

Examining the Effect of Course Content Difficulty on Students' Evaluation Scores of Professors Teaching Quality

Tahereh Zafaripour¹, Seyed Ahmad Madani²

1. PhD. Of Higher Education Curriculum Planning - National Organisation of Educational Testing, Tehran, Iran; (Corresponding Author), Email: zafaripour@yahoo.com

2. Assistant Professor, Department of Educational Sciences, Faculty of Humanities, University of Kashan, Kashan, Iran. Email: madani@kashanu.ac.ir

Article Info	ABSTRACT
Article Type: Research Article Received: 2025.04.02 Received in revised form: 2025.07.31 Accepted: 2025.09.06 Published online: 2025.09.27	Objective: This study examines the evaluation of teaching quality in higher education and its impact on improving student's educational and learning processes. Given the importance of teaching quality, this research investigates the effect of course difficulty on student's evaluation scores and strives to increase the accuracy of the professors' evaluation system. Methods: This correlational study employed a mixed-effects regression model. The statistical population included evaluation data of basic science and engineering students in eight semesters, which included over 100,000 evaluations and 31,598 sample evaluation scores from courses. The difficulty of the courses was measured by calculating the mean and standard deviation of the scores of students in each faculty and dividing the difference between the course GPA and the faculty average by the standard deviation of the scores. Results: The results showed that professors' evaluation scores are significantly affected by the difficulty of the courses; the effect size of this relationship was estimated to be between 0.04 and 0.07. After controlling for students' GPA, it was found that evaluation scores in the components of "scientific mastery," "creating motivation," and "effective teaching" were lower in difficult courses and higher in easy courses. Conclusion: Evaluating teaching quality is an important tool for improving higher education. Course difficulty is an influential factor in professors' evaluation scores, and ignoring this issue can lead to unfair and unequal evaluations in the higher education system. Therefore, student evaluation scores should not be considered the sole indicator of teaching quality. To establish a valid and scientific evaluation system, more attention should be paid to the comprehensive and multidimensional design of the evaluation process. Keywords: Higher education, Educational evaluation, Teaching evaluation, Content difficulty, Professors

Cite this article: Zafaripour Tahereh, Madani Seyed Ahmad (2025). Examining the Effect of Course Content Difficulty on Students' Evaluation Scores of Professors Teaching Quality. *Higher Education Letter*, 18(71), 139-164. DOI: 10.22034/hel.2025.2053889.2051



© The Author(s).

Publisher: Institute for Research & Planning in Higher Education & National Organization of Educational Testing (NOET)



بررسی اثر دشواری محتوای درس بر نمرات ارزشیابی دانشجویان از کیفیت تدریس استادان

طاهره ظفری پور^۱، سید احمد مدنی^۲

۱. دکتری برنامه ریزی درسی در آموزش عالی، سازمان سنجش آموزش کشور، تهران، ایران. رایانامه: zafaripour@noet.ir
۲. استادیار گروه علوم تربیتی، دانشکده علوم انسانی، دانشگاه کاشان، کاشان، ایران. رایانامه: madani@kashanu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله:	هدف: هدف از اجرای پژوهش حاضر روایت الگوی تحول دانشگاه شیراز به منظور تبیین و ترسیم تجربه تحول نسلی یا انقلابی این دانشگاه و فراهم سازی امکان تحقق عینی، بازسازی و اصلاح آن است.
مقاله پژوهشی	روش پژوهش: این پژوهش با بهره گیری از رویکرد روایت پژوهی انجام گرفت. داده های پژوهش شامل مصاحبه های روایتی، اسناد و مدارک و یادآورنویسی بود که در طی مراحل نظیر تعیین مضامین قالب، شناسایی مضامین، ساخت شبکه های مضمونی و یکپارچه سازی اکتشاف ها، به روایت تحول دانشگاه شیراز منجر شد.
دریافت: ۱۴۰۴/۰۱/۱۳	یافته ها: روایت تحول دانشگاه در پنج دوره توصیف شد: (۱) آموزش محوری و امپراتوری سازی (۱۳۷۰-۱۳۹۷)، (۲) پژوهش محوری و برنامه جهت دار پژوهشی (۱۳۹۷-۱۳۹۸)، (۳) دانشگاه کارآفرین و تغییرات پوسته ای (۱۳۹۸-۱۳۹۹)، (۴) حکمت بنیانی و سیاست گذاری فرارشته ای و میان رشته ای (۱۳۹۹-۱۴۰۱)، و (۵) سرگردانی تحول؛ استمرار یا تقلیل (۱۴۰۱-۱۴۰۳).
اصلاح: ۱۴۰۴/۰۵/۰۹	نتیجه گیری: تاریخت تحول دانشگاه شیراز نشان دهنده تنیدگی و رقابت میان دو نظام است: نظام بوروکراتیک با رویکرد سازمانی و قدرت محور و نظام علم و فناوری با رویکرد اجتماعی و تمدنی. بر اساس تجربه انباشته حاصل از مراحل پیشین، سه محور اصلی در حرکت تحولی دانشگاه عبارتند از: «تثبیت نظام های تحولی در بافتار»، «توسعه ظرفیت تصمیم های تحولی» و «توسعه حکمرانی علم و فناوری» که جهت دهنده دلان تحول خواهند بود.
پذیرش: ۱۴۰۴/۰۶/۱۶	واژه های کلیدی: تحول، روایت، روایت پژوهی، دانشگاه شیراز
انتشار: ۱۴۰۴/۰۷/۰۵	

استناد: ظفری پور طاهره، مدنی سید احمد (۱۴۰۴). بررسی اثر دشواری محتوای درس بر نمرات ارزشیابی دانشجویان از کیفیت تدریس

استادان. نامه آموزش عالی، ۱۸(۷۱)، ۱۳۹-۱۶۴. DOI: 10.22034/hel.2025.2053889.2051



ناشر: مؤسسه پژوهش و برنامه ریزی آموزش عالی و سازمان سنجش آموزش کشور حق مؤلف © نویسندگان.

مقدمه

ارزشیابی تدریس استادان یکی از ارکان اصلی سنجش کیفیت آموزش در دانشگاه‌هاست که به‌عنوان ابزاری برای بررسی عملکرد آموزشی و ارتقای مهارت‌های تدریس در نظر گرفته می‌شود. در این راستا، نظام‌های ارزشیابی تدریس معمولاً از دانشجویان، همکاران و مدیران آموزشی به‌عنوان منابع اصلی داده‌ها استفاده می‌کنند. ارزشیابی‌های دانشجویی، به‌ویژه در دانشگاه‌های مختلف، شاخصی مهم در سنجش کیفیت تدریس، ارتقای وضعیت استعدادی و حرفه‌ای استادان و برنامه‌ریزی‌های آموزشی محسوب می‌شوند (سارویان و آموندسن^۱، ۲۰۰۴). ارزیابی اثربخشی تدریس به‌عنوان الگویی برای ارزیابی عملکرد اعضای هیئت علمی، شامل رویکردی مبتنی بر ارزیابی راهبردهای آموزشی، مشارکت دانشجویان در فرایند ارزیابی و ارزیابی آموزش مؤثر است. این الگو از دیدگاهی جامع حمایت می‌کند تا آموزش مؤثر را به تصویر بکشد (براون و مور^۲، ۲۰۱۷؛ هاتی و مارش^۳، ۱۹۹۶). با این حال، آنچه در عمل مشاهده می‌شود این است که عوامل مختلفی می‌توانند بر نتایج ارزشیابی‌ها تأثیرگذار باشند. یکی از مهم‌ترین این عوامل «دشواری محتوای تدریس» است که نیاز به بررسی دقیق دارد. همچنین، نتایج پژوهش‌ها نشان از تأکید بر نقش عوامل فرهنگی در ارزیابی عملکرد استادان دارند. تفاوت‌های فرهنگی بر ادراک اثربخشی تدریس تأثیر گذاشته و موجب تغییر در شیوه‌های ارزشیابی در زمینه‌های مختلف دانشگاهی می‌شوند (بکرز و همکاران^۴، ۲۰۱۹؛ دیویس و فلیشر^۵، ۲۰۱۷). از این‌رو، درک تغییرات زمینه‌ای و ایجاد سازگاری در ارزیابی استادان ضروری است. این سازگاری شامل شناخت ویژگی‌های منحصر به فرد مؤسسات آموزشی و تنظیم ملاک‌های ارزیابی متناسب با آنهاست (دیویس و فلیشر^۵، ۲۰۱۷). افزون بر این، الگوهای ارزیابی عملکرد استادان نقش مهمی در شکل‌دهی محیط دانشگاهی دارند و تحقیقات اخیر نیز بر ضرورت بازنگری انتقادی در اقدامات ارزیابی و کاهش تفاوت‌های موجود در شیوه‌های ارزیابی استادان تأکید می‌کنند (پره^۶، ۲۰۱۷؛ دیویس و فلیشر^۵، ۲۰۱۷).

در میان چالش‌های استفاده از الگوهای ارزیابی سنتی، ضرورت تغییر پارادایم به‌طور فزاینده‌ای نمایان شده است. اتکای همیشگی بر ارزیابی اعضای هیئت علمی توسط دانشجویان به‌عنوان معیار اولیه بررسی اثربخشی تدریس، درک ماهیت چندوجهی مشارکت استادان را ناممکن می‌سازد (علیزاده، ۱۳۹۴؛ مهدوی و همکاران، ۱۳۹۳؛ فرهنگی و فراستخواه، ۱۳۹۳؛ درگاهی و محمدزاده، ۱۳۹۲؛ افشار و همکاران، ۱۳۸۹؛ صفری، ۱۳۸۹). علی‌رغم ارزش ارزیابی عملکرد اعضای هیئت علمی توسط دانشجویان، این محدودیت‌ها به دلیل سوگیری‌های بالقوه، نظارت بر سایر ابعاد اساسی مانند فعالیت‌های علمی و خدمات سازمانی اعضای هیئت علمی را دشوار

1. Saroyan & Amundsen
2. Brown & Moore
3. Hattie & Marsh
4. Beckers et al.
5. Davies & Fleisher
6. Paré

ساخته و بی‌توجهی به سایر الگوهای ارزیابی مانند بررسی هم‌تا را تشدید می‌کند (سنتر^۱، ۲۰۰۳؛ اسپورن و همکاران^۲، ۲۰۱۳). یکی از چالش‌های اصلی که در زمینه ارزشیابی تدریس به چشم می‌آید، سوگیری‌های احتمالی در فرایند ارزشیابی است. به‌عنوان مثال، دانشجویان ممکن است به دلایل مختلف مانند جذابیت یا نحوه ارتباط استاد با دانشجویان، نمرات متفاوتی به استادان اختصاص دهند، در حالی که این ارزشیابی‌ها لزوماً بر اساس عملکرد واقعی تدریس نباشد. در این زمینه، مطالعات مختلفی بر اهمیت تأثیر عوامل غیرآموزشی مانند ویژگی‌های شخصیتی استاد، میزان علاقه‌مندی دانشجویان به موضوع درس و همچنین نحوه ارائه محتوا تأکید کرده‌اند (تان و همکاران^۳، ۲۰۱۹). علاوه بر این، تحقیقات نشان می‌دهد بسیاری از گروه‌های آموزشی ارزشیابی تدریس استادان را مبنای تصمیم‌گیری برای استخدام رسمی قرار می‌دهند و در صورت پایین بودن نتایج، از تبدیل وضعیت آنان به استخدام رسمی جلوگیری می‌کنند (سارویان و آموندسن^۴، ۲۰۰۴).

در راستای کاهش اثرات سوگیری در ارزشیابی، برخی محققان پیشنهاد داده‌اند که استفاده از سیستم‌های جامع و متنوع می‌تواند دقت و اطمینان این فرایند را افزایش دهد. به‌عنوان نمونه، سولیس و همکاران^۴ (۲۰۱۸) نشان دادند که اگر ارزشیابی‌ها در طول چند نیمسال مختلف گردآوری و تحلیل شوند، اثر سوگیری‌های موقتی کاهش یافته و ارزیابی دقیق‌تری حاصل می‌شود. به‌ویژه آنکه هر نیمسال ویژگی‌های خاص خود را دارد و می‌تواند نتایج ارزشیابی را تحت تأثیر قرار دهد. همچنین یافته‌های پژوهش‌ها حاکی از آن است که در بسیاری از نظام‌های ارزشیابی، نمرات بالاتر معمولاً به استادانی تعلق می‌گیرد که کمتر از دانشجویان انتقاد می‌کنند و می‌کوشند تدریس خود را جذاب‌تر ارائه دهند (اسپورن و همکاران^۲، ۲۰۱۳).

مطالعات مختلف نشان داده‌اند که ارزشیابی‌ها نه تنها می‌توانند به بهبود کیفیت تدریس کمک کنند، بلکه موجب شفاف‌سازی نقاط ضعف و قوت استادان و ارتقای عملکرد آموزشی آنان نیز می‌شوند. به‌عنوان مثال، پژوهش‌هایی که به بررسی نظام‌های ارزشیابی در دانشگاه‌های مختلف پرداخته‌اند، نشان می‌دهند مهم‌ترین منابع گردآوری اطلاعات برای ارزشیابی استادان شامل دانشجویان، همکاران، مدیران گروه و مسئولان دانشکده است. به‌ویژه در دانشگاه‌های علوم پزشکی، ارزشیابی‌ها بیشتر به‌عنوان ابزاری برای تغییر وضعیت استخدامی، ارتقای علمی و سالیانه و همچنین اصلاح شیوه تدریس استادان به کار می‌روند (حقیقی و همکاران^۱، ۱۴۰۰). در کنار این چالش‌ها، برخی پژوهش‌ها به بررسی چگونگی بهینه‌سازی نظام‌های ارزشیابی پرداخته‌اند. به‌عنوان مثال، کشتکاران و همکاران (۱۳۹۸) با استفاده از رویکرد تصمیم‌گیری چندمعیاره فازی مبتنی بر توافق گروهی، مدل‌های نرم‌افزاری ارزشیابی را توسعه دادند که می‌تواند به بهبود دقت ارزشیابی‌ها و استفاده بهینه از شاخص‌های مختلف در ارزشیابی کمک کند. در واقع، پژوهش‌های این‌چنینی نشان می‌دهد استفاده از مدل‌های تحلیلی

1. Centra
2. Spooren et al.
3. Tan et al.
4. Sulis et al.

می‌تواند نتایج ارزشیابی را جامع‌تر و دقیق‌تر کند. شاخص‌های ارزشیابی استادان در بسیاری از دانشگاه‌ها صرفاً به توانمندی علمی و تدریس محدود نمی‌شود، بلکه عواملی همچون تعامل استاد با دانشجویان، رعایت انضباط در کلاس و نحوه ارائه مطالب درسی را نیز شامل می‌شود. در همین زمینه، پژوهش‌های متعددی به شناسایی شاخص‌های مؤثر پرداخته‌اند. برای مثال، عسگری و همکاران (۲۰۲۱) طی پژوهش خود در دانشگاه علوم پزشکی اصفهان نشان دادند که در درس‌های عملی، «توجه یادگیرنده در ابتدای آموزش» و در درس‌های نظری، «ارائه مطالب به صورت کاربردی» مهم‌ترین معیارها هستند. این شاخص‌ها به‌ویژه از دیدگاه دانشجویان نقش مهمی در ارزیابی کیفیت تدریس ایفا می‌کنند. همچنین بسیاری از تحقیقات بر اهمیت بازخوردهای مؤثر در فرایند ارزشیابی تأکید کرده‌اند. برای نمونه، پژوهش وینچستر و وینچستر^۱ (۲۰۱۱) نشان داد یکی از مشکلات عمده اعضای هیئت علمی، آگاهی ناکافی از چگونگی استفاده از نتایج ارزشیابی است؛ مسئله‌ای که می‌تواند مانع اصلاح روش‌های تدریس و فرایندهای آموزشی شود. با این حال، برخی پژوهش‌ها تصریح کرده‌اند که در صورت استفاده مؤثر از نتایج ارزشیابی، می‌توان کیفیت تدریس و عملکرد آموزشی استادان را ارتقا بخشید (یمنی و کمالی، ۲۰۱۸).

پژوهش‌های اخیر از ادغام طیف متنوعی از ملاک‌های ارزیابی، از جمله تدریس، فعالیت‌های علمی و خدمات به دانشگاه یا مؤسسه آموزش عالی حمایت می‌کنند (بانتا و پالومبا^۲، ۲۰۱۵). به‌طور کلی، از آنجا که آموزش عالی همواره با پیشرفت‌های فناورانه همراه بوده و تغییرات بسیاری را تجربه می‌کند، نیاز روزافزونی به الگوهای دقیق‌تر و سازگارتر برای ارزیابی عملکرد اعضای هیئت علمی احساس می‌شود (براون و مور، ۲۰۱۷؛ فلتون و همکاران^۳، ۲۰۲۰). این امر به‌ویژه برای دستیابی به ارزیابی جامع و عاری از سوگیری ضروری است. از این رو، در این مقاله و با توجه به چالش‌های مطرح‌شده و دغدغه جامعه دانشگاهی، بر اساس پژوهش‌های پیشین به بررسی تأثیر دشواری محتوای تدریس بر نمرات ارزشیابی دانشجویان از کیفیت تدریس استادان پرداخته می‌شود؛ موضوعی که یکی از محورهای مهم در مباحث ارزشیابی به شمار می‌آید. اهمیت این تحقیق در آن است که با شناسایی و بررسی اثر دشواری محتوای تدریس بر نتایج ارزشیابی، می‌توان گامی مؤثر در جهت بهبود دقت و اعتبار نظام‌های ارزشیابی استادان برداشت. به‌ویژه در نظر گرفتن این عامل به ما کمک می‌کند دریابیم چرا برخی استادان صرفاً به دلیل تدریس محتوای پیچیده‌تر و دشوارتر، نمرات پایین‌تری در ارزشیابی دریافت می‌کنند. در ادامه، با هدف تبیین دقیق‌تر فرایند ارزشیابی، به مفاهیم مرتبط اشاره خواهد شد.

کیفیت^۴: این واژه (Qualitas) ترجمه واژه‌ای یونانی است که افلاطون و ارسطو برای تفکیک ماهیت اشیاء به کار می‌بردند. کیفیت هر شیء شامل ویژگی‌هایی است که آن را از انواع دیگر اشیاء متمایز می‌سازد (محمدی

1. Winchester & Winchester
2. Banta & Palomba
3. Felton et al.
4. Quality

و همکاران، ۱۳۸۶). ففر^۱ و کوت^۲ نیز در بحث خود، کیفیت را در زمره «مفاهیم لغزشی» قرار داده‌اند؛ بدین معنا که دارای معانی گوناگون بوده و تعاریف آن بر موارد و مصادیق مختلف دلالت دارد. افزون بر این، جنبه‌های انگیزشی و اخلاقی کیفیت باعث می‌شود تعریف آن به‌سادگی امکان‌پذیر نباشد (سالیس^۳، ۲۰۰۱؛ ترجمه حدیقی، ۱۳۸۱). از این‌رو می‌توان کیفیت را به میزان مطابقت وضعیت موجود با استانداردها و معیارهای از پیش تعیین‌شده، رسالت‌ها، اهداف و انتظارات تعبیر کرد. بر اساس این تعریف، کیفیت نظام آموزشی را می‌توان با بهره‌گیری از الگوی عناصر سازمانی (عوامل تشکیل‌دهنده نظام) مورد ارزیابی قرار داد؛ به بیان دیگر، کیفیت بر پایه دروندادها، فرایندها، محصول، بروندادها و پیامدها تعریف می‌شود (محمدي و همکاران، ۱۳۸۶). بررسی مطلوبیت زیرسیستم‌های نظام آموزشی، از جمله «تدریس»، رویکردی است که به کمک آن می‌توان درکی جامع از کیفیت آموزشی به دست آورد؛ زیرا نگرش کل‌نگر حاکم بر این دیدگاه در نهایت مبنایی ارزشمند برای شناخت مفهوم کیفیت فراهم می‌آورد.

کیفیت در آموزش عالی: آموزش عالی در کشورهای در حال توسعه با مسئله جدی کیفیت و پاسخ‌گویی مواجه است. از این‌رو، برای مقابله با این چالش لازم است در سیستم‌ها و ابزارهای بهبود، سرمایه‌گذاری شود (الیویرا و فریرا^۴، ۲۰۰۹). به همین دلیل، ارتقا و بهبود کیفیت عملکردی نظام‌های دانشگاهی باید در اولویت رهبران و مدیران آموزش عالی قرار گیرد؛ زیرا بی‌توجهی به این امر حیاتی، منجر به عدم تحقق سیاست‌های توسعه پایدار و اتلاف منابع انسانی، مالی و کالبدی قابل توجه خواهد شد (محمدي و همکاران، ۱۳۸۷) و همواره بیم آن می‌رود که کیفیت فدای کمیت شود (منصوری، ۱۳۸۲). در پنجاه سال گذشته، کیفیت به‌عنوان یکی از دغدغه‌های اصلی مدیریت، همواره مورد توجه بوده است (بکفورد^۵، ۲۰۰۴). با این حال، نگرانی نسبت به کیفیت در آموزش عالی موضوعی تازه نیست. دانشگاه‌های قرون وسطی اجتماعی خودگردان بودند که در درون خود به بهبود کیفیت می‌پرداختند. این سنت تا زمانی ادامه یافت که دانشگاه‌ها همچنان به‌صورت مؤسسات خودگردان فعالیت داشتند. اما با گسترش جهانی شدن، دیدگاه جهانی نسبت به سازمان‌دهی، مدیریت و تأمین مالی آموزش عالی و همچنین پاسخ‌گو ساختن آن از طریق فرایندهای تضمین کیفیت، به‌طور بنیادین دگرگون شد (ویر^۶، ۲۰۰۹). آموزش و در پی آن آموزش عالی، مسیر رشد، بهبود و توسعه افراد و جوامع است. در این راستا، ارزشیابی یکی از مهم‌ترین و حساس‌ترین مؤلفه‌های فرایند برنامه‌ریزی به شمار می‌رود؛ چراکه کارایی و اثربخشی نظام را تضمین کرده و چارچوبی برای بهبود مستمر برنامه‌ها فراهم می‌آورد (ظفری‌پور و همکاران، ۱۴۰۰). تضمین کیفیت، شامل تمامی اقدامات برنامه‌ریزی‌شده و نظام‌مندی است که با هدف اطمینان از تحقق الزامات کیفیت عناصر

1. Naomi Pfeffer
2. Anna Coote
3. Sallis
4. Oliveira & Ferreira
5. Beckford
6. Weir

نظام آموزشی - از جمله محصول یا خدمت - انجام می‌شود (بروان^۱، ۲۰۲۴). در آموزش عالی، تضمین کیفیت فرایندی است که اهداف، ساختار، درونداها، فرایندها، محصولات، برونداهای نهایی و پیامدهای مورد انتظار نظام‌ها و برنامه‌های آموزش عالی را بررسی کرده و کیفیت را حفظ، بهبود و ارتقا می‌دهد (بازرگان، ۱۳۸۰). در نتیجه، تضمین کیفیت می‌تواند با شناسایی نقاط ضعف نظام‌های آموزشی، نقش ترغیب‌کننده و اصلاح‌گر ایفا کند (مک‌چورا^۲، ۲۰۰۷).

ارزشیابی کیفیت: انسان به‌طور معمول و در طول زندگی، با بهره‌گیری از عقل و تجربه‌های خود همواره در حال انتخاب و گزینش است و بالطبع در پس هر انتخاب و گزینشی نوعی ارزشیابی نهفته است. با این حال، آنچه در نظام آموزشی مورد توجه قرار می‌گیرد، ارزشیابی علمی و تخصصی است که بر پایه مبانی، اصول و فرایندی حساب‌شده و منطقی انجام می‌شود. نتایج معتبر و قابل اتکای این نوع ارزشیابی می‌تواند مبنایی برای برنامه‌ریزی علمی به منظور بهبود امور و تحقق رسالت‌ها و اهداف تعیین‌شده قرار گیرد (محمدی و همکاران، ۱۳۸۶). بنابراین، ارزشیابی فرایندی است که با هدف شناسایی نقاط قوت و ضعف نظام مورد بررسی انجام می‌شود و به ارتقای کیفیت آن کمک می‌کند. شناخت مفهومی این واژه، جهت‌گیری‌ها و ضرورت انجام آن را آشکار می‌سازد. برای ارزشیابی تعاریف گوناگونی ارائه شده است، به‌گونه‌ای که تقریباً به تعداد صاحب‌نظران و متخصصان این حوزه، تعریف وجود دارد. رالف تایلر نخستین کسی بود که تعریف ارزشیابی آموزشی را مطرح کرد: «عمل مقایسه اهداف و عملکردهای برنامه» (کیامنش، ۱۳۸۱). همچنین بیبای ارزشیابی را این‌گونه تعریف کرده است: «فرایند جمع‌آوری و تفسیر نظام‌مند شواهد که در نهایت به قضاوت ارزشی با چشمداشت به اقدامی معین منتهی می‌شود» (ولف^۳، ۱۹۸۴؛ ترجمه کیامنش، ۱۳۷۵).

با توجه به تعاریف فوق، می‌توان دریافت که در فرایند ارزشیابی عناصر مشخصی وجود دارد که آن را از دیگر روش‌های سنجش و بررسی کیفیت متمایز می‌سازد. این عناصر عبارت‌اند از:

- فرایند: ارزشیابی از طریق توالی گام‌های منظم و مرتبط با یکدیگر انجام می‌گیرد که در نهایت تصویری روشن از وضعیت موجود نظام مورد بررسی ارائه می‌دهد.
- گردآوری اطلاعات: در جریان ارزشیابی، داده‌هایی درباره نقاط قوت و ضعف، همچنین فرصت‌ها و تهدیدهای پدیده یا نظام مورد بررسی گردآوری می‌شود.
- تفسیر نظام‌دار شواهد: با توجه به اینکه ارزشیابی در نهایت به قضاوت ارزشی منجر می‌شود، تفسیر داده‌های گردآوری‌شده باید با دقت و حساسیت بالا صورت گیرد؛ زیرا هرگونه کاستی در این مرحله می‌تواند اعتبار نتایج را تحت تأثیر قرار دهد.

1. Bron
2. McChura
3. Wolf

- قضاوت ارزشی: این مرحله شامل داوری درباره میزان مطلوبیت و کیفیت نظام مورد بررسی است که سطح کیفیت آن را از دیدگاه مخاطبان آشکار می‌سازد.
- چشمداشت به اقدام معین: فرایند ارزشیابی باید به گونه‌ای طراحی و اجرا شود که نتایج آن به عنوان بازخورد به برنامه بازگردد و در جهت بهبود کیفیت، افزایش پاسخگویی و برنامه‌ریزی برای اقدامات آینده مورد استفاده قرار گیرد.

موارد فوق از جمله ملاحظات اساسی در هر برنامه ارزشیابی به شمار می‌آیند. با این حال، متناسب با شرایط و ویژگی‌های نظام یا برنامه آموزشی، می‌توان عوامل متعدد دیگری را نیز در فرایند ارزشیابی لحاظ کرد. بر این اساس، ارزشیابی را می‌توان به منزله جست‌وجویی منظم و سازمان‌یافته برای قضاوت یا دستیابی به توافق درباره ارزش یا اهمیت یک پدیده آموزشی (برنامه، فعالیت و...) تعریف کرد؛ جست‌وجویی که هدف آن بهبود وضعیت موجود و کاهش فاصله میان نتایج فعلی و نتایج مطلوب است (بازرگان، ۱۳۸۷).

ارزشیابی کیفیت در آموزش عالی: ارزشیابی در آموزش عالی فرایندی است که با هدف تصریح اهداف، پاسخگویی در برابر جامعه و ارتقای کیفیت عملکرد نظام آموزش عالی انجام می‌شود. سازوکارهای پایش و ارزشیابی در این حوزه باید به طور دقیق وضعیت نظام را در ابعاد مختلف، متناسب با وظایف و مراحل تحول آن، منعکس سازند (پازارگادی و همکاران، ۱۳۸۳). با توجه به اینکه ارزشیابی اساساً فرایندی همزمان و موازی با فعالیت‌های آموزشی است و هدف اصلی آن اصلاح و بهبود عملکرد جاری و در نهایت افزایش کارایی و اثربخشی نظام‌های آموزشی در عرصه‌های گوناگون علمی و پژوهشی - از جمله تربیت متخصصان، توسعه علمی، پاسخ‌گویی به نیازهای پژوهشی، تقویت و گسترش فرهنگ اجتماعی و ارائه خدمات مشاوره‌ای - است، می‌توان تعریف زیر را برای ارزشیابی در آموزش عالی ارائه کرد:

ارزشیابی در آموزش عالی فرایندی مستمر و منظم برای گردآوری، تجزیه و تحلیل داده‌های مرتبط با عملکرد تمامی عوامل درونی و بیرونی مؤثر بر نظام آموزش عالی و مقایسه آنها با اهداف و ملاک‌های متناسب با ویژگی‌ها و کارکردهای دانشگاه است. هدف نهایی این فرایند، پیشرفت و بهبود مستمر نظام آموزش عالی از طریق شناسایی به موقع نواقص و ایجاد تغییرات و اصلاحات لازم و ضروری است (قهرمانی، ۱۳۸۳).

ارزشیابی کیفیت تدریس: ارزشیابی کیفیت تدریس یکی از ارکان اصلی ارتقای سطح علمی و آموزشی دانشگاه‌ها و مؤسسات آموزش عالی است. این فرایند، به‌ویژه در نظام‌های آموزشی که بر کیفیت تدریس و یادگیری تأکید دارند، از اهمیت ویژه‌ای برخوردار است.

اعضای هیئت علمی به عنوان یکی از کلیدی‌ترین عوامل رشد علمی دانشگاه‌ها هستند. عملکرد مثبت آنان می‌تواند موجب رشد اقتصادی و پیشرفت فناوری شود و در مقابل، عملکرد ضعیف آنان می‌تواند به افت و حتی بحران‌های اقتصادی منجر گردد (محمدی و همکاران، ۲۰۱۳). از این رو، طراحی نظام ارزشیابی متناسب

برای عملکرد استادان از اهمیت بالایی برخوردار است. ارزشیابی در این حوزه به عنوان یک سازوکار و فرایند ساختاریافته برای گردآوری و تفسیر اطلاعات تعریف می شود؛ اطلاعاتی که تحقق اهداف آموزشی و میزان دستیابی به آنها را تعیین می نماید (امینی و هنردار، ۲۰۰۸). در این میان، بررسی توانایی های آموزشی استادان در آموزش به دانشجویان (نیکولایدیس و دیمتریادیس^۱، ۲۰۱۴) و بررسی نظرات، ادراکات و انتظارات دانشجویان از عملکرد استادان به منظور ارتقای کیفیت تدریس و برقراری ارتباط مؤثر میان آنها و تلاش در جهت برطرف کردن موانع، جایگاه ویژه ای دارد (داگلاس و داگلاس^۲، ۲۰۰۶). این نوع ارزشیابی، فرایندی مستمر و منظم برای اطمینان از کیفیت فعالیت های آموزشی استادان است (باستانی و همکاران، ۲۰۱۴).

به طور معمول، ارزشیابی کیفیت تدریس توسط دانشجویان انجام می شود و ابزاری مؤثر برای شناسایی نقاط قوت و ضعف استادان و بهبود فرایندهای آموزشی به شمار می آید. این امر از آنجا اهمیت می یابد که تدریس صرفاً به معنای انتقال دانش از استاد به دانشجو نیست؛ بلکه ایجاد ارتباط مؤثر، مدیریت کلاس، ارائه بازخورد مستمر نسبت به فعالیت ها و تکالیف نیز برابر با تدریس است. تدریس باکیفیت حاصل انگیزه و تمایل دانشجو، پشتیبانی از فرصت های فراوان برای یادگیری و به کارگیری عملکرد مؤثر از سوی استادان است (کینگ^۳، ۲۰۲۲). در ادامه، ارزشیابی کیفیت تدریس از جنبه های مختلف آن تشریح می شود.

۱. اهمیت و ابعاد ارزشیابی کیفیت تدریس

ارزشیابی کیفیت تدریس یکی از مهم ترین ابزارهای نظارتی در نظام های آموزشی است که به ویژه در آموزش عالی با هدف ارتقای سطح تدریس و یادگیری استفاده می شود. این نوع ارزشیابی می تواند به عنوان ابزاری برای ارائه بازخورد به استادان و نیز بهبود محتوای آموزشی عمل کند (نیومن^۴، ۲۰۰۳). در همین زمینه در مقاله «ارزشیابی کیفیت تدریس در آموزش عالی: بررسی برخی دیدگاه ها» (معروفی و همکاران، ۱۳۸۶)، نویسندگان بر ضرورت استفاده از ابزارهای معتبر برای ارزیابی تأکید کرده اند. آنها معتقدند که ارزشیابی تدریس باید به گونه ای انجام شود که بازخوردهای مفید و شفاف به استادان ارائه شود. این بازخوردها می توانند به استادان در شناسایی نقاط قوت و ضعف تدریس کمک کنند و آنان را در جهت بهبود فرایند تدریس هدایت نمایند. علاوه بر این، ارزشیابی باید مبتنی بر شاخص های مشخص و معتبری مانند توانایی های تدریس، کیفیت محتوای درسی، و تعامل استاد با دانشجویان باشد (مارشال^۵، ۲۰۰۴).

۲. عوامل مؤثر در ارزشیابی تدریس

عوامل بسیاری در ارزشیابی تدریس و نتایج آن تاثیر دارند. در مقاله «ارزشیابی کیفیت تدریس اعضای هیئت

1. Nikolaidis & Dimitriadis

2. Douglas & Douglas

3. King

4. Newman

5. Marshall

علمی دانشگاه تربیت مدرس» (حاتمی، ۱۳۹۶)، نویسندگان به تحلیل عوامل مختلفی پرداخته‌اند که بر نتایج ارزشیابی تأثیر می‌گذارند. یکی از این عوامل، مرتبه علمی استاد و وضعیت استخدامی وی است. این مقاله نشان می‌دهد که نمرات ارزشیابی استادان تنها به کیفیت تدریس وابسته نیست، بلکه تحت تأثیر عواملی مانند رتبه علمی استاد و نوع قرارداد استخدامی او نیز قرار دارد. این موضوع نشان می‌دهد که برای تجزیه و تحلیل دقیق نتایج ارزشیابی، باید به این عوامل توجه ویژه‌ای صورت گیرد (مارشال و رش^۱، ۲۰۱۱). همچنین، اهمیت در نظر گرفتن تفاوت‌های فردی و ویژگی‌های کلاس در تحلیل نمرات ارزشیابی از استادان مورد تأکید قرار گرفته است.

۳. تأثیر ارزشیابی بر کیفیت تدریس و یادگیری

ارزشیابی دانشجویان از تدریس می‌تواند تأثیرات مثبتی بر کیفیت تدریس و فرایند یادگیری داشته باشد. در این خصوص مقاله «ارتقای کیفیت تدریس در پرتو ارزشیابی از اساتید؛ تأملی بر دیدگاه دانشجویان» (قنبری و سلطانزاده، ۱۳۹۷) بر اهمیت ارزشیابی‌های مداوم تأکید دارد و بیان می‌کند که ارزشیابی‌های دقیق و به موقع می‌توانند موجب بهبود کیفیت تدریس و تعاملات استاد-دانشجو شوند. در این مقاله اشاره شده است که بازخوردهای ارزشیابی به استادان کمک می‌کند تا نقاط ضعف خود را شناسایی کرده و بهبودهای لازم را در فرایند تدریس اعمال کنند. بر اساس مطالعات مختلف، ارزشیابی‌های دانشجویی ابزاری مؤثر برای بهبود کیفیت تدریس هستند، زیرا می‌توانند بازتابی از نظرات واقعی و بی‌طرفانه دانشجویان ارائه دهند (بیکر و همکاران^۲، ۲۰۱۰).

۴. روش‌ها و ابزارهای ارزشیابی تدریس

در ارزشیابی کیفیت تدریس، استفاده از ابزارهای مختلف برای گردآوری داده‌ها اهمیت زیادی دارد. یکی از ابزارهای رایج در این زمینه، پرسشنامه‌های ارزشیابی است که معمولاً شامل پرسش‌هایی درباره کیفیت تدریس، مهارت‌های ارتباطی استاد، و محتوای درسی هستند. در مقاله «ارزیابی کیفیت تدریس اعضای هیئت علمی دانشگاه فرهنگیان از دیدگاه دانشجویان» (آتش‌روز و همکاران، ۱۳۹۷)، به استفاده از پرسشنامه‌ها و نظرسنجی‌های دانشجویی به عنوان یکی از روش‌های اصلی ارزشیابی اشاره شده است. این ابزارها می‌توانند به طور نظام‌مند داده‌هایی دقیق و معتبر گردآوری کنند که در تحلیل کیفیت تدریس و ارتقای آن مورد استفاده قرار گیرند. همچنین، استفاده از فناوری‌های نوین مانند سامانه‌های آنلاین ارزشیابی و تحلیل داده‌ها می‌تواند دقت و کارایی فرایند ارزشیابی را افزایش دهد (گراهام^۳، ۲۰۰۶).

۵. تأثیر ارزشیابی بر انگیزه استادان و دانشجویان

تحقیقات مختلف نشان داده‌اند که ارزشیابی کیفیت تدریس نه تنها بر بهبود تدریس تأثیر دارد، بلکه می‌تواند انگیزه استادان و دانشجویان را نیز تقویت کند. در مقاله «ارتقای کیفیت تدریس در پرتو ارزشیابی از اساتید؛

1. Marshall & Rush
2. Baker et al.
3. Graham

تأملی بر دیدگاه دانشجویان» (قنبری و سلطانزاده، ۱۳۹۷)، اشاره شده است که نظرات مثبت دانشجویان می‌تواند انگیزه استادان را برای ادامه بهبود تدریس افزایش دهد. همچنین، ارزشیابی‌های منفی می‌توانند فرصتی برای استادان فراهم کنند تا مشکلات و نقاط ضعف خود را شناسایی کرده و در جهت بهبود آنها گام بردارند. از سوی دیگر، برای دانشجویان نیز این امکان فراهم می‌شود که در فرایند تدریس مشارکت کنند و احساس مسئولیت بیشتری نسبت به کیفیت آموزش پیدا کنند (تامسون و همکاران^۱، ۲۰۰۷).

در مجموع، ارزشیابی کیفیت تدریس یکی از ارکان مهم در فرایندهای آموزشی است که می‌تواند تأثیرات زیادی بر بهبود کیفیت تدریس و یادگیری داشته باشد. این ارزیابی‌ها باید با استفاده از ابزارهای معتبر و مبتنی بر شاخص‌های مشخص انجام شود تا نتایج آن قابل اعتماد و مؤثر باشد. در این زمینه، عوامل مختلفی از جمله مرتبه علمی استادان، روش‌های تدریس و تعامل استاد-دانشجو بر نتایج ارزشیابی تأثیرگذارند. همچنین، بازخوردهای دقیق و مستمر به استادان می‌تواند موجب ارتقای تدریس و افزایش انگیزه استادان و دانشجویان شود.

روش پژوهش

این پژوهش از نوع همبستگی است و آزمون اصلی آن مبتنی بر رگرسیون مدل اثرات آمیخته^۲ انجام گرفت. این مدل برای بررسی رابطه میان متغیرها و کنترل واریانس ناشی از دیگر متغیرهای مرتبط با متغیر وابسته استفاده می‌شود (اسپیلمن و همکاران^۳، ۲۰۱۸).

جامعه آماری پژوهش شامل کلیه داده‌های ارزشیابی دانشجویان دانشگاه کاشان در حوزه‌های علوم پایه و فنی-مهندسی از تدریس استادان در هشت نیمسال تحصیلی پیش از همه‌گیری کرونا بود که در مجموع بالغ بر نتایج یکصد هزار ارزشیابی می‌شد. برای انتخاب نمونه، متناسب با فرضیه اصلی تحقیق، سه شرط بر داده‌های استخراج‌شده از سامانه آموزش اعمال شد: ۱) دوره کارشناسی، ۲) درس‌های نظری اصلی و الزامی، و ۳) ظرفیت کلاس بیش از ۱۰ نفر. در مرحله دوم، داده‌ها بر حسب معدل دانشجویان در درس و تعداد مشروطی دانشجویان بررسی و موارد پرت یا مفقود حذف شدند. در نهایت، نمونه‌ای شامل ۳۱۵۹۸ نمره ارزشیابی از درس‌ها به دست آمد.

برای سنجش میزان دشواری هر درس، ابتدا میانگین و انحراف معیار نمرات دانشجویان هر دانشکده در درس‌های نظری تخصصی محاسبه شد. به واسطه این تحلیل میانگین معتبری برای هر دانشکده به دست آمد. سپس تفاوت «معدل درس/کلاس» با «میانگین دانشکده» محاسبه و بر انحراف معیار نمرات دانشکده تقسیم شد. با توجه به این نحوه محاسبه، بالاتر بودن این نمره تفاضل از میانگین دانشکده به معنای دشواری درس بوده است. در ادامه، این نمرات مرتب شدند و بر حسب میزان تفاوت از میانگین دانشکده به پنج گروه «درس‌های بسیار دشوار»،

1. Thomson et al.

2. Mixed-Effects Regression Model

3. Speelman et al.

«درس‌های دشوار»، «درس‌های معمولی»، «درس‌های آسان» و «درس‌های بسیار آسان» طبقه‌بندی شدند. ابزار گردآوری داده‌ها فرم رسمی دانشگاه برای ارزشیابی کیفیت تدریس استادان بود که در پایان هر نیمسال و پیش از شروع امتحانات پایان‌ترم به صورت الکترونیکی تکمیل می‌شد. به طور معمول، این فرم‌ها در مؤسسات آموزش عالی، در «شورای آموزشی دانشگاه» بررسی و تصویب می‌شوند. فرم ارزشیابی در بازه زمانی پژوهش دارای ۱۷ آیتم و یک مقیاس نُه درجه‌ای (از نمره ۱۲ تا ۲۰) بود. همچنین در هر آیتم یک گزینه «نظری ندارم» نیز برای پاسخ‌دهنده‌ها پیش‌بینی شده بود.

با توجه به اینکه حدود ده درصد از فرم‌ها به طور کامل تکمیل نشده بود، شرط دیگری نیز اعمال شد؛ بدین صورت که فرم‌های دارای بیش از سه پاسخ «نظری ندارم» حذف شدند. در مواردی که تعداد بی‌پاسخ‌ها سه یا کمتر بود میانگین نمره کلی هر فرم جایگزین آیتم‌های بدون پاسخ شد. از آنجایی که فرم رسمی دانشگاه فاقد مؤلفه‌های از پیش تعیین‌شده بود، برای تحلیل دقیق‌تر ارزشیابی‌های دانشجویان، ابتدا مؤلفه‌های کیفیت تدریس با استفاده از تحلیل عامل اکتشافی به شرح جدول (۱) استخراج شدند.

جدول (۱) نتایج تحلیل‌های عامل اکتشافی بر روی آیتم‌های پرسشنامه ارزشیابی تدریس دانشجویان

سطح معنی‌داری	KMO ^۲	ESSL ^۱	پایایی	تعداد آیتم‌ها	مؤلفه‌های ارزشیابی از کیفیت تدریس استادان
۰/۰۰۰	۰/۸۵۵	۸۷/۲۷۶	۰/۹۵۱	۴	تسلط و توان علمی
۰/۰۰۰	۰/۹۵۶	۸۲/۰۴۹	۰/۹۶۴	۸	ایجاد انگیزش و احترام
۰/۰۰۰	۰/۹۲۲	۸۰/۸۸۲	۰/۹۵۳	۵	تدریس و مدیریت اثربخش

چنانکه رول و رولت^۳ (۲۰۲۱) اشاره کرده‌اند پدیده تک‌بعدی بودن^۴ مقیاس یکی از پدیده‌های رایج در تحلیل داده‌های ارزشیابی تدریس است؛ گویی فرم‌های ارزشیابی صرف‌نظر از تعداد آیتم‌ها، فقط یک خصیصه یا کیفیت واحد را اندازه‌گیری می‌کنند. با این حال، تفکیک و دسته‌بندی آیتم‌ها با توجه به مبانی نظری موجود در زمینه ویژگی‌های استادان اثربخش، منطقی و لازم به نظر می‌رسید. در واقع، با این روش متغیر وابسته کلی به سه متغیر جزئی دیگر تبدیل شد که این کار اطلاعات دقیق‌تری درباره ادراکات و برداشت‌های دانشجویان از تدریس استادان به دست می‌دهد. به بیان دیگر، مسئله این است که از تحلیل نمره کلی مشخص نمی‌شود که دانشجویان هر بخش رفتارهای حرفه‌ای استادان را چگونه ارزیابی کرده‌اند. به هر حال، پس از تحلیل عامل اکتشافی، انجام تحلیل عامل تأییدی نیز در دستور کار قرار گرفت که نتایج مؤید روایی سازه مطلوب ابزار اندازه‌گیری و مؤلفه‌های آن بود.

1. Extraction Sums of Squared Loadings
2. Kaiser-Meyer-Olkin Measure of Sampling Adequacy
3. Roll & Rollet
4. Unidimensionality

جدول (۲) نتایج تحلیل عامل تأییدی برای مؤلفه‌های ارزشیابی دانشجویان از تدریس

RMR	GFI	AGFI	TLI	CFI	شاخص‌های تحلیل عامل تأییدی
۰/۰۳۷	۰/۹۹۳	۰/۹۳۲	۰/۸۴۰	۰/۹۷۳	تسلط و توان علمی
۰/۰۲۴	۰/۹۹۶	۰/۹۸۹	۰/۹۵۲	۰/۹۸۰	ایجاد انگیزش و احترام
۰/۰۲۷	۰/۹۹۷	۰/۹۹۱	۰/۹۶۹	۰/۹۸۶	تدریس و مدیریت اثربخش

چنانکه آماره‌های جدول (۲) نشان می‌دهد وضعیت شاخص‌های تحلیل عامل تأییدی، مطلوب و نشانگر وجود روایی سازه ابزار اندازه‌گیری و مدل مفهومی دسته‌بندی مؤلفه‌های آیتم‌های ارزشیابی است. باید توجه کرد که برخی از شاخص‌ها^۱ نسبت به حجم بالای نمونه حساس هستند و نمی‌توان آنها را معیار قابل اعتنایی برای قضاوت درباره برازش مدل محسوب کرد (هولمز و همکاران^۲، ۲۰۱۶). به همین دلیل، شاخص‌ها بر اساس رهنمودهای منابعی نظیر مارش و بیلی^۳ (۱۹۹۴)، دیمیتروف^۴ (۲۰۱۱) و نیماند و مای^۵ (۲۰۱۸) انتخاب شده‌اند.

یافته‌ها

فرضیه اصلی پژوهش این است که ارزشیابی دانشجویان از تدریس استادان به سطح دشواری درس بستگی دارد. به بیان دقیق‌تر، میانگین کلی و نیز میانگین نمرات مؤلفه‌های ارزشیابی تدریس را می‌توان از روی میزان دشواری درس پیش‌بینی کرد. نتایج تحلیل رگرسیون مربوطه در جدول (۳) نشان داده شده است. لازم به ذکر است که در این تحلیل، متغیر دشواری درس از نوع فاصله‌ای یا پیوسته^۶ بوده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. RMSEA, CMIN
2. Holmes et al.
3. Marsh & Bailey
4. Dimitrov
5. Niemand & Mai
6. Scale

جدول (۳) نتایج تحلیل رگرسیون برای پیش‌بینی نمرات ارزشیابی تدریس از روی میزان دشواری درس

متغیر وابسته	متغیر مستقل	Coef.	Std. Err	t	P
تسلط و توان علمی	دشواری درس	۰/۰۷۳	۰/۰۰۵۶	۱۳/۰۲	۰/۰۰۰
	اندازه ثابت	۱۸/۰۲	۰/۰۳۳۰	۵۴۶/۶	۰/۰۰۰
ایجاد انگیزش و احترام	دشواری درس	۰/۰۵۹	۰/۰۰۵۷	۱۰/۳۵	۰/۰۰۰
	اندازه ثابت	۱۷/۹۸	۰/۰۳۳۲	۵۴۰/۶	۰/۰۰۰
تدریس و مدیریت اثربخش	دشواری درس	۰/۰۴۹	۰/۰۰۵۵	۸/۸۹	۰/۰۰۰
	اندازه ثابت	۱۸/۱۳	۰/۰۳۲۳	۵۶۱/۰	۰/۰۰۰
نمره کل ارزشیابی تدریس	دشواری درس	۰/۰۶۰	۰/۰۰۵۴	۱۱/۰۸	۰/۰۰۰
	اندازه ثابت	۱۸/۰۵	۰/۰۳۱۹	۵۶۵/۱	۰/۰۰۰

با توجه به اطلاعات جدول (۳) می‌توان ملاحظه کرد که ضرایب تعیین، اگر چه بسیار کوچک اما معنادار هستند. در بخش دوم جدول، ملاحظه ضرایب رگرسیونی نشان می‌دهد که با یک واحد افزایش در دشواری درس، شاهد افزایشی به اندازه ۷/۳٪ واحد در نمره ارزشیابی «تسلط و توان علمی»، ۵/۹٪ واحد در نمره «ایجاد انگیزش و احترام»، ۴/۴۹٪ واحد در نمره «تدریس و مدیریت اثربخش»، و ۰/۶٪ واحد در نمره کل ارزشیابی به وجود می‌آید. با توجه به اینکه فرم ارزشیابی دارای ۹ درجه (از ۱۲ تا ۲۰) بوده هر واحد برابر با ۲/۲۲۲ است. بنابراین، به بیان ملموس‌تر، در مورد «نمره کل ارزشیابی» می‌توان گفت که با یک واحد افزایش در متغیر دشواری درس، افزایش معناداری به اندازه ۰/۰۲۷ در نمره کل به وجود می‌آید. اما اگر این تحلیل بدون در نظر گرفتن تأثیر سایر متغیرها تفسیر شود نتایج گمراه‌کننده‌ای از آن قابل برداشت است. برای پیشگیری از بروز این خطا، در مرحله دوم تحلیل، متغیر دشواری درس از حالت فاصله‌ای به طبقه‌ای تبدیل شد به این صورت که بر اساس تفاوت معدل هر درس با میانگین دانشکده، پنج نوع درس تعریف شد: درس‌های عادی (کد ۰)، درس‌های بسیار دشوار (کد ۴)، درس‌های دشوار (کد ۳)، درس‌های آسان (کد ۲) و درس‌های بسیار آسان (کد ۱). برای نمونه، معدل «درس‌های عادی» $\pm 0/5$ انحراف معیار و معدل «درس‌های بسیار دشوار» ۲- انحراف معیار از میانگین دانشکده تفاوت داشته‌اند.

در ادامه، متغیر دشواری درس به عنوان متغیر مستقل و هر کدام از مؤلفه‌های ارزشیابی از تدریس به عنوان متغیر وابسته وارد مدل اثرات آمیخته چندسطحی^۱ شدند. با توجه به اینکه نمرات دانشجویان می‌تواند تحت تأثیر توان

1. Mixed Models Multilevel Regression

و استعداد عمومی آنها برای تحصیل قرار گیرد متغیر «معدل دانشجوی» نیز به عنوان پارامتر اثرات تصادفی وارد مدل شد. لازم به ذکر است که افزون بر معدل دانشجوی، متغیرهای دانشکده، اندازه کلاس، جنسیت دانشجوی و جنسیت استاد نیز به طور جداگانه به عنوان پارامتر اثرات تصادفی وارد مدل شدند اما به غیر از عامل «معدل دانشجوی» هیچ کدام از این متغیرها مقدار قابل توجه معناداری از واریانس نمرات ارزشیابی را تبیین نمی کرد. ذکر این نکته لازم است که درس های عادی به این دلیل به عنوان طبقه مرجع انتخاب شد که معدل نمرات دانشجویان در آنها به میانگین دانشکده بسیار نزدیک بود. در واقع، این درس ها نه آسان و نه دشوار بوده اند و بر اساس معیارهای انتخاب محتوای برنامه درسی می توان محتوای آنها را دارای سطح بهینه ای از دشواری دانست. ارنشتاین و هانکینز^۱ (۲۰۱۸) از این ویژگی به عنوان قابلیت یادگیری^۲ یاد کرده اند. نتیجه تحلیل مربوط به مؤلفه «تسلط و توان علمی» در جدول (۴) گزارش شده است.

جدول (۴) نتایج تحلیل رگرسیون برای مقایسه «تسلط و توان علمی» بر حسب انواع درس (طبقه مرجع: درس های عادی)

انواع درس	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
بسیار آسان	۰/۲۳۷	۰/۰۴۰	۵/۸۷	۰/۰۰۰	۰/۱۵۷	۰/۳۱۶
آسان	۰/۲۰۲	۰/۰۳۵	۵/۷۶	۰/۰۰۰	۰/۱۳۳	۰/۲۷۲
دشواری	۰/۰۸۶	۰/۰۳۲	۲/۶۳	۰/۰۰۹	۰/۰۲۲	۰/۱۵۱
بسیار دشوار	-۰/۱۶۲	۰/۰۳۸	-۴/۲۷	۰/۰۰۰	-۰/۲۳۶	-۰/۰۸۷
اندازه ثابت	۱۸/۴۴	۰/۰۳۴	۵۳۵/۴۱	۰/۰۰۰	۱۸/۳۷	۱۸/۵۰

پارامترهای اثرات تصادفی	برآورد	خطای معیار	فاصله اطمینان ۹۵٪	
معدل دانشجوی واریانس سطح ۲	۰/۳۴۸	۰/۰۲۶	۰/۲۹۹	۰/۴۰۵
واریانس خطا	۳/۹۵۴	۰/۰۳۱	۰/۸۹۲	۴/۰۱۷

با توجه به اطلاعات جدول (۴) وقتی درس های عادی مبنای مقایسه باشند می توان گفت که در مؤلفه «تسلط و توان علمی» نمره ارزشیابی درس های بسیار آسان، آسان و دشوار به طور معناداری بالاتر از درس های عادی است. در مقابل، ضریب رگرسیونی -۰/۱۶۲ برای درس های بسیار دشوار نشانگر آن است که ارزشیابی دانشجویان از «تسلط و توان علمی» استاد در این درس ها به طور معناداری از درس های عادی پایین تر است. اندازه اثر مربوط به این ضریب رگرسیونی نیز برابر با ۰/۰۷ محاسبه شد. در بخش دوم جدول (۴)، ضریب همبستگی درون رده ای

1. Ornstein & Hunkins
2. learnability

(حاصل تقسیم واریانس سطح دو بر مجموع واریانس سطح دو و واریانس خطا) معادل ۰/۰۸ است به این معنا که هشت درصد از واریانس نمرات در بُعد «تسلط و توان علمی» ناشی از معدل دانشجویان است که می توان آن را شاخص اساسی استعداد و توان تحصیلی محسوب کرد.

جدول (۵) نتایج تحلیل رگرسیون برای مقایسه «ایجاد انگیزش و احترام» بر حسب انواع درس (طبقه مرجع: درس های عادی)

انواع درس	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
بسیار آسان	۰/۲۱۲	۰/۰۴۱	۵/۲۵	۰/۰۰۰	۰/۱۳۳	۰/۲۹۲
آسان	۰/۱۸۵	۰/۰۳۵	۵/۲۷	۰/۰۰۰	۰/۱۱۶	۰/۲۵۵
دشوار	۰/۱۱۹	۰/۰۳۳	۳/۶۰	۰/۰۰۰	۰/۰۵۴	۰/۱۸۳
بسیار دشوار	-۰/۰۹۷	۰/۰۳۸	-۲/۵۷	۰/۰۱۰	-۰/۱۷۲	--۰/۰۲۳
اندازه ثابت	۱۸/۴۴	۰/۰۳۵	۵۱۴/۳۱	۰/۰۰۰	۱۸/۱۹	۱۸/۳۳

پارامترهای اثرات تصادفی	برآورد	خطای معیار	فاصله اطمینان ۹۵٪
معدل دانشجو واریانس سطح ۲	۰/۴۰۰	۰/۰۲۹	۰/۳۴۵ ۰/۴۶۳
واریانس خطا	۳/۹۶۸	۰/۰۳۱	۳/۹۰۵ ۴/۰۳۱

در بُعد «ایجاد انگیزش و احترام» نمره ارزشیابی درس های بسیار آسان، آسان و دشوار به طور معناداری از درس های عادی بالاتر است. چنانکه اطلاعات جدول (۵) نشان می دهد ضریب رگرسیونی ۰/۰۹۷- برای درس های بسیار دشوار معنادار است که نشان می دهد ارزشیابی دانشجویان از «ایجاد انگیزش و احترام» در این درس ها از درس های عادی پایین تر است. اندازه اثر مربوط به این ضریب رگرسیونی نیز برابر با ۰/۰۴ محاسبه شد. ضریب همبستگی درون رده ای معادل ۰/۰۹ و به این معناست که هشت درصد از واریانس نمرات در بُعد «ایجاد انگیزش و احترام» ناشی از معدل دانشجویان است که تأثیر آن در تحلیل ها کنترل شده است.

جدول (۶) نتایج تحلیل رگرسیون برای مقایسه «تدریس و مدیریت اثربخش» بر حسب انواع درس (طبقه مرجع: درس‌های عادی)

[95% Conf. Interval]		P> z	z	Std. Err.	Coef.	انواع درس
۰/۲۷۲	۰/۱۳۸	۰/۰۰۰	۴/۹۸	۰/۰۳۹	۰/۱۹۵	بسیار آسان
۰/۲۵۳	۰/۱۱۹	۰/۰۰۰	۵/۴۴	۰/۰۳۴	۰/۱۸۶	آسان
۰/۲۱۳	۰/۰۸۷	۰/۰۰۰	۴/۶۹	۰/۰۳۲	۰/۱۵۰	دشواری
--۰/۰۰۳	-۰/۱۴۸	۰/۰۴۱	-۲/۰۵	۰/۰۳۶	-۰/۰۷۵	بسیار دشوار
۱۸/۴۲	۱۸/۲۹	۰/۰۰۰	۵۲۹/۷۶	۰/۰۳۴	۱۸/۳۶	اندازه ثابت

فاصله اطمینان ۹۵٪		خطای معیار	برآورد	پارامترهای اثرات تصادفی
۰/۴۴۶	۰/۳۳۴	۰/۰۲۸	۰/۳۸۶	معدل دانشجو واریانس سطح ۲
۳/۸۰۰	۳/۶۸۲	۰/۰۳۰	۳/۷۴۰	واریانس خطا

با توجه به یافته‌های جدول (۶) در بُعد «تدریس و مدیریت اثربخش» نیز ارزشیابی درس‌های بسیار آسان، آسان و دشوار به‌طور معناداری بالاتر از درس‌های عادی است. در مقابل، ضریب رگرسیونی ۰/۰۷۵- برای درس‌های بسیار دشوار معنادار و نشان‌گر آن است که ارزشیابی دانشجویان از «تدریس و مدیریت اثربخش» در این درس‌ها از درس‌های عادی پایین‌تر است. اندازه اثر مربوط به این ضریب رگرسیونی نیز برابر با ۰/۰۵ محاسبه شد. در این تحلیل نیز ضریب همبستگی درون‌رده‌ای حدود ۰/۰۸ به دست می‌آید. سرانجام، نتایج تحلیل مربوط به نمره کل ارزشیابی تدریس در جدول (۷) ارائه شده است.

جدول (۷) نتایج تحلیل رگرسیون برای مقایسه «نمره کلی ارزشیابی تدریس» بر حسب انواع درس (طبقه مرجع: درس‌های عادی)

[95% Conf. Interval]		P> z	z	Std. Err.	Coef.	انواع درس
۰/۲۹۰	۰/۱۳۸	۰/۰۰۰	۵/۵۲	۰/۰۳۸	۰/۲۱۴	بسیار آسان
۰/۲۵۸	۰/۱۲۵	۰/۰۰۰	۵/۶۵	۰/۰۳۳	۰/۱۹۱	آسان
۰/۱۸۱	۰/۰۵۶	۰/۰۰۰	۳/۷۴	۰/۰۳۱	۰/۱۱۸	دشواری
--۰/۰۴۰	-۰/۱۸۳	۰/۰۰۲	-۳/۰۶	۰/۰۳۶	-۰/۱۱۱	بسیار دشوار
۱۸/۴۲	۱۸/۲۹	۰/۰۰۰	۵۳۵/۷۶	۰/۰۳۴	۱۸/۳۵	اندازه ثابت

فاصله اطمینان ۹۵٪		خطای معیار	برآورد	پارامترهای اثرات تصادفی
۰/۴۳۸	۰/۳۲۷	۰/۰۲۷	۰/۳۷۹	معدل دانشجو واریانس سطح ۲
۳/۷۱۴	۳/۵۹۹	۰/۰۲۹	۳/۶۵۶	واریانس خطا

در مورد نمره کل ارزشیابی دانشجویان از تدریس نیز یافته مشابهی به دست آمد. طبق جدول (۷) نمره کل ارزشیابی درس‌های بسیار آسان، آسان و دشوار به‌طور معناداری از درس‌های عادی بالاتر است اما با توجه به ضریب رگرسیونی ۰/۱۱۱- برای درس‌های بسیار دشوار، مشخص است که نمره ارزشیابی این درس‌ها به‌طور معناداری از درس‌های عادی پایین‌تر است. اندازه اثر مربوط به این ضریب رگرسیونی نیز برابر با ۰/۰۵ محاسبه شد.

لازم به ذکر است که برای اطمینان بیشتر از دقت نتیجه‌گیری‌ها تمامی تحلیل‌ها یک بار نیز با تغییر طبقه مرجع تکرار شد. در این راستا، درس‌های «دشوار» به‌عنوان طبقه مرجع انتخاب گردید که مشاهده شد همچنان ارزشیابی دانشجویان از تدریس استادان در درس‌های «بسیار دشوار» در تمام مؤلفه‌ها و نیز در نمره کل، به‌طور معناداری از درس‌های «دشوار» پایین‌تر است. این نتیجه را می‌توان نشان‌ر اعتبار درونی پژوهش و یافته‌های آماری محسوب کرد.

بحث و نتیجه‌گیری

نتایج این پژوهش نشان داد که نمرات ارزشیابی استادان به‌طور معناداری تحت تأثیر میزان دشواری محتوای درس قرار دارند. اندازه اثر ارتباط میان دشواری درس و نمرات ارزشیابی دانشجویان از تدریس استادان بین ۰/۰۴ تا ۰/۰۷ برآورد شد که نشان‌دهنده تأثیر قابل توجه این عامل است. پس از کنترل اثر معدل دانشجویان، مشخص شد که نمرات ارزشیابی استادان در مؤلفه‌های «تسلط و توان علمی»، «ایجاد انگیزش و احترام» و «تدریس و مدیریت اثربخش» در درس‌های بسیار دشوار به‌طور معناداری کمتر از درس‌های عادی است. در مقابل، نمرات ارزشیابی درس‌های آسان به‌طور معناداری از درس‌های عادی بالاتر است. این یافته‌ها نشان می‌دهد که نمرات ارزشیابی به‌طور نظام‌مند تحت تأثیر دشواری درس قرار می‌گیرند، در حالی که این عامل خارج از کنترل استادان است. برای مثال، درس‌های پایه و تخصصی در بسیاری از رشته‌ها ذاتاً دشوار هستند و حذف یا کاهش محتوای آنها می‌تواند پیامدهای منفی برای دانشجویان داشته باشد. با این حال، استنادی که این درس‌ها را تدریس می‌کنند، ممکن است صرفاً به دلیل سختی محتوای درس، نمرات پایین‌تری در ارزشیابی دریافت کنند که نشان‌دهنده نوعی سوگیری در این فرایند است.

یافته‌های پژوهش‌های مختلف نشان داده‌اند که ارزشیابی دانشجویان از استادان ممکن است تحت تأثیر عواملی

غیر از کیفیت واقعی تدریس قرار گیرد. از جمله ویژگی‌های شخصیتی استاد، شیوه مدیریت کلاس، میزان سخت‌گیری یا آسان‌گیری در نمره‌دهی و حتی جذابیت ظاهری آنان (سولیس، ۲۰۱۸). در واقع، بسیاری از دانشجویان ممکن است استاد را نه بر اساس توانایی تدریس، بلکه بر اساس واکنش‌های عاطفی یا ویژگی‌های غیرآموزشی ارزشیابی کنند. برای مثال، در پژوهشی مشخص شد که استادان آسان‌گیر معمولاً نمرات بالاتری در ارزشیابی دریافت می‌کنند (اسپورن و همکاران، ۲۰۱۳). همچنین، استادانی که جذابیت ظاهری دارند، ممکن است ارزیابی بهتری دریافت کنند، حتی اگر روش تدریس آنها تفاوتی با سایر همکاران نداشته باشد (وینچستر و وینچستر، ۲۰۱۱). این مسئله نشان می‌دهد که ارزشیابی دانشجویان نمی‌تواند تنها معیار قطعی برای سنجش کیفیت تدریس باشد. علاوه بر این، مطالعات دیگری نشان داده‌اند که عدم استفاده مؤثر از نتایج ارزشیابی باعث کاهش انگیزه دانشجویان برای مشارکت در این فرایند می‌شود (یمنی و کمالی، ۲۰۱۸). در برخی دانشگاه‌ها، دانشجویان با بی‌تفاوتی و بدون دقت، فرم‌های ارزشیابی را تکمیل می‌کنند و این امر دقت و اعتبار نتایج را کاهش می‌دهد.

بر همین اساس، برخی دانشگاه‌های معتبر جهان روش‌های دقیق‌تری را برای ارزشیابی اتخاذ کرده‌اند که علاوه بر نظرات دانشجویان، شامل خودارزیابی استادان، ارزیابی همکاران و تحلیل نتایج چندین نیمسال تحصیلی است (طاهری، ۲۰۲۳). از یافته‌های دیگر پژوهش می‌توان به این نکته اشاره کرد که برخی گروه‌های آموزشی در صورتی که ارزشیابی تدریس استاد «مطلوب» یا «رضایت‌بخش» نباشد، تمایلی به تبدیل وضعیت او به استخدام رسمی ندارند (سارویان و آموندسن، ۲۰۰۴). این مسئله نشان می‌دهد که نمرات ارزشیابی می‌توانند تأثیر مستقیم بر تصمیمات مدیریتی، ترفیع و آینده شغلی استادان داشته باشد. با این حال، اگر کاهش نمرات ارزشیابی ناشی از دشواری محتوای درس باشد، به قضاوت‌های غیرمنصفانه منجر می‌شود. بنابراین، یکی از دلالت‌های مهم این پژوهش برای مدیران دانشگاهی و هیئت‌های جذب آن است که هنگام تفسیر نمرات ارزشیابی، باید به ماهیت و سطح دشواری درس‌ها توجه شود. زیرا معمولاً ترکیب درس‌های هر عضو هیئت علمی در طول ترم‌های مختلف ثابت است. بنابراین، استادانی که مسئول تدریس درس‌های دشوار هستند، ممکن است در معرض ارزیابی‌های ناعادلانه قرار گیرند که بر ترفیع و آینده شغلی او تأثیرگذار خواهد بود.

برای بهبود فرایند ارزشیابی، پیشنهاد می‌شود که دانشگاه‌ها از مدل‌های جامع‌تری استفاده کنند که علاوه بر نظرات دانشجویان، شامل سایر معیارها مانند خودارزیابی استادان، ارزیابی همکاران و تحلیل نتایج چند نیمسال باشد. مطالعات نشان داده‌اند که گردآوری و تحلیل نتایج ارزشیابی در چندین نیمسال می‌تواند بخشی از سوگیری‌های مربوط به دشواری درس را کاهش دهد (سولیس و همکاران، ۲۰۱۸). علاوه بر این، برخی پژوهش‌ها پیشنهاد کرده‌اند که ابزارهای ارزشیابی به‌روز شوند و با روش‌های علمی مانند آنتروپی شانون وزن‌دهی شوند تا دقت و اعتبار نتایج افزایش یابد (عسگری و همکاران، ۲۰۲۱).

به طور کلی، نتایج این پژوهش نشان می‌دهد که نمرات ارزشیابی استادان نباید به عنوان تنها معیار سنجش کیفیت تدریس در نظر گرفته شود، بلکه باید در کنار سایر شاخص‌های آموزشی، پژوهشی و حرفه‌ای تحلیل شوند. از آنجا که نظام آموزش عالی یک نظام هدفمند و عامل توسعه همه‌جانبه کشور است، ارزیابی مستمر و دقیق فرایندهای آموزشی و تدریس، امری ضروری برای بهبود کیفیت آموزش و افزایش کارآمدی دانشگاه‌ها خواهد بود (ادوارد، ۱۳۸۰). با توجه به اهمیت این موضوع، پیشنهاد می‌شود که پژوهش‌های مشابه در سایر دانشگاه‌ها انجام شده و نتایج آنها با یکدیگر مقایسه شود. در نهایت، با انجام فراارزیابی‌های متعدد، می‌توان به اصلاح و بهینه‌سازی فرایند ارزشیابی استادان پرداخت و تصمیم‌گیری‌های مدیریتی را بر اساس داده‌های دقیق‌تر و منصفانه‌تر انجام داد.

نتایج این پژوهش به روشنی نشان می‌دهد که نمرات ارزشیابی استادان به طور نظام‌مند تحت تأثیر میزان دشواری محتوای درسی قرار می‌گیرند؛ این مسئله می‌تواند منجر به نوعی سوگیری در قضاوت نسبت به عملکرد استادان شود. در واقع، نمرات پایین‌تر استادان در درس‌هایی که ذاتاً دشوارتر هستند، لزوماً نشان‌دهنده کیفیت پایین تدریس نیست، بلکه بیانگر تأثیر یک عامل بیرونی است که از کنترل استاد خارج است. این یافته اهمیت بازنگری در روش‌های فعلی ارزشیابی عملکرد آموزشی را برجسته می‌سازد؛ به‌ویژه آنکه در بسیاری از دانشگاه‌ها، نمرات ارزشیابی نقش مستقیم در تصمیمات مهمی مانند ارتقا، ترفیع و استخدام رسمی استادان دارند. از این رو، اتکای صرف بر نمرات ارزشیابی دانشجویان می‌تواند به تصمیم‌گیری‌های ناعادلانه و ناکارآمد منجر شود، به‌ویژه برای استادانی که مسئول تدریس درس‌های تخصصی و دشوار هستند. به طور کلی موارد زیر در این پژوهش قابل تأکید هستند:

- نمرات ارزشیابی دانشجویان از استادان تحت تأثیر میزان دشواری درس قرار دارند.
 - استادانی که درس‌های دشوار را تدریس می‌کنند، معمولاً نمرات ارزشیابی پایین‌تری دریافت می‌کنند، در حالی که استادانی که درس‌های آسان‌تری دارند، نمرات بالاتری کسب می‌کنند.
 - این مسئله ممکن است به ارزیابی‌های ناعادلانه و تصمیمات مدیریتی نادرست منجر شود.
- این پژوهش با برخی محدودیت‌ها مواجه بوده است؛ از جمله اینکه داده‌ها تنها از یک دانشگاه و تعداد محدودی از رشته‌های تحصیلی گردآوری شده‌اند؛ بنابراین، تعمیم‌پذیری یافته‌ها به سایر دانشگاه‌ها یا رشته‌های مختلف ممکن است با محدودیت‌هایی همراه باشد. همچنین، با وجود تلاش برای کنترل برخی متغیرها مانند معدل دانشجویان، همچنان امکان تأثیر سایر عوامل غیر قابل کنترل مانند ویژگی‌های شخصیتی استادان یا فرهنگ سازمانی گروه‌های آموزشی وجود دارد.

به عنوان نتیجه‌گیری پایانی، یافته‌های این پژوهش به روشنی نشان داد که دشواری درس عامل مؤثری بر نمرات ارزشیابی استادان است و بی‌توجهی به این عامل می‌تواند منجر به ارزیابی‌های غیرمنصفانه و نابرابر در نظام

- آموزش عالی شود. بنابراین، نمرات ارزشیابی دانشجویان نباید به عنوان تنها شاخص ارزیابی کیفیت تدریس در نظر گرفته شوند. برای دستیابی به یک نظام ارزشیابی معتبر و علمی، باید فرایند ارزشیابی با رویکردی جامع و چندبُعدی طراحی و اجرا شود. در همین راستا پیشنهادهای زیر ارائه می‌شود:
- توسعه مدل‌های جامع ارزشیابی: طراحی نظام‌های ارزشیابی ترکیبی که علاوه بر نظر دانشجویان، شامل خودارزیابی استادان، نظرات همکاران، مدیران گروه و تحلیل روند عملکرد در چندین نیمسال باشد.
 - وزن‌دهی علمی به آیت‌های ارزشیابی: استفاده از روش‌های آماری مانند آنتروپی شانون برای تعیین وزن دقیق هر آیت در ابزار ارزشیابی و به‌روزرسانی این ابزارها به‌طور منظم.
 - آموزش دانشجویان درباره ارزشیابی دقیق و مسئولانه: برگزاری کارگاه‌های آموزشی برای دانشجویان درباره اهمیت و نحوه صحیح ارزشیابی استادان
 - توجه به ماهیت درس‌ها در تفسیر نمرات: هنگام تحلیل نمرات ارزشیابی، باید دشواری درس، سطح درس (پایه یا تخصصی) و ویژگی‌های محتوایی آنها به‌صورت دقیق در نظر گرفته شود.
 - فرارزیابی و بررسی روندی: گردآوری و تحلیل نمرات ارزشیابی در طول چند ترم متوالی برای افزایش دقت در تصمیم‌گیری‌های مدیریتی مرتبط با استادان.
- در مجموع، برای دستیابی به نظام آموزش عالی منصفانه و کارآمد، بازنگری علمی در نحوه ارزشیابی استادان ضرورتی اجتناب‌ناپذیر است.

References

- Afshar, M., Hassanzadeh Taheri, M., Riasi, H., & Naseri, M. (1389). Evaluation of faculty members by students with different levels of academic achievement. *Scientific Journal of Birjand University of Medical Sciences*, 17(2), 118-126. [In Persian]
- Alizadeh, A. (2014). The degree of reliability of the evaluation scores of undergraduate students on the educational performance of faculty members. *Educational Measurement & Evaluation Quarterly*, 5(12), 27-47. [In Persian]
- Amini, M., & Honardar, M. (2008). The view of faculties and medical students about evaluation of faculty teaching experiences. *Koomesh*, 9(3), 171-178. [In Persian]
- Asgari, I., & Khosravi, M. (2021). Item Weighting of the Faculty Member's Evaluation Form Regarding Shannon Entropy Technique from Dental Students' View. *Iranian Journal of Medical Education*, 20(48), 415. [In Persian]
- Atash-Rouz, M., Pour-Moghaddasian, A., & Marashi, S. M. (2017). Evaluating the Teaching Quality of Faculty Members at the University of Education from Students' Perspectives.

- Journal of Educational Research*, 4(2), 23–34. <https://doi.org/10.22034/jedu.2017.126257>
[In Persian]
- Baker, J., Larson, M., & Colleagues. (2010). Impact of student evaluations on faculty improvement. *Journal of Educational Research*, 104(3), 223–233. <https://doi.org/10.1080/00220670903383031>
- Bazargan, A. (2002). *Educational Evaluation*. Tehran: SAMT Press. [In Persian]
- Bazargan, A. (2008). *Educational Evaluation (Concepts, Models, and Operational Processes)* (6th Edition). Tehran: SAMT Press. [In Persian]
- Banta, T. W., & Palomba, C. A. (2015). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. John Wiley & Sons.
- Bastani, P., Rouhollahi, N., & Tahernejad, A. (2015). Validity and reliability of teachers' evaluation questionnaires from students' point of view in Tehran University of Medical Sciences. *Journal of Medical Education*, 3(1), 7–14. <https://doi.org/10.18502/jmededu.v3i1.7073>
- Beckford, G. (2004). *Quality Management in Higher Education*. London: Routledge.
- Beckers, J. J., Belder, R., & Van Den Berg, B. A. (2019). Faculty performance appraisal: An exploration of content and implementation in practice. *Assessment & Evaluation in Higher Education*, 44(3), 366–381.
- Brown, G. T., & Moore, D. (2017). Conceptualizing and validating sources of teaching self-efficacy: A phenomenographic study of pre-service elementary teachers. *Teaching & Teacher Education*, 63, 356–369.
- 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.
- Douglas, J., Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher Education*, 12(1), 3–13.
- Davies, A., & Fleisher, C. (2017). Peer observation of teaching: A case for engaging in critical conversations about teaching and learning. *Higher Education Research & Development*, 36(6), 1142–1156.
- Dargahi, H., & Mohammadzadeh, N. (2012). Evaluation of faculty members by students: valid or invalid. *Iran Journal of Education in Medical Sciences*, 13(1), 39–48. [In Persian]
- Ghanbari, S., & Soltanzadeh, V. (2018). Improving the Quality of Teaching in the Light of Evaluation of Professors: Reflecting on Students' Perspective. *Journal of Research in*

- Teaching*, 6(2), 15-31. [In Persian]
- Graham, J. (2006). The use of online surveys to evaluate teaching quality. *Teaching in Higher Education*, 11(2), 203-217.
- Ghahremani, M. (2004). *Evaluation of the internal efficiency of graduate programs in universities under the Ministry of Science, Research, and Technology during the academic years 1994-1995 to 1999-2000*. Tehran: Faculty of Educational Sciences and Psychology, Shahid Beheshti University; Institute for Research and Planning in Higher Education. Panel Discussion on Improving the Quality of University Education. [In Persian]
- Farhangi, M., Farasatkah, M. (2013). Examining the competency model of faculty members in the virtual environment in terms of faculty members and students (based on Armaner's model). *Iran Information & Communication Technology Quarterly*, 7(21), 1-14. [In Persian]
- Felton, P., Koper, R., Pallof, R., & Pratt, K. (2020). *The distance education evolution: Issues and case studies*. Routledge.
- Haghighi, M., Cherabin, M., Akbari, A., & Karimi, M. (2021). Necessity, position, and view of educational evaluation of faculty members in higher education. *Journal of Zabol Medical School*, 4(2), 83-90. <https://doi.org/10.18502/jzms.v4i2.7073> [In Persian]
- Hatami, J. (2017). Evaluation of Teaching Quality of Faculty Members at Tarbiat Modares University. *Journal of Educational Measurement & Evaluation Studies*, 7(19), 163-184. [In Persian]
- Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, 66(4), 507-542.
- Holmes, A., Imex, T., & French, P. (2016). *Quality Assurance in University Teaching*. Cambridge: Cambridge University Press.
- Keshtkaran, Z., Moeini, Z., Shoshtarian, M. R., et al. (2019). *Developing a Software Model for Faculty Evaluation Using a Fuzzy Multi-Criteria Decision-Making Approach Based on Group Consensus*. [In Persian]
- Kiamanesh, A., Mehrmohammadi, M., & Asgari, M. (2007). Evaluation of Teaching Quality in Higher Education: A Review of Some Views. *Curriculum Studies*, 2(5). [In Persian]
- Kiamanesh, A. (2002). *Educational Evaluation Methods (Educational Sciences)*. 9th Edition. Tehran: Payame Noor University Press. [In Persian]
- King, H. (2022). The Characteristics of Expertise for Teaching in Higher Education. In *Developing Expertise for Teaching in Higher Education*, 15-28. Routledge.

- Mahdavi, S., Zare, S., & Naimi, N. (2015). Comparison of academic performance evaluation by students with their self-evaluation. *Research in Medical Science Education*, 6(2), 51-61. [In Persian]
- Mansouri, R. (2003). *Scientific development in Iran* (2nd Ed.). Tehran: Ettelaat Publishing. [In Persian]
- Marshall, K. (2004). *Evaluating Teaching Effectiveness*. Jossey-Bass.
- Marshall, K., & Rush, D. (2011). *Student Feedback in Higher Education*. Routledge.
- Marsh, H. W., & Bailey, M. (1994). Multidimensional Students' Evaluations of Teaching Effectiveness: A Test of Alternative Higher-Order Structures. *Journal of Educational Psychology*, 86(4), 631-645.
- McChura, T. (2007). *Evaluation Methods in Higher Education*. Oxford: Blackwell.
- Mohammadi, R., Fath-Abadi, J., Yadgarzadeh, G., Mirzamohammadi, M. H., & Parand, K. (2007). *Evaluation of Quality in Higher Education: Concepts, Principles, Methods, and Criteria*, Tehran: National Organization for Educational Testing. [In Persian]
- Mohammadi, R., Shariati, S., Mokhtarian, F., & Karamzadeh, S. (2008). The Evolution of the Implementation of the Performance Evaluation System in the Ministry of Science, Research, and Technology. *Higher Education Letter*, 1(1), 133-167. [In Persian]
- Mahmoudi, M., Nasei, S., Gholiniya, H. (2013). Identifying Criteria for Evaluating Teaching Performance of Faculty Members with Emphasizing on Engineering Teaching. *Journal of Technology of Education*, 7(4), 324-315. [In Persian]
- Newman, F. (2003). *The Future of Higher Education: Rhetoric, Reality, and the Risks of the Market*. San Francisco: Jossey-Bass.
- Nikolaidis, Y., Dimitriadis, S.G. (2014). On the student evaluation of university courses and faculty members' teaching performance. *European Journal of Operational Research*, 238, 199-207.
- Oliveira, O. J., & Ferreira, E. C. (2009). Adaptation and Application of the Servqual Scale in Higher Education. *POMS 20th Annual Conference*, Orlando, Florida.
- Ornstein, A., & Hunkins, F. (2018). *Curriculum: Foundations, Principles and Issues*. Boston: Pearson.
- Pazargadi, M., Mirani Ashtiani, R., & Vaezzadeh, S. (2004). *Macro and micro evaluation of higher education in the Islamic Republic of Iran*. In *Proceedings of the Conference on Higher Education and Sustainable Development* (Vol. 2). Tehran: Institute for Research

- and Planning in Higher Education, Farhang Dehkhoda Publishing. [In Persian]
- Paré, A. (2017). Tracing the work of policy in translational research: A case study of the development of a tool for assessing research impact. *Research Evaluation*, 26(3), 157-169.
- Qanbari, S., & Sultan-Zadeh, V. (2018). Improving Teaching Quality through Evaluations from Faculty; Reflection on Students' Views. *Teaching Research Journal*, 6(2), Summer. [In Persian]
- Roll, D., & Rollet, S. (2021). *Global Trends in Higher Education Quality Assessment*. London: Palgrave Macmillan.
- Safari, Th. (1389). The role of various information sources in the educational evaluation of academic staff members. *Research & Planning Quarterly in Higher Education*, 55, 69-85. [In Persian]
- Saroyan, A., & Amundsen, C. (2004). The Final Step: Evaluation of Teaching. In *Rethinking Teaching in Higher Education: From a Course Design Workshop to a Faculty Development Framework*. Virginia: Stylus Publishing, LLC.
- Sallis, E. (2001). *Total Quality Management in Education*. Translated by Hadiqi, A. Tehran: University Publishing Center. [In Persian]
- Speelman, D., Heylen, K., & Geeraerts, D. (2018). *Mixed-Effects Regression Models in Linguistics*. Cham, Switzerland: Springer International Publishing.
- 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.
- Sulis, I., Porcu, M., & Capursi, V. (2019). On the Use of Student Evaluation of Teaching: A Longitudinal Analysis Combining Measurement Issues and Implications of the Exercise. *Social Indicators Research*, 142(3), 1305-1331.
- Tan, S., Lau, E., Ting, H., et al. (2019). How Do Students Evaluate Instructors' Performance? Implication of Teaching Abilities, Physical Attractiveness, and Psychological Factors. *Social Indicators Research*, 146(1/2), 61-76.
- Thomson, D., Robinson, S., & Colleagues (2007). Motivation and Student Evaluations of Teaching. *Journal of College Teaching*, 55(3), 123-134.
- Weir, J. (2009). *Globalization and Higher Education: Quality Assurance in a Changing World*. New York: Routledge.
- Winchester, T. M., & Winchester, M. (2011). Exploring the Impact of Faculty Reflection on Weekly Student Evaluations of Teaching. *International Journal for Academic Development*,

- 16(2), 119 – 131.
- Wolf, R. M. (1984). *Educational evaluation: The measurement of ability and the assessment of programs* (A. Kiamanesh, Trans., 1999). Tehran: Markaz Nashr. [In Persian]
- Yamani, N., & Kamali, F. (2018). *Improving the Faculty Evaluation System in Isfahan University of Medical Sciences: An Action Research* [Dissertation]. Isfahan, Iran: Isfahan University of Medical Sciences. [In Persian]
- Zafaripour, T., Arefi, M., Fathi Vajargah, K., & Mohammadi, R. (2021). Analysis of the curriculum revision process in Iran's higher education system. *Journal of Educational Measurement & Evaluation Studies*, 11(33), 140–183. <https://doi.org/10.22034/emes.2021.247581>. [In Persian]

