning fifth grader, received a G. E. (grade equivalent score) of 8.0 on the Reading Comprehension subtest of a standardized test. This score should be interpreted to mean Frank
- can read and understand 8th grade reading level material.
  - scored as well as a typical beginning 8th grader scored on this test.
  - is performing in Reading Comprehension at the 8th grade level.
  - will probably reach maximum performance in Reading Comprehension at the beginning of the 8th grade.
15. When the directions indicate each section of a standardized test is timed separately, which of the following is **acceptable** test-taking behavior?
- John finishes the vocabulary section early; he then rechecks many of his answers in the section.
  - Mary finishes the vocabulary section early; she checks her answers on the previous test section.
  - Jane finishes the vocabulary section early; she looks ahead at the next test section but does not mark her answer sheet for any of those items.
  - Bob did not finish the vocabulary section; he continues to work on the section when the testing time is up.
16. Ms. Camp is starting a new semester with a factoring unit in her Algebra I class. Before beginning the unit, she gives her students a test on the commutative, associative, and distributive properties of addition and multiplication. Which of the following is the **most** likely reason she gives this test to her students?
- The principal needs to report the results of this assessment to the state testing director.
  - Ms. Camp wants to give the students practice in taking tests early in the semester.
  - Ms. Camp wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
  - Ms. Camp wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.
17. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed for third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third- grade norm group. Why is this an **incorrect** application of standardized test norms?
- The norms are not reliable for first graders.
  - The norms are not valid for first graders.
  - Third grade mathematics items are too difficult for first graders.
  - The time limits are too short for first graders.
18. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have **more** potential to be helpful?
- norm-referenced information*: describes each student's performance relative to other students in a group (e.g., percentile ranks, stanines), or

*criterion-referenced information*: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective)

- Norm-referenced information.
  - Criterion-referenced information.
  - Both types of information are equally useful in helping to plan for instruction.
  - Neither, test information is not useful in helping to plan instruction.
19. Students' scores on standardized tests are sometimes inconsistent with their performances on classroom assessments (e.g., teacher tests or other in-class activities). Which of the following is **not** a reasonable explanation for such discrepancies?
- Some students freeze up on standardized tests, but they do fine on classroom assessments.
  - Students often take standardized tests less seriously than they take classroom assessments.
  - Standardized tests measure only the recall of information while classroom assessments measure more complex thinking.
  - Standardized tests may have less curriculum validity than classroom assessment.
20. Elementary school teachers in the Baker School system collectively designed and developed new curricula in Reading, Mathematics, and Science based on locally developed objectives and objectives in state curriculum guides. The new curricula were not matched directly to the content of the fourth-grade standardized test. A newspaper reports the fourth-grade students in Baker Public Schools are among the lowest scoring districts in the State Assessment Program. Which of the following would **invalidate** the comparison between Baker Public Schools and other schools in the state?
- The curriculum objectives of the other districts may more closely match those of the State Assessment.
  - Other school systems did not design their curriculum to be consistent with the State Assessment test.
  - Instruction in Baker schools is poor.
  - Other school systems have different promotion policies than Baker.
21. Which of the following choices typically provides the **most** reliable student-performance information a teacher might consider when assigning a unit grade?
- Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
  - Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
  - Oral responses to questions asked in class of each student over the course of the unit.
  - Daily grades designed to indicate the quality of in-class participation during regular instruction.
22. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be **closest** to equal?
- Number of items.
  - Number of students taking each test.
  - Average scores.
  - Variation (range) of scores.

23. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should
- explain the grades are assigned fairly, based on the student's performance and other related factors.
  - ask the parents what they think should be the basis for the child's grade.
  - explain exactly how the grade was determined and show the parent samples of the student's work.
  - indicate the grading scale is imposed by the school board and the teachers have no control over grades.
24. Which of the following grading practices results in a grade **least** reflecting students' achievement?
- Mr. Jones requires students to turn in homework; however, he only grades the odd numbered items.
  - Mrs. Brown uses weekly quizzes and three major examinations to assign final grades in her class.
  - Ms. Smith permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
  - Miss Engle deducts 5 points from a student's test grade for disruptive behavior.
25. During the most recent grading period, Ms. Johnson graded no homework and gave only one end-of-unit test. Grades were assigned only on the basis of the test. Which of the following is the **major** criticism regarding how she assigned the grades?
- The grades probably reflect a bias against minority students existing in most tests.
  - Decisions like grade assignment should be based on more than one piece of information.
  - The test was too narrow in curriculum focus.
  - There is no significant criticism of this method providing the test covered the unit's content.
26. In a routine conference with Mary's parents, Mrs. Estes observed Mary's scores on the state assessment program's quantitative reasoning tests indicate Mary is performing better in mathematics concepts than in mathematics computation. This probably means
- Mary's score on the computation test was below average.
  - Mary is an excellent student in mathematics concepts.
  - the percentile bands for the mathematics concepts and computation tests do not overlap.
  - the mathematics concepts test is a more valid measure of Mary's quantitative reasoning ability.
27. Many states are revising their school accountability programs to help explain differences in test scores across school systems. Which of the following is **not** something to be considered in such a program?
- The number of students in each school system.
  - The average socio-economic status of the school systems.
  - The race/ethnic distribution of students in each school system.
  - The drop-out rate in each school systems.
28. The following standardized test data are reported for John. **Subject -- Stanine Score**  
*Vocabulary – 7*  
*Mathematics Computation – 7*

*Social Studies -- 7*

Which of the following is a **valid** interpretation of this score report?

- John answered correctly the same number of items on each of the three tests.
  - John's test scores are equivalent to a typical seventh grader's test performance.
  - John had the same percentile rank on the three tests.
  - John scored above average on each of the three tests.
29. Mr. Klein bases his students' grades mostly on graded homework and tests. Mr. Kaplan bases his students' grades mostly on his observation of the students during class. A major difference in these two assessment strategies for assigning grades can **best** be summarized as a difference in
- formal and informal assessment.
  - performance and applied assessment.
  - customized and tailored assessment.
  - formative and summative assessment.
30. John scored at the 60th percentile on a mathematics concepts test and scored at the 57th percentile on a test of reading comprehension. If the percentile bands for each test are five percentile ranks wide, what should John's teacher do in light of these test results?
- Ignore this difference.
  - Provide John with individual help in reading.
  - Motivate John to read more extensively outside of school.
  - Provide enrichment experiences for John in mathematics, his better performance area.
31. In some states testing companies are required to release items from prior versions of a test to anyone who requests them. Such requirements are known as
- open-testing mandates.
  - gag rules.
  - freedom-of-information acts.
  - truth-in-testing laws.
32. Mrs. Brown wants to let her students know how they did on their test as quickly as possible. She tells her students their scored tests will be on a chair outside of her room immediately after school. The students may come by and pick out their graded test from among the other tests for their class. What is wrong with Mrs. Brown's action?
- The students can see the other students' graded tests, making it a violation of the students' right of privacy.
  - The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
  - Mrs. Brown will have to rush to get the tests graded by the end of the school day; hence, the action prevents her from using the test to identify students who need special help.
  - The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat.
33. A state uses its statewide testing program as a basis for distributing resources to school systems. To establish an equitable distribution plan, the criterion set by the State Board of Education provides additional resources to every school system with student achievement test scores above the state average. Which cliché **best** describes the likely outcome of this regulation?
- Every cloud has its silver lining.
  - Into each life some rain must fall.

- The rich get richer and the poor get poorer.
  - A bird in the hand is worth two in the bush.
34. In a school where teacher evaluations are based in part on their students' scores on a standardized test, several teachers noted one of their students did not reach some vocabulary items on a standardized test. Which teacher's action is considered **ethical**?
- Mr. Jackson darkened circles on the answer sheet at random. He assumed Fred, who was not a good student, would just guess at the answers, so this would be a fair way to obtain Fred's score on the test.
  - Mr. Hoover filled in the answer sheet the way he thought Joan, who was not feeling well, would have answered based on Joan's typical in-class performance.
  - Mr. Stover turned in the answer sheet as it was, even though he thought George, an average student, might have gotten a higher score had he finished the test.
  - Mr. Lund read each question and darkened in the bubbles on the answer sheet representing what he believed Felicia, a slightly below average student, would select as the correct answers.
35. Mrs. Overton was concerned her students would not do well on the State Assessment Program to be administered in the Spring. She got a copy of the standardized test form to be used. She did each of the following activities to help increase scores. Which activity was **unethical**?
- Instructed students in strategies on taking multiple choice tests, including how to use answer sheets.
  - Gave students the items from an alternate form of the test.
  - Planned instruction to focus on the concepts covered in the test.
  - None of these actions are unethical.

پیوست «ج»: تطابق معیارها با گویه‌های پرسشنامه‌ی سواد ارزشیابی کلاسی

جدول ۱. گویه‌های مرتبط با معیارها در فهرست سواد ارزشیابی کلاسی

شماره گویه‌های سواد ارزشیابی کلاسی	معیارهای توانش معلم
۱، ۸، ۱۵، ۲۲، ۲۹	معیار اول: انتخاب روش‌های ارزشیابی مناسب
۲، ۹، ۱۶، ۲۳، ۳۰	معیار دوم: طراحی روش‌های ارزشیابی مناسب
۳، ۱۰، ۱۷، ۲۴، ۳۱	معیار سوم: اجراء، نمره دهی و تفسیر نتایج ارزشیابی
۴، ۱۱، ۳۲، ۲۵، ۱۸	معیار چهارم: استفاده از نتایج ارزشیابی برای تصمیم‌گیری
۵، ۱۲، ۱۹، ۲۶، ۳۳	معیار پنجم: طراحی رویه‌های نمره‌دهی معتبر
۶، ۱۳، ۲۰، ۲۷، ۳۴	معیار ششم: بیان نتایج ارزشیابی
۷، ۱۴، ۲۱، ۲۸، ۳۵	معیار هفتم: تشخیص روش‌های غیراخلاقی یا غیرقانونی

**استناد به این مقاله:** استاجی، معصومه. (۱۴۰۰). بررسی سواد ارزشیابی معلمان در کلاسهای آموزش زبان انگلیسی

به عنوان یک زبان خارجی: تأثیر پیشینه تحصیلی، فصلنامه اندازه‌گیری تربیتی، ۱۲(۴۵)، ۷-۳۸.

doi: 10.22054/jem.2022.59105.2153



Educational Measurement is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.