



Psychometric Properties of the “Authentic Assessment” Questionnaire in the Teaching-Learning Process: A Case Study of Faculty Members at the Kurdistan University

Naser Shirbagi¹, Jamal Salimi², Abdullah Azizi³

1. Department of Educational Sciences, Faculty of Humanities and Social Sciences, University of Kurdistan, Sanandaj, Iran; (Corresponding Author), Email: nshirbagi@uok.ac.ir

2. Department of Educational sciences, Faculty of Humanities and Social Sciences, University of Kurdistan, Sanandaj, Iran. Email: j.salimi@uok.ac.ir

3. Department of Educational Sciences, Faculty of Humanities and Social Sciences, University of Kurdistan, Sanandaj, Iran. Email: abduallahazizi829@gmail.com

Article Info	ABSTRACT
Article Type: Research Article	Objective: The purpose of this study was to investigate the extent to which faculty members utilize authentic assessment components in the teaching-learning process. Methods: This research followed a quantitative approach using a survey method. The statistical population included all faculty members of Kurdistan university, 194 of whom were randomly selected using the Cochran formula. Data were collected using a researcher-made questionnaire based on Wiggins' eight-dimensional framework. The questionnaire's face and content validity was confirmed by experts, and its reliability was assessed through Cronbach's alpha and confirmatory factor analysis.
Received: 2024.12.29	Results: The model results showed that the original measurement questionnaire with eight main factors has high validity. Evaluation of the measurement model with indicators such as Cronbach's alpha, composite reliability, and convergent and divergent validity indicated the suitability and appropriateness of the framework. Analysis of the structural model based on Z-scores also showed that this framework has a strong and stable structure. In addition, a significant difference was observed between the level of use of authentic assessment and variables such as faculty and academic rank, such that experienced professors focused more on self-assessment and reflective assignments, and assistant professors focused more on modeling. Also, faculty members from the faculties of engineering and basic sciences used authentic assessment the most, while faculty members from the Faculty of Language and Literature used it the least.
Received in revised form: 2025.05.05	Conclusion: By validating assessment tools and identifying evaluation patterns, this research has contributed to the enrichment of scientific literature in the field of educational evaluation and can be used to improve empowerment programs and educational policies.
Accepted: 2025.06.02	Keywords: Authentic assessment, psychometric evaluation of the questionnaire, faculty members, thoughtful tasks
Published online: 2025.06.25	

Cite this article: Shirbagi, Naser; Salimi, Jamal; Azizi, Abdullah (2025). Psychometric Properties of the “Authentic Assessment” Questionnaire in the Teaching-Learning Process: A Case Study of Faculty Members at the Kurdistan University. *Educational Measurement and Evaluation Studies*, 15(50), 121-172.

DOI: 10.22034/emes.2025.2059237.2652



© The Author(s).

Publisher: National Organization of Educational Testing (NOET)



شاخصه‌های روان‌سنجی پرسشنامه «سنجش اصیل» در فرایند یاددهی-یادگیری مطالعه موردی اعضای هیئت علمی دانشگاه کردستان

ناصر شبیربگی^۱، جمال سلیمی^۲، عبدالله عزیزی^۳

۱. گروه علوم تربیتی، دانشکده علوم انسانی و اجتماعی، دانشگاه کردستان، سنندج، ایران؛ (نویسنده مسئول)، رایانامه: uok.ac.ir@nshirbagi
۲. گروه علوم تربیتی، دانشکده علوم انسانی و اجتماعی، دانشگاه کردستان، سنندج، ایران، رایانامه: j.salimi@uok.ac.ir
۳. گروه علوم تربیتی، دانشکده علوم انسانی و اجتماعی، دانشگاه کردستان، سنندج، ایران، رایانامه: abduallahazizi829@gmail.com

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	هدف: هدف از اجرای این پژوهش، بررسی میزان بهره‌گیری اعضای هیئت علمی از شاخصه‌های سنجش اصیل در فرایند یاددهی یادگیری بود.
دریافت: ۱۴۰۳/۱۰/۰۹	روش پژوهش: این مطالعه با رویکرد کمی و راهبرد پیمایشی انجام گرفت. جامعه آماری شامل تمامی اعضای هیئت علمی دانشگاه کردستان بود که با استفاده از فرمول کوکران، ۱۹۴ نفر به صورت تصادفی طبقه‌ای انتخاب شدند. ابزار گردآوری داده‌ها، پرسشنامه‌ای محقق ساخته بر پایه چارچوب هشت‌گانه ویگینز بود. روایی صوری و محتوایی پرسشنامه با نظر متخصصان تأیید شد و پایایی آن با آلفای کرونباخ و تحلیل عاملی تأییدی ارزیابی گردید.
اصلاح: ۱۴۰۴/۰۲/۱۵	یافته‌ها: نتایج مدل نشان داد که پرسشنامه سنجش اصیل با هشت عامل اصلی از اعتبار بالایی برخوردار است. ارزیابی مدل اندازه‌گیری با شاخص‌هایی مانند آلفای کرونباخ، پایایی ترکیبی و روایی همگرا و واگرا، بیانگر مطلوبیت و تناسب چارچوب بود. تحلیل مدل ساختاری بر اساس شاخص‌های Z نیز نشان داد که این چارچوب از ساختاری قوی و پایدار برخوردار است. افزون بر این، بین میزان بهره‌گیری از سنجش اصیل و متغیرهایی مانند دانشکده و مرتبه علمی تفاوت معنی‌داری مشاهده شد؛ به گونه‌ای که استادان باتجربه بیشتر بر خودارزیابی و تکالیف فکورانه و استادیاران بر الگودهی تمرکز داشتند. همچنین اعضای هیئت علمی دانشکده‌های مهندسی و علوم پایه بیشترین و اعضای هیئت علمی دانشکده زبان و ادبیات کمترین بهره‌گیری از سنجش اصیل را داشتند.
پذیرش: ۱۴۰۴/۰۳/۱۲	نتیجه‌گیری: این پژوهش با اعتبارسنجی ابزار سنجش و شناسایی الگوهای ارزیابی، به غنای ادبیات علمی در حوزه ارزشیابی آموزشی کمک کرده و می‌تواند در بهبود برنامه‌های توانمندسازی و سیاست‌گذاری‌های آموزشی مورد استفاده قرار گیرد.
انتشار: ۱۴۰۴/۰۴/۰۴	واژه‌های کلیدی: سنجش اصیل، روان‌سنجی پرسشنامه، اعضای هیئت علمی، تکالیف فکورانه

استاد: شبیربگی، ناصر؛ سلیمی، جمال؛ عزیزی، عبدالله (۱۴۰۴). شاخصه‌های روان‌سنجی پرسشنامه «سنجش اصیل» در فرایند یاددهی-یادگیری: مطالعه موردی اعضای هیئت علمی دانشگاه کردستان. مطالعات اندازه‌گیری و ارزشیابی آموزشی، ۱۵(۵۰)، ۱۲۱-۱۵۲.

DOI: 10.22034/emes.2025.2059237.2652



ناشر: سازمان سنجش آموزش کشور حق مؤلف © نویسندگان.

مقدمه

در دهه‌های اخیر، توجه به کیفیت آموزش عالی و بهبود عملکرد تحصیلی دانشجویان به یکی از مهم‌ترین دغدغه‌های پژوهشگران و سیاست‌گذاران آموزشی تبدیل شده است (ویگینز و مک‌تای^۱، ۲۰۰۵؛ بلک و ویلیام^۲، ۲۰۱۸). ارزیابی عملکرد تحصیلی دانشجویان نه تنها برای بررسی پیشرفت فردی آنها بلکه برای ارتقای سطح کلی آموزش نیز ضروری است (ساسکی^۳، ۲۰۱۸). یکی از جنبه‌های کلیدی در این زمینه، بررسی دیدگاه‌های استادان در مورد سنجش دانشجویان است؛ دیدگاه‌هایی که می‌تواند اطلاعات ارزشمندی درباره نقاط قوت و ضعف دانشجویان ارائه دهد و نقش مهمی در بهبود فرایندهای آموزشی ایفا کند (فلتچر و همکاران^۴، ۲۰۱۲). با وجود اهمیت سنجش اصیل^۵، همچنان نیاز به تحقیقات بیشتری در خصوص نگرش‌ها و نظرات استادان وجود دارد. بسیاری از مطالعات به برنامه‌ریزی، اجرا و بهبود ارزشیابی و عملکرد دانشجویان (بانتا و پالومبا^۶، ۲۰۱۴؛ هایل و ایفنتالر^۷، ۲۰۲۳؛ دری^۸، ۲۰۲۲) پرداخته‌اند، اما کمتر به این مسئله توجه شده است که استادان به‌عنوان افراد نزدیک به فرایند آموزشی، چه دیدگاه‌هایی درباره نقاط قوت و ضعف عملکرد دانشجویان دارند و این نظرات چگونه می‌تواند به بهبود روش‌های آموزشی و ارزیابی کمک کند. چنان‌که فلتچر و همکاران (۲۰۱۲) بیان کرده‌اند، سنجش در آموزش عالی می‌تواند اطلاعاتی درباره یادگیری دانشجویان، پیشرفت آنان، کیفیت تدریس و پاسخگویی نهادی فراهم آورد. با این حال، داده‌های اندکی درباره نگرش استادان نسبت به جنبه‌های مختلف ارزیابی وجود دارد؛ داده‌هایی که می‌تواند تأثیرات گسترده‌ای بر سیاست‌ها و روش‌های آموزشی در مؤسسات آموزش عالی داشته باشد. افزون بر این، عوامل متعددی همچون فناوری‌های نوین آموزشی، روش‌های ارزیابی پویا و تعاملات فردی و گروهی میان استادان و دانشجویان بر نحوه ارائه بازخورد و ارزیابی تأثیر گذاشته و زمینه را برای ارزیابی عمیق‌تر عملکرد تحصیلی دانشجویان فراهم کرده است (عزیزی و سلیمی، ۱۴۰۲).

تاریخچه مفاهیم «ارزیابی عملکرد تحصیلی» و «بازخورد تحصیلی» به اوایل قرن بیستم بازمی‌گردد؛ زمانی که پژوهشگران حوزه آموزش به اهمیت سنجش عملکرد و ارائه بازخورد به منظور بهبود فرایند یادگیری پی بردند (ثورندایک و هاگن^۹، ۱۹۲۱). سنجش اصیل به تدریج به‌عنوان ابزاری برای اندازه‌گیری میزان موفقیت دانشجویان در دستیابی به اهداف آموزشی معرفی شد. در دهه‌های بعد،

1. Wiggins & McTighe
2. Black & William
3. Suskie
4. Fletcher et al.
5. Authentic assessment
6. Banta & Palomba
7. Heil & Ifenthaler
8. Dery
9. Thorndike & Hagen

به‌ویژه پس از جنگ جهانی دوم، روش‌های بازخورد تحصیلی و ارزیابی دانشجویان توسعه بیشتری یافت و ابزارهایی مانند آزمون‌های استاندارد و گزارش‌های عملکرد به وجود آمدند (سدلر^۱، ۱۹۸۹). با ورود به قرن بیست و یکم، این مفاهیم تحت تأثیر فناوری‌های نوین و روش‌های یادگیری دیجیتالی گسترش یافته و امروزه جزء اساسی نظام‌های آموزشی مدرن محسوب می‌شوند (القرشی^۲، ۲۰۱۹؛ هوانگ و چانگ^۳، ۲۰۲۰). در حالی که تحقیقات گسترده‌ای به بررسی سنجش تدریس از دیدگاه دانشجویان پرداخته‌اند و عواملی مانند سهولت در نمره‌دهی، حجم کار و بازخورد را تحلیل کرده‌اند (مارش و روش^۴، ۲۰۰۰؛ اسپورن و همکاران^۵، ۲۰۱۳؛ سنترا^۶، ۲۰۰۳)، شکاف قابل توجهی در درک دیدگاه‌های استادان نسبت به این ارزیابی‌ها وجود دارد. هنوز روشن نیست که استادان چگونه این ارزیابی‌ها را در محیط آموزش عالی تفسیر کرده و به آنها واکنش نشان می‌دهند. به‌ویژه، اطلاعات محدودی درباره دیدگاه‌های استادان در خصوص انصاف، دقت و کارایی ارزیابی‌های دانشجویی در بازتاب عملکرد تدریس و نتایج یادگیری دانشجویان در دسترس است. همچنین، پژوهش‌های اندکی بررسی کرده‌اند که این ارزیابی‌ها چگونه بر تصمیمات استادان در زمینه روش‌های تدریس، شیوه‌های نمره‌دهی و رشد حرفه‌ای آنها تأثیر می‌گذارد.

بدون درک دیدگاه‌های استادان، سنجش اصیل به‌عنوان ابزاری برای ارزیابی و بهبود کیفیت تدریس ممکن است محدود یا ناهماهنگ با اهداف واقعی آموزش باشد. دیدگاه استادان اهمیت زیادی دارد، زیرا می‌تواند تعصبات، چالش‌ها و محدودیت‌های موجود در ارزیابی‌های دانشجویی را آشکار سازد؛ عواملی که ممکن است اثربخشی و انصاف این ارزیابی‌ها را تضعیف کند. نادیده گرفتن این دیدگاه‌ها می‌تواند به ارزیابی‌هایی منجر شود که رشد حرفه‌ای معنی‌داری برای استادان ایجاد نمی‌کند و حتی موجب نارضایتی یا کاهش انگیزه آنان می‌شود. پرداختن به این شکاف، زمینه‌ساز ایجاد سیستمی جامع و متعادل در ارزیابی‌های تدریس است؛ سیستمی که دیدگاه استادان را منعکس کرده و بهبود مستمر و بازخورد دوطرفه را در محیط آموزش عالی ترویج می‌دهد. با گنجاندن دیدگاه‌های استادان در فرایند ارزیابی، دانشگاه‌ها می‌توانند چارچوب‌هایی قابل اعتماد، منصفانه و سازنده طراحی کنند؛ چارچوب‌هایی که نه تنها اثربخشی تدریس را تقویت می‌کنند، بلکه باعث افزایش روحیه و انگیزه استادان نیز می‌شوند. پر کردن این شکاف برای دستیابی به رویکردی جامع در ارزیابی و ارتقای کیفیت آموزشی در دانشگاه‌ها ضروری است.

1. Sadler
2. Alqurashi
3. Hwang & Chang
4. Marsh & Roche
5. Spooren et al.
6. Centra

یافته‌های این پژوهش می‌توانند برای دانشگاه‌ها و سیاست‌گذاران آموزش عالی ارزشمند باشند و آنها را در بهبود چارچوب‌های ارزیابی تدریس یاری رسانند؛ به گونه‌ای که این ارزیابی‌ها به جای ایجاد تعصبات، به ابزاری برای ارتقای کیفیت تدریس تبدیل شوند. به کارگیری نتایج این مطالعه می‌تواند به شکل‌گیری فضای آموزشی مثبت‌تر و پشتیبانی‌کننده‌تر برای استادان و دانشجویان منجر شود و انگیزه، روحیه و رضایت شغلی استادان را تقویت کند. از این طریق، این تحقیق نه تنها به بهبود نظام ارزیابی در دانشگاه‌ها کمک می‌کند، بلکه می‌تواند بر کیفیت کلی آموزش عالی و تجربه یادگیری دانشجویان نیز اثرگذار باشد. افزون بر این، این مطالعه فرهنگ بازخورد و رشد متقابل را در دانشگاه‌ها تقویت کرده و دیدگاهی جامع‌تر از عملکرد تدریس ارائه می‌دهد که دانشجویان و استادان را به طور هم‌زمان در فرایند یادگیری دخیل می‌سازد. در نهایت، نتایج این پژوهش می‌توانند به سیاست‌گذاری‌های آموزشی مؤثرتری منجر شوند که به ارتقای کیفیت آموزش در سطح جهانی و بهبود تجربه یادگیری دانشجویان در دانشگاه‌ها کمک کند.

با وجود تأکید روزافزون بر استفاده از سنجش اصیل به‌عنوان رویکردی نوین در ارزشیابی دانشجویان، شواهد نشان می‌دهد که در نظام آموزش عالی ایران همچنان شیوه‌های سنتی ارزشیابی غالب هستند. این امر موجب شده است توانایی دانشجویان در به کارگیری دانش در موقعیت‌های واقعی و نیز پرورش مهارت‌های کلیدی همچون تفکر انتقادی، خلاقیت، حل مسئله و همکاری کمتر مورد توجه قرار گیرد. مرور مطالعات داخلی (عبداللهی گل و همکاران، ۱۳۹۴؛ احمدی و همکاران، ۱۴۰۲) آشکار می‌سازد که بهره‌گیری از مؤلفه‌های سنجش اصیل در دانشگاه‌های ایران کمتر از حد مطلوب است و ابزارهای بومی و معتبر برای سنجش این پدیده وجود ندارد. در سطح بین‌المللی نیز اگرچه ابزارهایی برای سنجش اصیل طراحی و اعتبارسنجی شده‌اند (ویلاروئل و همکاران^۱، ۲۰۱۹؛ ولسانت و همکاران^۲، ۲۰۲۴)، اما بیشتر آنها متناسب با زمینه‌های آموزشی کشورهای غربی هستند و به‌سختی می‌توان آنها را در محیط‌های فرهنگی و آموزشی ایران به کار گرفت. در چنین شرایطی، مسئله اصلی پژوهش حاضر آن است که شاخصه‌های روان‌سنجی پرسشنامه سنجش اصیل تا چه میزان از روایی و پایایی کافی برای استفاده در بستر آموزش عالی ایران برخوردارند و این ابزار تا چه حد می‌تواند برای ارزیابی دیدگاه اعضای هیئت علمی دانشگاه کردستان به کار رود. روشن‌سازی این موضوع می‌تواند خلأ موجود در نبود ابزار معتبر را پر کرده و زمینه‌ساز ارتقای کیفیت یاددهی-یادگیری در دانشگاه‌ها شود.

۱. پرسشنامه «سنجش اصیل» در میان اعضای هیئت علمی دانشگاه کردستان از چه میزان روایی و پایایی برخوردار است؟

1. Villarroel et al.
2. Welsandt et al.

۲. آیا شاخصه‌های روان‌سنجی پرسشنامه «سنجش اصیل» بر اساس ویژگی‌های جمعیت‌شناختی اعضای هیئت علمی (دانشکده و مرتبه علمی) تفاوت معناداری دارد؟

مبانی نظری و پیشینه پژوهش

نظریه‌های کلیدی مرتبط با سنجش اصیل دانشجویان شامل مروری بر نظریه‌های ارزشیابی و بازخورد است. در حوزه سنجش تحصیلی، نظریه‌های متعددی وجود دارند که هرکدام با تأکید بر ابعاد گوناگون، به بهبود یادگیری و رشد دانشجویان کمک می‌کنند. (بلوم^۱، ۱۹۶۸) نظریه یادگیری مبتنی بر تسلط^۲ را در دهه ۱۹۶۰ ارائه کرد. او تأکید داشت که یادگیرندگان باید قبل از پیشروی به مباحث جدید، بر محتوای درسی کاملاً مسلط شوند. این نظریه بعدها توسط افرادی همچون (گوسکی^۳، ۱۹۸۵) و (کارول^۴، ۱۹۶۳) گسترش یافت که به کاربردهای عملی و ارزشیابی پیوسته توجه ویژه‌ای داشتند. توماس گوسکی اهمیت ارزشیابی‌های مستمر و پیوسته را برای عمیق‌تر شدن درک دانش‌آموزان از مفاهیم مطرح کرد و بر استفاده از ارزشیابی به‌عنوان ابزاری برای تقویت یادگیری تأکید داشت. همچنین جان کارول مدلی ارائه داد که ارتباط بین زمان مورد نیاز و زمان صرف‌شده برای یادگیری را توضیح می‌داد. در ادامه این جریان، پژوهش‌های جدیدتر مانند پانادرو و همکاران^۵ (۲۰۱۸) نشان داده‌اند که خودارزیابی و ارزیابی همتایان نقش کلیدی در درونی‌سازی فرایند یادگیری دارند و موجب افزایش خودکارآمدی و مسئولیت‌پذیری دانشجویان می‌شوند.

در همین راستا می‌توان به نظریه بازخورد تأملی^۶ اشاره کرد که به معنای تأمل و نگاه نقادانه دانشجو بر عملکرد خود است که به آنان کمک می‌کند تا اشتباهات خود را شناسایی کرده و در طول فرایند یادگیری به بهبود و اصلاح مهارت‌های خود بپردازند. برخلاف بازخوردهای پایانی و نهایی که پس از اتمام فعالیت ارائه می‌شوند، این بازخوردها دانشجویان را ترغیب می‌کنند تا به تدریج و با دقت بیشتر یادگیری خود را ارتقا دهند (بروکهارت^۷، ۲۰۰۸). این دو نظریه هر دو به اهمیت بازخوردهای پیوسته و هدفمند اشاره دارند که دانشجویان را به یادگیری مؤثرتر هدایت می‌کنند.

سنجش اصیل به گفته ویگینز^۸ (۱۹۹۸) بر ارزیابی دانش‌آموزان از طریق انجام وظایف واقعی تمرکز دارد؛ وظایفی که نیازمند به‌کارگیری معنادار دانش و مهارت‌ها هستند. این رویکرد در تضاد با روش‌های سنتی ارزشیابی قرار دارد که بیشتر بر حفظ کردن مطالب تأکید می‌کنند. وظایف ارزیابی اصیل

1. Bloom

2. Mstery Leaning

3. Guskey

4. Carroll

5. Panadero et al.

6. Reflective feedback

7. Brookhart

8. Wiggins

به‌گونه‌ای طراحی شده‌اند که پیچیدگی‌ها و چالش‌های موقعیت‌های دنیای واقعی را منعکس کنند و دانش‌آموزان را به استفاده عملی از آموخته‌های خود در زمینه‌های مرتبط تشویق کنند. برای مثال، به جای آزمون چندگزینه‌ای در مورد نوشتار اقناعی، ارزیابی اصیل ممکن است شامل نگارش نامه‌ای به یک روزنامه محلی درباره یک موضوع جاری باشد که نیازمند استفاده از تکنیک‌های قانع‌کننده و برقراری ارتباط با مخاطب واقعی است. ویگینز معتقد بود که این ارزیابی‌ها نشان‌دهنده میزان توانایی دانش‌آموزان در انتقال آموخته‌هایشان به موقعیت‌های ناآشناست، نه صرفاً به یادآوری اطلاعات. امروزه نیز چارچوب‌های نوبن مانند مرور نظام‌مند اسکالی^۱ (۲۰۱۹) بر اهمیت سنجش اصیل در آموزش عالی تأکید دارند و نشان می‌دهند که فناوری‌های دیجیتال فرصت‌های تازه‌ای برای طراحی تکالیف معنادار و واقع‌گرایانه فراهم کرده‌اند.

ارزیابی معنادار و منصفانه به رهبری لیندا ساسکی بر اهمیت ارزیابی‌هایی تأکید می‌کند که به دانشجویان فرصت می‌دهند تا توانایی‌های واقعی خود را نشان دهند و در محیطی منصفانه و متنوع ارزیابی شوند. این رویکرد، با تکیه بر شفافیت و تنوع روش‌های ارزیابی مانند پروژه‌ها و ارائه‌ها، اطمینان می‌دهد که دانشجویان با توانایی‌های مختلف فرصت موفقیت دارند و می‌توانند در شرایط منصفانه‌تری رقابت کنند (ساسکی، ۲۰۰۰). از این منظر، اهمیت عدالت و انصاف در ارزیابی آشکار می‌شود.

ارزیابی دانشجو محور، نظریه‌ای دیگر از (استیگینز^۲، ۲۰۰۵) با تأکید بر مشارکت دانشجویان در فرایند ارزیابی و یادگیری، به آنها اجازه می‌دهد تا مالک فرایند یادگیری خود شوند. این رویکرد، دانشجویان را به خودارزیابی و ارزیابی همتایان تشویق کرده و موجب افزایش انگیزه و مشارکت آنان در یادگیری می‌شود. این نظریه در کنار نظریه ساسکی، دانشجویان را فردی فعال و مشارکت‌کننده در فرایند یادگیری می‌داند. این دیدگاه با نظریه خودتعیین‌بخشی^۳ (دسی و رایان^۴، ۲۰۰۰؛ ریو، ۲۰۱۲) نیز هم‌راستا است. در این نظریه، یادگیرندگان دارای سه نیاز بنیادین روان‌شناختی معرفی می‌شوند: خودمختاری^۵ (احساس کنترل و انتخاب در فرایند یادگیری)؛ شایستگی^۶ (تجربه موفقیت و توانایی در انجام تکالیف یادگیری)؛ ارتباط^۷ (احساس تعلق و تعامل مثبت با دیگران). هنگامی که فرایند ارزیابی به صورت اصیل طراحی می‌شود، مثلاً با سپردن مسئولیت بخشی از ارزشیابی به خود دانشجویان یا فراهم آوردن

1. Scully
2. Stiggins
3. Self-Determination Theory
4. Deci & Ryan
5. Autonomy
6. Competence
7. Relatedness

فرصت ارزیابی همتایان، این سه نیاز به شکل مؤثری تقویت می‌گردد. برای مثال، تکالیفی که در بستر واقعی یا شبه‌واقعی انجام می‌شوند، به دانشجویان خودمختاری می‌بخشند زیرا آن‌ها در تصمیم‌گیری درباره چگونگی حل مسئله آزادی عمل دارند. بازخوردهای سازنده و مستمر، سطح شایستگی آنها را ارتقا داده و به آنان اطمینان می‌دهد که در مسیر درست قرار دارند. همچنین، فعالیت‌های گروهی و مشارکتی در چارچوب ارزیابی اصیل، احساس ارتباط و تعلق اجتماعی را افزایش داده و باعث می‌شود دانشجویان خود را عضوی فعال در جامعه یادگیری بدانند. بنابراین، می‌توان گفت ارزیابی اصیل نه تنها ابزاری برای سنجش یادگیری است، بلکه سازوکاری برای انگیزش درونی^۱ یادگیرندگان محسوب می‌شود. پژوهش‌ها نشان داده‌اند که وقتی نیازهای سه‌گانه نظریه خودتعیین‌گری ارضا شوند، دانشجویان انگیزه پایدار، مشارکت فعال‌تر و یادگیری عمیق‌تری از خود نشان می‌دهند (دسی و رایان، ۲۰۰۰؛ ریو، ۲۰۱۲).

از دیگر انواع ارزشیابی آموزشی می‌توان به ارزیابی‌های تکوینی^۲ و پایانی^۳ اشاره کرد. بلوم به‌عنوان یک روان‌شناس آموزشی، نقش بزرگی در ترویج ارزیابی‌های تکوینی برای بهبود روش‌های آموزشی داشت. اسکرینون، فیلسوف و ارزیاب آموزشی، در دهه ۱۹۶۰ تفاوت بین ارزیابی‌های تکوینی و پایانی را در چارچوب ارزیابی برنامه‌ها مطرح کرد. ارزیابی تکوینی به ارزیابی‌های مداوم برای بهبود فرایند یادگیری اشاره دارد، در حالی که ارزیابی پایانی برای سنجش اثربخشی کلی در پایان فرایند یادگیری است. چارچوب اسکرینون، نقش‌ها و اهداف متفاوت ارزیابی‌های تکوینی و پایانی را در آموزش مشخص کرد و بر چگونگی استفاده از ارزیابی‌ها در جهت پیشبرد یادگیری تأکید نمود (اسکرینون^۴، ۱۹۶۷). همچنین ارزیابی پایدار که توسط بود و فالچیکوف^۵ (۲۰۰۷) ارائه شد بر ایجاد مهارت‌های خودارزیابی و یادگیری مادام‌العمر در دانشجویان تأکید دارد. بود معتقد است که ارزیابی‌ها نباید تنها به سنجش دانش موجود بپردازند، بلکه باید مهارت‌هایی را در دانشجویان ایجاد کنند که در مسیر یادگیری آینده نیز مؤثر باشند. این دیدگاه مشابه نظریات استیگینز است که بر اهمیت فراگیری مهارت‌های یادگیری و خودارزیابی تأکید می‌کند و دانشجویان را برای پیشرفت پایدار در زندگی آموزشی و حرفه‌ای خود آماده می‌سازد. در تأیید همین نظریات می‌توان به نظریه ارزشیابی برای یادگیری^۶ اشاره کرد؛ رویکردی آموزشی که توسط بلک و ویلیام (۱۹۹۸) معرفی شده و بر استفاده از ارزشیابی به‌عنوان ابزاری برای بهبود یادگیری، نه صرفاً برای سنجش عملکرد، تأکید دارد. این نظریه بر این اصل استوار

1. Intrinsic motivation
2. Formative assessment
3. Summative Assessment
4. Scriven
5. Boud & Falchikov
6. Assessment for Learning

است که فرایند ارزشیابی باید به دانش‌آموزان کمک کند تا درک بهتری از مسیر یادگیری خود داشته باشند، نقاط ضعف و قوت خود را شناسایی کنند و از طریق بازخورد مستمر و سازنده، یادگیری خود را اصلاح و تقویت نمایند. برخلاف ارزشیابی‌های سنتی که معمولاً بر نمره‌دهی و رتبه‌بندی تمرکز دارند، این نظریه یادگیرندگان را در فرایند یادگیری فعال‌تر می‌کند و آنها را تشویق می‌کند که به طور مداوم عملکرد خود را ارزیابی کرده و پیشرفت کنند.

در نهایت ارزشیابی‌های اصیل که بر کاربردهای عملی و واقعی دانش تأکید دارند، به دانشجویان این امکان را می‌دهند که مهارت‌های خود را در زمینه‌های مختلف توسعه داده و از یادگیری خود در موقعیت‌های واقعی استفاده کنند. به‌طورکلی، تلفیق این نظریه‌ها نشان می‌دهد که ارزشیابی اصیل نه تنها ابزاری برای سنجش دانشجویان، بلکه راهکاری برای ایجاد یادگیری عمیق، پایدار و معنادار است که به ارتقای توانایی‌های آنها در مسیر حرفه‌ای و تحصیلی منجر می‌شود.

در خصوص نقش و جایگاه سنجش اصیل در فرایند یاددهی / یادگیری، مطالعاتی در داخل و خارج از کشور صورت گرفته است که به برخی از مهم‌ترین و مرتبط‌ترین آنها اشاره می‌شود.

در سال‌های اخیر، پژوهش‌هایی در داخل کشور میزان به‌کارگیری ارزیابی اصیل در نظام آموزشی را بررسی کرده‌اند. عبداللهی گل و همکاران (۱۳۹۴) در پژوهشی با هدف بررسی میزان استفاده از اصول سنجش اصیل در آموزش عالی ایران، دریافتند که استفاده از مؤلفه‌هایی چون سنجش تکوینی، تکالیف معنادار و یادگیری در جریان ارزشیابی کمتر از حد مطلوب است؛ با این حال، تنوع ابزارها و پاسخ‌های صحیح وضعیت نسبتاً مطلوب‌تری داشت. همچنین در پژوهشی دیگر، احمدی و همکاران (۱۴۰۲) به بررسی درک و میزان بهره‌گیری معلمان دبیرستانی شهر سنج از ارزیابی اصیل پرداختند. نتایج نشان داد که اگرچه سطح ادراک و استفاده معلمان از ارزیابی اصیل بالاتر از حد متوسط بود، اما این متغیر با ویژگی‌های فردی معلمان رابطه معناداری نداشت. این یافته‌ها حاکی از آن است که با وجود آگاهی نسبی نسبت به ارزیابی اصیل، چالش‌هایی در اجرای اثربخش آن در محیط‌های آموزشی ایران وجود دارد.

در حوزه مطالعات بین‌المللی، ولسانت و همکاران (۲۰۲۴) به طراحی و اعتبارسنجی یک ابزار ارزیابی اصیل مبتنی بر فناوری برای سنجش سواد اقتصادی دانش‌آموزان پایه هشتم در آلمان پرداختند. نتایج نشان‌دهنده آن بود که میزان اصالت آزمون تأثیر معناداری بر دشواری سؤالات نداشت، اما توانست ابعاد مختلف سواد اقتصادی را با موفقیت پوشش دهد. در همین راستا، گوش و همکاران^۱ (۲۰۲۰) در پژوهشی تطبیقی، تأثیر ارزیابی اصیل و سنتی را بر عملکرد تحصیلی دانشجویان رشته

¹ Ghosh et al.

دریانوردی در کالج دریایی استرالیا مقایسه کردند. یافته‌ها نشان داد که استفاده از ارزیابی اصیل به‌طور معناداری عملکرد تحصیلی را ارتقا داده است. همچنین کوه و همکاران^۱ (۲۰۱۹) در بررسی نقش سنجش اصیل در دوره‌های مبتنی بر یادگیری مسئله‌محور دریافتند که این سنجش می‌تواند مهارت‌هایی مانند یادگیری خودمحور و حل مسئله را در دانشجویان تقویت کند.

ویلاروئل و همکاران (۲۰۱۸) با مرور نظام‌مند متون علمی، سه مؤلفه اصلی ارزیابی اصیل یعنی واقع‌گرایی، چالش‌شناختی و قضاوت ارزیابانه را شناسایی کردند. آنان در پژوهش بعدی خود (۲۰۱۹) راهکارهایی برای به‌کارگیری این مؤلفه‌ها در طراحی آزمون‌های کتبی ارائه دادند. در نهایت، سخنور (۲۰۲۱) در یک مرور نظام‌مند بر ۲۶ مقاله پژوهشی نشان داد که ارزیابی اصیل تأثیر مثبتی بر تجربه یادگیری، مشارکت، رضایت و رشد مهارت‌های اشتغال‌پذیری دانشجویان از جمله ارتباطات، همکاری، تفکر انتقادی و حل مسئله دارد.

تحقیقات گسترده‌ای درباره دیدگاه‌های دانشجویان نسبت به عملکرد استادان و ارزشیابی‌های آموزشی، به‌ویژه در ایران، انجام شده است. این مطالعات عمدتاً بر ارزشیابی استادان و تأثیر آن بر کیفیت تدریس و آموزش تمرکز داشته‌اند. با این حال، در زمینه دیدگاه استادان نسبت به ارزشیابی دانشجویان و تأثیر آن بر عملکرد تحصیلی آنان، خلأ قابل توجهی مشاهده می‌شود. این در حالی است که استادان نقش کلیدی در فرایند آموزش و ارزشیابی دارند و نگرش و روش‌های آنان می‌تواند تأثیر قابل توجهی بر تجربه یادگیری دانشجویان داشته باشد.

یکی از یافته‌های اصلی این پژوهش‌ها، اهمیت ویژگی‌های استاد در بهبود عملکرد دانشجویان است. کیفیت تدریس و بهره‌وری پژوهشی استادان نقش مثبتی در پیشرفت تحصیلی دانشجویان ایفا می‌کند. در مقابل، عواملی مانند محدودیت زمانی در تدریس یا وضعیت استخدامی موقت می‌توانند به کاهش نتایج تحصیلی دانشجویان منجر شوند. این امر نشان می‌دهد که سرمایه‌گذاری بر بهبود مهارت‌های تدریس و حمایت از توسعه حرفه‌ای استادان می‌تواند به‌طور مستقیم بر موفقیت تحصیلی دانشجویان تأثیرگذار باشد. در تنها پژوهش انجام‌شده درباره نگرش استادان نسبت به ارزشیابی دانشجویان، مشخص شد که بسیاری از استادان ارزشیابی را ابزاری معنادار و مفید می‌دانند، اما این ابزار بیشتر برای بهبود کلی تدریس استفاده می‌شود و کمتر برای اصلاح جنبه‌های خاص دوره، مانند منابع آموزشی یا امتحانات، به کار می‌رود. با این حال، استادانی که نگرش مثبتی نسبت به ارزشیابی دارند، از نتایج آن برای بهبود جنبه‌های مختلف تدریس خود استفاده بیشتری می‌کنند. این یافته‌ها بر اهمیت آموزش و توانمندسازی استادان در استفاده بهینه از ابزارهای ارزشیابی تأکید می‌کند.

1. Koh et al.

افزون بر این، یکی از عناصر کلیدی در بهبود عملکرد تحصیلی دانشجویان، ارتباط مثبت میان استاد و دانشجو است. پژوهش‌ها نشان می‌دهد که روابط نزدیک و حمایتی بین استادان و دانشجویان نه تنها موجب افزایش مشارکت دانشجویان می‌شود، بلکه تأثیر مثبتی بر روش‌های یادگیری عمیق نیز دارد. این تعاملات، هم در کلاس و هم در خارج از آن، می‌توانند نقش مهمی در ایجاد انگیزه و بهبود تجربه یادگیری دانشجویان ایفا کنند. در نهایت، یافته‌های این پژوهش‌ها نشان می‌دهد که تقویت فرصت‌های حرفه‌ای برای توسعه مهارت‌های تدریسی استادان، فراهم کردن ابزارهای معتبرتر ارزشیابی، و تشویق تعاملات میان همکاران می‌تواند به ارتقای کیفیت آموزش کمک کند. تمرکز بر استانداردهای ارزشیابی‌ها، تقویت روابط استاد و دانشجو، و تطبیق برنامه‌های درسی با نیازهای واقعی دانشجویان، از مهم‌ترین اقدامات برای بهبود نتایج تحصیلی است.

روش پژوهش

در این پژوهش با بهره‌گیری از رویکرد کمی و راهبرد پیمایشی، میزان بهره‌گیری استادان از سنجش اصیل در فرایند یاددهی-یادگیری بررسی می‌شود. رویکرد کمی در این مطالعه به پژوهشگر این امکان را می‌دهد که داده‌های عددی و قابل اندازه‌گیری را گردآوری و تحلیل کند تا روابط میان متغیرها شناسایی شوند. استفاده از راهبرد پیمایشی نیز به دلیل ماهیت توصیفی پژوهش و نیاز به گردآوری داده‌های تجربی از یک نمونه گسترده، مناسب است، زیرا این روش به پژوهشگر اجازه می‌دهد که نگرش‌ها، باورها، و میزان بهره‌گیری از سنجش اصیل را در بین گروهی از استادان بررسی و تحلیل کند. ابزار گردآوری داده‌ها در این پژوهش شامل داده‌های ثانویه و اولیه است. در گام نخست، برای بررسی مبانی نظری و پیشینه پژوهش، از کتاب‌ها، مقاله‌های علمی و پایان‌نامه‌های فارسی و انگلیسی مرتبط با موضوع استفاده شد. این منابع مبنای تدوین ابزار گردآوری داده‌های اولیه را فراهم کرده و به شناسایی مؤلفه‌های اصلی سنجش اصیل کمک کردند. در گام بعد، برای گردآوری داده‌های اولیه، پرسشنامه‌ای محقق‌ساخته طراحی شد که بر اساس چارچوب نظری هشت‌گانه ویگینز (۱۹۹۸)، شامل معیارهای شفاف، خودارزیابی، یادگیری اصیل، تکالیف فکورانه، الگودهی، استانداردهای شناخته‌شده و توجیه یادگیری، تکالیف جایگزین، اقتباس شده است. این پرسشنامه دارای ۵۹ سؤال بوده و هشت مؤلفه اصلی سنجش اصیل را پوشش می‌دهد. پرسشنامه در دو بخش تنظیم شده است: بخش نخست شامل سؤالات جمعیت‌شناختی مانند جنسیت، سابقه خدمت، مرتبه علمی و دانشکده؛ و بخش دوم شامل سؤالاتی برای سنجش میزان بهره‌گیری استادان از سنجش اصیل در فرایند تدریس است. پاسخ‌دهندگان در این بخش میزان موافقت خود را بر اساس یک مقیاس پنج‌سطحی لیکرت از «خیلی کم (۱)» تا «خیلی زیاد (۵)» مشخص می‌کنند.

برای اعتبارسنجی پرسشنامه، از اعتبار محتوایی (اعتبار صوری) استفاده شد. در این مرحله، پرسشنامه پس از تدوین اولیه در اختیار استادان متخصص در حوزه علوم تربیتی قرار گرفت و پس از دریافت بازخوردهای آنها، اصلاحات لازم در متن و گویه‌های پرسشنامه اعمال شد تا از روایی مناسب ابزار اطمینان حاصل شود. برای بررسی پایایی پرسشنامه نیز از آلفای کرونباخ استفاده گردید که میزان همسانی درونی گویه‌ها را مشخص می‌کند. به این منظور، در مرحله پیش‌آزمون، ۵۰ پرسشنامه میان اعضای نمونه اولیه توزیع و تکمیل شد. پس از گردآوری داده‌ها، میزان آلفای کرونباخ برای کل پرسشنامه و همچنین برای هر یک از مؤلفه‌ها محاسبه گردید. در صورت مشاهده گویه‌های نامناسب، تغییرات و اصلاحات لازم انجام شد تا ابزار از سطح پایایی مطلوبی برخوردار شود. همچنین، برای اطمینان بیشتر از اعتبار سازه‌ای پرسشنامه، از تحلیل عاملی تأییدی استفاده شد. در این روش، میزان بار عاملی گویه‌ها بررسی گردید تا مشخص شود که هر سؤال به‌درستی در راستای سنجش مؤلفه‌های مورد نظر عمل می‌کند. یافته‌های بخش تحلیل عاملی تأییدی نشان داد که تمامی گویه‌ها دارای بار عاملی مناسب بوده و ساختار عاملی پرسشنامه از روایی کافی برخوردار است. بدین ترتیب، فرایند طراحی و اعتبارسنجی این پرسشنامه بر اساس مبانی علمی دقیق انجام شده و ابزار مورد استفاده از روایی و پایایی مطلوبی برخوردار است. این پرسشنامه می‌تواند به‌عنوان یک ابزار معتبر برای سنجش میزان بهره‌گیری استادان از سنجش اصیل در فرایند تدریس و یادگیری مورد استفاده قرار گیرد.

از آنجایی که واحد تحلیل این پژوهش دانشگاه کردستان است، انتخاب این جامعه آماری بر مبنای رویکرد مطالعات موردی انجام شده است. دلیل اصلی انتخاب دانشگاه کردستان، جایگاه آن به‌عنوان یکی از دانشگاه‌های منطقه‌ای معتبر در غرب کشور و برخورداری از تنوع رشته‌ها و دانشکده‌هاست که بستری مناسب برای بررسی روایی و پایایی پرسشنامه «سنجش اصیل» در بافت آموزش عالی ایران فراهم می‌سازد. در چنین پژوهش‌هایی تمرکز بر یک سازمان یا پدیده با مرزهای مشخص ضروری است و دانشگاه کردستان به‌عنوان یک «مورد طبیعی» با این ویژگی‌ها در نظر گرفته شد. جامعه آماری شامل همه اعضای هیئت علمی این دانشگاه بوده و به‌دلیل تنوع رشته‌ها و دانشکده‌ها، روش نمونه‌گیری طبقه‌ای متناسب با حجم جامعه به کار گرفته شد تا نمایندگی دقیق‌تری از کل جامعه حاصل شود. حجم نمونه نیز بر اساس فرمول کوکران برآورد گردید و در نهایت ۱۹۴ نفر از اعضای هیئت علمی دانشگاه کردستان به‌عنوان نمونه پژوهش انتخاب شدند.

در این پژوهش، داده‌های حاصل از پرسشنامه در دو سطح آمار توصیفی و آمار استنباطی تحلیل شدند. در بخش آمار توصیفی، از شاخص‌های آماری نظیر فراوانی، میانگین، و انحراف معیار برای توصیف و تحلیل داده‌ها استفاده گردید. علاوه بر این، از نمودارها و جدول‌های آماری برای نمایش ویژگی‌های

جمعیت‌شناختی پاسخ‌دهندگان و روند توزیع داده‌ها بهره گرفته شد. از آنجاکه بسیاری از آزمون‌های آماری پارامتریک بر فرض نرمال بودن توزیع داده‌ها استوار هستند، پیش از انجام تحلیل‌های آماری، توزیع داده‌ها بررسی گردید. برای این منظور، از آزمون کلموگروف-اسمیرنوف استفاده شد تا مشخص شود که آیا داده‌های گردآوری‌شده از توزیع نرمال پیروی می‌کنند یا خیر. این آزمون امکان تعیین مناسب‌ترین روش آماری برای تجزیه و تحلیل داده‌ها را فراهم می‌کند. پس از بررسی نوع توزیع داده‌ها، برای پاسخ به سؤال‌های پژوهش از روش‌های آمار استنباطی بهره گرفته شد. در این راستا، برای بررسی تفاوت‌های میان دانشکده‌های مختلف و مرتبه علمی استادان از تحلیل واریانس یک طرفه آنوا استفاده شد. از آنجایی که استادان زن در دانشگاه کردستان اندک بودند محققان از مقایسه دیدگاه‌ها بر حسب جنسیت خودداری کردند. علاوه بر این، به منظور بررسی روابط بین متغیرهای پژوهش و ارزیابی صحت مدل مفهومی، از مدل‌یابی معادلات ساختاری بهره گرفته شد. در این بخش، از روش حداقل مربعات جزئی با استفاده از نرم‌افزار SmartPLS برای تحلیل داده‌ها استفاده شد. این روش به پژوهشگر امکان بررسی میزان تأثیرگذاری هر یک از متغیرها بر متغیر وابسته را داده و به ارزیابی صحت مدل مفهومی کمک می‌کند. از نرم‌افزار SPSS نیز برای انجام تحلیل‌های آماری پایه و آزمون‌های توصیفی و استنباطی بهره گرفته شد.

یافته‌ها

در این پژوهش، ۱۹۴ نفر از اعضای هیئت علمی دانشگاه‌ها شرکت کرده‌اند. از نظر جنسیت، اکثریت شرکت‌کنندگان را مردان با ۱۷۷ نفر (۹۱/۲ درصد) بودند، در حالی که زنان تنها ۱۷ نفر (۸/۸ درصد) از نمونه را شامل می‌شوند. این ترکیب نشان‌دهنده حضور گسترده‌تر مردان در میان اعضای هیئت علمی بررسی شده است. از نظر مرتبه علمی، بیشترین فراوانی مربوط به دانشیاران با ۱۰۰ نفر (۵۱/۵ درصد) است، در حالی که استادیاران ۷۰ نفر (۳۶/۱ درصد) و استادان ۲۴ نفر (۱۲/۴ درصد) را تشکیل می‌دهند. این توزیع نشان می‌دهد که بخش عمده‌ای از شرکت‌کنندگان در سطح میان‌رتبه‌ای از مسیر علمی خود قرار دارند. در رابطه با دانشکده محل خدمت، بیشترین تعداد اعضای هیئت علمی در دانشکده‌های علوم انسانی و اجتماعی (۴۹ نفر معادل ۲۵/۳ درصد)، مهندسی (۳۹ نفر معادل ۲۰/۱ درصد) و علوم پایه (۳۸ نفر معادل ۱۹/۶ درصد) فعالیت دارند. سایر شرکت‌کنندگان در دانشکده‌های کشاورزی (۲۸ نفر)، زبان و ادبیات (۲۱ نفر) و منابع طبیعی (۱۸ نفر) مشغول به کار هستند. تنها یک نفر اطلاعات مربوط به دانشکده خود را ارائه نکرده است. از نظر سابقه خدمت، بیشتر شرکت‌کنندگان بین ۹ تا ۱۲ سال سابقه کاری دارند (۸۸ نفر معادل ۴۵/۴ درصد)، در حالی که گروه‌های با سابقه ۴ تا ۸ سال (۳۰ نفر)، ۱۳ تا ۱۶ سال (۳۹ نفر)، ۱۷ تا ۲۰ سال (۲۶ نفر) و ۲۱ تا

۲۴ سال (۱۱ نفر) نیز در پژوهش حضور دارند. این ترکیب نشان‌دهنده مشارکت بیشتر اعضای هیئت علمی با تجربه متوسط در مطالعه است. به‌طور کلی، داده‌های جمعیت‌شناختی این پژوهش نشان می‌دهد که اکثریت پاسخ‌دهندگان مردان، دانشیاران و اعضای هیئت علمی با سابقه کاری متوسط هستند و بیشترین تعداد آنان در دانشکده‌های علوم انسانی، مهندسی و علوم پایه مشغول به تدریس و پژوهش هستند.

میانگین و انحراف معیار مؤلفه‌های پرسشنامه سنجش اصیل: در جدول (۱) میانگین نظرات جامعه آماری در ابعاد پرسشنامه سنجش اصیل درج شده است. نتایج نشان می‌دهد که از دیدگاه اعضای هیئت علمی بیشترین میانگین ابعاد آموزش ارزشیابی اصیل به ترتیب شامل مؤلفه‌های «استانداردهای شناخته‌شده» با میانگین ۴/۸۶ و انحراف معیار ۰/۰۹ و «یادگیری اصیل» با میانگین ۴/۰۲ و انحراف معیار ۰/۱۲ است. میانگین سایر مؤلفه‌های دیگر در جدول (۱) درج شده است.

جدول (۱) میانگین و انحراف معیار مؤلفه‌های پرسشنامه سنجش اصیل

ابعاد	نمونه‌ها	میانگین	انحراف معیار
معیارهای شفاف	استادان	۲۲/۳	۴۳/۰
خودارزیابی	استادان	۲۸/۳	۱۱/۰
یادگیری اصیل	استادان	۰۲/۴	۱۲/۰
تکالیف فکورانه	استادان	۰۱/۳	۸۲/۰
الگودهی	استادان	۰۷/۴	۴۳/۰
استانداردهای شناخته‌شده	استادان	۸۶/۴	۰۹/۰
توجیه یادگیری	استادان	۴۳/۱	۸۷/۱
تکالیف جایگزین	استادان	۶۹/۲	۳۲/۱

برای بررسی نرمال بودن توزیع داده‌های حاصل از پرسشنامه‌ها، از آزمون کولموگروف-اسمیرنوف در نرم‌افزار SPSS نسخه ۲۷ استفاده شد. این آزمون یک روش آماری است که به‌منظور مقایسه توزیع داده‌های نمونه با یک توزیع نرمال (یا هر توزیع نظری دیگر) طراحی شده است. فرض صفر در این آزمون بیان می‌کند که توزیع داده‌ها نرمال است. اگر نتیجه آزمون نشان دهد که مقدار آماره آزمون (Sig) بیشتر از ۰/۰۵ است، فرض صفر تأیید می‌شود؛ به این معنی که می‌توان با اطمینان ۹۵ درصد گفت که داده‌های مورد بررسی به‌طور نرمال توزیع شده‌اند. در این صورت، استفاده از آزمون‌های

پارامتریک که بر پایه فرض نرمال بودن توزیع داده‌ها هستند، مناسب خواهد بود. اما اگر مقدار Sig کمتر از $0/05$ باشد، فرض نرمال بودن توزیع داده‌ها رد می‌شود و ممکن است نیاز باشد از روش‌های آماری ناپارامتریک که نیازی به فرض نرمال بودن توزیع داده‌ها ندارند، استفاده شود. این آزمون به محقق این امکان را می‌دهد تا از نرمال بودن یا نبودن توزیع داده‌های گردآوری‌شده اطمینان حاصل کند و بر اساس نتایج آن، روش‌های مناسب آماری را برای تحلیل داده‌ها انتخاب کند.

جدول (۲) نتایج آزمون کولموگروف-اسمیرنف برای پرسشنامه سنجش اصیل

مؤلفه	آماره آزمون	سطح معناداری	نتیجه آزمون
معیارهای شفاف	۰/۱۵۴	۰/۰۶۵	نرمال است
خودارزیابی	۰/۱۷۶	۰/۰۸۵	نرمال است
یادگیری اصیل	۰/۰۸۰	۰/۱۳۵	نرمال است
تکالیف فکورانه	۰/۱۱۴	۰/۱۲۱	نرمال است
الگودهی	۰/۲۳۴	۰/۰۷۴	نرمال است
استانداردهای شناخته‌شده	۰/۱۴۳	۰/۱۱۸	نرمال است
توجه یادگیری	۰/۱۰۹	۰/۱۵۷	نرمال است
تکالیف جایگزین	۰/۰۹۸	۰/۰۵۹	نرمال است

همان‌طور که از داده‌های جدول (۲) مشخص است، سطح معناداری آزمون کلموگروف-اسمیرنوف برای مؤلفه‌های پژوهش بزرگ‌تر از مقدار $0/05$ است. در نتیجه مؤلفه‌های مورد بررسی در پژوهش حاضر دارای توزیع نرمال هستند. بنابراین برای بررسی سؤالات پژوهش از آزمون‌های پارامتری استفاده می‌شود. افزون بر این، با توجه به حجم بالای نمونه و بهره‌گیری از قضیه حد مرکزی، که بیان می‌کند با افزایش حجم نمونه، ویژگی‌های نمونه به ویژگی‌های جامعه نزدیک می‌شود می‌توان از مدل تحلیل عاملی تأییدی برای هنجاریابی پرسشنامه آموزش آکنده از حیرت استفاده کرد. برای این منظور، مدل‌یابی معادلات ساختاری با روش حداقل مربعات جزئی (PLS) در نرم‌افزار Smart PLS^۳ به کار گرفته شد. این روش شامل دو مرحله اساسی است.

اعتبارسنجی پرسشنامه میزان بهره‌گیری استادان از سنجش اصیل: اعتبارسنجی پرسشنامه میزان بهره‌گیری استادان از سنجش اصیل با استفاده از نرم‌افزار PLS و روش تحلیل عامل تأییدی انجام شد. برای بررسی روایی سازه، بارهای عاملی، میانگین واریانس استخراج‌شده و پایایی ترکیبی

مورد ارزیابی قرار گرفتند. همچنین، ضریب آلفای کرونباخ برای بررسی پایایی استفاده شد که مقادیر به‌دست آمده نشان‌دهنده برازش مناسب مدل اندازه‌گیری و اعتبار مطلوب پرسشنامه بودند. در ادامه به‌صورت جزئی به گزارش آن پرداخته می‌شود.

برازش مدل اندازه‌گیری: مدل اندازه‌گیری به‌عنوان بخشی از مدل کلی معادلات ساختاری شامل یک متغیر و شاخص‌های آن است. برای بررسی این مدل در روش حداقل مربعات جزئی، سه مرحله اصلی شامل بررسی پایایی، روایی همگرا و روایی واگرا انجام می‌گیرد که در ادامه به هر یک از این مراحل پرداخته می‌شود.

آلفای کرونباخ به‌عنوان یک معیار سنتی برای سنجش پایایی و ارزیابی سازگاری درونی به کار می‌رود. مقدار مطلوب برای آلفای کرونباخ بیشتر از 0.7 است. با توجه به جدول (۳)، تمامی متغیرهای مکنون در این پژوهش دارای آلفای کرونباخ بالاتر از 0.5 هستند که نشان‌دهنده پایایی مناسب پژوهش است. از آنجا که پایایی ترکیبی به‌عنوان یک معیار جدیدتر و دقیق‌تر نسبت به آلفای کرونباخ در مدل‌سازی معادلات ساختاری شناخته می‌شود، نتایج جدول (۳) نشان می‌دهد که تمامی متغیرهای مکنون دارای پایایی ترکیبی بالاتر از 0.5 هستند که به برازش مناسب مدل اندازه‌گیری اشاره دارد. روایی همگرا با استفاده از میانگین واریانس استخراج‌شده (AVE) و بارهای عاملی متغیرهای مکنون محاسبه می‌شود. معیار میانگین واریانس استخراج‌شده نشان‌دهنده میزان واریانس به اشتراک گذاشته‌شده بین هر سازه و شاخص‌های مربوط به آن است. مقدار بالاتر از 0.5 برای این معیار، روایی همگرای قابل قبول را تأیید می‌کند.

جدول (۳) مقادیر بارعاملی، آماره معناداری، آلفای کرونباخ، پایایی ترکیبی و AVE برای پرسشنامه تحقیق

سازه	گویه	بارعاملی	میانگین واریانس (AVE)	پایایی ترکیبی	آلفای کرونباخ
پهلوهای شفاف	دانشجویان معیارهای نمره‌دهی را از قبل می‌دانند	۰/۵۶۴	۰/۶۵۴	۰/۵۹۹	۰/۶۳۴
	دانشجویان الگوهای برای نوشتن مقاله یا اجرای ضبط‌شده از قبل دیده‌اند	۰/۴۳۲			
	نظام سنجش و ارزیابی من برایشان شفاف است	۰/۶۳۴			
	می‌دانند که نظام ارزشیابی من وزارتی و غیر اختیاری است	۰/۴۷۶			
	دانشجویان امکان استفاده از روش‌های جایگزین برای انجام یک تکلیف را دارند	۰/۶۴۵			
	فهرستی از سؤالاتی که احتمال دارد برای امتحان انتخاب شوند در اختیار دارند	۰/۴۵۶			
	دانشجویان بر پایه معیارها و استانداردهای شناخته‌شده ارزیابی می‌شوند	۰/۵۳۲			
	ارزیابی از عملکرد دانشجویان باید دارای معیارهای کاملاً شفاف باشد	۰/۵۷۶			
	ارزیابی باید با شفاف‌سازی انتظارات معلم قبل از شروع امتحان انجام شود	۰/۴۹۸			
پهلوهای شفاف	ارزیابی دانشجویان هسته اصلی کارم را تشکیل می‌دهد	۰/۶۵۴	۰/۵۳۲	۰/۵۴۳	۰/۵۴۳
	خودارزیابی دانشجویان رویدادی متداول و عادی است	۰/۶۳۲			
	دانشجویان باید یک خودارزیابی برای بخش‌های مهم کارشان ارائه دهند	۰/۶۵۴			
	به دانشجو شیوه خودارزیابی و تطبیق با معیارهای عملکرد آموزش داده می‌شود	۰/۵۹۸			
	دانشجو باید خودارزیابی را در همه تکالیف اصلی خود ادغام کند	۰/۷۶۵			
	به‌درستی به خودارزیابی دانشجو نمره اختصاص داده می‌شود	۰/۶۵۴			

سازه	گویه	بارعاملی	میانگین واریانس (AVE)	پایایی ترکیبی	آلفای کرونباخ
دکتری اصل	تکالیف پیوندی قوی با دنیای واقعی دارند	۰/۴۳۲	۰/۶۵۴	۰/۵۴۳	۰/۷۶۵
	همواره در محیط شبیه‌سازی شده واقعی انجام می‌شوند	۰/۶۳۴			
	همواره به تولید محصول یا اثر هنری منجر می‌شوند	۰/۵۴۳			
	نیازمند استفاده از دانش در دنیای واقعی هستند	۰/۶۵۴			
	شبیه مسائل و مشکلاتی هستند که افراد در محیط واقعی با آن مواجه می‌شوند	۰/۴۳۲			
	دربرگیرنده کارهای ضروری تکرار شونده هستند	۰/۴۷۶			
	بر اساس مهارت‌ها و موقعیت‌های قابل پیش‌بینی تنظیم می‌شوند	۰/۶۵۴			
	کاربرد ذهنی مهارت‌های گوناگون آموخته شده به یک محیط جدید باشد	۰/۶۵۴			
	قابل اجرا بودن در دنیای واقعی «یک ویژگی کلیدی باشد»	۰/۶۵۴			
	دانشجویان را آگاه می‌سازم به ماهیت نسبی و غیرقطعی علم و نظام‌های دانش	۰/۶۳۴			
	راهنمایی‌شان می‌کنم از مطلق‌گرایی هر رشته علمی خاص اجتناب کنند	۰/۴۷۶			
	با تاریخچه رشته علمی آشنا می‌سازم	۰/۵۳۴			
با موضوعات مورد اختلاف در رشته‌های علمی آشنا می‌سازم	۰/۴۷۶				
تکالیف فکروانه	دانشجویان اجراها و کارکردهای هنری، سیاسی یا علمی را نقد کنند	۰/۵۴۳	۰/۶۵۴	۰/۷۶۵	۰/۷۸۹
	دانشجویان بر آفرینش یک محصول، اجرای یک پروژه یا فعالیت توسط خودشان تأکید کنند	۰/۸۷۶			
	دانشجویان به کمک خودشان تکالیف انجام شده را خوانده و نقد می‌کنند	۰/۶۵۴			
	به خاطر اشتباهات دستوری یا روش شناختی آنها را برمی‌گردانم	۰/۵۴۳			
	اگر تکالیف‌شان کسل‌کننده، نامرتب یا عاری از خلاقیت باشند، آنها را برمی‌گردانم	۰/۵۹۸			
	پرسش‌های مطرح شده در کلاس توسط دانشجویان مورد ارزیابی قرار گیرد	۰/۴۹۰			
	نتیجه ارزیابی باید به شکل عملکرد یا خلق یک اثر باشد	۰/۵۲۳			
	ارزیابی باید چالش برانگیز باشد	۰/۵۶۷			

سازه	گویه	بارعاملی	میانگین واریانس (AVE)	پایایی ترکیبی	آلفای کرونباخ
الگوها و راه‌ها	دانشجویان به‌عنوان الگو به افراد موفق نیاز دارند	۰/۶۴۵	۰/۶۳۶	۰/۷۶۵	۰/۶۵۴
	به فرصتی برای تقلید و سرمشق‌گیری از الگوهای موفق نیاز دارند	۰/۳۴۵۶			
	باید سطح مناسبی از موفقیت را نشان دهند	۰/۵۴۳			
	باید محصولاتی تولید کنند که از فعالیت‌های یادگیری قبلی آنان الگوبرداری شده است	۰/۵۴۳			
	باید براساس دسترسی به الگوها و بازخورد در یادگیری و ارزیابی با آنان برخورد شود	۰/۴۳۲			
استانداردهای شناخته‌شده	باید دانشجویان از تولیدات خود (مقاله یا گزارش آزمایشگاهی) دفاع کنند	۰/۵۴۳	۰/۶۵۴	۰/۶۵۳	۰/۷۴۳
	باید به انتقادات من و هم‌کلاسی‌هایشان از تکالیف‌شان پاسخ دهند	۰/۴۲۳			
	هر نوع نتایج و محصولاتی که تولید می‌کنند، باید سطح مناسبی از موفقیت را نشان دهد	۰/۵۴۳			
	دانشجویان باید پروژه‌ها را از همان ابتدا مستندسازی و به نام خود ثبت کنند	۰/۴۳۲			
	باید صداقت علمی خود را در انجام تکالیف و پروژه‌ها نشان می‌دهد	۰/۸۷۶			
	آزمون‌هایی که من از دانشجویان می‌گیرم، بر تولید اثر و معیارهای شناخته‌شده با کیفیت بالا متمرکز است	۰/۷۶۵			
	ارزیابی از عملکرد دانشجویان باید معیاری باشد که در عمل مورد نیاز است	۰/۴۸۷			
تجربه یادگیری	دانشجویان باید آنچه را که یاد گرفته‌اند توجیه کنند	۰/۵۴۳	۰/۶۵۴	۰/۵۴۳	۰/۴۷۶
	فقط دیدگاه‌های متداول را که آموخته‌اند بازگو کنند	۰/۴۶۵			
	فنونی را که آموخته‌اند بدون تفکر و خارج از بافت به کار گیرند	۰/۵۳۴			
	آزمون‌هایی که من از دانشجویان می‌گیرم، شامل بیان دوباره و به‌خاطر سپردن دانش است	۰/۸۶۲			
	ارزیابی باید فرصت دریافت بازخورد، اصلاح عملکرد و تولیدات را فراهم کند	۰/۴۷۶			
	آزمون‌هایی که من از دانشجویان می‌گیرم، باید سطح بالای تفکر و تلفیق مهارت‌های بنیادین را ارزیابی کند	۰/۵۴۳			

سازه	گویه	بارعاملی	میانگین واریانس (AVE)	پایایی ترکیبی	آلفای کرونباخ
توانایی حل مسئله	من این امکان را برای دانشجویان فراهم می‌کنم که از روش‌های جایگزین برای انجام تکالیف خود استفاده کنند	۰/۷۶۵	۰/۶۵۳	۰/۵۴۳	۰/۵۲۳
	دانشجویان باید بتوانند مسائل و مشکلات دنیای واقعی را در تکالیف خود شبیه‌سازی و تحلیل کنند	۰/۴۶۵			
	ارزشیابی در کلاس‌های من محدود به پاسخ‌های صحیح نیست، بلکه به میزان تحلیل و استدلال دانشجویان بستگی دارد	۰/۵۶۷			
	از دانشجویان انتظار دارم که در آزمون‌ها نه فقط به خاطر سپاری پاسخ‌ها، بلکه به شیوه‌ای انتقادی و تحلیلی پاسخ دهند.	۰/۵۴۳			

روایی همگرا زمانی وجود دارد که پایایی ترکیبی از $0/5$ و AVE از $0/5$ بزرگ‌تر باشند. همچنین پایایی ترکیبی باید از AVE بزرگ‌تر باشد. در این صورت شرط روایی همگرا وجود خواهد داشت. با توجه به جدول (۳) هر سه شرط فوق برقرار بوده، بنابراین پرسشنامه از روایی همگرا برخوردار است. چنانچه جذر میانگین واریانس استخراج شده (\sqrt{AVE}) برای یک متغیر بزرگ‌تر از همبستگی آن متغیر با سایر متغیرها باشد، روایی تشخیصی برای آن متغیر وجود دارد. در جدول زیر اعداد روی قطر اصلی جذر میانگین واریانس استخراج شده می‌باشند. همان‌طور که از داده‌های جدول (۴) مشخص است، جذر میانگین واریانس استخراج شده برای هر متغیر بیشتر از همبستگی آن متغیر با سایر متغیرها است. بنابراین روایی تشخیصی پرسشنامه تحقیق مورد تأیید است.

ردیف	شاخص	۱	۲	۳	۴	۵	۶	۷	۸
۱	معیارهای شفاف	۰/۷۳۴							
۲	خودارزیابی	۰/۷۰۹	۰/۶۷۵						
۳	یادگیری اصیل	۰/۶۹۹	۰/۶۰۱	۰/۶۹۸					
۴	تکالیف فکورانه	۰/۶۴۵	۰/۶۰۴	۰/۶۳۳	۰/۷۴۵				
۵	الگودهی	۰/۶۳۲	۰/۶۰۷	۰/۶۳۲	۰/۶۵۴	۰/۷۶۴			
۶	استانداردهای شناخته‌شده	۰/۵۹۹	۰/۵۶۷	۰/۵۳۴	۰/۵۴۳	۰/۶۵۴	۰/۸۹۰		
۷	توجه یادگیری	۰/۶۹۶	۰/۵۳۲	۰/۴۹۸	۰/۶۵۴	۰/۵۸۷	۰/۶۷۸	۰/۸۵۶	
۸	تکالیف جایگزین	۰/۵۳۵	۰/۵۴۲	۰/۴۲۹	۰/۵۴۳	۰/۶۴۳	۰/۶۴۳	۰/۶۵۴	۰/۸۶۵

برازش مدل ساختاری: نخستین شاخص برای بررسی معناداری مدل، ضریب معناداری Z یا همان مقادیر T است. برای تأیید برازش مدل ساختاری، این ضرایب باید بیش از ۱/۹۶ باشند تا بتوان نتیجه گرفت که مدل با سطح اطمینان ۹۵ درصد معنادار است. همان‌طور که در شکل (۱) دیده می‌شود، مدل در وضعیت معناداری ضرایب (T-Values) قرار دارد. از آنجا که مقدار آماره T برای متغیر مورد نظر بیشتر از ۱/۹۶ محاسبه شده است، می‌توان نتیجه گرفت که سؤالات و روابط بین متغیرها در سطح اطمینان ۹۵ درصد معنادار هستند.

شکل (۱) مدل در حالت ضرایب معناداری (T-value)

الف) معیار R^2 : معیار R^2 مربوط به متغیرهای پنهان درون‌زای (وابسته) مدل است. R^2 معیاری است که مصداق تأثیر یک متغیر برون‌زا بر یک متغیر درون‌زا دارد و سه مقدار ۰/۱۹، ۰/۳۳ و ۰/۶۷ به‌عنوان مقدار ملاک برای مقادیر ضعیف، متوسط و قوی R^2 در نظر گرفته می‌شود. مقدار R^2 در جدول زیر تقریباً مناسب بودن برازش مدل ساختاری را تأیید می‌کند.

جدول (۵) مقادیر R^2 متغیرهای تحقیق

ردیف	متغیر	R^2
۱	معیارهای شفاف	۰/۶۹۷
۲	خودارزیابی	۰/۷۵۴
۳	یادگیری اصیل	۰/۶۴۳
۴	تکالیف فکورانه	۰/۶۷۸
۵	الگودهی	۰/۶۹۹
۶	استانداردهای شناخته‌شده	۰/۷۶۵
۷	توجیه یادگیری	۰/۶۵۴
۸	تکالیف جایگزین	۰/۶۵۴

ب) معیار Q^2 : معیار Q^2 قدرت پیش‌بینی مدل را مشخص می‌کند و در صورتی که مقدار آن در مورد یک سازه درون‌زا سه مقدار ۰/۰۲، ۰/۱۵ و ۰/۳۵ را کسب نماید، به ترتیب نشان از قدرت پیش‌بینی ضعیف، متوسط و قوی سازه یا سازه‌های برون‌زای مربوط به آن را دارد. مقدار Q^2 در جدول زیر نشان از قدرت پیش‌بینی مناسب مدل در خصوص سازه‌های درون‌زای پژوهش دارد و برازش مناسب مدل ساختاری را تأیید می‌کند.

جدول (۶) مقادیر Q^2 متغیرهای تحقیق

ردیف	متغیر	Q^2
۱	معیارهای شفاف	۰/۴۳۲
۲	خودارزیابی	۰/۴۵۴
۳	یادگیری اصیل	۰/۳۹۰
۴	تکالیف فکورانه	۰/۴۱۲
۵	الگودهی	۰/۳۷۶
۶	استانداردهای شناخته‌شده	۰/۴۹۹
۷	توجیه یادگیری	۰/۲۹۸
۸	تکالیف جایگزین	۰/۳۴۵

پ) معیار GOF: یکی دیگر از شاخص‌های برازش که توسط تنن‌هاوس و همکاران معرفی شده است، معیار کلی برازش است که از طریق میانگین هندسی میانگین اشتراک محاسبه می‌شود. این شاخص عملکردی مشابه شاخص‌های برازش مدل در نرم‌افزار لیزرل دارد و مقدار آن بین صفر تا یک متغیر است، به طوری که مقادیر نزدیک به یک نشان‌دهنده کیفیت مناسب مدل هستند. با این حال، برخلاف شاخص‌های مبتنی بر خی دو در مدل‌های لیزرل که میزان برازش مدل نظری با داده‌های گردآوری شده را ارزیابی می‌کنند، شاخص GOF بر توانایی پیش‌بینی کلی مدل تمرکز دارد. این شاخص بررسی می‌کند که آیا مدل آزمون‌شده در پیش‌بینی متغیرهای مکنون درون‌زا عملکرد موفقی داشته است یا خیر.

جدول (۷) نتایج برازش کلی مدل با معیار GOF

\bar{R}^2	<i>Communalities</i>	$GOF = \sqrt{\text{Communalities} \times \bar{R}^2}$
۰/۶۴۳	۰/۳۶۵	۰/۵۰۲

همانطور که در جدول (۷) مشاهده می‌شود، مقدار میانگین مقادیر اشتراکی مقدار ۰/۳۶۲ و میانگین مقادیر R^2 برابر ۰/۶۴۳ به دست آمده است و با توجه به فرمول، مقدار معیار GOF معادل ۰/۵۰۲ به دست آمد که بزرگ‌تر از مقدار ملاک ۰/۳ بوده و نشان از توان مناسب مدل در پیش‌بینی متغیر مکنون درون‌زای مدل دارد. تفاوت میان نظرات استادان در مورد سنجش اصیل بر حسب متغیرهای جمعیت شناختی (دانشکده و مرتبه علمی): برای مقایسه مؤلفه‌های سنجش اصیل بر حسب دانشکده، از آزمون تحلیل واریانس یک طرفه استفاده شده است. نتایج مربوط به این آزمون در جدول (۸) داده شده است.

جدول (۸) نتایج آزمون تحلیل واریانس یک طرفه برای مقایسه مؤلفه‌های سنجش اصیل بر حسب دانشکده

متغیر	منبع	مجموع مربعات	میانگین مربعات	F مقدار آماره	سطح معناداری
معیارهای شفاف	بین گروهی	۲۵۲/۴۲۴	۲۸/۰۴۷	۲۹/۴۹۰	۰/۰۷۱
	درون گروهی	۳۷۰۷/۳۲۶	۰/۹۵۱		
	جمع کل	۳۹۵۹/۷۵۰			
خودارزیابی	بین گروهی	۱۴۲/۱۵۱	۱۵/۷۹۵	۲۰/۸۸۹	۰/۰۰۳
	درون گروهی	۲۹۴۷/۳۹۸	۰/۷۵۶		
	جمع کل	۳۰۸۹/۵۴۹			
یادگیری اصیل	بین گروهی	۱۱۰/۸۱۵	۱۲/۳۱۳	۱۶/۵۴۵	۰/۰۰۱
	درون گروهی	۲۹۰۰/۹۶۹	۰/۷۴۴		
	جمع کل	۳۰۱۱/۷۸۵			
تکالیف فکورانه	بین گروهی	۲۱۷/۳۰۷	۲۴/۱۴۵	۲۷/۴۷۴	۰/۰۰۱
	درون گروهی	۳۴۲۳/۹۹۵	۰/۸۷۹		
	جمع کل	۳۶۴۱/۳۰۲			
الگودهی	بین گروهی	۱۵۰/۴۷۳	۱۶/۷۱۹	۲۲/۱۴۹	۰/۰۶۱
	درون گروهی	۲۸۱۳/۱۰۷	۰/۷۲۲		
	جمع کل	۲۹۶۳/۵۸۰			
استانداردهای شناخته شده	بین گروهی	۱۳۲۴/۲۸۷۶	۱۵/۲۳۴	۲۱/۳۰۹	۰/۰۰۱
	درون گروهی	۱۶۷/۴۳۲۱	۰/۶۵۴		
	جمع کل	۱۴۹۱/۷۱۹۷			
توجه یادگیری	بین گروهی	۲۳۳۳/۱۶۳۲	۱۵/۵۴۳	۱۹/۴۷۶	۰/۰۰۱
	درون گروهی	۶۷۸/۳۲۱	۰/۶۷۸		
	جمع کل	۴۸۴۲/۳۰۱۱			
تکالیف جایگزین	بین گروهی	۵۴۳/۱۴۵۶	۱۴/۲۵	۲۳/۳۰۹	۰/۰۸۷
	درون گروهی	۳۴۵/۲۴۶	۰/۹۷۶		
	جمع کل	۸۸۸/۷۰۲			

نتایج تحلیل‌های آماری نشان داد که میزان استفاده و بهره‌گیری استادان دانشکده‌های مختلف از مؤلفه‌های سنجش اصیل تفاوت معناداری دارد. برای بررسی این تفاوت‌ها، از آزمون تحلیل واریانس یک طرفه استفاده شد. یافته‌ها نشان داد که استادان برخی دانشکده‌ها، به‌ویژه دانشکده‌های علوم پایه، کشاورزی، منابع طبیعی و مهندسی، به مؤلفه‌های خودارزیابی، یادگیری اصیل، تکالیف فکورانه،

استانداردهای شناخته‌شده و توجیه یادگیری توجه بیشتری دارند. به‌طور خاص، میزان استفاده از خودارزیابی در میان دانشکده‌ها تفاوت معناداری داشت ($p=0.003$)، به این معنا که استادان برخی دانشکده‌ها اهمیت بیشتری به ارزیابی دانشجویان از عملکرد خودشان می‌دهند. یادگیری اصیل نیز با مقدار $p=0.001$ نشان داد که در برخی دانشکده‌ها، توجه بیشتری به ارتباط یادگیری با دنیای واقعی وجود دارد. در مورد تکالیف فکورانه ($p=0.001$)، مشخص شد که برخی دانشکده‌ها بر تفکر انتقادی و تحلیل‌گرایانه دانشجویان در تکالیف خود تأکید بیشتری دارند. همچنین، در مؤلفه استانداردهای شناخته‌شده ($p=0.001$) تفاوت معناداری مشاهده شد، که بیانگر میزان متفاوت استفاده از معیارهای معتبر و پذیرفته‌شده در فرایند سنجش دانشجویان در دانشکده‌های مختلف است. در نهایت، مؤلفه توجیه یادگیری ($p=0.001$) نیز نشان داد که میزان انتظار از دانشجویان برای تبیین و توجیه آموخته‌هایشان در برخی دانشکده‌ها بیشتر است. در این میان، استادان دانشکده زبان و ادبیات کمترین میزان بهره‌گیری از این مؤلفه‌ها را نشان دادند، که نشان‌دهنده اهمیت کمتر این دانشکده به برخی اصول سنجش اصیل نسبت به سایر دانشکده‌ها است. این تفاوت‌ها می‌تواند ناشی از ماهیت رشته‌های مختلف، نیازهای آموزشی خاص آنها و رویکردهای متفاوت استادان در ارزیابی دانشجویان باشد.

جدول (۹) آمار توصیفی (آزمون تعقیبی) مربوط به مؤلفه‌های سنجش اصیل بر حسب دانشکده

نام دانشکده	خودارزیابی		یادگیری اصیل		تکالیف فکورانه		استانداردهای شناخته‌شده		توجیه یادگیری	
	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار
علوم انسانی	۳/۷۶	۰/۷۵۸	۲/۴۲	۰/۵۶۷	۳/۳۵	۱/۰۳۴	۳/۳۴	۱/۶۲۳	۲/۷۶	۱/۲۳۴
زبان و ادبیات	۳/۰۱	۰/۷۶۷	۱/۳۵	۱/۹۷۳	۳/۲۹	۱/۰۳۷	۳/۰۳	۱/۰۸۶	۱/۰۱	۱/۳۱۷
علوم پایه	۴/۲۱	۱/۰۷۷	۴/۸۰	۱/۷۵۱	۳/۰۱	۰/۹۴۶	۴/۳۲	۱/۳۲۱	۳/۲۱	۱/۹۴۶
کشاورزی	۳/۹۹	۱/۰۴۴	۳/۵۹	۱/۷۷۰	۳/۱۶	۱/۰۱۵	۳/۹۸	۱/۴۳۲	۳/۹۹	۱/۴۱۵
منابع طبیعی	۴/۰۱	۰/۸۳۹	۴/۶۷	۰/۵۴۳	۳/۲۰	۰/۸۸۹	۳/۶۵	۱/۱۹۸	۳/۰۱	۰/۵۹۸
مهندسی	۴/۵۹	۰/۸۱۲	۴/۰۱	۰/۷۷۲	۳/۶۲	۰/۹۵۶	۴/۶۳	۱/۵۴۳	۴/۳۲	۱/۳۲۱

بر اساس تحلیل‌های توصیفی، استادان دانشکده زبان و ادبیات، کمترین میزان استفاده و بهره‌گیری از مؤلفه‌های خودارزیابی، یادگیری اصیل و استانداردهای شناخته‌شده را نشان داده‌اند. در مقابل، استادان دانشکده‌های علوم پایه، کشاورزی، منابع طبیعی و مهندسی به میزان بیشتری از این مؤلفه‌ها در فرایند سنجش و ارزیابی خود بهره می‌برند. به‌عنوان نمونه، دانشکده مهندسی بالاترین میانگین را در تمامی این مؤلفه‌ها، به‌ویژه در خودارزیابی (۴.۵۹)، یادگیری اصیل (۴.۰۱) و استانداردهای شناخته‌شده (۴.۶۳) داشته است. در حالی که استادان دانشکده زبان و ادبیات کمترین میانگین را در این زمینه‌ها کسب کرده‌اند. این نتایج نشان می‌دهد که میزان توجه به اصول سنجش اصیل در دانشکده زبان و ادبیات نسبت به سایر دانشکده‌ها کمتر بوده و این تفاوت‌ها می‌تواند ناشی از ماهیت رشته‌های مختلف و نیازهای آموزشی خاص آنها باشد.

همچنین برای مقایسه مؤلفه‌های سنجش اصیل بر حسب مرتبه علمی، از آزمون تحلیل واریانس یک‌طرفه استفاده شده است. نتایج مربوط به این آزمون در جدول (۱۰) ارائه شده است.

جدول (۱۰) نتایج آزمون تحلیل واریانس یک‌طرفه برای مقایسه مؤلفه‌های سنجش اصیل بر حسب مرتبه علمی

متغیر	منبع	مجموع مربعات	میانگین مربعات	F مقدار آماره	سطح معناداری
معیارهای شفاف	بین گروهی	۲۱۶۵/۳۱۲	۱۹/۰۳۴	۱۹/۵۲۳	۰/۳۷۱
	درون گروهی	۱۴۳۲/۷۶۰	۰/۷۳۴		
	جمع کل	۳۵۹۸/۰۷۲			
خودارزیابی	بین گروهی	۱۹۸۷/۱۰۹	۱۶/۴۳۲	۲۳/۶۵۴	۰/۰۰۳
	درون گروهی	۱۷۶۵/۰۰۱	۰/۷۵۶		
	جمع کل	۳۷۵۲/۱۱			
یادگیری اصیل	بین گروهی	۳۱۲۳/۱۰۴	۱۰/۳۲۱	۳۲/۵۴۳	۰/۶۲۰
	درون گروهی	۱۷۶۵/۴۷۶	۰/۵۴۳		
	جمع کل	۴۸۸۸/۵۸			
تکالیف فکورانه	بین گروهی	۳۱۵۷/۲۱۳	۱۳/۳۴۵	۲۲/۳۲۱	۰/۰۰۹
	درون گروهی	۳۲۱۵/۲۵۴	۰/۶۹۸		
	جمع کل	۶۳۷۲/۴۶۷			

متغیر	منبع	مجموع مربعات	میانگین مربعات	F مقدار آماره	سطح معناداری
الگودهی	بین گروهی	۱۹۸/۵۴۳	۱۴/۴۲۱	۱۹/۴۳۲	۰/۰۳۱
	درون گروهی	۲۸۱۳/۳۱۲	۱/۵۴۳		
	جمع کل	۳۰۱۱/۸۵۵			
استانداردهای شناخته‌شده	بین گروهی	۳۵۴۳/۱۵۴۳	۱۴/۳۷۶	۲۷/۴۳۲	۰/۰۲۱
	درون گروهی	۱۴۳۲/۶۵۴	۰/۹۶۷		
	جمع کل	۴۹۷۵/۸۰۸۳			
توجیه یادگیری	بین گروهی	۱۸۷۵/۳۷۶۵	۱۴/۵۴۳	۲۵/۴۳۲	۰/۰۶۵۱
	درون گروهی	۶۷۸/۵۴۳	۰/۵۴۳		
	جمع کل	۲۵۵۳/۹۱۹۵			
تکالیف جایگزین	بین گروهی	۲۷۶۵/۳۴۸	۱۴/۵۹۸	۲۵/۲۰۱	۰/۵۸۷
	درون گروهی	۱۲۳۴/۲۴۶	۰/۵۴۳		
	جمع کل	۲۷۶۵۱/۵۸۲			

تحلیل داده‌ها نشان داد که در برخی مؤلفه‌ها، تفاوت معناداری میان گروه‌های مورد بررسی وجود دارد. به‌عنوان مثال، مؤلفه خودارزیابی ($p=0.003$) نشان داد که میزان اهمیت و استفاده از این روش در میان اعضای هیئت علمی با مرتبه‌های علمی مختلف تفاوت قابل توجهی دارد. همچنین، در مؤلفه تکالیف فکورانه ($p=0.009$)، نتایج بیانگر تفاوت معنادار میان گروه‌های مورد مطالعه بود، به این معنا که برخی اعضای هیئت علمی نسبت به سایرین بر تفکر انتقادی و تحلیل‌گرایانه در تکالیف دانشجویان تأکید بیشتری داشتند. در مؤلفه الگودهی ($p=0.031$)، مشخص شد که تفاوتی قابل توجه بین گروه‌ها وجود دارد، که نشان می‌دهد برخی استادان در نقش الگو برای دانشجویان خود مؤثرتر عمل می‌کنند. همچنین، مؤلفه استانداردهای شناخته‌شده ($p=0.021$) دارای تفاوت معناداری بین مرتبه‌های علمی مختلف بود، که می‌تواند به رویکردهای متفاوت استادان در تأکید بر معیارهای علمی پذیرفته‌شده مرتبط باشد.

اما در برخی مؤلفه‌ها مانند معیارهای شفاف ($p=0.371$)، یادگیری اصیل ($p=0.620$)، توجیه یادگیری ($p=0.651$) و تکالیف جایگزین ($p=0.587$)، تفاوت معناداری میان گروه‌های مورد بررسی مشاهده نشد، که نشان می‌دهد این مؤلفه‌ها به صورت یکنواخت‌تری در میان استادان با مرتبه‌های علمی مختلف مورد استفاده قرار گرفته‌اند.

جدول (۹) آمار توصیفی مربوط به مؤلفه‌های سنجش اصیل بر حسب مرتبه علمی

مرتبه علمی	خودارزیابی		تکالیف فکورانه		الگودهی		استانداردهای شناخته‌شده
	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	
استادیار	۲/۳۳	۱/۴۷۶	۳/۹۹	۱/۴۳۲	۴/۴۱	۱/۰۹۴	۴/۳۴
دانشیار	۳/۱۱	۰/۶۵۴	۳/۳۵	۰/۴۳۲	۳/۹۹	۱/۰۰۷	۳/۰۳
استاد	۴/۰۶	۱/۰۹۷	۴/۸۰	۱/۲۱۳	۳/۰۱	۲/۱۴۶	۲/۶۵

بررسی میانگین و انحراف معیار مؤلفه‌های مختلف بر اساس مرتبه علمی نشان می‌دهد که در مؤلفه خودارزیابی، استادان با مرتبه علمی بالاتر (استادان) بیشترین میزان استفاده را داشته‌اند (۴.۰۶)، در حالی که استادیاران کمترین مقدار را گزارش کرده‌اند (۲.۳۳). این یافته نشان می‌دهد که استادان با سابقه، اهمیت بیشتری به خودارزیابی دانشجویان می‌دهند. در مؤلفه تکالیف فکورانه نیز استادان با میانگین ۴.۸۰ در مقایسه با دانشیاران (۳.۳۵) و استادیاران (۳.۹۹) عملکرد بهتری داشته‌اند، که نشان می‌دهد آنها بیشتر بر تفکر انتقادی و تحلیل گرایانه در تکالیف دانشجویان تأکید دارند. در مقابل، مؤلفه الگودهی در میان استادیاران بالاترین مقدار (۴.۴۱) را نشان می‌دهد، در حالی که استادان کمترین میانگین (۳.۰۱) را داشته‌اند. این تفاوت می‌تواند بیانگر آن باشد که استادان جوان‌تر تلاش بیشتری برای ایفای نقش الگو در کلاس‌های خود دارند. در مؤلفه استانداردهای شناخته‌شده، استادیاران بیشترین میزان استفاده را گزارش کرده‌اند (۴.۳۴)، اما با افزایش مرتبه علمی، این مقدار کاهش یافته و در میان استادان به ۲.۶۵ رسیده است. این نتیجه می‌تواند نشان‌دهنده کاهش تأکید استادان با مرتبه علمی بالاتر بر معیارهای سخت‌گیرانه ارزیابی باشد.

بحث

هدف این پژوهش، بررسی میزان بهره‌گیری اعضای هیئت علمی از شاخصه‌های سنجش اصیل در فرایند یاددهی-یادگیری، با تأکید بر تعیین پایایی و روایی سازه ابزار سنجش مورد استفاده است. با این حال، سنجش اصیل به‌عنوان رویکردی که بر ارزیابی واقعی، معنادار و مبتنی بر عملکرد دانشجویان تأکید دارد، نیازمند بررسی دقیق میزان پذیرش و اجرای آن در محیط‌های دانشگاهی است. از این رو، این پژوهش تلاش می‌کند تا میزان آگاهی، نگرش و میزان به‌کارگیری این رویکرد توسط اعضای هیئت علمی را شناسایی کرده و از طریق بررسی ویژگی‌های روان‌سنجی ابزار مورد استفاده، به ارائه داده‌های معتبر در این زمینه بپردازد.

نتایج پرسش اول پژوهش نشان داد که پرسشنامه میزان بهره‌گیری استادان از سنجش اصیل از پایایی و روایی بالایی برخوردار است و می‌تواند به‌عنوان ابزاری معتبر در بررسی شیوه‌های ارزیابی استادان مورد استفاده قرار گیرد. یافته‌ها تأیید کردند که این ابزار دارای همسانی درونی قوی است، به‌گونه‌ای که ضریب آلفای کرونباخ و پایایی ترکیبی آن بالاتر از ۰.۵ بوده و از ساختار عاملی تأییدشده‌ای برخوردار است. این نتایج اهمیت ویژه‌ای دارد، زیرا سنجش اصیل به‌عنوان یک روش ارزیابی نوین، نقش مؤثری در بهبود تفکر انتقادی، حل مسئله و کاربرد دانش در موقعیت‌های واقعی دارد (ویگینز، ۱۹۹۸؛ بلک و ویلیام، ۲۰۱۸).

چنانکه یافته‌های این پژوهش با مطالعات پیشین درباره اهمیت نقش استادان در فرایند سنجش و یادگیری دانشجویان همخوانی دارد. پژوهش‌های متعدد نشان داده‌اند که استادان تأثیر مستقیمی بر تجربه یادگیری دانشجویان دارند و نحوه ارزیابی آنها می‌تواند نقش کلیدی در بهبود فرایند یادگیری ایفا کند (بانتا و پالومبا، ۲۰۱۴؛ فلنچر و همکاران، ۲۰۱۲). با این حال، بیشتر مطالعات قبلی بر ارزیابی تدریس از دیدگاه دانشجویان متمرکز بوده‌اند و کمتر به نگرش استادان درباره روش‌های ارزیابی پرداخته‌اند. این پژوهش این خلأ را پوشش داده و ابزاری معتبر برای سنجش نگرش و عملکرد استادان در زمینه سنجش اصیل ارائه کرده است. علاوه بر این، یافته‌ها تأیید می‌کنند که خودارزیابی و شفافیت معیارهای سنجش نقش مهمی در بهبود یادگیری دانشجویان دارد، موضوعی که در نظریات گوسکی (۱۹۸۵) درباره بازخورد یادگیری نیز مورد تأکید قرار گرفته است. از این رو، تأیید روایی این پرسشنامه نشان می‌دهد که مؤسسات آموزش عالی می‌توانند از این ابزار برای اصلاح روش‌های ارزیابی، افزایش آگاهی استادان نسبت به اصول سنجش اصیل و تطبیق آن با استانداردهای آموزشی بهره ببرند.

در ارتباط با پرسش دوم پژوهش، نتایج نشان داد که نظرات استادان درباره سنجش اصیل بر اساس متغیرهای جمعیت‌شناختی (مانند دانشکده و مرتبه علمی) دارای تفاوت‌های معناداری است. استادان بر حسب دانشکده، دیدگاه‌های متفاوتی نسبت به روش‌های سنجش اصیل دارند؛ به‌طور خاص، استادان دانشکده‌های علوم پایه، کشاورزی و مهندسی بیشترین میزان استفاده از سنجش اصیل را گزارش کرده‌اند، در حالی که استادان دانشکده‌های علوم انسانی و زبان و ادبیات کمترین بهره‌گیری را نشان داده‌اند. این یافته‌ها بیانگر تفاوت‌های بنیادین در ماهیت رشته‌های تحصیلی و رویکردهای ارزیابی در آنها است؛ رشته‌های فنی و تجربی اغلب به ارزیابی‌های عملی، پروژه‌محور و مبتنی بر حل مسئله متکی هستند، در حالی که در رشته‌های علوم انسانی، روش‌های ارزیابی سنتی‌تر مانند امتحانات کتبی و مقالات بیشتر رایج است. یافته‌های این پژوهش در این زمینه با مطالعات پیشین

همچون پژوهش‌های کوه و همکاران (۲۰۱۹) و ویگینز (۱۹۹۸) همخوانی دارد که تأکید کرده‌اند ماهیت رشته‌های تحصیلی و الزامات آموزشی آن‌ها تأثیر مستقیمی بر نوع ارزیابی مورد استفاده دارد. در زمینه مرتبه علمی، یافته‌ها نشان داد که استادان با مرتبه علمی بالاتر بیشترین میزان بهره‌گیری از خودارزیابی و تکالیف فکورانه را داشته‌اند، در حالی که استادیاران و دانشیاران بیشتر بر الگودهی و استانداردهای شناخته‌شده تأکید داشتند. این یافته‌ها بیانگر این است که استادان باتجربه‌تر بر پرورش تفکر انتقادی دانشجویان تمرکز بیشتری دارند، در حالی که استادان جوان‌تر به نقش خود به‌عنوان الگو و ارائه استانداردهای مشخص در سنجش توجه بیشتری دارند. یافته‌ها در این حوزه نیز با مطالعات گوسکی (۱۹۸۵) و ساسکی (۲۰۰۰) سازگار است که نشان داده‌اند استادان باتجربه‌تر گرایش بیشتری به استفاده از شیوه‌های یادگیری عمیق مانند خودارزیابی و تفکر انتقادی دارند، در حالی که استادان کم‌تجربه‌تر بیشتر به استانداردهای شناخته‌شده و الگودهی توجه می‌کنند. این موضوع می‌تواند بیانگر تغییر نگرش استادان در طول مسیر حرفه‌ای‌شان و تأثیر تجربه آموزشی بر نوع ارزیابی مورد استفاده باشد.

نتیجه‌گیری

یافته‌های این پژوهش پیامدهای مهمی برای نظام ارزیابی در آموزش عالی دارد. نتایج بر ضرورت توانمندسازی استادان در زمینه سنجش اصیل و ضرورت تطبیق راهبردهای ارزیابی با ویژگی‌های رشته‌های مختلف دانشگاهی تأکید می‌کند. همچنین، اعتبارسنجی پرسشنامه «سنجش اصیل» ابزاری معتبر و بومی شده برای سنجش و بهبود میزان بهره‌گیری استادان از این رویکرد در اختیار دانشگاه‌ها قرار می‌دهد؛ ابزاری که می‌تواند در نهایت به ارتقای کیفیت تدریس و بهبود یادگیری دانشجویان منجر شود. بدین ترتیب، این پژوهش نه تنها شکاف موجود در نبود ابزار معتبر را پر می‌کند، بلکه زمینه‌ساز توسعه سیاست‌ها و برنامه‌های آموزشی مبتنی بر رویکردهای نوین ارزشیابی در دانشگاه‌هاست.

با این حال، این مطالعه محدودیت‌هایی نیز دارد. نخست، ماهیت خودگزارشی پرسشنامه می‌تواند به سوگیری پاسخ‌ها منجر شود، زیرا برخی استادان ممکن است میزان استفاده خود از سنجش اصیل را بیش از حد واقعی گزارش کنند؛ دوم، عدم همکاری کامل برخی از اعضای هیئت علمی به‌ویژه در دانشکده‌های علوم پایه و منابع طبیعی، دامنه داده‌های گردآوری شده را محدود ساخت؛ سوم، تمرکز پژوهش بر یک محیط آموزشی خاص (دانشگاه کردستان) باعث می‌شود قابلیت تعمیم نتایج به سایر دانشگاه‌ها و مؤسسات آموزشی با زمینه‌های فرهنگی و سازمانی متفاوت کاهش یابد. برای رفع این محدودیت‌ها، پیشنهاد می‌شود در پژوهش‌های آتی از طراحی‌های طولی استفاده شود تا تغییر نگرش

استادان نسبت به سنجش اصیل در گذر زمان آشکار شود. همچنین، استفاده از روش‌های کیفی همچون مشاهده کلاسی و مصاحبه عمیق می‌تواند درک غنی‌تری از نحوه به‌کارگیری سنجش اصیل و تأثیر آن بر فرایند یاددهی-یادگیری به دست دهد.

تقدیر و تشکر

بدین‌وسیله از تمامی اعضای هیئت علمی گرامی دانشگاه کردستان که در فرایند گردآوری داده‌ها ما را یاری کردند، صمیمانه قدردانی و سپاسگزاری می‌کنیم.

References

- Abdollahi Gol, T., Karami, M., & Karshki, H. (2016, April 14). Authentic assessment: A strategy for improving the quality of higher education system. *National Congress on Higher Education in Iran*. <https://profdoc.um.ac.ir/paper-abstract-1056447.html> [In Persian]
- Ahmadi, H., Shirbagi, N., & Shirbagi, S. (2023). Dark va mizan-e bahregiri-ye mo'alleman az "Sanjesh-e Asil" dar farayand-e yad-dehi-yad-giri [The concept and extent of teachers' use of "Authentic Assessment" in the teaching-learning process]. *Tadris Pazhuhi (Teaching Research)*, 11(4), 170–197. [In Persian]
- Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, 40(1), 133–148.
- Azizi, S., & Salimi, J. (2023). A study of undergraduate curricula in basic sciences, engineering, and social sciences based on the integrated curriculum. *Curriculum Planning Studies*, 12(24), 148–174. [In Persian]
- Banta, T. W., & Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education* (2nd ed.). Jossey-Bass.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. ASCD.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723–733.
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495–518.

- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Dery, L. (2022). Forming a ranking from tied evaluations: A case of an online, interactive student peer assessment system. *arXiv*. <https://arxiv.org/abs/2212.13774>
- Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., Brown, S., & Chen, J. (2012). Faculty and students’ conceptions of assessment in higher education. *Higher Education, 64*, 119–133. <https://doi.org/10.1007/s10734-011-9484-1>
- Ghosh, S., Brooks, B., Ranmuthugala, D., & Bowles, M. (2020). Authentic versus traditional assessment: An empirical study investigating the difference in seafarer students’ academic achievement. *The Journal of Navigation, 73*(4), 797–812.
- Guskey, T. R. (1985). *Implementing mastery learning*. Wadsworth.
- Heil, J., & Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning, 27*(1), 187–218.
- Hwang, G., & Chang, H. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*, 1023–1031. <https://doi.org/10.1016/j.compedu.2010.12.002>
- Koh, K., Delanoy, N., Thomas, C., Bene, R., Chapman, O., Turner, J., ... & Hone, G. (2019). The role of authentic assessment tasks in problem-based learning. *Papers on Postsecondary Learning & Teaching, 3*, 17–24.
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students’ evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*(1), 202–228.
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher, 45*(1), 13-31.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.
- Scully, D. (2019). *Constructing multiple-choice items to measure higher-order thinking*. *Practical Assessment, Research & Evaluation, 24*(1), 1–13. <https://doi.org/10.7275/2mc9-4f52>

- Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation, 70*, 101030.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598–642.
- Stiggins, R. J. (2005). *Student-involved assessment for learning*. Pearson.
- Suskie, L. (2000). Fair assessment practices: Giving students equitable opportunities to demonstrate learning. *AAHE Bulletin, 52*(9), 7–9.
- Suskie, L. (2018). *Assessing student learning: A common sense guide*. John Wiley & Sons.
- Thorndike, R. L., & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education*. John Wiley & Sons.
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840–854.
- Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2019). Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education & Teaching International, 57*(1), 38–49. <https://doi.org/10.1080/14703297.2018.1564882>
- Welsandt, N. C. J., Fortunati, F., Winther, E., & Beck, K. (2024). Constructing and validating authentic assessments: The case of a new technology-based assessment of economic literacy. *Empirical Research in Vocational Education & Training, 16*, 4. <https://doi.org/10.1186/s40461-024-00158-0>
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. ASCD.