

تحلیلی بر منابع سوگیری و خطا در ارزشیابی دانشجویان از اثربخشی تدریس استادان

An Analysis of the Sources of Bias and Error in Students' Evaluation of Teaching

تاریخ پذیرش مقاله: ۹۲/۰۴/۱۷

تاریخ دریافت مقاله: ۹۲/۰۱/۰۶

Naser Shirbagi

ناصر شیربگی *

Abstract: To evaluate faculty members' teaching effectiveness, the most common method of data collection is using student assessment scales. The potential for false results and the ambiguity of the source evaluation alone can lead to bias and error. This paper by adapting an analytic method using data collected from literature review, articles and books examined this question: "which errors threaten students' perception of their teachers teaching?" Results showed that there are at least eight sources of bias and error in perception of students as follows: error in instrument content, error in interpretation of the meaning of ratings, showmanship, error in instrument reliability, mixed purposes of evaluation, inconsistent methods of instrument administration, common rating error effects, and errors in data implementation. Therefore, to reflect a border continuum of the teaching characteristics of all sorts of teaching, current evaluation systems need to be reviewed and changed. Finally, the paper proposes a model to solve this problem.

چکیده: در زمینه ارزشیابی اثربخشی تدریس اعضای هیأت علمی، معمولترین شیوه گردآوری داده‌ها به کارگیری مقیاس‌های ارزیابی توسط دانشجویان است. پتانسیل‌های موجود برای نتایج نادرست و ابهام در این منبع به تنهایی می‌تواند به سوگیری و خطای مختلف ارزشیابی منجر شود. این مقاله، با روش توصیفی تحلیلی و با استفاده از داده‌های گردآوری شده از پیشینه تحقیقاتی، مقالات و کتب موجود به بررسی این سؤال پژوهشی پرداخته است که چه خطاهایی ادراک دانشجویان از تدریس استادانشان را تهدید می‌کند. نتایج نشان داد که حداقل هشت منبع سوگیری و خطا در ارزشیابی دانشجویان از تدریس وجود دارد که عبارتند از: خطا در محتوای ابزار ارزشیابی، خطا در تفسیر معانی و نتایج ارزشیابی، نمایش، خطا در پایایی ابزار، اهداف چندگانه ارزشیابی، ناهماهنگی در به کارگیری ابزارها، تأثیرات ارزشیابی، خطا در کاربرد داده‌ها. از این رو، نظام‌های ارزشیابی کنونی برای انعکاس پیوستار گسترده‌تری از صفات تدریس در تمام گونه‌های موجود تدریس، نیازمند تغییر و بازنگری هستند. این مقاله در انتها الگویی برای حل این مشکل پیشنهاد می‌نماید.

Key Words: Evaluation, teaching effectiveness, bias and error, Students

واژگان کلیدی: ارزشیابی، اثربخشی تدریس، سوگیری و خطا، دانشجو

مقدمه

تدریس در فرهنگ دانشگاه‌ها و مؤسسات آموزش عالی دارای اهمیت و ارزش بالایی است، اما ضروری است که رؤسای دانشگاه‌ها، مدیران گروه‌ها و هم‌تایان اعضای هیأت علمی دربارهٔ اثربخشی تدریس داوری نمایند. این داوری‌ها بر بررسی نتایج عملکرد سالانه، فرایند ارتقا و اشتغال تأثیر می‌گذارد.

ارزشیابی دانشجویان از عملکرد استادان، یک عنصر جدایی‌ناپذیر از فعالیت‌های نظام آموزش عالی بیشتر کشورهاست و هستهٔ اصلی بیشتر فعالیت‌های تضمین کیفیت در دانشگاه‌ها را تشکیل می‌دهد، و کمتر مؤسسه آموزش عالی وجود دارد که توجه کافی به نتایج آن نداشته باشد. معمولاً از نتایج چنین ارزشیابی‌هایی به منظور نیل به سه هدف عمدهٔ ذیل استفاده می‌شود:

۱. کمک به سیستم ارزشیابی مدیریتی از طریق اندازه‌گیری اثربخشی تدریس که معیار مهمی برای تصمیمات در زمینهٔ استخدام اعضای هیأت علمی، اعطای پایه سالیانه، افزایش حقوق و ارتقای رتبه آنان است.

۲. کمک به دانشجویان برای انتخاب دروس و استاد درس: بیشتر دانشجویان تازه وارد، نتایج ارزیابی دانشجویان قدیمی‌تر از اساتید را به عنوان منابع با ارزشی برای چنین تصمیم‌گیری‌هایی در نظر می‌گیرند (ویلهم، ۲۰۰۴).

۳. کمک به پیشرفت حرفه‌ای خود اعضای هیأت علمی تا آنها بتوانند شیوه‌های تدریستان را با فراهم کردن بازخوردهای حاصل شده بهبود ببخشند.

کمیت تحقیق در خصوص ارزیابی دانشجویان از تدریس مدرسان نشان دهنده اهمیت و توجه روز افزون به آن در نظام‌های آموزش عالی است. هرچند ارزشیابی تدریس و موضوعات مرتبط با آن در کشورمان تاریخچهٔ کوتاهی دارد، اما به نظر می‌رسد، در دههٔ گذشته کیفیت تدریس، همواره یکی از نگرانی‌های اصلی دست‌اندرکاران نظام آموزش عالی بوده است. تأکید آیین‌نامه‌های وزارت علوم، تحقیقات و فناوری (۱۳۸۷، ۱۳۸۹) بر به کارگیری نتایج حاصل شده از این راهبرد، در فرم‌های ویژه ارتقا و تبدیل وضعیت اعضای هیأت علمی، مبین این موضوع است.

موضوعاتی مانند روایی و پایایی ارزشیابی‌ها و عناصر تشکیل‌دهندهٔ تدریس اثربخش، مباحث چندان جدیدی نیستند اما پاسخگویی به شبهات مربوط به آن در کشور هنوز ارزشمند است. از لحاظ سابقه، مقیاس ارزیابی مدرسان توسط دانشجویان

به عنوان ابزار اصلی اندازه‌گیری اثربخشی تدریس در چهار دهه گذشته غالب شده است. برخی محققان اذعان می‌کنند که استفاده از نظرات دانشجویان عموماً روش معتبری برای ارزیابی اثربخشی تدریس اساتید است؛ اما تعدادی دیگر دریافته‌اند که تردیدهایی هنوز وجود دارد. با وجود استدلال‌های منتقدان و نگرانی‌های مؤسسات آموزش عالی درباره روایی و پایایی این روش، هنوز معمول‌ترین ابزار اندازه‌گیری اثربخشی تدریس در نظام‌های آموزش عالی جهان استفاده از نظرات دانشجویان است. در یک دهه گذشته، تمایلاتی برای ترکیب نظرات دانشجویان با سایر منابع اطلاعاتی دیگر به منظور غنای شواهد به کار رفته در ارزیابی دروس و کیفیت تدریس ظهور پیدا کرده است (ناپر و کرانتون، ۲۰۰۱). اگر هدف آن باشد که دست‌اندرکاران امر قضاوت در تدریس به نظام ارزشیابی عملکرد آموزشی اطمینان داشته و مدرسان نیز عملکرد تدریس خود را بر اساس بازخورد فراهم شده از نظام‌های ارزشیابی بهبود بخشند، باید خطاها در انجام ارزیابی دانشجویان از تدریس استادان به حداقل برسد. هدف این مقاله، ترویج این ایده است که به دلیل پیچیدگی فرایند تدریس باید با بهره‌گیری از منابع مختلف، تصویری جامع‌تر و معتبرتر از اثربخشی تدریس در مقایسه با تکیه صرف بر ارزیابی دانشجویان فراهم کرد؛ اما باید تنها آن دسته از منابع با هم ترکیب شوند که شواهد موثقی برای روایی آنها در دسترس باشد. نگارنده امیدوار است که تحلیل جوانب مختلف ارزشیابی دانشجویان از تدریس اساتید، به تصمیم‌گیرندگان و خود مدرسان کمک نماید تا با چالش‌های مربوط به خطاها و سوگیری‌های احتمالی در ارزشیابی‌ها به نحو شایسته‌ای کنار آیند.

با توجه به هدف فوق سؤالات پژوهشی عبارتند از:

۱- چه خطاها یا سوگیری‌هایی ادراک دانشجویان از تدریس استادان را تهدید می‌کند؟

۲- چگونه می‌توان تأثیر خطاها و سوگیری‌های احتمالی را به حداقل رساند؟

روش‌شناسی تحقیق

طرح تحقیق کمی با روش توصیفی تحلیلی است. گردآوری داده‌ها از طریق جست‌وجوی مقالات انگلیسی و فارسی با موضوعات ارزشیابی از تدریس و خطاهای ارزشیابی در سایت‌های معتبر علمی و پایگاه‌های اطلاعاتی مربوط به حوزه تعلیم و تربیت و تحلیل داده‌ها با استفاده از روش تحلیل محتوای کیفی صورت گرفت.

منابع خطا در ارزشیابی

راهبرد ارزیابی دانشجویان از استادان ضرورتاً باید با ایده‌های مختلف در زمینه ویژگی‌های تدریس خوب و الگوهای یاددهی- یادگیری در ارتباط باشد. شاید پیچیدگی مفهوم تدریس در دانشگاه به بهترین وجهی در تعریف ذیل از کاسی^۱، جنتیل^۲ و بیگر^۳ (۱۹۹۷) ارائه شده باشد: "یک فعالیت خلاق که به منظور تسهیل یادگیری دانشجویان، پروراندن توانایی‌ها و تمایلاتشان برای انجام کار علمی و رشد و پیشرفتشان به عنوان یک شخص به صورت کل طراحی می‌شود" (ص ۶۱). تلاش‌ها برای بررسی متغیرهایی که نمرات ارزیابی دانشجویان از اثربخشی تدریس استادان را توضیح می‌دهند بیش از چند دهه سابقه دارد. به عقیده مک فرسون^۴ (۲۰۰۶) متأسفانه روش‌های آماری به کار رفته در تحقیقات اولیه در این زمینه دارای کاستی‌های جدی بوده‌اند. مشکل دیگر بیشتر تحقیقات انجام شده، عدم دسترسی به داده‌های بیش از دو یا سه ترم تحصیلی متوالی بوده است. سؤال مهمی که مطرح می‌شود این است که آیا مهم‌ترین متغیرهای اثربخشی تدریس ارزیابی می‌شوند یا اینکه تنها متغیرهایی که قابل اندازه‌گیری‌اند با اهمیت تلقی می‌شوند؛ چرا که در ارزیابی دانشجویان از عملکرد مدرسان، تنها یک نوع داده برای داوری در مورد اثربخشی تدریس فراهم می‌شود. تحلیل‌ها نشان می‌دهد که پتانسیل‌های موجود برای نادرستی و ابهام در این منبع به تنهایی می‌تواند به حداقل هشت منبع خطای ارزشیابی که در ذیل به آنها پرداخته می‌شود، منجر شود.

۱- خطا در محتوی ابزار ارزشیابی

بیشتر ابزارهای ارزشیابی تهیه شده برای ارزشیابی آموزش بر اساس تدریس کلاسی بنا نهاده شده‌اند. و اغلب ممکن است تدریس در محیط‌های آزمایشگاهی، عملی یا صحرایی ارزشیابی نشود؛ علاوه بر این، امکان دارد که حیطه‌های رفتاری تدریس کلاسی و گویه‌های مربوط به آن، به صورت نامناسب برای ارزشیابی تدریس در چنین موقعیت‌هایی به کار گرفته شود. بدون وجود یک پیوند منطقی بین رفتار تدریس کلاسی و رفتارهایی که در سایر موقعیت‌های غیرکلاسی به کار می‌رود، نمی‌توان فرض کرد که ابزارهای

1. Casey
2. Gentile
3. Bigger
4. McPherson

ارزشیابی تدریس قابلیت تعمیم دارند. از آن جایی که حیطه‌های تدریس متفاوتند، از این رو، به مقیاس‌های ارزشیابی جداگانه‌ای برای هر موقعیت تدریس، از جمله سخنرانی در کلاس، محیط آزمایشگاهی و کار عملی، نیاز است.

جدول (۱) مشهورترین حیطه‌های تدریس برگرفته شده از پژوهش‌های اصیل و مبانی نظری حوزه علوم تربیتی و روان‌شناسی را خلاصه نموده است. وتروبا^۱ و رایت^۲ (۱۹۷۵) نتایج ۲۱ پژوهش منتشر شده در خصوص ارزشیابی دانشجویان از تدریس را مرور کرده‌اند. از ۴۰ معیاری که آنها فهرست کرده‌اند ۹ مورد توسط محققان دیگر هم ذکر شده است (سترا^۳، ۱۹۷۶).

جدول (۱) حیطه‌های تدریس برگرفته شده از مبانی نظری

حیطه‌ها	منابع و حیطه‌ها			
	دیکنسون ^۴	وتروبا ^۱	سترا ^۳	داس ^۵
سازمان‌دهی دوره	-	موضوع و دوره	سازمان- ساختار- شفافیت	رئوس مطالب- ساخت‌دهی
توانایی تدریس	روش‌های تدریس	سازمان‌دهی- بیان و تفسیر	توانایی خطابه	سبک و روش‌های تدریس
نمره‌دهی و بازخورد	آزمون به عنوان تجارب یادگیری	انصاف در نمره‌دهی	نمره‌دهی و امتحان	ارزشیابی عینی، بازخورد.
تعامل استاد و دانشجو	تعامل استاد- دانشجو	انعطاف‌پذیر	تعامل- سازگاری دانشجو	حساسیت، دسترسی، پاسخگویی
بارکاری و سختی دوره	-	-	بارکاری، سختی دوره	-
اشتیاق و انگیزش	اشتیاق	اشتیاق- تشویق تفکر	-	پویایی، اشتیاق
تسلط بر موضوع.	شایستگی.	دانش موضوع	-	-

1. Wotruba
2. Wright
3. Centra
4. Dickinson
5. Wotruba
6. Das
7. Hilderbrand

هفت حیطه از مواردی که معمولاً مورد توافق بیشتر صاحب نظران علوم تربیتی و روان شناسی است در ستون اول جدول آمده است. ستون های بعدی حیطه هایی را که توسط مطالعات تحلیل عاملی مورد تأیید قرار گرفته اند نشان می دهد. آن حیطه هایی از تدریس که در بین موقعیت های مختلف تدریس مشابه هستند می تواند به صورت مشترک در مقیاس های ارزشیابی مورد استفاده قرار گیرد.

در جدول (۲) توصیف مختصری از مقیاس هایی که برای تدریس در سه موقعیت مختلف می توان مورد استفاده قرار داد آمده است. شورا های نظارت و ارزشیابی آموزشی یا هیأت های ممیزه دانشگاه ها ممکن است وزن های مختلفی را برای حیطه های گوناگون و مقیاس ها تعیین نمایند تا به یک ابزار ارزشیابی کلی برای فرایند آموزش دست یابند.

جدول (۲) محتوای مقیاس ارزشیابی برای حیطه های تدریس در سه محیط

محیط تدریس			حیطه های تدریس
عملیات میدانی	لابراتوار	کلاس	
-	-	X	۱ توانایی تدریس - سخنرانی
-	X	-	۲ توانایی تدریس - آزمایشگاه
X	-	-	۳ توانایی تدریس - عملیات میدانی
X	X	X	۴ سازماندهی دوره
-	X	X	۵ انتخاب و استفاده از رسانه
X	X	X	۶ ارزشیابی عملکرد دانشجویان
X	X	X	۷ تعامل استاد و دانشجو
X	X	X	۸ اشتیاق / انگیزش
X	X	X	۹ تسلط بر موضوع

خطا در محتوای ابزار ارزشیابی ممکن است در ارزشیابی موقعیت ها یا شرایطی رخ دهد که خارج از کنترل مدرس است. سؤال کردن از دانشجویان در موارد ذیل از آن جمله است: ارزیابی مسائل مربوط به برنامه درسی (مانند تناسب پیش نیاز دروس)؛ ارزیابی محیط یادگیری (مانند روشنایی کلاس، دما و تهویه)؛ یا گزارش خود-ارزیابی علایق یادگیری و نتایج عملکرد آنان. هرچند ممکن است که موارد بیان شده حیطه های مفیدی برای ارزشیابی برنامه های درسی باشد، اما آنها بر مبنای متغیرهای مربوط به دانشکده ها و دانشجویان نهاده شده اند و نه بر عملکرد مدرسان؛ از این رو،

ناظران آموزشی باید از تفاوت مقیاس‌های رفتاری تدریس مانند «مدرس از فهرست مطالب درس پیروی می‌کند» با سایر رفتارهای مربوط به دانشجو مانند «من درخواست کمک می‌کنم وقتی که مطالبی را نمی‌فهمم» یا توصیف کننده‌های فضای فیزیکی کلاس مانند «محیط کلاس یک فضای مناسب را فراهم ساخت» آگاه باشند.

۲- خطا در تفسیر معانی و نتایج ارزشیابی‌ها

استفاده از هنجارها در تفسیر ارزشیابی دانشجویان از تدریس برای مقایسه‌های معنادار و منصفانه، اساسی است؛ با این حال، ماهیت هنجارهای گروهی ممکن است در خطای تفسیرها مشارکت داشته باشد؛ برای مثال، چون مطالعات نشان داده است که مدرسان در کلاس‌های دروس علوم پایه ارزشیابی‌های کمتری نسبت به آنهایی که در علوم انسانی و اجتماعی تدریس می‌کنند دریافت می‌کنند؛ لذا معنی هنجارهای دانشگاه مورد سؤال است (گروسینگ^۱، ۱۹۹۴). علاوه بر متغیرهای مربوط به محتوای برنامه درسی متغیرهای متعدد دیگری ممکن است در خطاهای تفسیر هنجاری مشارکت داشته باشند که سطح پختگی دانشجویان، مقطع تحصیلی، نوع دروس (اختیاری/پایه) و متغیرهای آموزشی از آن جمله‌اند (براسکمپ^۲، براندنبرگ^۳ و اوری^۴، ۱۹۸۴). هنجارهای مقیاس ارزشیابی باید در چندین نرم نرم گروهی تهیه شوند تا مقایسه‌ها معنادارتر شوند؛ به عنوان مثال: نرم‌های برای دروس پایه/اختیاری، دروس نظری/آزمایشگاهی و عملی؛ مقطع تحصیلی و نرم‌هایی برای هر گروه، دانشکده و دانشگاه. پیامدهای هر نوع خطایی ممکن خیلی گسترده‌تر از تفسیرهای فردی ارزشیابی‌ها باشد. تجارب منفی از نظام ارزشیابی ممکن است باعث کاسته شدن اطمینان از فرایند ارزشیابی دانشکده شود و اعضای هیأت علمی و مدیران را دل سرد نماید.

۳- نمایش یا مجری‌گری

تأثیر نحوه ارائه درس توسط یک مدرس بر یادگیری دانشجویان و ادراک آنان از رضایت از مدرس در تحقیقات مختلف اثبات شده است (وار^۵ و ویلیامز^۶، ۱۹۷۵). ارائه‌هایی که به دلیل ویژگی‌های اغواگری و سرگرم‌کنندگی به خاطر می‌مانند ممکن

-
1. Grussing
 2. Braskamp
 3. Brandenburg
 4. Ory
 5. Ware
 6. Williams

است خطاهایی را در ارزشیابی برخی از جنبه‌های اثربخشی تدریس مانند «تحت پوشش قرار دادن محتوای برنامه درسی» ایجاد نمایند؛ اما جنبه‌های از شخصیت و سبک مدرس همانند میزان مراقبت و پشتیبانی تلقینی از یادگیری دانشجویان و همچنین رفتارهای انگیزشی که قابل مشاهده است واقعاً در ایجاد یک محیط یادگیری مهیج و شاداب مشارکت می‌کند؛ بنابراین، ممکن است به عنوان جزئی از عوامل مرتبط با تدریس اثربخش به نظر برسد (هیلدبراند^۱، ویلسون^۲ و دینت^۳، ۱۹۷۱). مقیاس‌های تهیه شده باید از ارزشیابی دانشجویان از شخصیت، جذبه (کاریزما) یا صفات مشابه با آن خودداری کند و تنها آن دسته از ویژگی‌های مدرسان (مانند: تعامل معلم و دانشجویان آموز یا دغدغه برای یادگیری آنها) که ثابت شده با تدریس اثربخش مرتبط هستند باید مورد تأکید قرار گیرند.

۴- خطا در پایایی ابزار

وقتی که گویه‌های متعددی در پرسشنامه‌های معمولی - که به طور عددی نمره‌گذاری شده‌اند - ظاهر می‌شوند، گویه‌های مشابه یا آنهایی که معانی مشابه دارند به عنوان مقیاس‌های مشخص اندازه‌گیری حیطه‌های عملکرد تدریس جمع بسته می‌شوند و یا میانگین آنها محاسبه می‌گردد؛ لذا باید همبستگی‌های بالایی از ارزشیابی دانشجویان را نشان دهد. برآوردهای مربوط به سازگاری درونی پرسشنامه‌ها باید به استفاده‌کنندگان ابزارها گزارش شود تا پایایی چنین مقیاس‌های چند گویه‌ای را اثبات کند؛ به علاوه، استفاده از ابزارهای درجه‌بندی باید ثبات را در به‌کارگیری‌های متوالی نشان دهد. داده‌های به دست آمده از آزمون پایایی به روش بازآزمایی اجازه می‌دهد که تضمین‌های بیشتری از کیفیت ابزار ارائه شود؛ بویژه، هنگامی که داده‌ها نشان داده‌اند که صفات یا رفتارهای اندازه‌گیری شده در طول زمان ثابت می‌مانند. مقیاس‌های ارزشیابی‌های دانشجویی باید نشان دهند که همبستگی بالایی با سایر ابزارهای استاندارد و مشهور که تهیه‌کنندگان آن ادعا می‌کنند همان حیطه‌های عملکرد تدریس را اندازه‌گیری می‌کنند داشته باشند. این نوع تخمین پایایی به شکل فرم‌های موازی، با مفهوم ایجاد روایی هم‌زمان از معانی نمرات حاصل شده از مقیاس‌های مشابه مرتبط است. افزایش اطمینان از معنی ارزشیابی می‌تواند با توسعه مقیاس‌ها (خوشه‌هایی از گویه‌های ارزیابی مرتبط) بر اساس حیطه‌های عملکرد تدریس که روایی آنها به اثبات رسیده تقویت شود.

1. Hildebrand
2. Wilson
3. Dienst

۵- اهداف چندگانه ارزشیابی

طبیعی است که اغلب مدرسان به دنبال خود-بهبودی و کاربرد روش‌های ارزشیابی برای دستیابی به بازخورد دربارهٔ اثربخشی نوآوری‌ها یا سایر جرح و تعدیل‌ها در تدریس باشند. این رویکرد تکوینی^۱ ممکن است مورد تأکید مدیران گروه‌ها یا اساتید برجسته باشد. چنین خود-ارزیابی‌هایی حساس بوده و یک فرایند شخصی است. استقلال در ارائه ابتکارات در تدریس، مستلزم آن است که مدرسان آزادی داشته باشند که نتایج چنین ارزیابی‌هایی را در میان‌ترم یا پایان‌ترم انتخاب و تفسیر نمایند. همچنین، مدرسان در ارزشیابی‌هایی که به صورت رسمی لازم است انجام شود، ماهیت تراکمی^۲ دارد و برای بررسی‌های دوره‌ای عملکرد سازمان یا ارتقای افقی و عمودی اعضای هیأت علمی و سایر تصمیمات استخدامی به کار می‌رود، شرکت می‌کنند اگر داده‌های تکوینی و تراکمی به وسیله مدیران در هم ادغام شده و هر دو بر مبنای داده‌های تراکمی در تصمیم‌های استخدامی و کارگزینی در نظر گرفته شوند، به دلیل استفاده نامناسب و پریشان از داده‌های شخصی تکوینی خطاهایی در ارزشیابی رخ می‌دهد و شاید ارزشیابی عملکرد کلی را تنزل دهد (مانینگ^۳، ۱۹۸۸). نظام‌های ارزشیابی آموزشی باید برای هر دو هدف فراهم شوند: یک مجموعه ثابت از گویه‌ها و مقیاس‌ها برای ارزشیابی تراکمی و یک مجموعه از گویه‌ها به انتخاب مدرسان برای اهداف خودکاو شخصی در ارزشیابی تکوینی.

۶- روش‌های ناهماهنگ در به کارگیری ابزارها

عملیات اجرایی متعددی در نادرستی و اشتباهات ارزشیابی دانشجویان از تدریس و عدم اطمینان در گردآوری داده‌ها از جمله: نبود استانداردهای در مدیریت زمان (مانند انجام ارزشیابی‌ها زود هنگام در طول ترم، خیلی نزدیک به دوره آزمون نهایی، بدون زمان کافی برای تکمیل ابزار) و انجام ارزشیابی در زمان‌های متفاوت در ترم برای هر مدرس. همچنین از آن جا که خطاها در ارزشیابی ممکن است به عدم بیان اهداف فرایند ارزشیابی نسبت داده شود (تکوینی در مقابل تراکمی)، لذا باید هم برای ارزیابان و هم ارزشیابی

۱. ارزشیابی تکوینی به ارزشیابی‌های یک فرایند یا محصول برای فراهم ساختن بازخورد به منظور انجام اصلاحات میان فرایندی اشاره می‌کند.

۲. ارزشیابی تراکمی یا پایانی به این منظور انجام می‌شود که کیفیت یا اثر یک فرایند تکمیل شده یا محصول نهایی را بیازماید.

شوندگان توضیح داده شود که چه استفاده ویژه‌ای از داده‌ها می‌شود و چه کسانی از داده‌ها استفاده می‌کنند (ستترا، ۱۹۷۹). اگر ابزارهای ارزشیابی به وسیله خود مدرسان به دانشجویان معرفی شود و پخش و تکمیل آنها در حضور مدرسان یا نظارت آنها انجام شود پرسش‌هایی در مورد تأثیرات نامناسب ممکن است مطرح شود. به‌کارگیری ابزارهای ارزشیابی توسط خود دانشجویان به عنوان راه‌حل چنین تأثیرات نامناسبی از طرف برخی صاحب‌نظران سفارش شده است (براسکمپ، براندنبرگ و اوری، ۱۹۸۴). نهایتاً چون ابزارها در یک حالت نظام‌دار و امن جمع‌آوری نشده‌اند خطاهای اندازه‌گیری رخ خواهد داد. دانشجویان گزارش داده‌اند که در یک فعالیت ارزشیابی مشارکت کرده‌اند و بعد از آن کلاس را ترک کرده‌اند، قبل از آنکه ابزار را به یک دانشجو که به عنوان نماینده تعیین شده باز گردانند. هنگامی که تکمیل فرم ارزشیابی در یک محیط کلاسی استاندارد شده به انجام نرسد، صحت آنها مورد سوءظن قرار می‌گیرد. به دلیل این که فاقد نمایندگی کل دانشجویان است؛ چرا که تنها بخشی از دانشجویان آنها را تکمیل می‌کنند و بر می‌گردانند. هنگامی که چند مدرس یک کلاس با یک ابزار یکسان ارزشیابی می‌شوند یک تهدید نهایی در ادغام ارزشیابی‌ها رخ می‌دهد. برای چنین کلاس‌های تدریس شده توسط چند استاد، مدت ناکافی در معرض دید قرار گرفتن مدرس برای دانشجویان و دوره زمانی ناسازگار بین زمان ارائه درس توسط مدرس و ارزشیابی توسط دانشجویان در خطای ارزشیابی از تدریس مشارکت می‌کند. خطاهای نسبت داده شده به روش‌های اجرایی ارزشیابی می‌تواند از طریق استفاده از دستورالعمل‌های استاندارد و به‌کارگیری کارشناسان توجیه شده که هدف و کاربرد فرایند ارزشیابی را هنگام به‌کارگیری مقیاس‌ها در زمان‌های معمول و منظم پایان دوره آموزش در تمام کلاس‌ها بیان کنند، به حداقل برسد.

۷- خطاهای معمول تأثیرات ارزشیابی

در صورتی که دانشجویان دیدگاه مثبتی نسبت به یک مدرس داشته باشند، ممکن است که این دید، به ارزیابی مثبت‌تر بدون توجه به سطح واقعی اثربخشی تدریس او گردد. پژوهش‌ها نشان داده‌اند که صفات افراد بر قضاوت‌های آنها مؤثر بوده است. پشتیبانی از این ایده از نظریه شخصیت ضمنی آچ^۱ (۱۹۶۶) و برونر^۲ و تاگیوری^۳

1. Asch
2. Bruner
3. Tagiuri

(۱۹۵۴) ناشی می‌شود. همچنین، کلی^۱ (۱۹۵۰) در مطالعات خود دریافت که دست کاری ویژگی‌های دو قطبی مانند گرم- سرد تأثیر فراوانی در قضاوت‌های دانشجویان از اساتید دارند (به نقل از شیربگی، ۱۳۸۹).

حداقل پنج نوع از تأثیرات ارزشیابی ممکن است باعث خطای ارزشیابی شود. اصطلاح اثرات *هاله‌ای* (ورنون^۲، ۱۹۶۴) به موقعیت‌هایی اطلاق می‌شود که در آن ارزشیابی‌های ویژه بر اساس مشاهده عملکرد مربوطه است بلکه مبنای آن مشاهده تجارب یا نگرش‌های پیشین است (گروسینگ، ۱۹۹۴). این تأثیر معمولاً زمانی رخ می‌دهد که تعبیر عملکردهای موفقیت‌آمیز در یک حیطة، به مبالغه در حیطة دیگری منجر می‌شود. در مقابل *تأثیر هاله‌ای معکوس* ممکن است وقتی رخ دهد که دانشجویان متمایل به ارائه ارزیابی پایین از عملکرد مدرس در یک حیطة به دلیل آنکه ممکن است آنها نگرش‌های نامناسبی درباره حیطة دیگری از عملکرد وی داشته متمایل باشند؛ مثلاً سخت‌گیری گروه آموزشی مربوطه *تأثیر آسان‌گیری* نیز در خطای ارزشیابی سهیم است که در مقابل آن *تأثیر سخت‌گیری* وجود دارد. این نوع تأثیرات، به ارزیابی‌های یکنواخت پایین یا بالای دانشجویان که بدون در نظر گرفتن تفاوت‌ها، عملکرد مشخصی مورد ارزیابی قرار می‌گیرد، اشاره می‌کنند دانشجویان ممکن است احساس کنند که وادار شده‌اند تا مدرسانشان را به صورت بالا ارزشیابی کنند یا چنین دریافته‌اند که دادن ارزشیابی‌های پایین بر شخصیت آنها تأثیر منفی خواهد گذاشت؛ چرا که با عملکرد یادگیری یا میزان توانایی درک آنان مرتبط است. دادن ارزشیابی پایین به صورت ثابت، وقتی که بر اساس معیارهای مقایسه‌ای معتبری نباشد، ممکن است که ریشه در باورهای دانشجویان درباره سخت‌گیری آموزشی داشته باشد یا از اثرات انگیزشی ناشی شود. مدرسان همچنین شواهدی از *تأثیر تمایل به حد وسط* مشاهده کرده‌اند که نتیجه آن، ارزشیابی‌های نه چندان بالا و نه پایین است (اسمیت^۳، ۱۹۸۳). چنین رویکرد ارزشیابی ممکن است با موضوع ترس دانشجویان حمایت شود که ارزشیابی‌های پایین یا بالا ممکن است نیازمند توجهات اضافی باشد یا اینکه ارزشیابی‌های پایین بر روابط حرفه‌ای شخصی مدرس با دانشجو تأثیر بگذارد. ارائه دستورالعمل‌های استاندارد مناسب به دانشجویانی که ارزیابی می‌کنند می‌تواند خطاهای تأثیرات را در ارزشیابی به کمترین حد ممکن برساند.

1. Kelley
2. Vernon
3. Smith

۸- خطا در کاربرد داده‌ها

یک مشکل دیگر، به روایی نتایج حاصل شده از داده‌های ارزیابی دانشجویان از تدریس استادان باز می‌گردد. به دلیل فقدان تبحر آماری برخی از اعضای شوراها یا هیأت‌های ارزیابی، اطلاعات به دست آمده به صورت نادرست در تصمیم‌گیری‌ها مورد استفاده قرار می‌گیرد (مک کیچی^۱، ۱۹۹۷). برخی از مدیران مراکز آموزش عالی وقتی به اعضای هیأت علمی بازخورد می‌دهند ممکن است احساس ناراحتی کنند (ایوانسویچ^۲، ۱۹۹۲). شاید چنین سرپرستانی به این دلیل از استفاده و تفسیر داده‌های مربوط به عملکرد آموزشی بیمناک هستند که از تربیت یا تجربه کافی مدیریتی برخوردار نیستند و یا مقیاس‌های ارزشیابی (یا کل فرایند ارزشیابی عملکرد اعضای هیأت علمی) را به صورت عینی در نظر نمی‌گیرند؛ بنابراین، از ارزش فرایند ارزیابی و داده‌ها می‌کاهند. خطاها ممکن است در یک بازخورد ناقص یا اشتباه، یا در نبود هر بازخورد مفید خود را در کل نشان دهند. بخشی از فرایند بهسازی مدیران کم تجربه باید شامل آموزش آمادگی نحوه ارائه بازخورد عملکرد به کارکنان و استفاده از داده‌های عملکردی روا و پایا باشد.

بحث و نتیجه‌گیری

روی هم رفته، نتایج تحقیقات به عمل آمده روی اثرات متغیرهای نامربوط بر روایی ارزیابی دانشجویان از مدرسان نشان می‌دهد که در تفسیر این داده‌ها لازم است که شرط احتیاط رعایت شود. چالش کاهش خطا برای همه مدرسان، مدیران گروه‌ها، مدیران عالی و مسئولان ارزشیابی، مشابه است. علاقه ذاتی مدرسان برای خود-بهبودی نیازمند آن است که بازخوردهای تکوینی عملکرد روا، پایا و محرمانه باشند. اعضای هیأت علمی از تقویت عینیت در فرایند کنترل و نظارت همگنان سود می‌برند. دست‌اندرکاران در گروه‌ها و دانشکده‌ها نیازمند نوعی ابزارهای ارزشیابی عملکرد هستند که استانداردهای قانونی حرفه‌ای برای روایی و پایایی را تأمین کند؛ علاوه بر این، مرکز خدمات دانشگاه که درگیر فرایند ارزشیابی آموزشی هستند مجبورند که رواترین و پایاترین ابزارهای موجود را فراهم کنند.

1. McKeachie
2. Ivanceovich

در جدول (۳) ویژگی‌های برجسته راهبردهای تهیه مستندات اثربخشی تدریس که در مؤسسات مختلف آموزش عالی به کار می‌روند ذکر شده و تصمیم‌گیری در مورد تهیه ترکیبی مناسب از این راهبردها قبل از آغاز فرایند ارزیابی در هر دانشگاه یا مؤسسه توصیه شده است. پس از فراهم شدن این ترکیب، گام بعدی، طراحی، اجرا و گزارش نتایج آن به شکل صحیح است. صحت تصمیم‌های گرفته شده در خصوص اعضای هیأت علمی، به ادغام راهبردهای ارزیابی، روایی و پایایی شواهد جمع‌آوری شده بستگی دارد؛ البته باید در نظر داشت که این راهبردها دارای سهم مساوی در ارزشیابی نهایی اثربخشی تدریس نیستند. سهم هریک از این راهبردها را می‌توان با روش‌های آماری پیشرفته تعیین کرد.

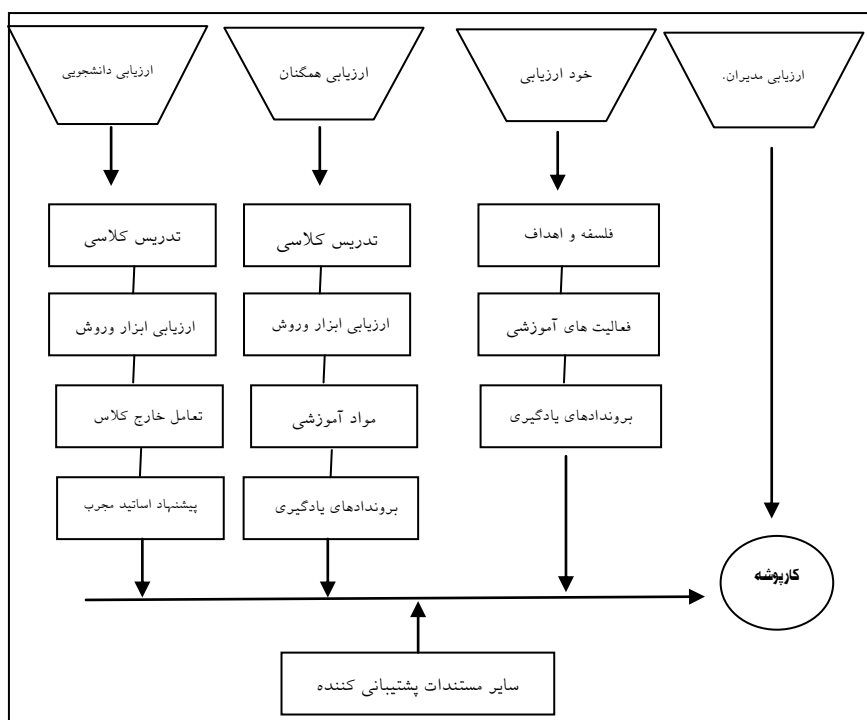
جدول (۳) ویژگی‌های منابعی که برای اثبات اثربخشی تدریس به کار می‌روند

منبع	نوع ابزار	ارائه کننده مستندات	افراد ذی نفع	نوع تصمیم‌گیری
ارزیابی دانشجویان	مقیاس ارزیابی	دانشجویان	مربیان/ مدیران	بهبود/ امور استخدامی/ ارزشیابی برنامه
ارزیابی هم‌تایان	مقیاس ارزیابی	هم‌تایان	مربیان	بهبود/ امور استخدامی
خود ارزیابی	مقیاس ارزیابی	مربیان	مربیان/ مدیران	بهبود/ امور استخدامی
ویدئوها	مقیاس ارزیابی	مربیان/ هم‌تایان	مربیان/ هم‌تایان	بهبود/ امور استخدامی
مصاحبه با دانشجویان	پرسشنامه	دانش آموزان	مربیان/ مدیران	بهبود/ امور استخدامی
ارزیابی	مقیاس ارزیابی	دانش‌آموختگان	مربیان/ مدیران	بهبود/ به امور استخدامی
فارغ‌التحصیلان	مقیاس ارزیابی	کارفرمایان	مربیان/ مدیران	بهبود/ امور استخدامی/ ارزشیابی برنامه
ارزیابی کارفرمایان	مقیاس ارزیابی	مدیران	مدیران	بهبود/ امور استخدامی
ارزیابی مدیران	بررسی داوران	مربیان	مدیران	بهبود/ امور استخدامی
پژوهشگری در تدریس	بررسی داوران	مربیان	مدیران	بهبود/ امور استخدامی
جوایز سرآمدی	آزمون‌ها	دانشجویان	شورای برنامه	بهبود/ امور استخدامی
ابزارهای نتایج یادگیری	بیشتر موارد	بیشتر موارد	درسی	بهبود/ ارزشیابی برنامه
کارپوشه تدریس	فوق	فوق	هیأت ممیزه	بهبود/ ارزشیابی برنامه

اقتباس از: برک (۲۰۰۶)

یکی از راه‌حل‌های کم کردن خطاهای ارزیابی، استفاده از راهبردهای متنوع تهیه مستندات و شواهد اثربخشی تدریس است. در این خصوص، چندین مدل جامع

ارزشیابی پیشنهاد شده است (برک، ۲۰۰۶)، همه این مدل‌ها در برگیرنده منابع چندگانه‌ای برای تهیه مستندات با وزن بخشیدن بیشتر به ارزیابی دانشجویان و هم‌تایان، و در مقابل، دادن وزن کمتری به سمت روش‌های خود ارزیابی، نظرسنجی از دانش‌آموختگان، مدیران و کارفرمایان است (سنتر،^۱ ۱۹۹۹). نمونه‌ای از یک نظام جامع ارزشیابی (برنت^۲ و فدلر^۳، ۲۰۰۴) با استفاده از منابع مختلف در نمودار (۱) به تصویر کشیده شده است. اگر چه راهبردهای متعددی برای تهیه مستندات و شواهد اثربخشی تدریس وجود دارد، اما برای اجرایی و عملی شدن آن در مؤسسات آموزشی، پنج مورد از آنها در نظام جامع پیشنهادی ذیل به کار گرفته شده است.



نمودار (۱) نمونه‌ای از یک مدل جامع ارزشیابی عملکرد تدریس اقتباس از، برنت و فدلر (۲۰۰۴)

1. Centra
2. Brent
3. Felder

همچنان که نمودار (۱) نشان می‌دهد، از پنج راهبرد مختلف استفاده شده است. پیکان‌ها نشان دهنده ترکیب و جمع‌بندی مستندات مختلف به منظور ارزشیابی از اثربخشی تدریس هستند. محتوای هر راهبرد ارزشیابی می‌تواند دربرگیرنده فلسفه تدریس، ابزار و روش‌ها، مواد آموزشی یا بروندهای یادگیری باشد. از مزیت‌های این نظام جامع ارزشیابی، نخست آن است که مبنای خوبی را برای ارزشیابی تکوینی فراهم می‌نماید، و دوم آنکه اگر این الگو در چند سال اولیه مسیر شغلی اعضای هیأت علمی به کار گرفته شود، امکان این که به صورت معناداری ارزشیابی‌های تراکمی مناسبی را فراهم آورد بسیار زیاد خواهد بود.

باید به خاطر داشت که اگر دو یا چند مجموعه داده گردآوری شده مربوط به ارزشیابی تدریس یک مدرس نتایج کاملاً متفاوتی به دست دهد، پس احتمالاً مواردی در دست کم یکی از منابع ایراد دارد یا تدریس مدرس در کلاس ارزشیابی شده نماینده تدریس واقعی او نبوده است؛ لذا باید بررسی‌های بیشتری انجام شود (برنت و فدلر، ۲۰۰۴). یک شیوه برای افزایش پایایی ارزشیابی‌هایی که بر اساس چند منبع انجام می‌شود آن است که اطمینان حاصل شود که برخی از همپوشانی‌ها در اطلاعات که منابع فراهم کرده‌اند وجود دارد؛ به عنوان مثال، اگر برگه مشاهده کلاس توسط همگنان شامل گویه‌هایی درباره آمادگی برای سخنرانی، وضوح بیان، و احترام به دانش‌آموزان باشد، پس برگه‌های ارزشیابی که توسط دانشجویان تکمیل می‌شود باید همان ویژگی‌ها را مورد پرسش قرار دهد.

در گام نخست انتخاب یک نظام جامع ارزشیابی، باید بر مفهوم واحدی از اثربخشی تدریس توافق حاصل شود تا بر مبنای آن از طیفی وسیع از منابع مختلف مستندات جمع‌آوری و سازه مربوط به آن تعریف شود و نهایتاً برای دستیابی به آن، تصمیم‌گیری به عمل آید. مطالب فراوانی درباره مزیت‌ها و کاستی‌های منابع متفاوتی که اخیراً به کار می‌روند نوشته شده است. هر کدام از منابع می‌توانند اطلاعات منحصر به فردی را عرضه کنند، اما این اشتباه‌آمیز است که معمولاً به شیوه متفاوتی نسبت به سایر منابع دیگر عمل شود؛ به عنوان مثال، ناپایایی یا سوگیری‌های موجود در ارزشیابی هم‌تایان با محدودیت‌های ارزشیابی دانشجویان، مشابه نیست. ارزشیابی دانشجویان دارای نقاط ضعف دیگری است. با بهره‌گیری از چند منبع مختلف، محاسن هر منبعی می‌تواند جبران‌کننده ضعف سایر منابع دیگر باشد، و به تبع آن، همگرایی بر تصمیمی

درباره اثربخشی تدریس خواهد بود که بسیار معتبرتر از تصمیمی است که تنها بر اساس یک منبع استوار است (برک، ۲۰۰۶).

دورنمای تدریس همواره در حال تغییر است؛ علاوه بر این، ارزشیابی دانشجویان از عملکرد مدرسان، تنها یک نوع داده را برای داوری درباره اثربخشی تدریس فراهم می‌کند؛ از یک سو، این ابزارهای سنتی ارزیابی توانایی تدریس، شاید در موقعیت‌هایی که فرایند حل مسأله یا رویکردهای نوین تدریس مدنظر باشند، پاسخگو نباشند؛ از این رو، نظام‌های ارزشیابی کنونی برای انعکاس پیوستار گسترده‌تری از صفات تدریس در تمام گونه‌های موجود تدریس، نیازمند تغییر و بازنگری هستند.

منابع

شیربگی، ناصر (۱۳۸۹). راهبردهای اندازه‌گیری اثربخشی تدریس در نظام‌های آموزش عالی. *نامه آموزش عالی*. دوره جدید، سال سوم، شماره ۱۰، ص ۳۳-۹.

شیربگی، ناصر و همتی، آزاد (۱۳۸۹). مدل یابی تأثیر جذبه شخصیتی اساتید بر ارزیابی دانشجویان از تدریس اثربخش در دانشگاه کردستان. چهارمین همایش سالانه ارزیابی درونی کیفیت در نظام دانشگاهی. دانشگاه تهران.

وزارت علوم، تحقیقات و فناوری (۱۳۸۷). *آیین‌نامه ارتقای مرتبه اعضای هیأت علمی دانشگاه‌ها و مؤسسه‌های آموزش عالی، پژوهشی و فناوری* (مصوب ۱۳۸۷/۸/۱۸).

وزارت علوم، تحقیقات و فناوری (۱۳۸۹). *آیین‌نامه ارتقای مرتبه اعضای هیأت علمی دانشگاه‌ها و مؤسسه‌های آموزش عالی، پژوهشی و فناوری*، (مصوب ۱۳۸۹/۱۲/۳).

- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41 (2), 258-290.
- Berk, R. A. (2006). *Thirteen strategies to measure college teaching*, Stylus Publishing, Richmond, Virginia.
- Braskamp, L. A., Brandenburg, D. C. & Ory, J. C., (1984). *Evaluating Teaching Effectiveness*, Sage Publications, Newbury Park CA.
- Brent, R. & Felder, R. M. (2004). A Protocol for Peer Review of Teaching, *Proceedings of the American Society for Engineering Education Annual Conference & Exposition*, North Carolina State University
- Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of Social Psychology*. Addison Wesley, London.
- Casey, R. J., Gentile, P. & Bigger, S. W. (1997). Teaching appraisal in higher education: an Australian perspective, *Higher Education*, 34:549-482
- Centra, J. A. & Creech, F. R., (1976). The Relationship between Student, Teachers, and Course Characteristics and Student Ratings of Teacher Effectiveness, *Educational Testing Service, Project Report I-76*, Princeton NJ
- Centra, J. A. (1979). *Determining Faculty Effectiveness*, Jessy-Bass. San Francisco CA.
- Centra, J. A. (1999). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. Jessy-Bass San Francisco

- Das, H., Frost, P. J. & Barnowe, J. T., (1979). Behaviorally anchored scales for assessing behavioral science teaching, *Canadian Journal of Behavioral Science*, 11, 79-88.
- Dickinson, T. L. & Zelliger, P. M., (1980). A Comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology*, 65, 147-154.
- Grussing, P. G., Valuck, R. & Williams, R. G., (1994). "Development and validation of behaviorally-anchored rating scales for student evaluation of pharmacy instruction", *American Journal of Pharmacia Education*., 58, 25-37.
- Hildebrand, M., Wilson, R. C. & Dienst, E. R., (1971). *Evaluating University Teaching*, Center for Research and Development in Higher Education, Berkely, CA.
- Ivancevich, J. M., (1992). *Human Resource Management*, Irwin, Homewood
- Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality and Social Psychology*, 18 (3), 431-439.
- Knapper, c., & Cranton, P. (2001). *Fresh approaches to the evaluation of teaching*. Jossey-Bass, San Francisco.
- Manning, R. C., (1988). *The Teacher Evaluation Handbook*, Prentice-Hall, Englewood Cliffs NJ.
- McKeachie, W. J. (1997). Student ratings: the validity of use. *American Psychologist*, 52 (11), 1218-1225.
- McPherson, M. A. (2006). Determinants of how Students Evaluated Teachers. *Journal of Economic Education*, 37 (1), 13-20.
- Smith, P. C., (1983). Behaviors, results and organizational effectiveness: The problem of criteria, in M. Dunnette, (ed.) *Handbook of Industrial Psychology*, Wiley & Sons, New York, NY.
- Vernon, P. E. (1964). *Personality Assessment: a critical survey*. London,: Methuen.
- Ware, J. E. & Williams, R. G., (1975). The Dr. Fox effect: A study of lecture effectiveness and ratings of instruction, *Journal of Medical Education*, 50, 149-156.
- Wilhelm, W. B. (2004). The relative influence of published teaching evaluations and other instructor attributes on course choice. *Journal of Marketing Education*, 26, 17-30.
- Wotruba, T. R., & Wright, P. L., (1975). How to develop a teacher-rating instrument: A research approach, *Journal of Higher Education*, 46, 653-670