



Analyzing the Relationship Between Grade Inflation and Students' Evaluation of University Professors' Teaching¹

Seyyed Ahmad Madani², Ali Ghasempour^{3*}

(Received: 2022.12.17 - Accepted: 2023.05.08)

1. This article is the independent research report of the authors.
 2. Assistant Professor, Department of Educational Sciences, Faculty of Humanities, University of Kashan, Kashan, Iran.
 3. Assistant Professor, Department of Educational Sciences, Faculty of Humanities, University of Payam Noor, Tehran, Iran.
- * Corresponding author: Email: alighasempoor@pnu.ac.ir

Abstract

The purpose of this research was to investigate the relationship between students' academic progress scores and their evaluation of the university professors' teaching. The longitudinal design of this research is a panel study in which the necessary measurements are made over time and at specific intervals. The students studying in the three semesters were the statistical population of this research. The sample size was determined by GPower software and based on the type of statistical tests required. After screening more than one hundred thousand pairs of "student grade-teacher evaluation", the final analysis was performed on the data of 1970 students who had studied with a certain professor in three consecutive semesters. The tool for collecting the evaluation data of the professors was the official survey form of the institution in the intended semesters. The analysis showed that with the increase of the student's final grade, his/her evaluation grade from the professor in the next semester increases significantly. Supplementary analysis showed that the effect of this initial score remains stable until two semesters late. It can be said that with a one-unit increase in students' final grade, after two semesters, a significant increase between 1 and 4% in their evaluation grade from professors can be seen. The regression coefficients of the relationship between the grade point average of the whole class and the teaching evaluation of professors were also significant in the case of two consecutive semesters. Finally, the analyzes on the grades of the specialized and main courses of the students showed that compared to the past semesters, there is a grade inflation in the groups of humanities, literature, architecture and art. Based on this evidence, it is suggested that the phenomenon of grade inflation be considered from the point of view of its possible connection and interrelation with the way student evaluations are used in the employment and academic decisions of faculty members.

Keywords: Evaluation, teaching, teaching evaluation, students, university teachers, academic achievement



واکاوی ارتباط میان تورم نمرات با ارزشیابی دانشجویان از تدریس اساتید^۱

سید احمد مدنی^۲، علی قاسم‌پور^{۳*}

(دریافت: ۱۴۰۱/۰۹/۲۶ - پذیرش: ۱۴۰۲/۰۲/۱۸)

چکیده

هدف این پژوهش واکاوی ارتباط میان نمرات پیشرفت تحصیلی دانشجویان با ارزشیابی آنها از تدریس اساتید بود. طرح طولی تحقیق از نوع مطالعات پنلی است که در آن طی فواصل زمانی مشخص اندازه‌گیری‌های لازم انجام می‌شود. دانشجویان شاغل به تحصیل در سه نیمسال پیش از شرایط آموزش مجازی جامعه آماری این تحقیق بوده‌اند. حجم نمونه با نرم‌افزار جی‌پاور و بر اساس آزمون‌های آماری تعیین شد. پس از شرط‌گذاری و غربال بیش از یکصد هزار جفت «نمره دانشجو-ارزشیابی استاد»، تحلیل نهایی روی داده‌های مربوط به ۱۹۷۰ دانشجو انجام شد که در سه نیمسال متوالی با یک استاد درس گرفته بودند. ابزار گردآوری داده‌های ارزشیابی اساتید، فرم رسمی مؤسسه در نیمسال‌های مورد نظر بوده است. تحلیل‌ها نشان داد با افزایش نمره پایان‌ترم دانشجو، نمره ارزشیابی او از استاد در نیمسال بعد افزایش معناداری پیدا می‌کند. تحلیل تکمیلی مشخص ساخت که اثر نمره اولیه تا دو نیمسال بعد نیز پایدار می‌ماند، به نحوی که می‌توان گفت با یک واحد افزایش در نمره پایان‌ترم دانشجویان، پس از دو نیمسال افزایش معناداری بین ۱ تا ۴ درصد در نمره ارزشیابی آنها از استادان قابل مشاهده است. ضرایب رگرسیونی ارتباط معدل کل کلاس با ارزشیابی تدریس استادان نیز در حالت دو نیمسال متوالی معنادار بود. سرانجام، تحلیل‌ها بر روی نمرات دروس تخصصی و اصلی دانشجویان نشان داد، در مقایسه با گذشته، در گروه رشته‌های علوم انسانی، ادبیات و معماری و هنر تورم نمره وجود دارد. بر اساس این شواهد پیشنهاد شده است که پدیده تورم نمرات از منظر ارتباط احتمالی آن با نحوه بهره‌برداری از ارزشیابی‌های دانشجویی در تصمیم‌گیری‌های استخدامی و دانشگاهی اعضای هیات علمی مورد تأمل قرار گیرد.

واژگان کلیدی: ارزشیابی، تدریس، ارزشیابی تدریس، دانشجویان، استادان، پیشرفت تحصیلی

۱- این مقاله، گزارش پژوهش مستقل نویسندگان است.

۲- استادیار گروه علوم تربیتی، دانشکده علوم انسانی، دانشگاه کاشان، کاشان، ایران.

۳- استادیار گروه علوم تربیتی، دانشگاه پیام‌نور، تهران، ایران.

* نویسنده مسئول: alighasempoor@pnu.ac.ir

مقدمه

ارزشیابی دانشجویان از تدریس استادان، یکی از فرایندهای رایج برای بهبود کیفیت آموزش در دانشگاه‌ها است (رادماکر و مارتین^۱، ۲۰۰۱؛ دانلیسون و مک‌گریل، ۲۰۰۰؛ لیک‌من^۲ و همکاران، ۲۰۲۱؛ مارشال^۳، ۲۰۲۲). ارزشیابی‌های دانشجویی کاربردهای مختلفی دارند (مارش و روکه، ۱۹۹۳). غایت نظری این ارزشیابی‌ها، بهبود یادگیری دانشجویان و ارتقای تدریس اساتید است (بولر^۴، ۲۰۱۲؛ گالزیو^۵، ۲۰۱۱). در عمل اما از نتایج آنها برای تصمیم‌گیری‌های سازمانی مهمی نظیر تمدید قرارداد^۶، تبدیل وضعیت استخدامی به رسمی^۷ و ارتقای مرتبه علمی نیز استفاده می‌شود (لینزی^۸، ۲۰۱۷). واقعیت این است که ارزشیابی، یکی از ابزارهای اصلی برای بهبود عملکرد شغلی در حرفه‌های مختلف است (مارشال^۹، ۲۰۰۹). اما از یک سو، ماهیت حرفه تدریس و رفتارهای تدریس اثربخش منحصر به فرد هستند (آدامز و انگلمان^{۱۰}، ۱۹۹۵) و نمی‌توان ارزشیابی آن را به عنوان مثال، با ارزشیابی یک فروشنده توسط مشتریان مقایسه کرد. از سوی دیگر، در زمینه ارزشیابی استادان توسط دانشجویان، یافته‌های تحقیقاتی مهمی وجود دارد که نگاه‌ها را در زمینه چگونگی بهره‌گیری از نتایج ارزشیابی‌ها تغییر داده است.

یکی از مهم‌ترین یافته‌ها این است که بین نمرات ارزشیابی دانشجویان از اساتید و مقدار یادگیری واقعی دانشجویان همبستگی ناچیزی وجود دارد (کلایسون^{۱۱}، ۲۰۰۹). در واقع، نتایج ارزشیابی‌ها به میزان یادگیری واقعی دانشجو در درس ربطی ندارند (آتل، وایت و گونزالز^{۱۲}، ۲۰۱۷). ارزشیابی‌ها می‌توانند تحت تأثیر متغیرهایی نظیر جنسیت (منگل، سورمان و زولیتز^{۱۳}، ۲۰۱۷) و سطح توانایی و ویژگی‌های شخصی دانشجویان قرار گیرند (مک‌میلان، سولانلز و راجرز^{۱۴}، ۲۰۲۱). یافته‌های مهم و جالب دیگری نیز وجود دارد. از جمله اینکه دانشجویان متناسب با جنسیت استاد او را ارزشیابی می‌کنند به این معنا که از استاد زن انتظار یک سبک زنانه و از استاد مرد انتظار یک سبک مردانه دارند. آنها طبق آمایه جنسیتی^{۱۵} خود، استاد را ارزشیابی می‌کنند (باکن، مک‌لافلین و گارسیا^{۱۶}، ۱۹۹۹). یافته جالب دیگر این است که ارزشیابی‌ها بیش

-
- 1- Radmacher & Martin
 - 2- Lakeman
 - 3- Marshall
 - 4- Buller
 - 5- Golzio
 - 6- Retention
 - 7- Tenure
 - 8- Linse
 - 9- Marshall
 - 10- Adams & Engelmann
 - 11- Clayson
 - 12- Uttl, White & Gonzalez
 - 13- Mengal, Sauermann & Zolitz
 - 14- McMillan, Solanelles & Rogers
 - 15- Gender schema
 - 16- Bachen, McLoughlin & Garcia

از هر چیز به شخصیت کلی استاد بستگی دارد و «شاید مدرسان برون‌گرایی که ارتباط‌گرا، مشتاق و شوخ‌طبع هستند، معلمان اثربخش‌تری باشند» (ص، ۲۶۶) (رادماکر و مارتین، ۲۰۰۱، ص ۲۶۶). اینها بخشی از عوامل اثرگذار بر ارزشیابی‌های دانشجویی هستند. اما مسأله ابعاد دیگری نیز دارد. یکی از نکات قابل تأمل این است که ادراک دانشجو از نظام ارزشیابی، روی ارزشیابی‌های وی از استاد تأثیر دارد. پژوهش اسپورن و کریستیانز^۱ (۲۰۱۶) نشان می‌دهد دانشجویانی که نظام ارزشیابی تدریس در مؤسسه را کارآمد و ارزشمند می‌پندارند، نمرات بالاتری به استادان می‌دهند. حتی نکات فنی بسیار جزئی نظیر تعداد درجات مقیاس یا همان گزینه‌های نمره‌دهی نیز می‌توانند ارزشیابی نهایی استاد را تحت تأثیر قرار دهند (لاندرام و برایتمن^۲، ۲۰۰۸). سرانجام، لیک‌من و همکاران (۲۰۲۱) مشاهده کردند که پنهان بودن هویت دانشجویان به هنگام ارزیابی می‌تواند به بیان نظرات تهاجمی، تنبیهی و آبروریزی استاد منجر شود.

بنابر نکاتی که گفته شد حداقل به لحاظ مبانی پژوهشی تردیدی نیست که منظومه‌ای از عوامل تصادفی یا غیرتصادفی می‌توانند داده‌های ارزشیابانه را در معرض سوگیری قرار دهند. بی دلیل نیست که اعضای هیأت علمی ملاحظات و دغدغه‌هایی جدی دربارهٔ چگونگی بهره‌برداری و تفسیر نتایج ارزشیابی‌ها دارند (بویسن، کلی، ریزلی و کاسنر^۳، ۲۰۱۴). افزون بر این، آنها از دانشجویان نیز انتظار دارند که اهمیت و ضرورت درس را درک کنند و این قدرشناسی در ارزشیابی آنها از درس انعکاس داشته باشد. اما در واقع، ارزشیابی دانشجویان از اساتید، ربطی به سطح تعهد و علاقه آنها نسبت به موضوع و درس ندارد (بران و وایولاتو^۴، ۲۰۰۹) و بیش از آن، انعکاسی از تصورات دانشجویان دربارهٔ میزان دشواری درس است (آدیسون، بست و وارینگتون^۵، ۲۰۰۶). این تعارض که در نتایج ارزشیابی‌ها نیز منعکس خواهد شد می‌تواند اثرات مخربی برای سبک تدریس استاد داشته باشد.

یکی از پیامدهای احتمالی نظام ارزشیابی، این است که استادان بر اساس ارزشیابی‌های دانشجویی، تغییراتی را در سبک و رویهٔ تدریس خود ایجاد می‌کنند. اما در عمل تضمینی برای ثمربخشی این تغییرات وجود ندارد. یکی از دلایل مهم به فرم‌های ارزشیابی از تدریس مربوط می‌شود. بسیاری از فرم‌های ارزشیابی، بازخوردهای مفیدی را به مدرسان ارئه نمی‌دهند (مارشال، ۲۰۰۹). شاید به همین علت باشد که برخی از استادان به مرور زمان یاد می‌گیرند که اساساً ارزشیابی‌های دانشجویی را نادیده بگیرند! اما آیا باید نتیجه گرفت که ارزشیابی‌های دانشجویی هیچ اعتباری ندارند؟

1- Spooen & Christiaens
2- Landrum & Braitman
3- Boysen, Kelly, Raesly & Casner
4- Beran & Violato
5- Addison, Best & Warrington

این در حالی است که استادان صرف نظر از ابهامات مربوط به چگونگی اجرای این ارزشیابی‌ها، نگرانی‌هایی نیز درباره چگونگی بهره‌برداری و تفسیر نتایج آنها دارند (ضیایی، میری، حاجی آبادی و آذرکار، ۱۳۸۹). علی‌رغم وجود این نگرانی‌ها مدیران مؤسسات آموزشی به دلیل الزام قانونی برای تکمیل پرونده اساتید موظفانند امتیازات آموزشی آنها را محاسبه و تفسیر کنند. بر اساس آئین‌نامه ارتقا امتیازات آموزشی دارای بخش‌های مختلفی نظیر فعالیت‌های برنامه‌ریزی درسی، نظم و شؤون مدرسی و کمیت تدریس است. طبق گزارش حقیقی، اکبری، چرابین و کریمی (۱۳۹۹) از میان این مؤلفه‌ها به ارزشیابی‌های دانشجویی بین ۴۰ تا ۵۰ درصد وزن داده می‌شود. نکته اساسی اما اینست که اهداف ارزشیابی‌های دانشجویان از اساتید فقط به تکمیل پرونده استادان محدود نیست.

پژوهش‌های کلاسیک اولیه نشان‌گر پایایی و روایی ارزشیابی‌های دانشجویی بود (مارش، ۱۹۷۷، ۱۹۸۲). اما به مرور انتقادات مهمی به آنها وارد شد از جمله اینکه این پایایی و روایی ندارند؛ بیش از اندازه با نمرات همبستگی دارند، نزد اساتید محبوبیت ندارند و از عواملی نظیر ویژگی‌های درس، اندازه کلاس، مقطع درس، الزامی یا اختیاری بودن درس، ساعت کلاس در طول روز، و اینکه آیا درس برای با اول ارائه گردیده یا نه تأثیر می‌پذیرند (اشملکین، اسپنسر و گل‌مان، ۱۹۹۷). بولر^۱ (۲۰۱۲) با عنایت به این قبیل یافته‌ها اعتقاد دارد که با جستجو در منابع علمی می‌توان هر فرضیه‌ای را در زمینه ارزیابی‌های دانشجویان از اساتید تأیید یا رد کرد. نظر بولر این است که از ارزشیابی‌های دانشجویی، اطلاعات چندانی به دست نمی‌آید. زیرا شاید تعداد دانشجویان ارزیابی‌کننده کم باشد، یا اینکه دانشجویان به دلایلی به استاد علاقه ویژه داشته باشند و ارزیابی‌های او را بالا بزنند و در مقابل، نظرات کسانی که در طول ترم حذف درس کرده‌اند بررسی نشده باشد. بولر همچنین به این مسأله اشاره می‌کند که فرم‌های ارزشیابی دانشجویان از اساتید بیشتر به برداشت کلی دانشجو از نگرش‌ها و باورها و شخصیت استاد متکی هستند و به همین دلیل باید بر رفتارهای قابل مشاهده و نتایج قابل اثبات استاد تمرکز کرد.

از دیدگاه مدیریت دانشگاهی، دو پرسش اساسی وجود دارد که پاسخ به آنها سیاست‌ها و مقررات دانشگاهی در این زمینه را تحت تأثیر قرار خواهد داد. پرسش اول، این است که آیا احتمالی وجود دارد که استادان به صورت هشیارانه یا ناهشیارانه سبک تدریس خود را برای دریافت نمره ارزشیابی بالاتر تغییر دهند؟ مارش (۱۹۸۰) چهار دهه قبل مسأله را چنین تشریح کرد که به نظر منتقدان ارزشیابی‌های دانشجویی، اگر استادان می‌خواهند در تمام آیت‌ها ارزشیابی مطلوب‌تر داشته باشند فقط کافی است نمرات بالاتری بدهند و کار کمتری از دانشجویان بخواهند. مارش با این طرز فکر موافق نبود زیرا از نظر او درصد بالایی از واریانس نمرات ارزشیابی به متغیرهای زمینه‌ای بستگی دارد. مارش (۲۰۰۱) دو دهه بعد نیز به این مسأله اشاره کرد که گویی استادان می‌پندارند یکی از راه‌های به دست آوردن نمرات بالا در ارزشیابی‌های دانشجویی، تدریس

1- Buller

ساده و آرام محتوایی آسان است که یادگیری آن به تلاش کمی نیاز دارد و دانشجویان را به چالش نمی‌کشد. او با استناد به شواهد این باور را غلط و مخرب می‌داند. بنابراین، مسأله مهم برای مدیران دانشگاهی این است که از نتایج ارزشیابی‌های دانشجویی به گونه‌ای استفاده شود که استادان هشیارانه یا ناهشیارانه سبک تدریس خود را به دلیل موضوعات غیرآکادمیک تغییر ندهند. پرسش دوم این است که آیا امکان دارد کل دانشجویان کلاس به صورت آگاهانه یا ناآگاهانه با استاد درس وارد مبادله‌ای پنهان بشوند؟ این پرسش به ویژه در شرایطی اهمیت پیدا می‌کند که استادان به دلایلی نظیر کسب افتخارات آموزشی در پی نمره عالی در ارزشیابی‌های دانشجویی باشند یا اینکه دانشگاه‌ها مقررات سختگیرانه‌ای در مورد نتایج ارزشیابی‌های دانشجویی داشته باشند. در چنین شرایطی بعید نیست به مرور زمان معامله هشیارانه یا ناهشیارانه‌ای میان اساتید و دانشجویان کلاس شکل بگیرد تا هر دو گروه به نمرات مورد نیاز خود دست پیدا کنند.

در قسمت‌های قبل به عوامل متعدد اثرگذار بر ارزشیابی‌های دانشجویی اشاره شد. نقطه مشترک بسیاری از متغیرهای اثرگذار بر ارزشیابی‌های دانشجویی «نمره» است. نمره‌ها محصول و نتیجه مستقیم سبک نمره‌دهی خود استادان محسوب می‌شوند. در واقع، سبک نمره‌دهی استاد، چه به شکل هشیارانه و چه ناهشیارانه، ارزشیابی دانشجویان از تدریس را تحت تأثیر قرار می‌دهد. اما چرا در اینجا به دو شکل اثرگذاری هشیارانه یا ناهشیارانه تأکید می‌شود؟ واقعیت این است که مدرسان به هنگام ارزشیابی و تعیین نمره پایانی دانشجو به ابعادی غیر از یادگیری (نظیر میزان تلاش، میزان پیشرفت، تکمیل و تحویل کار، مشارکت، میزان توجه در کلاس، و تا حدی به شخصیت و رفتارهای دانشجو) نیز نمره می‌دهند. نمره‌دهی استادان در هر دوره یا مقطع آموزشی نیز متفاوت است. سرانجام، حتی در درون یک کلاس نیز نحوه نمره‌دهی آنها برای افراد متفاوت است (بروک‌هارت^۱، ۲۰۱۶). این تفاوت‌ها نه به دلیل میل یا سلیقه شخصی بلکه مبتنی بر تشخیص آکادمیک استاد می‌باشند. جالب اینکه این رویکرد نمره‌دهی مورد تأکید برخی از متخصصان سنجش و اندازه‌گیری نیز هست. به عنوان مثال، مارک ویلسون^۲ (۲۰۱۹) موافق است که فهرستی از ویژگی‌های شخصی که برای دستاوردها و موفقیت کلی در زندگی مهم‌اند و خود دانشجویان، والدین، مؤسسات آموزشی و کارفرمایان آینده نیز برای آنها ارزش قائل‌اند در نمره‌دهی مورد توجه قرار گیرند. اینکه استادان تا چه اندازه به هر کدام از عوامل متفرقه نمره می‌دهند به دیدگاه‌ها و رویکرد دانشگاهی آنها بستگی دارد. اما آنچه مسلم اینکه دانشجویان این تفاوت‌ها را درک می‌کنند. همان گونه که استادان در مورد کاهش «غیرمنصفانه» ارزشیابی‌های خود حساس‌اند دانشجویان نیز از افت نمره خود ناراحت می‌شوند. اما در اینجا طبق نظریه اسناد^۳ (نیوید^۴، ۲۰۱۸) پیش‌بینی می‌شود دانشجویانی که «اسناد بیرونی» دارند

1- Brookhart

2- Mark Wilson

3- Attribution theory

4- Nevid

علت کاهش نمره خود را در عوامل بیرونی جستجو کنند. طبیعی است که برای آنها استاد درس می‌تواند یکی از مقصرهای اصلی باشد! در چنین وضعیتی، این امر محتمل به نظر می‌رسد که کیفیت گذراندن تجربه درس با استاد، آمایه ذهنی ویژه‌ای را در مورد آن استاد در ذهن دانشجوی ایجاد کرده باشد. این آمایه ذهنی می‌تواند قضاوت‌های بعدی دانشجو را تحت تأثیر قرار دهد. بعید نیست دانشجو از ابتدای شروع نیمسال تحصیلی، کلیه رفتارها و اقدامات استاد را متناسب با آمایه‌ای که از قبل دارد تفسیر و برداشت کند. بنابراین، به لحاظ نظری در صورتی که نمره دانشجو از سطح مورد انتظار وی کمتر بود شاید به اسناد موقعیتی^۱ روی آورد و تدریس استاد را ضعیف‌تر ارزیابی کند.

این در حالی است که نمره دانشجویان به عنوان شاخص یادگیری‌های آنها تحت تأثیر عوامل بی‌نهایت متغیری است. به قول جان هتی^۲ (۲۰۰۹) تقریباً هر عاملی بر پیشرفت تحصیلی تأثیر دارد. از همین نکته و یافته ساده، دو اصل اساسی در زمینه ارزشیابی اساتید قابل استنتاج است. یکی اینکه مدیران دانشگاهی نباید استادان را بر اساس میزان پیشرفت تحصیلی دانشجویان قضاوت کنند. دوم اینکه مدیران دانشگاهی نباید با تصویب مقررات سختگیرانه در مورد ارزشیابی‌های دانشجویی، استادان را به نادیده انگاشتن واقعیت‌های محیطی، معیارهای آکادمیک و رفتارهای آموزشی اصیل سوق دهند. نولان و هوور^۳ (۲۰۰۸) از متخصصان ارزشیابی آموزشی هشدار می‌دهند که هدف غایی ارزشیابی بهبود تدریس است. طبیعتاً برای بهبود تدریس، مدرسان باید خطرپذیری کرده و رفتارها و روش‌های تازه‌ای را امتحان کنند. متأسفانه یکی از کارکردهای پنهان فرم‌های ارزشیابی دانشجویی اینست که مدرسان در گذار زمان به مجموعه‌ای از رفتارهای یکسان که در فرم ارزشیابی آمده توجه می‌کنند.

مسئله مرتبط دیگر با ارزشیابی‌های دانشجویی، پدیده تورم نمرات^۴ در سطح دانشگاه‌ها است. نمره همیشه برای دانشجویان جذاب بوده است و دانشگاه‌ها همواره کوشیده‌اند دانشجوی عالی را از خوب، متوسط و ضعیف متمایز کنند. در دانشگاه‌های ایران «معدل الف بودن» دانشجو نوعی اعتبار و ارزش اجتماعی دارد. در سنت بریتانیا و آمریکا نیز ساز و کارهایی برای تعیین دانشجویان معدل/الف وجود دارد (مکینون و کاشیف^۵، ۲۰۲۲). البته مسئله تورم نمرات لزوماً به تعداد بالای معدل الف‌ها مربوط نیست. معنای دقیق‌تر آن این است که در مؤسسه دانشجویان به صورت نظام‌مند نمرات بالاتری بگیرند در حالی که شایستگی‌های تحصیلی بیشتری کسب نکرده‌اند. چنین پدیده‌ای در همه جا به یک صورت وجود ندارد. به عنوان مثال، در دانشگاه‌های تراز اول تورم نمره کمتری وجود دارد (جفکات، میدلند و لیگو-بیکر^۶، ۲۰۲۰). اما به هر

1- Situational attribution

2- John Hattie

3- Nolan & Hoover

4- Grade inflation

5- Mackinnon & Kashif

6- Jephcote, Medland & Lygo-Baker

حال، اکثر نظام‌های دانشگاهی این پدیده را به طور مداوم رصد می‌کنند (مولر بن‌دیکت و گینزا، ۲۰۲۰). دلایل مختلفی نیز برای تورم نمرات قید شده است. دشواری ذاتی مباحث در یک رشته، سطح بالای کیفیت آموزش، سطح بالای استعداد تحصیلی دانشجویان در یک مؤسسه، و رقابت دانشجویان برای دستیابی به اهداف دور از جمله مواردی است که یورکی^۲ (۲۰۰۸) بدان‌ها اشاره کرده است. برخی از محققان نیز مدرسان مدعو یا پاره‌وقت را عامل وخیم‌تر شدن تورم نمرات در مؤسسه دانسته‌اند (چواری^۳، ۲۰۱۸). اما هرگز نباید مسأله تورم نمرات را به مدعو بودن مدرسان فرو کاست. تورم نمرات نه برای دروس استادان رسمی مؤسسه و نه برای مدعویان آن قابل توجیه نیست. مدیران دانشگاهی باید به نحوی از استادان خود حمایت روان‌شناختی کنند که فقط طبق معیارهای آکادمیک نمره‌دهی را انجام دهند. باید دقت کرد که در تورم نمره، دانشجویانی که شایستگی کمتری دارند نمرات بالاتری می‌گیرند و این امر می‌تواند بی‌انگیزگی دانشجویان اخلاق‌مدار را به دنبال داشته باشد و رفتارهای انتخاب درس و استاد را نیز در دانشجویان تحت تاثیر قرار دهد به گونه‌ای که به سوی استادان خوش‌نمره گرایش پیدا کنند.

مجموعه دغدغه‌ها و اشکالات فوق، برخی از مدرسان، گروه‌های آموزشی و دانشگاه‌ها را به اجرای روش‌های جایگزین نمره‌دهی سوق داده است. چانگ و کاتینو^۴ (۲۰۲۱) نمره‌دهی دانشجویان مهندسی به تکالیف خود را روش نسبتاً کارآمدی ارزیابی کرده‌اند که مزایایی از لحاظ بازخورد دهی به دانشجویان و افزایش شناخت استاد از دانشجویان دارد. دانشگاه کرنل^۵ نیز زمانی در برابر پدیده تورم نمرات دو خط‌مشی اتخاذ کرد: ۱) انتشار میانگین نمرات هر درس بر روی اینترنت؛ ۲) درج میانگین نمرات دانشجویان در درس بر روی کارنامه هر دانشجو. این کار صرفاً در مورد کلاس‌های بالای ۱۰ نفر در مقطع کارشناسی و بدون انتشار نام استاد انجام شد. البته، پیامد طرح مذکور استقبال دانشجویان از دوره‌های خوش‌نمره بود (بار، کادیلالی و زوسمان^۶، ۲۰۰۹). چنین یافته‌هایی به خوبی نشان می‌دهند که اجرای هر گونه طرح جدید باید توأم با احتیاط باشد. البته باید به این نکته نیز اشاره کرد که حداقل بر اساس برخی از شواهد، ممکن است آن گونه که برخی می‌پندارند «نمره» اولویت نخست دانشجویان نباشد. لستوساری، لاکانن و مورتون^۷ (۲۰۱۹) دریافته‌اند که قوی‌ترین عامل ماندگاری دانشجویان در یکی از گرایش‌های شیمی و عدم تغییر رشته، رویکرد عمیق به یادگیری بود. یافته‌های ایشان نشان می‌داد که بیش از خود نمره، ایجاد جو و فضای آموزشی حمایت‌گرایانه بر روی یادگیری‌های دانشجویان تأثیر می‌گذارد.

1- Müller-Benedict & Gaens

2- Yorke

3- Chowdhury

4- Chang & Coutinho

5- Cornell

6- Bar, Kadiyali & Zussman

7- Lastusaari, Laakkonen & Murtonen

مسأله «نمره» از این دیدگاه نیز قابل تحلیل است که عده‌ای از استادان در زمینه سنجش و ارزشیابی دانشجویان به انعطاف‌پذیری و سنجش مجزا^۱ قائل نیستند و انتظار دارند دانشجویان محتوای تخصصی دروس را به طور کامل یاد بگیرند و یادگیری‌های خود را روی برگه امتحان نشان بدهند. برخی از این استادان متعهد، نمره خالصی را که دانشجو دریافت کرده در کارنامه او درج می‌کنند و معتقدند که به عنوان مثال، بی هیچ ارفاقی نباید نمره ۷ را در کارنامه به ۹ تبدیل کرد. معدودی از استادان نیز وجود دارند که گمان می‌کنند توزیع نمرات دانشجویان کلاس باید «نرمال» باشد. این در حالی است که توزیع نمرات قبول‌شدگان باید نرمال باشد (یورکی، ۲۰۰۸) نه اینکه استاد بکوشد اجباراً کل نمرات دانشجویان کلاس را به صورت نرمال در آورد. در هر صورت، پیامدهای سبک نمره‌دهی خود استادان می‌تواند مخرب باشد. باید توجه کرد عنصر اساسی انصاف^۲ در سنجش نهایی دانشجویان این است که استاد فرصت‌های جایگزین را برای افرادی که شرایط خاص دارند تدارک ببیند به نحوی که استاندارد یا هدف محتوایی مخدوش نشود، همه به سمت آن هدف تلاش کنند و این هدف/استاندارد برای همه یکسان باشد (هرمان و کوک^۳، ۲۰۱۹). از آنچه گفته شد می‌توان نتیجه گرفت که ارتباط تنگاتنگی میان نمره‌دهی استادان، واکنش عاطفی و روان‌شناختی دانشجویان به نمره پایان‌ترم خود، و نیز نحوه ارزشیابی دانشجویان از اساتید وجود دارد. با توجه به ابعاد مختلف مسأله، هدف پژوهش حاضر بررسی روابط احتمالی میان نمره‌دهی استاد و ارزشیابی دانشجویان از استادان می‌باشد. به طور مشخص، تحلیل داده‌ها برای آزمودن فرضیه‌های زیر انجام شده است:

- ۱) با افزایش نمره پایان‌ترم دانشجو، نمره ارزشیابی او از آن استاد در نیمسال بعد به طور معناداری افزایش می‌یابد.
- ۲) با افزایش نمره پایان‌ترم دانشجو، نمره ارزشیابی او از آن استاد در دو نیمسال بعد به طور معناداری افزایش می‌یابد.
- ۳) با افزایش معدل نمرات کل کلاس درس، نمره ارزشیابی استاد در نیمسال بعد به طور معناداری افزایش می‌یابد.
- ۴) با افزایش معدل نمرات کل کلاس درس، نمره ارزشیابی استاد در دو نیمسال بعد به طور معناداری افزایش می‌یابد.
- ۵) در سنجش پیشرفت تحصیلی دانشجویان در کلاس‌های درس دانشگاهی، تورم نمرات وجود دارد. با توجه به اینکه در داده‌های تحقیق اطلاعات متغیرهایی نظیر جنسیت دانشجو، جنسیت استاد و تعداد ثبت‌نامی در کلاس نیز موجود می‌باشد فرضیه‌های آماری مربوط به این متغیرها نیز آزمون شده است.

1- Differentiated assessment

2- Fairness

3- Herman & Cook

روش پژوهش

پژوهش حاضر از نوع پنلی^۱ است. در این روش افراد یکسان در گذر زمان و در نقاط متوالی مورد مطالعه قرار می‌گیرند (جانسون و کریستنسون^۲، ۲۰۱۶). مدل‌های پنلی و روش‌های تحلیل داده‌های آن در منابعی نظیر رید و یی^۳ (۲۰۱۲) تشریح و در پژوهش‌هایی نظیر بریزاوی، لوکاتز و مولونتای^۴ (۲۰۲۰)، بین‌استاک و فلدمن^۵ (۲۰۱۶)، کاپلان، لیو و کاپلان^۶ (۲۰۱۴) و فلیکس^۷ و همکاران (۲۰۱۸) به کار رفته است. جامعه آماری پژوهش حاضر را کلیه دانشجویان شاغل به تحصیل در سه نیمسال آموزش پیش از پاندمی کرونا و در شرایط آموزش حضوری تشکیل داده‌اند. گزارش داده‌های مربوط به ارزشیابی دانشجویی و نمرات پایان‌ترم آنها از سامانه آموزش دانشگاه استخراج گردید. در فایل استخراج شده اطلاعات نمره‌ای که دانشجوی در پایان‌ترم گرفته بود، نمره‌ای که در ارزشیابی تدریس به استاد داده بود، به همراه کد درس و چند متغیر قابل تحلیل نظیر اندازه کلاس، میانگین نمره دانشجویان کلاس و جنسیت دانشجو موجود بود. با توجه به اینکه هر دانشجو در هر نیمسال به طور متوسط بیش از پانزده واحد درس دارد فایل داده‌های اصلی یا جامعه آماری شامل بیش از یکصد هزار جفت «نمره دانشجو-ارزشیابی استاد» بود. شرط انتخاب افراد در نمونه تحقیق اینگونه تعریف شد که دانشجو در سه نیمسال متوالی با یک استاد مشخص درس گرفته باشد. حجم نمونه مورد نیاز بر انجام تحلیل‌های اصلی با نرم‌افزار جی‌پاور^۸ برآورد شد. با توجه به اینکه داده‌ها از نوع پنل طولی^۹ بود و برای تعیین تأثیر نمره دانشجو در هر ترم نیز کاربرد مدل‌های اثرات ترکیبی چندسطحی^{۱۰} لازم بود لذا از خانواده آزمون^{۱۱} گزینه رگرسیون چندگانه^{۱۲} به عنوان روش تحلیل آماری انتخاب شد. اندازه اثر^{۱۳} متغیر مستقل برابر با ۰/۰۲، آلفا برابر با ۰/۰۱، توان آزمون برابر با ۰/۹۹، و تعداد متغیرهای پیش‌بین ۵ عدد در نظر گرفته شد. نمونه برآوردی نرم‌افزار ۱۶۸۳ نفر بود اما پس از شرط‌گذاری، غربال اولیه و حذف سطرهایی که حتی یکی از سلول‌های اساسی آن مفقود بود نمونه‌ای به تعداد ۱۹۷۰ نفر به دست آمد که خوشبختانه بیشتر از حداقل حجم برآوردی نرم‌افزار بود. ابزار گردآوری داده‌ها در این تحقیق، فرم رسمی ارزشیابی دانشجویان از اساتید در بازه چهار نیمسال بود که در هفده آیتم در مقیاسی نه درجه‌ای (از ۱۲ تا ۲۰) به همراه گزینه «نظری ندارم» در تمامی دانشکده‌ها اجرا شده بود. برای محاسبه

-
- 1- Panel study
 - 2- Johnson & Christensen
 - 3- Reed & Ye
 - 4- Berezvai, Lukats & Molontay
 - 5- Beenstock & Feldman
 - 6- Kaplan, Liu & Kaplan
 - 7- Felix
 - 8- GPower
 - 9- longitudinal/panel data
 - 10- Multilevel mixed-effects models
 - 11- Test family
 - 12- Multiple regression
 - 13- Effect size

نمره‌ای که دانشجوی در ارزشیابی به استاد داده بود میانگین ساده آیت‌های هدفه‌گانه نظرسنجی، بدون اعمال ضرایب در نظر گرفته شد.

یافته‌ها

آیا نمره‌ای که دانشجوی در یک نیمسال از استاد می‌گیرد در نیمسال یا نیمسال‌های آتی، ارزشیابی او از کیفیت تدریس استاد را تحت تأثیر قرار می‌دهد؟ پیش از انجام آزمون آماری لازم برای پاسخ به این سؤال، نمایش ساختار اولیه داده‌ها در جدول شماره ۱ لازم به نظر می‌رسد.

جدول ۱: ساختار داده‌های سه نیمسال نمرات دانشجو و نمرات ارزشیابی همان دانشجو از استاد

Table 1
students' academic progress scores and evaluation of the university professors' Framework of three semester's teaching

متغیرها Variables	نمره دانشجو students' scores	ارزشیابی استاد evaluation of professors'	اندازه کلاس class size	تعداد مردودی fails	ساعت امتحان time	معدل کلاس mean
نیمسال‌ها Semesters	1	3 2 1 3 2	3 2 1 3 2 1	3 2 1 3 2 1	3 2 1 3 2	3 2 1 3 2

کد یکتای دانشجو*استاد Unique ID of Student*Teacher

چنانکه در جدول شماره ۱ مشخص است ساختار داده‌ها از نوع طولی است. البته اطلاعات متغیرهای ثابتی نظیر جنسیت، وضعیت تأهل، رشته، دانشکده، و سهمیه کنکور دانشجو نیز وجود داشته که در این جدول درج نشده‌اند. برای انجام آزمون ابتدا ساختاردهی لازم انجام گردید و داده‌ها به صورت پنلی در آمد. سپس در محیط نرم‌افزار استاتا^۱ رگرسیون مدل اثرات ثابت^۲ اجرا شد. نمره ارزشیابی دانشجویان از اساتید به عنوان متغیر وابسته و نمره دانشجو به عنوان متغیر مستقل در نظر گرفته شد. نتیجه این تحلیل در جدول شماره ۲ ارائه شده است.

جدول ۲: مدل اثرات تصادفی برای پیش‌بینی نمره ارزشیابی استاد از روی نمره پایان‌ترم دانشجو

Table 2
Random effects model for estimate evaluation of university professors' teaching by students' academic progress scores

متغیر مستقل Independent variable	ضریب Coef.	خطای استاندارد میانگین Std. Err.	آزمون تی t test	مقدار احتمال P> t	فاصله اطمینان ۹۵٪ [95% Conf. Interval]
نمره دانشجو students' scores	0.029	0.0076	3.85	0.000	0.0145 0.0446
اندازه ثابت Fixed size	18.30	0.1205	151.81	0.000	18.070 18.543

ضریب همبستگی درون-رده‌ای = 0.6261 F=150.68 سطح معنی‌داری = 0.000

1- STATA

2- Fixed Effects (within) Regression

در جدول شماره ۲ ضریب رگرسیونی «نمره دانشجو» نشان می‌دهد که با یک واحد افزایش در نمره پایان‌ترم دانشجو نمره ارزشیابی استاد در نیمسال بعد حدود ۳٪ افزایش می‌یابد. بر این اساس با یک واحد افزایش در نمره پایان‌ترم دانشجو با اطمینان ۹۵ درصد می‌توان گفت افزایشی به بین ۱ تا ۴ درصد در نمره ارزشیابی دانشجو از استاد در دو نیمسال بعد به وجود آمده است. نکته مهم در مورد جدول شماره ۲ مقدار ضریب همبستگی درون-رده‌ای است که نشان می‌دهد ۶۲/۶۱ درصد از تغییرات در نمرات ارزشیابی استادان به علت تفاوت میان پنل‌ها (در اینجا نیمسال‌ها) می‌باشد. با توجه به تغییر در درس‌ها، تعداد ثبت‌نامی در درس‌ها و سایر شرایط متفاوت نیمسال‌ها و کلاس‌های درس، این مقدار قابل انتظار است.

در تحلیل قبل مشخص شد که نمره دانشجو در ترم قبل با ارزشیابی دانشجو از استاد در ترم بعد ارتباط دارد. به لحاظ منطقی این سؤال پیش می‌آید که آیا نمره تا نیمسال‌های بعدی نیز روی ارزشیابی دانشجو از استاد تأثیر می‌گذارد؟ به بیان دیگر و به لحاظ روان‌شناختی، سؤال این است که تأثیر آمایه‌ای که به سبب دریافت نمره‌ای خاص در ذهن دانشجو ایجاد گردیده چقدر یا تا چه مدتی پایدار است؟ برای پاسخ به این سؤال از رگرسیون چندسطحی اثرات ترکیبی^۱ از نوع اثرات تصادفی^۲ استفاده شده است. استفاده از این روش به این دلیل لازم بود که باید برای تعیین تأثیر نمره ترم ۱ روی نمره ارزشیابی استاد در ترم ۳، باید تأثیر نمرات نیمسال ۲ کنترل می‌شد. مدل اثرات تصادفی برای کنترل تأثیر متغیرهایی نظیر رشته، دانشکده، و سهمیه کنکور دانشجو نیز که کدهای متعددی داشته‌اند انجام شد اما در اینجا مهم‌ترین و اصلی‌ترین تحلیل که مرتبط با موضوع مقاله حاضر است گزارش شده است. به هر حال، در تمام این تحلیل‌ها نمره نیمسال ۱ به عنوان متغیر مستقل و نمره نیمسال ۳ به عنوان متغیر وابسته در مدل وارد شد. نتیجه تحلیل در جدول شماره ۳ گزارش شده است.

جدول ۳: پیش‌بینی ارزشیابی استاد در نیمسال ۳ از روی نمره دانشجو در نیمسال ۱

Table 3

Predicting evaluation of university professors' teaching in 3rd semester by student academic progress score in 1rd semester

متغیر مستقل Independent variable	ضریب Coef.	خطای استاندارد میانگین Std. Err.	آزمون z z test	مقدار احتمال P> z	فاصله اطمینان ۹۵٪ [95% Conf. Interval]
نمره نیمسال ۱ Semester 1 score	0.024	0.012	1.88	0.060	-0.001
اندازه ثابت Fixed measure	18.27	0.201	90.49	0.000	17.881

لگاریتم درست‌نمایی: -3995.92 کای‌اسکوئر والد: 3.53 سطح معنی‌داری: 0.0602

1- Mixed-effects ML regression

2- Random effects

فاصله اطمینان 95% Confidence interval %95	خطای معیار Standard error	برآورد Estimate	پارامترهای اثرات تصادفی random effects parameters
0.435	0.0679	0.0814	نمره نیمسال ۲ Semester 2 scor
		0.1719	واریانس سطح ۲ Variance at level 2
3.733	3.281	0.1154	واریانس خطا Error Variance
سطح معنی داری = 0.000			مقدار آماره کای اسکوتر = 23.84

بر اساس اطلاعات جدول شماره ۳ ضریب همبستگی درون رده‌ای که از نسبت واریانس سطح دو به مجموع واریانس سطح دو و واریانس خطا به دست می‌آید برابر با ۰/۰۴ و معنادار است که نشان می‌دهد ۴ درصد از واریانس ارزشیابی اساتید در نیمسال ۳ ناشی از اثر نمرات دانشجویان در نیمسال ۲ است. به هر حال، پس از کنترل تأثیر نمره نیمسال دوم، البته در سطح اطمینان ۹۰ درصد می‌توان گفت که با یک واحد افزایش در نمره دانشجو در نیمسال ۱ افزایش معناداری به اندازه ۰/۰۲۴ واحد (طبق فرم ارزشیابی نه درجه‌ای، معادل ۰/۰۵) در نمره ارزشیابی استاد در نیمسال ۳ به وجود آمده است. اما آیا متغیرهایی نظیر تعداد افتاده‌ها، اندازه کلاس (تعداد ثبت نام) نیز روی ارزشیابی استاد تأثیر معنادار دارند؟ برای پاسخ به این سؤال ابتدا کلیه متغیرهای وارد مدل گردید و به ترتیب، تأثیر متغیرهایی نظیر جنسیت دانشجو، وضعیت تأهل، تعداد واحد و غیره بررسی شد. کلیه متغیرهای غیراثرگذار طی چند مرحله از مدل خارج شدند و سرانجام عوامل اثرگذار بر ارزشیابی استاد در نیمسال سوم، پس از کنترل تأثیر متغیر «دانشکده»، به شرح جدول ۴ شناسایی شدند.

جدول ۴: پیش‌بینی ارزشیابی استاد در نیمسال سوم از روی نمره دانشجو، تعداد افتاده‌ها، اندازه کلاس و معدل

Table 4

Predicting evaluation of university professors' teaching in 3rd semester by students' academic progress scores, failes, class size and mean of class

فاصله اطمینان ۹۵% [95% Conf. Interval]	مقدار احتمال P> z	آزمون z z test	خطای استاندارد میانگین Std. Err.	ضریب Coef.	نیمسال‌ها Semesters	متغیرهای مستقل Independent varible
0.024	-0.364	0.706	-0.38	0.015	-0.005	نیمسال ۱ Semeter1
0.111	0.049	0.000	4.74	0.017	0.084	نیمسال ۲ Semeter2
0.060	-0.004	0.093	1.68	0.016	0.028	نیمسال ۳ Semeter3
0.076	0.010	0.011	2.55	0.0169	0.043	نیمسال ۱ Semeter1
0.022	-0.070	0.294	-1.05	0.024	-0.026	نیمسال ۲ Semeter2

0.010	-0.924	0.120	-1.55	0.026	-0.040	نیمسال ۳ Semeter3	
0.001	-0.284	0.078	-1.76	0.007	-0.013	نیمسال ۱ Semeter1	
0.028	-0.004	0.145	1.46	0.008	0.012	نیمسال ۲ Semeter2	اندازه کلاس Class size
0.027	-0.004	0.175	1.36	0.008	0.011	نیمسال ۳ Semeter3	
18.163	15.614	0.000	25.96	0.650	16.889	اندازه ثابت Fixed size	

لگاریتم درست‌نمایی: -3937.82 کای اسکوئر والد: 68.52 سطح معنی‌داری: 0.000

فاصله اطمینان 95% Confidence internal %95	خطای معیار Standard error	برآورد Estimate	پارامترهای اثرات تصادفی random effects parameters
1.422	0.171	0.266	0.493
			دانشکده: Faculty
			واریانس سطح ۲ Variance at level 2
5.562	3.140	0.107	3.334
			واریانس خطا Error Varriance
مقدار آماره کای اسکوئر=71.69 سطح معنی‌داری=0.000			

طبق اطلاعات جدول شماره ۴ در اینجا نیز ضریب همبستگی درون‌رده‌ای معادل $0/13$ است به این معنا که سیزده درصد از واریانس نمرات ارزشیابی اساتید در نیمسال سوم ناشی از تفاوت‌های دانشکده‌ای نمره‌دهی دانشجویان به اساتید است. طبیعتاً عوامل دانشکده‌ای شامل تفاوت‌هایی نظیر ماهیت تدریس استادان، سطح هوش و استعداد تحصیلی دانشجویان و انگیزش آنها می‌باشد. پس از کنترل تأثیر واریانس متغیر دانشکده ملاحظه می‌گردد که دیگر تأثیر نمره دانشجوی در نیمسال اول معنادار نیست. دلالت‌های این یافته در قسمت نتیجه‌گیری تبیین شده‌اند. اما چنانکه مشخص است تأثیر ضریب رگرسیونی نمره نیمسال ۲ (نزدیک‌ترین نیمسال) همچنان معنادار است و می‌توان گفت با یک واحد افزایش در نمره پایان‌ترم دانشجوی در نیمسال ۲ افزایش معناداری به اندازه $0/08$ واحد (طبق فرم ارزشیابی نه‌درجه‌ای، معادل $0/18$) در نمره ارزشیابی دانشجوی از استاد در نیمسال بعدی به وجود آمده است. یافته دیگر این است که ضریب رگرسیونی مربوط به نمره نیمسال سوم معنادار نیست. این یافته از این نظر منطقی است که موعد ارزشیابی دانشجویان از اساتید، قبل از برگزاری امتحانات پایان‌ترم است و طبیعتاً ارزشیابی دانشجویان از استاد، از نمره‌ای که هنوز مشخص نیست تأثیر نگرفته است.

ضریب رگرسیونی معنادار دیگر برابر با 0.43 و مربوط به «تعداد افتاده‌ها» است. این ضریب معنادار است اما باید به این نکته اشاره کرد که تعداد 70 نمره مساوی یا زیر $25/0$ در میان نمرات نیمسال اول وجود داشت که وقتی این نمره‌ها از تحلیل خارج شدند دیگر ضریب رگرسیونی «تعداد افتاده‌ها» معنادار نبود. فرضیه مهم دیگر در پژوهش حاضر این بود که با افزایش معدل نمرات کل کلاس درس، نمره ارزشیابی استاد در نیمسال‌های بعد به طور معناداری افزایش می‌یابد. برای آزمودن این فرضیه دو تحلیل لازم بود. در تحلیل اول معدل کل کلاس در نیمسال اول به عنوان متغیر مستقل و نمره ارزشیابی استادان در نیمسال دوم به عنوان متغیر وابسته در نظر گرفته شد. نتیجه در جدول شماره ۵ گزارش شده است.

جدول ۵: پیش‌بینی ارزشیابی استاد در نیمسال دوم از روی معدل کل کلاس در نیمسال اول

Table 5

Predicting evaluation of university professors' teaching in 2rd semester by total mean of class in 1rd semester					
متغیر مستقل Independent variable	ضریب Coef.	خطای استاندارد میانگین Std. Err.	آزمون z z test	مقدار احتمال P> z	فاصله اطمینان ۹۵٪ [95% Conf. Interval]
معدل کل کلاس در نیمسال ۱ total mean of class in 1rd semester	0.089	0.021	4.27	0.000	0.048
اندازه ثابت Fixed measure	17.283	0.328	52.62	0.000	16.64
لگاریتم درست‌نمایی: -3883.18	کای اسکوتر والد: 18.20	سطح معنی‌داری: 0.000			
پارامترهای اثرات تصادفی random effects parameters	برآورد Estimate	خطای معیار Standard error	فاصله اطمینان ۹۵٪ Confidence interna%95		
دانشکده faculty	0.134	0.088	0.036	0.478	
واریانس سطح ۲ Variance at level 2					
واریانس خطا Error Varriance	3.101	0.099	2.912	3.302	

مقدار آماره کای اسکوتر = 38.04 سطح معنی‌داری = 0.000

طبق اطلاعات جدول شماره ۵ ضریب همبستگی درون‌رده‌ای برای اثر «دانشکده» برابر با 0.4 می‌باشد و با توجه به معناداری آن به خوبی نشان می‌دهند که بخشی از واریانس نمرات ارزشیابی اساتید متأثر از تفاوت‌های دانشکده‌ای است. تأثیرگذاری متغیر دانشکده به دلایل مختلفی نظیر ماهیت متفاوت درس‌ها، سبک متفاوت تدریس استادان و نیز ساختمان و کیفیت تجهیزات کلاس‌ها می‌باشد. به هر حال، پس از کنترل تأثیرات مربوط به دانشکده، ضریب رگرسیونی مربوط «معدل کل کلاس» در نیمسال ۱ که برابر با 0.489 است در سطح اطمینان ۹۹ درصد معنادار می‌باشد که نشان می‌دهد تورم نمرات واحد کلاس نیز با افزایش نمرات ارزشیابی استادان همراه است.

در تحلیل دوم، معدل کل کلاس در نیمسال ۲ به عنوان متغیر مستقل و نمره ارزشیابی استادان در نیمسال ۳ به عنوان متغیر وابسته در نظر گرفته شد. نتیجه تحلیل در جدول شماره ۶ گزارش شده است.

جدول ۶: پیش‌بینی ارزشیابی استاد در نیمسال سوم از روی معدل کل کلاس در نیمسال دوم

متغیر مستقل Independent variable	ضریب Coef.	خطای استاندارد Std. Err. میانگین	آزمون z z test	مقدار احتمال P > z	فاصله اطمینان ۹۵٪ [95% Conf. Interval]
معدل کل کلاس در نیمسال ۲ Total mean of class in 2nd semester	0.07	۰/۰۲۵	2.81	0.005	0.021
اندازه ثابت Fixed size	17.28	0.449	38.44	0.000	16.399
لگاریتم درست‌نمایی: -3967.56		کای اسکوتر والد: 7.87		سطح معنی‌داری: 0.005	
پارامترهای اثرات تصادفی random effects parameters	برآورد Estimate	خطای معیار Standard error	فاصله اطمینان ۹۵٪ [95% Conf. Interval]		
دانشکده faculty	0.523	0.279	0.183	1.491	
واریانس سطح ۲ Variance at level 2					
واریانس خطا Error Variance	3.448	0.111	3.237	3.673	

مقدار آماره کای اسکوتر = 71.35 سطح معنی‌داری = 0.000

طبق اطلاعات جدول شماره ۶ ضریب همبستگی درون‌رده‌ای برای اثر «دانشکده» برابر با ۰/۱۳۱ می‌باشد و با توجه به معناداری آن به خوبی نشان می‌دهد که درصد قابل توجهی از واریانس نمرات ارزشیابی اساتید متأثر از تفاوت‌های دانشکده‌ای است. شاید یکی از دلایل افزایش سه برابری همبستگی درون‌رده‌ای در این تحلیل این باشد که برخلاف داده‌های دو نیمسال قبلی که تعطیلات سه ماهه تابستان میان آن قرار گرفته بود در اینجا فاصله‌ای بین دو نیمسال نبوده است. در اینجا نیز ضریب رگرسیونی مربوط «معدل کل کلاس» در نیمسال ۲ برابر با ۰/۰۷ و معنادار است که نشان می‌دهد تورم نمرات واحد کلاس پس از گذشت یک نیمسال با افزایش نمرات ارزشیابی استادان در نیمسال بعد همراه بوده است.

پنجمین فرضیه مهم در پژوهش پیش رو این بود که در زمینه ارزشیابی و سنجش پیشرفت تحصیلی دانشجویان در کلاس‌های درس دانشگاهی، تورم نمرات وجود دارد. برای آزمودن این فرضیه، وضعیت معدل نمرات کلاس‌های درس دانشجویان به تفکیک دانشکده بررسی شد. برای این منظور ابتدا گزارش معدل کلیه دروس ۱۰ نیمسال اخیر استخراج شد. در مرحله اول، داده‌ها شرط‌گذاری و غربال شد. شروط

اصلی برای غربال معدل کلاس‌ها این بود که تعداد ثبت‌نامی در هر کلاس ۱۰ نفر و بالاتر باشد و نوع درس‌ها نیز به دروس اصلی و تخصصی محدود شد. معدل کلاس‌های دانشکده‌ها به شکل جدول شماره ۷ دسته‌بندی و آزمون شده‌اند.

جدول شماره ۷: تعداد دروس، میانگین و انحراف معیار نمرات کلاس‌های دانشگاه در ده نیمسال

Table 7

Number of courses, mean and standard deviation of classes in 10 semesters

سطح معنی‌داری Sig.	آزمون t t test	تفاوت میانگین	نمره ملاک Criterion score	انحراف معیار Standard deviation	میانگین میانگین‌ها population mean (μ)	تعداد دروس Number of course	آماره‌ها Statistic دانشکده‌ها facultys
0.480	0.711	0.104	15	1.194	15.104	8790	ریاضی، فیزیک، شیمی، مهندسی‌ها شیمی، مهندسی، Mathemathis Physic Chemistry Engineering علوم انسانی، ادبیات، معماری و هنر
0.0001	8.527	0.554	16	0.373	16.554	7154	Human Science Litrature Architecture And Ar

* این اطلاعات مربوط به ۱۰ نیمسال (از نیمسال دوم سال ۱۳۹۵ تا نیمسال اول سال ۱۴۰۰) می‌باشد.

** معدل کل کلاس/درس بدون در نظر گرفتن نمرات زیر ۰/۲۵ مبنای محاسبه بوده است. همچنین فقط دروس اصلی و تخصصی در نظر گرفته شده‌اند.

مقایسه نمرات کلاس‌ها (دروس) به یک نمره ملاک معتبر نیاز بود. برای این منظور، پس از ملاحظه معدل کلی دانشگاه در دهه گذشته برای گروه دانشکده‌های ریاضی، فیزیک، شیمی و مهندسی‌ها نمره معیار ۱۵ و برای گروه‌های علوم انسانی، ادبیات و معماری و هنر نمره ملاک ۱۶ به عنوان میانگین جامعه^۱ در نظر گرفته شد. چنانکه از اطلاعات و آزمون‌های تی تک‌نمونه‌ای در جدول شماره ۷ بر می‌آید در دانشکده‌های علوم انسانی، ادبیات و معماری و هنر میانگین معدل کلاس‌ها به طور معناداری از نمره ملاک بالاتر است. اما در دانشکده‌های ریاضی، فیزیک، شیمی و دانشکده‌های مهندسی تفاوت معناداری مشاهده نمی‌شود. این تحلیل به خوبی نشان می‌دهد که حداقل در برخی از رشته‌ها نشانه‌هایی از وجود پدیده تورم نمرات وجود دارد که دلالت‌های آن در قسمت بعد تبیین و تشریح شده است.

1- population mean (μ)

بحث و نتیجه‌گیری

محققان و متخصصان آموزش عالی ایران پدیده‌های مختلفی را از سطح کلان تا خرد مورد بررسی قرار می‌دهند. شناسایی راهکارهای بهبود کیفیت آموزش (محسن‌زاده، ادیب و حسینی، ۱۳۹۱؛ هادی، رضاپور و سلمان‌پور، ۱۳۹۵)، واکاوی الگوی تدریس اثربخش در کلاس‌های دانشگاهی (محمدی خانقاهی و حسین‌زاده، ۱۳۹۴)، تعیین ویژگی‌های استاد خوب (کمالی و پورچناری، ۱۳۹۴)، و شناسایی شایستگی‌های اختصاصی استادان (کریمی و احمدی، ۱۴۰۰) همگی در راستای بهبود فرایندهای یاددهی-یادگیری در کلاس‌های درس است. در مورد برخی از پدیده‌ها، وجود ارتباط میان سیاست‌های سطح کلان و عملکرد مجریان در سطح خرد حساس‌تر است. در پژوهش حاضر وجود ارتباط احتمالی میان سیاست‌های ارزشیابی استادان (سطح کلان) و رویه‌های سنجش و نمره‌دهی استادان (سطح خرد) مد نظر بوده است. این فرض محتمل است که قوانین و مقررات مربوط به ارزشیابی کیفیت تدریس به تغییر معیارها و رویه‌های اساتید برای سنجش پیشرفت تحصیلی دانشجویان منجر شده باشد. ارتباط میان نمرات پایان‌ترم دانشجویان و ارزشیابی آنها از تدریس اساتید و نیز تورم نمرات را می‌توان به عنوان شواهدی برای وجود این مسأله تفسیر کرد.

ضرایب رگرسیونی مربوط به نمره دانشجو و نیز معدل کلاس به خوبی نشان می‌دهد که به افزایش نمرات چه در واحد دانشجو و چه در واحد کلاس درس، شاهد افزایش معناداری در نمرات ارزشیابی دانشجویان از اساتید طی دو نیمسال بعدی هستیم. با در نظر گرفتن مجموع ضرایب رگرسیونی می‌توان ادعا کرد که پس از کنترل اثر تفاوت‌های میان پنل‌ها (نیمسال‌ها) با یک واحد افزایش در نمره پایان‌ترم دانشجو، ارزشیابی او از استاد در نیمسال بعدی حدود پنج درصد افزایش می‌یابد. همسو با یافته‌های تحقیق حاضر مبنی بر ارتباط میان ارزشیابی‌های دانشجویی از استادان و نمرات پیشرفت تحصیلی دانشجویان، می‌توان به نتایج برخی از پژوهش‌ها اشاره کرد. مارشال (۲۰۲۲) در پژوهشی به این نتیجه رسید که مثبت یا منفی بودن نظرات کیفی دانشجو با نتیجه ارزشیابی کمی او از استاد همبستگی معناداری دارند. در واقع، دانشجو بر مبنای تصور مثبت یا منفی ویژه‌ای که از استاد دارد او را ارزیابی می‌کند. در پژوهش حاضر نیز می‌توان گفت ضرایب رگرسیونی معنادار برای نمره دانشجو و معدل کلاس نشان می‌دهند که نگاه مثبت یا منفی دانشجو یا واحد کل که متأثر از نمره بالا یا پائین است، در گذر زمان ثابت مانده و حتی می‌تواند تا دو نیمسال بعد روی ارزیابی از استاد تأثیر بگذارد. در همین راستا، پژوهش پارک و چو (۲۰۲۲) نیز معلوم شد که اگر دانشجویان به دلایلی میان نمره ایدآل و واقعی خود شکاف یا اختلافی را حس کنند استاد را تنبیه کرده یا پاداش می‌دهند. البته این محققان گزارش کرده‌اند که در صورت اختلاف مثبت یا همان تورم نمره، پاداش کاهش می‌یابد اما تنبیه کاهش نخواهد داشت. تبیین محققان این است که تورم نمرات، انتظارات

دانشجویان را تغییر خواهد داد. ایراد روش شناختی پژوهش مزبور این است که معدل کلی دانشجویان را به عنوان متغیر مستقل در نظر گرفته است. در واقع، در پژوهش مذکور بیشتر به نقش سطح و توان تحصیلی دانشجویان توجه شده است.

از سوی دیگر، تحلیل نمرات دروس اصلی و تخصصی دانشجویان نشان داد که در دانشکده‌های علوم انسانی، ادبیات و معماری و هنر پدیده تورم نمرات وجود دارد. شکی نیست که دلایل پدیده تورم نمرات بسیار متعدد و گوناگون‌اند. وجود استادان مدعو در مؤسسه (چواری، ۲۰۱۸)، تفاوت‌های بین‌دانشگاهی در زمینه کیفیت آموزش (بن‌استاک و فلدمن، ۲۰۱۶)، و انگیزش بالا و پیشرفت تحصیلی مطلوب دانشجویان (گرزلی^۱، ۲۰۱۱) همگی از عوامل طبیعی افزایش میانگین نمرات در سطح مؤسسه‌اند. همچنین باید اشاره کرد بر خلاف پیامدهای مخرب گسترده‌ای که شیخ‌الاسلامی، یونسی و دلاور (۱۴۰۰) برای تورم نمرات گزارش کرده‌اند در برخی شرایط، تورم نمره مطلوب و دارای پیامدهای مثبت است. به عنوان مثال، کری و کاریفو (۲۰۱۲) دریافتند نمراتی که فراگیران در ترم نخست می‌گیرند نقش مؤثری در بهبود نمرات بعدی، وضعیت روانی و نگرشی آنها در درس‌ها دارد. با چنین استدلالی بود که در دانشگاه ام‌آی‌تی^۲ از دهه ۱۹۸۰ در گروه مهندسی برق و علوم کامپیوتر از نمره محرمانه^۳ استفاده شد (هایلپرین^۴، ۱۹۸۱). نمره‌ای کیفی که در نیمسال اول به اطلاع دانشجویان رسانده می‌شود اما در کارنامه او صرفاً به صورت قبول/عدم ثبت درج می‌شود. هدف از اجرای این طرح کاستن از اضطراب نودانشجویان و فشار کاری آنها به هنگام انتقال از دبیرستان به دانشگاه و تشویق آنها به کسب دانش پایه، رشد مهارت‌های اجتماعی، آمادگی برای انطباق با تفاوت‌های میان دانشگاه و مدرسه، و دادن حق انتخاب بیشتر برای شرکت دوره‌ها و درس‌ها بود (دانشگاه ام‌آی‌تی، ۲۰۰۰). پس از حذف نمرات محرمانه در نیمسال دوم، افزایش معناداری در معدل افراد مشاهده شد. تأثیرات مثبت دیگری نیز به ویژه در مورد دانشجویانی که با نمرات ضعیف‌تر وارد دانشگاه ام‌آی‌تی شده بودند و آمادگی مطالعاتی کمتری داشتند مشاهده شد (هریس^۵، ۲۰۱۱). سرانجام، باید اشاره کرد که امکان گرفتن نمرات بالاتر در دانشگاه، حداقل به گروه‌هایی از دانشجویان انگیزه می‌دهد که با هدف جذب و استخدام موفق در آینده برای مطالعه و یادگیری، پشتکار بیشتری داشته باشند. در آموزش عالی ایران مسأله موفقیت در آزمون‌های استخدامی یا کنکورهای تحصیلات تکمیلی نیز قابل طرح است، زیرا به عنوان مثال، معدل دانشجویان دوره کارشناسی در قبولی آنها طی کنکور کارشناسی ارشد اثرگذار است.

1- Gerzeli

2- MIT

3- hidden grade

4- Hailperin

5- Harris

اگر چه تورم نمرات ابعاد مثبتی دارد که بدان‌ها اشاره شد اما مسأله ارتباط میان نمرات دانشجویان و ارزشیابی آنها از تدریس اساتید قابل چشم‌پوشی نیست زیرا دانشگاه‌های ایران از نتایج ارزشیابی‌های دانشجویی برای تصمیم‌گیری‌های پرسنلی اساتید استفاده می‌کنند و متأسفانه رویه‌ها و دستورالعمل‌های آنها در این زمینه نیز بسیار ساده و کمیت‌گرایانه است. به عنوان مثال، در اکثر دانشگاه‌ها میانگین ساده ارزشیابی‌های استاد در دروس مختلف مد نظر قرار می‌گیرد. برخی دانشگاه‌ها ارزشیابی درس را بر اساس تعداد واحد آن وزن‌دهی می‌کنند. برخی دانشگاه‌ها در ارزشیابی دروس تحصیلات تکمیلی، حد نصاب تعداد دانشجویان شرکت‌کننده در ارزشیابی را ۲ در نظر می‌گیرند و این نمره ارزشیابی دوفره به عنوان یک نمره مستقل روی میانگین نهایی استاد تأثیر می‌گذارد. در برخی دانشگاه‌ها معیار ثابتی برای مقایسه نمره ارزشیابی استادان وجود دارد. در دانشگاه‌هایی که دقت نظر بیشتری وجود دارد ارزشیابی استاد با میانگین دانشکده یا گروه مقایسه می‌شود اما باز نحوه این مقایسه در هاله‌ای از ابهام است. به عنوان مثال، هرگز مشخص نمی‌کنند که چقدر فاصله از میانگین گروه، قابل توجیه و قابل پذیرش است. در واقع، قضاوت مراجعی نظیر هیأت جذب دانشگاه بیشتر بر اساس برداشت کلی و یا تصویری است که از استاد مورد نظر وجود دارد. در برخی دانشگاه‌ها، به ویژه دانشگاه‌های علوم پزشکی، فرم‌های ارزشیابی از تدریس بر اساس ماهیت درس‌ها (نظری، عملی، بالینی و غیره) متفاوت است. اما در دانشگاه‌های وزارت علوم فرم‌ها تنوع کمتری دارند. سرانجام، در برخی از دانشگاه‌ها فرم ارزشیابی واحدی در تمام دانشکده‌ها به اجرا در می‌آید! هیچ کدام از این رویه‌ها و نمونه‌هایی که گفته شد با روش‌های مورد تأکید متخصصان ارزشیابی اعضای هیأت علمی (بولر، ۲۰۱۲؛ گالزیو، ۲۰۱۱) سنخیت و سازگاری کامل ندارند.

بنابر آنچه گفته شد از آنجا که اعتبار و پایایی ارزشیابی‌های دانشجویی به عوامل بسیاری بستگی دارد حداقل در تفسیر و بهره‌برداری از نتایج این ارزشیابی‌ها باید جانب احتیاط را رعایت کرد. همچنین نباید فرایندها و رویه‌های ارزشیابی به گونه‌ای باشد که ابعاد اساسی و بنیادین آموزش عالی نظیر سنجش پیشرفت تحصیلی دانشجویان در معرض تغییر قرار گیرد. اولین دلالت عملی یافته‌های این تحقیق، ضرورت سیاست‌گذاری مناسب در زمینه ارزشیابی اساتید، با هدف پیشگیری از تورم نمرات در مؤسسه است. در پژوهش حاضر شواهدی برای وجود ارتباط میان نمره دانشجو و ارزشیابی استاد ارائه شد. چنانکه پارک و چو (۲۰۲۲) نیز تصریح کرده‌اند بعید نیست استادان با دادن نمرات بالاتر به دانشجویان پاداش (نمره ارزشیابی بالاتر) دریافت کنند. در واقع، امکان دارد حداقل در مورد طیف خاصی از گروه‌ها، معامله‌ای هشیارانه یا ناهشیارانه میان استاد و دانشجویان جریان پیدا کند. این در حالی است که طبق قواعد صریح سنجش و اندازه‌گیری پیشرفت تحصیلی، اگر چه در نظر گرفتن ویژگی‌هایی به غیر از پیشرفت تحصیلی قابل توجیه است (ویلسون، ۲۰۱۹) اما نمره باید بیش از هر چیز انعکاسی از یادگیری‌های واقعی دانشجویان در آن درس باشد. متأسفانه در دانشگاه‌های ایران، به هنگام بررسی پرونده تمدید پیمانی، تبدیل وضعیت

استخدامی یا ارتقای مرتبه علمی، صرفاً به خود نمرات ارزشیابی تدریس توجه می‌شود و همین امر می‌تواند به بروز رفتارهای آموزشی غیرسازنده منجر شود. صرف نظر از این مسأله، همان‌طور که تحلیل‌های ناماری و ناکاک^۱ (۲۰۲۲) نشان می‌دهد تأکید بیش از اندازه مؤسسه روی نمرات ارزشیابی دانشجویان از اساتید می‌تواند به تشدید مسأله تورم نمرات منجر شود. تورم نمرات می‌تواند آفت بسیار مهلک و مخربی برای نظام دانشگاهی ایران باشد. به ویژه با در نظر گرفتن اینکه در کارنامه‌های دانشگاهی (ریز نمرات دانشجو) رتبه یا جایگاه وی در کلاس درج نمی‌شود. دومین دلالت عملی یافته‌های این تحقیق ضرورت تدوین و تصویب دستورالعمل‌های دقیق‌تر در شورای آموزشی دانشگاه‌ها می‌باشد. آنگونه که مارتینز، شوینگ و گولداشمیت^۲ (۲۰۱۶) می‌گویند بهره‌گیری از اندازه‌گیری‌های چندگانه‌ای باشد که بر مبنای منطقی و مورد توافق ذینفعان، وزن‌دهی شده باشند. در هر صورت، چنانکه بویسن، کلی، ریزلی و کاسنر^۳ (۲۰۱۴) تذکر داده‌اند تفاوت‌های ناچیز در میانگین ارزشیابی دانشجویی اساتید به قضاوت‌های جدی و سرنوشت‌ساز منجر گردیده و به همین دلیل، برای برداشت‌های صحیح از این داده‌ها باید تحلیل‌های آماری دقیقی انجام داد. این ابعاد مسأله باید از منظر آثار احتمالی منفی ارزشیابی‌ها در تخریب حس امنیت شغلی اساتید و فرهنگ سازمانی دانشگاه در شوراهای آموزشی و هیأت‌های جذب مد نظر قرار گیرد. مسأله آثار ارزشیابی‌ها بر امنیت روانی اساتید به قدری اهمیت دارد که بسیاری از دانشگاه‌های جهان حتی نظرات کیفی (نوشتاری) دانشجویان را قبل از ارسال برای استاد درس، هم با استفاده از فناوری هوش مصنوعی و هم به صورت دستی، غربال و پالایش می‌کنند و حتی در صورت نیاز برای اصلاح به دانشجو باز می‌گرداند (کانینگهم^۴ و همکاران، ۲۰۲۲). البته، در مطالعه اخیر کمتر از پنج درصد نظرات، غیرسازنده و نیازمند اصلاح بودند. اما به هر حال، باید توجه کرد که حتی یک نظر منفی غیرمنصفانه می‌تواند آثار مخرب جبران‌ناپذیر داشته باشد. در پایان مجدداً به نظر هربرت مارش (مارش، ۱۹۷۷؛ مارش، اورال و کسلر^۵، ۱۹۷۹؛ مارش، ۲۰۰۱) اشاره می‌شود که بیش از سه دهه به پژوهش در این زمینه پرداخته است. او معتقد است نظام‌های ارزشیابی نباید به گونه‌ای طراحی شوند که استادان به این نتیجه برسند که بهترین روش، ارائه نمرات بالا و کاستن از بار کاری دانشجویان است. مورین و مارش^۵ (۲۰۱۴) وجود الگوی مناسب در این زمینه را ضروری می‌دانند. مولر^۶ (۲۰۱۲) با عنایت به در هم تنیدگی عوامل اثرگذار بر ارزشیابی‌های دانشجویی تأکید می‌کند که کارگروه‌های بررسی صلاحیت باید فقط بر ابعادی از نمرات ارزشیابی‌های دانشجویی تمرکز کنند که دانشجو در بهترین موقعیت قضاوت قرار دارد و هرگز از این نمرات به عنوان تنها روش تعیین کیفیت تدریس استفاده نکنند. سرانجام، بولر توصیه می‌کند که نمرات ارزشیابی دانشجویی با در نظر گرفتن و

1- Maamari & Naccache

2- Martínez, Schweig & Goldschmidt

3- Cunningham

4- Marsh, Overall & Kesler

5- Morin

مقایسه عواملی نظیر گروه آموزشی، اندازه کلاس، درس‌های مشابه، و روند تغییرات در گذر زمان تفسیر کنند. علی‌رغم بلوغ نظام دانشگاهی ایران، هنوز اقدامات بسیاری در این زمینه لازم است. از محدودیت‌های این پژوهش می‌توان به دو مورد اشاره کرد. (۱) در غربال داده‌ها برای انتخاب نمونه آماری، از میان ده‌ها هزار مورد فقط مواردی انتخاب شدند که دانشجو در سه نیمسال متوالی با یک استاد مشخص درس گذرانده بود. (۲) در زمینه ارتباط میان نمره دانشجویان و ارزشیابی آنها از اساتید باید به ویژگی‌های مهم شخصیتی استادان نظیر برون‌گرایی و وجدان کاری آنها (یارمحمدزاده، فیضی، محمدیان و آق‌آتابای، ۱۳۹۶) توجه شود. به بیان آماری، این احتمال نیز هست که معناداری ضرایب رگرسیونی به پایداری همزمان ویژگی‌های شخصیتی استادان و سبک و استعداد تحصیلی دانشجویان مربوط باشد.

References

منابع

- شیخ الاسلامی، ندا، یونسی، جلیل، و دلاور، علی. (۱۴۰۰). پیامدهای تورم نمره در آموزش عالی و راهکارهای مقابله با آن. *اندیشه‌های نوین تربیتی*، ۱۷(۱)، ۷-۳۸.
- ضیایی، مسعود، میری، محمدرضا، حاجی‌آبادی، محمدرضا، آذرکار، سیده قدسیه، و پویا اشیک (۱۳۸۶). دیدگاه اساتید و دانشجویان دانشگاه علوم پزشکی بیرجند نسبت به ارزشیابی اساتید توسط دانشجو. *مجله علمی دانشگاه علوم پزشکی بیرجند*، ۱۳(۴)، ۹-۱۵.
- کریمی، منوچهر، و احمدی، سیدعلی اکبر. (۱۴۰۰). آرایه مدل شایستگی اختصاصی استادان دانشگاه. *آموزش و ارزشیابی*، ۱۴(۵۳)، ۱۸۰-۱۵۹.
- محمدی خوانقاهی، محمد، و حسین‌زاده، امیدعلی. (۱۳۹۴). تدوین و اعتباریابی مدل تدریس اثربخش برای اساتید دانشگاه تبریز. *آموزش و ارزشیابی*، ۸(۳۱)، ۷۱-۹۱.
- کمالی، یحیی، و پورچناری، عبدالله. (۱۳۹۴). بررسی ویژگی‌های استاد خوب از دیدگاه دانشجویان علوم سیاسی دانشگاه شهید باهنر کرمان. *آموزش و ارزشیابی*، ۸(۳۰)، ۱۱۷-۱۴۰.
- محسن‌زاده کریمی، ساناز، ادیب، یوسف، و حسینی، محمدحسین. (۱۳۹۱). مقایسه عوامل آموزشی موثر بر کیفیت آموزشی دانشکده علوم انسانی دانشگاه آزاد اسلامی واحد تبریز از دیدگاه استادان و دانشجویان. *فصلنامه علوم تربیتی*، ۵(۱۸)، ۱۳۶-۱۲۱.
- هادی، رضا، رضاپور، یوسف، و سلمان‌پور، صدیقه. (۱۳۹۵). شاخص‌های سنجش کیفیت آموزش عالی بر اساس مدل کانو در دانشگاه ارومیه. *آموزش و ارزشیابی*، ۹(۳۴)، ۳۵-۵۰.
- یارمحمدزاده، پیمان، فیضی، ایوب، محمدیان، کویستان، آق‌آتابای، آمنه. (۱۳۹۶). رابطه ویژگی‌های شخصیت و سبک رهبری کلاس درس با کیفیت تدریس اساتید دانشگاه به روش خودارزیابی، ارائه یک مدل ساختاری. *آموزش و ارزشیابی*، ۱۰(۳۹)، ۲۹-۵۱.

- Adams, Gary. L., & Engelmann, Siegfried. (1996). *Research on Direct Instruction: 25 Years Beyond DISTAR*. Seattle: Educational Achievement Systems.
- Addison, W. E., Best, J., & Warrington, J. D. (2006). Students' perceptions of course difficulty and their ratings of the instructor. *College Student Journal*, 40 (-): 409-416.
- Alauddin, M., & T. Kifle. (2014). Does the Student Evaluation of Teaching Instrument Really Measure Instructors' Teaching Effectiveness? An Econometric Analysis of Students' Perceptions in Economics Courses. *Economic Analysis and Policy*, 44(2): 156-168. doi:10.1016/j.eap.2014.05.009.
- Atanasio, M., & Capursi, V. (Editors), (2011) *Statistical Methods for the Evaluation of University Systems*. Berlin, Heidelberg: Springer.
- Bachen, Christine. M., McLoughlin, Moira. M., & Garcia, Sara. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48(3): 193-210.
- Bar, T., Kadiyali, V., & Zussman, A. (2009). Grade Information and Grade Inflation: The Cornell Experiment. *The Journal of Economic Perspectives*, 23(3): 93-108.
- Beenstock, M., & Feldman, D. (2016). Decomposing university grades: a longitudinal study of students and their instructors. *Studies in Higher Education*, DOI: 10.1080/03075079.2016.1157858
- Beran, T., & Violato, C. (2009). Student ratings of teaching effectiveness: Student engagement and course characteristics. *Canadian Journal of Higher Education*, 39(1): 1-13.
- Berezvai, Z., Lukats, Gergely. D., & Molontay, R. (2020). Can professors buy better evaluation with lenient grading? The effect of grade inflation on student evaluation of teaching. *Assessment & Evaluation in Higher Education*, DOI:10.1080/02602938.2020.1821866
- Boysen, Guy. A., Kelly, Timothy. J., Raesly, Holly. N., & Casner, Robert. W. (2014). The (mis)interpretation of teaching evaluations by college faculty and administrators, *Assessment & Evaluation in Higher Education*, 39(6): 641-656.
- Buller, Jeffrey. L. (2012). *Best practices in faculty evaluation: A Practical guide for academic leaders*. San Francisco. Jossey-Bass.
- Brookhart, Susan. M. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, Vol. XX, No. X, pp. 1-46.
- Carey, Th., & Carifio, J. (2012). The Minimum Grading Controversy: Results of a Quantitative Study of Seven Years of Grading Data from an Urban High School. *Educational Researcher*, 41(6): 201-208.
- Chowdhury, F. (2018). Grade Inflation: Causes, Consequences and Cure. *Journal of Education and Learning*, 7(6): 86-92.
- Chang, K., & Coutinho, Lucas de Lemos. (2021). *Administering Homework Self-Grading for an Engineering Course*. College Teaching, DOI: 10.1080/87567555.2021.1957666

- Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, 31(1): 16-30.
- Coffey, M., & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education. *Assessment & Evaluation in Higher Education*, 26(1): 89-93.
- Cunningham, Samuel., Laundon, Melinda., Cathcart, Abby., Bashar, Md Abul., & Nayak, Richi. (2022). First, do no harm: automated detection of abusive comments in student evaluation of teaching surveys. *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2022.2081668
- Danielson, Charlotte., & McGreal, Thomas. L. (2000). *Teacher evaluation to enhance professional practice*. Princeton, NJ: Association for Supervision and Curriculum Development.
- Felix, E. D., Binmoeller, C., Sharkey, J. D., Dowdy, E., Furlong, M. J., & Latham, N. (2018). The influence of different longitudinal patterns of peer victimization on psychosocial adjustment. *Journal of School Violence*, 1-15.
- Gerzeli, Simone. (2011). *The Impact of Teaching Evaluation: Factors that Favour Positive Views from Student Representatives*. In: Massimo Attanasio & Vincenza Capursi (Editors), *Statistical Methods for the Evaluation of University Systems*. Berlin, Heidelberg: Springer.
- Golzio, Luigi. Enrico. (2011). *The Assessment of University Teaching by Students: The*
- Hadi, R., Rezapoor, Y., & salmanpoor, S. (2016). Evaluation Indexes in Higher Education Based on Kano's Model at Urmia University. *Journal of Instruction and Evaluation*, 9(34): 35-50 [In Persian].
- Hailperirn, Mlax. (1982). SCC worth jumps despite cash logs. *The Tech*, 102(1): 1-24.
- Harris, Gregory. A. (2011). *The Impact of Hidden Grades on Student Decision-Making and Academic Performance: An Examination of a Policy Change at MIT*. Toronto, Ontario: AIR Forum.
- Hattie, John. (2009). *Visible learning: A synthesis of over 800 meta-analysis relating to achievement*. New York: Routledge.
- Herman, Joan., & Cook, Linda. (2019). *Fairness in Classroom Assessment*. In: Susan Brookhart & James McMillan (Editors): *Classroom Assessment and Educational Measurement*. New York: Routledge.
- Jephcote, C., Medland, E., & Lygo-Baker, S. (2020). Grade inflation versus grade improvement: Are our students getting more intelligent? *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2020.1795617
- Johnson, R. Burke., & Christensen, Larry. (2016). *Educational research: Quantitative, qualitative, and mixed approaches*. Los Angeles: Sage.

- Kaplan, D. S., Liu, X., & Kaplan, H. B. (1995). Explaining Dropout Behavior: A Longitudinal Study of Direct and Indirect Psychosocial Effects of Seventh and Eighth Grade Experiences. *Educational Research and Evaluation*, 1(3):205-225.
- Karimi, M., and Ahmadi S. (2021) Providing Exclusive Competency Model for University Faculty Members. *Journal of Instruction and Evaluation*, 14(53): 159-180. [In Persian].
- Kamali, Y., Pourchenary, A. (2015). The Characteristics of a Good Teacher from the Political Sciences Students' Point of view at Shahid Bahonar University of Kerman. *Journal of Instruction and Evaluation*, 8(30): 21179-140. [In Persian].
- Lakeman, Richard., Coutts, Rosanne., Hutchinson, Marie., Lee, Megan., Massey, Debbie., Nasrawi, Dima., & Fielden, Jann. (2021). Appearance, insults, allegations, blame and threats: an analysis of anonymous non-constructive student evaluation of teaching in Australia, *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2021.2012643
- Landrum, R. Eric., & Braitman, Keli. A. (2008). The effect of decreasing response options on students' evaluation of instruction. *College Teaching*, 56(4): 215-217.
- Lastusaari, Mika., Laakkonen, Eero., & Murtonen, Mari. (2019). Persistence in studies in relation to learning approaches and first-year grades: a study of university chemistry students in Finland. *Chem Educ. Res. Pract*, 20(3): 452-467.
- Linse, Angela. R. (2016). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54: 94-106.
- Maamari, B. E., & Naccache, H. S. (2022). The impact of grade inflation on teachers' evaluation: A quantitative study conducted in the context of five Lebanese universities. *Journal of Global Education and Research*, 6(2): 192-205.
- Mackinnon, Sean. P., & Kashif, Shazia. (2022). *A Short History of Grading Practices at Dalhousie University (1901-2021)*. PeerRef, 10.35542/osf.io/fmgv8
- Marsh, Herbert. W. (1977). The Validity of Students' Evaluations: Classroom Evaluations of Instructors Independently Nominated as Best and Worst Teachers by Graduating Seniors. *American Educational Research Journal*, 14(4): 441.
- Marsh, Herbert. W., & Overall, Jesse. U., & Kesler, Steven. P. (1979). Long-Term Stability of Students' Evaluations: A Note on Feldman's "Consistency and Variability among College Students in Rating Their Teachers and Courses. *Research in Higher Education*, 10(2): 139-147.
- Marsh, Herbert. W. (1982). Factors Affecting Students' Evaluations of the Same Course Taught by the Same Instructor on Different Occasions. *American Educational Research Journal*, 19(4): 485-497.
- Marsh, Herbert. W. (2001). Distinguishing between Good (Useful) and Bad Workloads on Students' Evaluations of Teaching. *American Educational Research Journal*, 38(1): 183-212.
- Marsh, Herbert. W. (1977). The Validity of Students' Evaluations: Classroom Evaluations of Instructors Independently Nominated as Best and Worst Teachers

- by Graduating Seniors. *American Educational Research Journal*, 14(4): 441-447.
- Marsh, Herbert. W., & Roche, Lawrence. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal*, 30(1): 217-251.
- McEwan, Patrick. j., & et al. (2006). Roommate effects on grades: evidence from first-year housing assignments. *Research in higher education*, 47(3): 347-370.
- McMillan, Amy., Solanelles, Pol., & Rogers, Bryan. (2021). Bias in student evaluations: Are my peers out to get me? *Studies in Educational Evaluation*, 70 (101032): 1-11.
- Marshall, Kim. (2009). *Rethinking teacher supervision and evaluation: How to work smart, build collaboration, and close the achievement gap*. San Francisco: Jossey-Bass.
- Marshall, Pablo. (2022). Contribution of open-ended questions in student evaluation of teaching. *Higher Education Research & Development*, 4(6): 1992-2005.
- Martínez, José. Felipe., Schweig, Jonathan., & Goldschmidt, Pete. (2016). Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy. *Educational Evaluation and Policy Analysis*, 38(4): 738-756.
- Mengel, Friederike., Sauermann, Jan., & Zolitz, Uif. (2019). *Gender Bias in Teaching Evaluations*. Germany: IZA Institute of Labor Economics.
- MIT. (2000). *Freshman pass/no record grading and advanced placement policy: Report of a special subcommittee on the Committee on the Undergraduate Program*. Cambridge, Massachusetts: The Massachusetts Institute of Technology.
- Mohammadi khanghahi, M., & Hosinzadeh, O. (2015) Presenting and Verifying an Effective Teaching Model for Teachers at Tabriz University. *Journal of Instruction and Evaluation*, 8(31): 71-91 [In Persian].
- Mohsenzade karimi, S., Adib, y., & Hoseini, M. (2012). The comparison study of educational factors affecting the educational quality of human science faculty from professors and student's points of view, Tabriz Islamic Azad University, *Journal of Instruction and Evaluation*, 5(18): 121-136. [In Persian].
- Morin, A. J. S., & Marsh, H. W. (2014). Disentangling Shape from Level Effects in Person-Centered Analyses: An Illustration Based on University Teachers' Multidimensional Profiles of Effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1): 39-59.
- Muller-Benedict, Volker., & Gaens, Thomas. (2020). A new explanation for grade inflation the long-term development of German university grades, *European Journal of Higher Education*, DOI: 10.1080/21568235.2020.1718516
- Nevid, Jeffrey. S. (2018). *Essentials of Psychology: Concepts and Applications*. Australia: CENGAGE Learning.

- Nolan, James. H., & Hoover, Linda. A. (2008). *Teacher supervision & Evaluation: Theory into Practice*. Wiley.
- Park, Byungjin., & Cho, Joonmo. (2022). How does grade inflation affect student evaluation of teaching? *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2022.2126429
- Radmacher, Sally. A., & Martin, David. J. (2001). Identifying Significant Predictors of Student Evaluations of Faculty Through Hierarchical Regression Analysis. *The Journal of Psychology: Interdisciplinary and Applied*, 135(3): 259-268.
- Reed, W. R., & Ye, H. (2011). Which panel data estimator should I use? *Applied Economics*, 43(8): 985-1000.
- Schmelkin, Liora. Pedhazur., Spencer, Karin. J., & Gellman, Estelle. S. (1997). Faculty Perspectives on Course and Teacher Evaluations. *Research in Higher Education*, 38(5): 575-592.
- Sheikholeslami, N., Younesi, J., & Delavar. Ali (2021). The consequences of grade inflation in higher education and the strategies to deal with it. Quarterly Journal of New Thoughts on Education. *Journal of Instruction and Evaluation*, 14(1): 7-38. [In Persian].
- Spooren, P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in Educational Evaluation*, 54: 43-49.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54: 22-42.
- Wilson, Mark. (2019). *Emerging Issues in Classroom Assessment*. In: Susan Brookhart & James McMillan (Editors): Classroom Assessment and Educational Measurement. New York: Routledge.
- Yarmohammadzadeh, P., Feyzi, A., Mohamadian, K., & Aghatabay, A. (2017). Investigating the Relationship among University Teachers' personality Traits, Classroom Leadership Style, and Efficacy through Self-Evaluation: Providing a Structural Model. *Journal of Instruction and Evaluation*, 10(39): 29-51 [In Persian].
- Yorke, Mantz. (2008). *Grading Student Achievement in Higher Education: Signals and Shortcomings*. London: Routledge.
- Ziaee, M., Miri, M., Shima. Hajiabadi, M., Azarkar, S., & Ashik, p. (2008). ZiaSurveying Birjand faculty members', paramedical and nursing students' viewpoints on educational evaluation plan. *Journal of Instruction and Evaluation*, 13(4): 9-15 [In Persian].

