

Construction of a Rubric to Evaluate Fair Classroom Assessment at the University of Tehran: An Exploratory Mixed-Method Study

Ali Baniasadi*
Keyvan Salehi**
Ebrahim Khodaie***
Khosro Bagheri****
Balal Izanloo*****

Introduction

Fairness is an important quality of classroom assessment as a continuous process that affects learners' beliefs, mentality, motivation, satisfaction, and performance. The main purpose of this study was to construct a rubric for evaluating fairness in classroom assessment.

Method

For this purpose, based on an exploratory sequential mixed methods design, two studies were conducted. In the first study, using purposive and non-proportional quota sampling, 29 students and 10 faculty members of the University of Tehran were selected, respectively. A group interview and a Focus group interview were also conducted and in the second study, 511 students from the University of Tehran were selected through three-stage cluster sampling. Qualitative data analysis led to the identification of a conceptual model including the four main themes of "procedural fairness", "interactional fairness", "nature of assessment" and "fairness in the paper-pencil test" as the basis for the development of criteria and making Rubric. In the second study, the psychometric properties of Rubric were investigated using content validity, item

* Ph.D. Student, Faculty of Psychology and Education, University of Tehran, Tehran, Iran.

** Assistant Professor, Faculty of Psychology and Education, University of Tehran, Iran.
Corresponding Author: keyvansalehi@ut.ac.ir

*** Associate Professor, Faculty of Psychology and Education, University of Tehran, Tehran, Iran.

**** Professor, Faculty of Psychology and Education, University of Tehran, Tehran, Iran.

***** Associate Professor, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran.

analysis, construct validity and reliability determination.

Results

The content validity of each criterion was higher than the standard value of 0.78 and the validity of the overall scale was 0.94. The results of exploratory factor analysis showed the extraction of three factors of procedural fairness, the nature of measurement and interactional fairness, which explained 57.08% of the total variance of variables. The results of factor analysis along with the study of discrimination and difficulty coefficients of items led to the elimination of three weak criteria and the final rubric was compiled with 17 criteria. The calculated Cronbach's alpha for procedural fairness, the nature of assessment and interactional fairness were 0.87, 0.84 and 0.79, respectively. Also, the retest coefficient for them was calculated at the desired level and 0.65, 0.63 and 0.54, respectively.

Discussion

According to the desirable psychometric properties of the rubric, this rubric can help to increase the literacy of professors, increase the quality of teaching and assessment and strengthen the civic and educational behaviors of students by identifying the strengths and weaknesses of professors in the field of fairness.

Keywords: Rubric, Exploratory Mixed Method, Classroom Assessment, Assessing Student Performance, Fair Assessment

Author Contributions: In the present study, the second author, as a supervisor, was responsible for supervising and strategizing the overall research process and compiling and finalizing the corrections of the article. In developing the research plan, the first author has been in charge of the process of gathering, analyzing and interpreting the findings and writing the text of the article, and in general, drawing conclusions from the findings and expanding and interpreting them jointly and with the discussion and exchange of opinions of all colleagues and with the accompaniment of the third, fourth and The fifth was done as consultant professors.

Conflict of interest: The authors acknowledge that there is no conflict of interest in this article.

Funding: The current research did not receive financial support from any institution and institution, and all expenses during the research implementation process were borne by the researchers.

Acknowledgments: The present research would not have been possible without the cooperation of the participants; we hereby acknowledge and thank all the participants.

تاریخ دریافت مقاله: ۱۴۰۰/۰۵/۲۷
تاریخ پذیرش مقاله: ۱۴۰۱/۰۷/۱۹

مجله‌ی علوم تربیتی دانشگاه شهید چمران اهواز
پاییز و زمستان ۱۴۰۱، دوره‌ی ششم، سال ۲۹
شماره‌ی ۲، صص: ۲۳۴-۲۱۱

مقاله پژوهشی

ساخت روبریک برای ارزشیابی سنجش کلاسی عادلانه در دانشگاه تهران: مطالعه به روش آمیخته اکتشافی

علی بنی‌اسدی*
کیوان صالحی**
ابراهیم خدایی***
خسرو باقری****
بلال ایزانلو*****

چکیده

عدالت یک کیفیت مهم سنجش کلاسی است که بر باورها، ذهنیت، انگیزش، رضایتمندی و عملکرد یادگیرندگان اثر می‌گذارد. هدف اصلی این پژوهش ساخت روبریکی برای ارزشیابی عدالت در سنجش کلاسی بود. بدین‌منظور مبتنی بر روش پژوهش آمیخته اکتشافی از نوع ساخت ابزار دو مطالعه به انجام رسید که در مطالعه اول با استفاده از نمونه‌گیری هدفمند از نوع ملاکی ۲۹ مصاحبه با دانشجویان دانشگاه تهران و با استفاده از نمونه‌گیری سهمیه‌ای غیرنمونه‌ای ۱۰ مصاحبه با اعضای هیأت علمی دانشگاه تهران انجام شد. همچنین در این مرحله یک مصاحبه گروهی و یک مصاحبه گروه کانونی نیز انجام شد. در مطالعه دوم، از میان دانشجویان دانشگاه تهران تعداد ۵۱۱ نفر از طریق نمونه‌گیری خوشه‌ای سه‌مرحله‌ای انتخاب شدند. تحلیل داده‌های کیفی به شناسایی مدلی مفهومی شامل چهار مضمون اصلی «عدالت رویه‌ای»، «عدالت تعاملی»، «ماهیت سنجش» و «عدالت در آزمون کتبی» مبنای تدوین ملاک‌ها و ساخت روبریک انجامید. در مطالعه دوم، ویژگی‌های روان‌سنجی روبریک با استفاده از روایی محتوایی، تحلیل گویه‌ها، روایی سازه و تعیین اعتبار به هر دو شیوه همسانی درونی و ثبات بررسی شد. روایی محتوایی هر یک از ملاک‌ها بالاتر از مقدار معیار ۰/۷۸ و

* دانشجوی دکتری سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی دانشگاه تهران، تهران، ایران.

** استادیار دانشکده روان‌شناسی و علوم تربیتی؛ دانشگاه تهران، تهران، ایران (نویسنده مسئول)

keyvansalehi@ut.ac.ir

*** دانشیار دانشکده روان‌شناسی و علوم تربیتی دانشگاه تهران، تهران، ایران

**** استاد، دانشکده روان‌شناسی و علوم تربیتی دانشگاه تهران، تهران، ایران

***** استادیار، دانشگاه خوارزمی، تهران، ایران

روایی مقیاس کلی ۰/۹۴ به دست آمد. نتایج تحلیل عاملی اکتشافی نشان‌دهنده استخراج سه عامل عدالت رویه‌ای، ماهیت سنجش و عدالت تعاملی بود که در مجموع ۵۷/۰۸ کل واریانس متغیرها را تبیین می‌کردند. نتایج تحلیل عاملی به همراه بررسی ضرایب تمیز و دشواری گویه‌ها منجر به حذف سه ملاک ضعیف شد و روبریک با ۱۷ ملاک نهایی گردید. آلفای کرونباخ محاسبه شده برای عدالت رویه‌ای، ماهیت سنجش و عدالت تعاملی به ترتیب ۰/۸۷، ۰/۸۴ و ۰/۷۹ به دست آمد. همچنین ضریب بازآزمایی برای آن‌ها در حد مطلوب و به ترتیب ۰/۶۵، ۰/۶۳ و ۰/۵۴ محاسب شد. بررسی ویژگی‌های فنی روبریک نشان داد که می‌تواند به عنوان ابزاری روا و معتبر برای ارزشیابی کیفیت سنجش عادلانه در کلاس درس استفاده شود.

واژه‌های کلیدی: روبریک، پژوهش آمیخته اکتشافی، سنجش کلاسی، سنجش عملکرد دانشجویان، سنجش عادلانه

مقدمه

رویکردهای نوین آموزش با هدف ارتقاء کیفیت آموزشی، پژوهشی و یادگیری در نظام آموزشی توسعه یافته‌اند (Vaghoor Kashani et al., 2019) و روایی، اعتبار و عدالت کیفیت‌هایی هستند که در نظریه اندازه‌گیری برای سنجش آموزشی^۱ بهینه به آن‌ها اشاره شده است (Tierney, 2014) و عدالت به عنوان یک مفهوم انتزاعی در دامنه گسترده سنجش‌های آموزشی، از شناخت نیازهای یادگیری دانشجویان تا آزمون‌های بین‌المللی گسترده مقیاس مورد توجه و بررسی قرار می‌گیرد (Green et al., 2007; Tierney, 2013). فرضیه‌های نظریه اندازه‌گیری با سنجش کلاسی به طور کامل برآزش ندارند؛ چراکه سنجش کلاسی وابسته به محیط آموزشی است که در یک زمینه روانی اجتماعی انجام می‌شود (Brookhart, 2003)؛ بنابراین، سنجش کلاس فراتر از کاربرد فنون قابل اندازه‌گیری بوده و شامل نگرانی‌های فردی و اجتماعی است (Allal, 2013). عدالت یک ساختار چندبُعدی است و نظریه عدالت سازمانی و نظریه اندازه‌گیری اصول سنجش عادلانه را بررسی کرده‌اند. پژوهش‌ها در مورد عدالت سازمانی و عدالت آموزشی چهار نوع عدالت را تعریف می‌کند که شامل عدالت تعاملی^۲، رویه‌ای^۳، آموزشی^۴ و توزیعی^۵ است. عدالت توزیعی اولین نوع عدالت است که در این پژوهش‌ها به آن

-
- 1- Educational Assessment
 - 2- interactional
 - 3- procedural
 - 4- instructional
 - 5- distributive

اشاره شده است. در اواسط دهه ۱۹۷۰، تمرکز بر پیامدهای تصمیمات اتخاذ شده به سمت تمرکز بر عدالت رویه‌های تخصیص این پیامدها تغییر یافت (Duplaga & Astani, 2010). عدالت توزیعی به تخصیص عادلانه نمره‌ها گفته می‌شود که در آن افراد نتایج را بر اساس دروندادها قضاوت می‌کنند (Schmidt et al., 2003; Bempchat et al., 2013). عدالت رویه‌ای به سیاست‌ها و رویه‌های مربوط به درجه‌بندی و نمره‌دهی اشاره دارد (Bempchat et al., 2013). عدالت تعاملی اولین بار توسط Bies and Moag (1986) مطرح شد. Greenberg (1990) اظهار داشت که عدالت تعاملی شامل دو بُعد عدالت اطلاعاتی^۱ و عدالت بین فردی^۲ است (cited in Colquitt, 2001)؛ در همین حال، Gilliland (1993) عدالت بین فردی را زیرمجموعه عدالت رویه‌ای می‌داند (Pepper & Pathak, 2008). عدالت تعاملی بر ماهیت رابطه استاد و دانشجو تأکید دارد و بر کیفیت رفتار استاد و دانشجو هنگام اجرای رویه‌ها یا تخصیص پیامدها متمرکز است. بی‌عدالتی آموزشی، به‌عنوان "بخشی از موضوع گسترده‌تر بی‌عدالتی در فرصت‌های آموزشی" (ص. ۱۸۰)، اولین بار توسط بمپچت و همکاران با تجزیه و تحلیل مصاحبه‌های شرکت‌کنندگان تدوین یافت (Bempchat et al., 2013). عدالت آموزشی به عدم موفقیت استاد در ایفای نقش خود به‌عنوان مربی، با تمرکز بر فعالیت‌های تدریس استاد اشاره دارد. آموزش بی‌اثر، انگیزه دانشجویان را کاهش می‌دهد و باعث می‌شود آن‌ها فکر کنند تلاش‌هایشان به نتایج مورد انتظار منجر نمی‌شود.

اخیراً از روبریک‌ها به‌عنوان ابزاری که سنجش عادلانه‌تر را ممکن می‌کنند برای ارزیابی شایستگی‌های حرفه‌ای و مهارت‌های تحصیلی (Usman- González-Chordá et al., 2016; Stillman et al., 2018)، مهارت‌های هیجانی-اجتماعی (Pancorbo et al., 2020) و نگرش‌ها (Daas et al., 2019) استفاده می‌شود. روبریک‌ها که به‌عنوان «نوعی ماتریس که سطوح دستیابی یا ادراک را برای مجموعه‌ای از ملاک‌ها یا ابعاد کیفیت عملکردی معین فراهم می‌کنند»؛ تعریف می‌شوند (Allen & Tanner, 2006, 197) دو قسمت دارند، ملاک‌ها که همان انتظارها و ویژگی‌هایی هستند که در عملکرد باید بررسی شوند و توصیف سطوح عملکرد که سطوح کیفیت مختلف را برای ملاک‌ها توصیف می‌کنند. روبریک‌ها هماهنگی قضاوت بین درجه‌بندی‌کنندگان

1- informational fairness

2- interpersonal fairness

مختلف را افزایش می‌دهند و نمره‌دهی و ارزیابی عادلانه را امکان‌پذیر می‌کنند (Brame, 2019; Jonsson & Svingby, 2007). چراکه با استفاده از آن‌ها ارزیابان به‌جای مفروضات بر اساس واقعیت‌های توصیفی قضاوت می‌کنند (De la Rosa Gómez et al., 2019; Pancorbo et al., 2020). روبریک‌ها به دو نوع کلی و تحلیلی^۱ تقسیم‌بندی می‌شوند در نوع کلی همه ملاک‌ها هم‌زمان در نظر گرفته شده و همه ویژگی‌های کار عالی در یک توصیف ارائه می‌شود در حالی که در روبریک‌های تحلیلی توصیف سطوح عملکرد برای هر ملاک به تفکیک ارائه می‌شود و این نوع روبریک برای بازخورد عملکرد افراد مناسب‌ترند (Brookhart, 2003; Brame, 2019).

پژوهش‌هایی که موضوع عدالت را در کلاس مورد بررسی قرار داده‌اند ابزارهای مختلفی برای سنجش عدالت به‌کار برده‌اند. این ابزارها بعضاً برای سنجش حوزه‌های محدودتر سنجش کلاسی تدوین شدند؛ به‌عنوان مثال (Gordon and Fay (2010) برای سنجش ادراک دانشجویان از عدالت نمره‌دهی اساتید پرسشنامه‌ای شامل ۱۴ آیتم (مانند «دانشجویانی که یادگیری بیشتری داشتند نمره‌های بهتری گرفتند») که عدالت توزیعی، رویه‌ای و تعاملی را ارزیابی می‌کرد و دانشجویان در قالب طیف ۵ نقطه‌ای لیکرت از بسیار موافقم تا بسیار مخالفم به آن پاسخ می‌دادند، تدوین کردند. در موارد دیگری ابزارهای ساخته شده برای سنجش عدالت در حوزه‌های وسیع‌تری مانند ارزشیابی عدالت در دوره تحصیلی یا در کلاس درس تدوین شده‌اند. به‌عنوان مثال (Chory-Assad (2002 پرسشنامه‌ای شامل ۴ آیتم خودگزارش‌دهی برای سنجش عدالت توزیعی (مانند «نمره‌ها در آخرین آزمون») و ۱۴ آیتم برای سنجش عدالت رویه‌ای تدوین کرد که شرکت‌کنندگان نظراتشان را روی طیف ۵ نقطه‌ای لیکرت از بسیار عادلانه تا بسیار ناعادلانه مشخص می‌کردند. همچنین (Rodabaugh (1994 برای سنجش ادراک بی‌عدالتی در کلاس درس پرسشنامه‌ای شامل ۱۸ آیتم (مانند «بی‌عدالتی در نمره‌دهی، دادن نمره‌های دلخواهی یا تغییر سیاست‌ها در طول دوره») تدوین کرد که دانشجویان اهمیت رفتارهای نادرست اساتید را درجه‌بندی کردند. از پژوهش‌هایی که به‌طور خاص در حوزه ارزشیابی کلاسی انجام شده می‌توان به پژوهش (Johnson et al. (2008 اشاره کرد که اخلاقی بودن عملکردها در ارزشیابی دانشجویان را در قالب ۳۶ سناریوی مرتبط با ارزشیابی کلاسی (مانند «یک استاد با استفاده از ابزارهای مختلف سنجش مانند آزمون‌های چندگزینه‌ای، مقاله،

1- holistic & analytic

پروژه، نمونه کار و نظایر آن یادگیری دانشجویان را می‌سنجد» از دیدگاه رهبران آموزشی بررسی کردند.

پژوهش‌هایی نیز در ایران در حوزه سنجش کلاسی در دانشگاه انجام شده است و چالش‌های ارزشیابی دانشجویان را مورد بررسی قرار داده‌اند که می‌توانیم به پژوهش Seraji et al. (2013) اشاره کنیم که مواردی مانند عدم آشنایی مدرسان به فلسفه و هدف ارزشیابی، مشکلات موجود در ساختار نظام آموزش عالی، عدم توجه به پیامدهای غیرتریبیتی شیوه‌های ارزشیابی، نبود بستر لازم برای ارزشیابی مستمر، عدم توجه به کسب مهارت‌های سطح بالا و عدم توجه به حیطه‌های مختلف یادگیری را از مصاحبه با اساتید استخراج کردند. همچنین Ghanbari et al. (2015) تأثیر این چالش‌ها بر رویکرد مطالعه عمل فکورانه دانشجویان را بررسی کردند که چالش مربوط به عدم توجه به تفاوت‌های فردی بیشترین و چالش عدم استفاده از ارزشیابی‌های غیرعینی کمترین بار عاملی را داشتند. شاید تنها پژوهشی که به‌طور خاص ادراک دانشجویان از عدالت در سنجش کلاسی را مورد بررسی قرار داده است پژوهشی است که توسط Baniasadi et al. (2021) و با عنوان «بازنمایی ادراک دانشجویان از سنجش کلاسی عادلانه: یک مطالعه کیفی» انجام شد. پژوهشگران در این پژوهش ۴۴۹ واحد معنایی به‌دست‌آمده را در قالب ۱۶ زیرمضمون و چهار مضمون اصلی «عدالت تعاملی»، «عدالت آموزشی»، «ماهیت سنجش» و «انصاف» ارائه کردند. همچنین این پژوهشگران در یک مطالعه مروری منظم، ۱۱۰ نشانگر برای سنجش کلاسی عادلانه در قالب ۷ مضمون «سنجش عادلانه تساوی گرایانه»، «سنجش عادلانه منصفانه»، «روایی»، «عدالت توزیعی»، «عدالت رویه‌ای»، «عدالت تعاملی» و «عدالت آموزشی» معرفی کردند (Baniasadi et al., 2022).

هر گام از فرایند ارزشیابی همچون جمع‌آوری داده‌ها، پیش‌نویس گزارش و نتایج می‌تواند طعمه رفتارهای نامناسب و سوء استفاده باشد (Ghorbankhan et al., 2020) و سنجش عادلانه از این جهت که نتایج سنجش زمینه‌های تحصیلی آینده و انتخاب‌های شغلی را تحت تأثیر قرار می‌دهد مهم است و ماهیت چنین سنجش‌هایی روی تمرین‌های کلاسی، روی انگیزش و خودتنظیمی فراگیران در یادگیری اثر داشته و می‌تواند پیامدهای اجتماعی، هیجانی و تحصیلی بلندمدتی برای ایشان به همراه داشته باشد (DeLuca, 2012) این مکانیسم به‌گونه‌ای است که عناصر موقعیتی و ادراکی نظیر عدالت در آغاز بر اخلاق تحصیلی تأثیر گذاشته و سپس اخلاق

تحصیلی موجب کاهش رفتارهای منفی و تقویت رفتارهای مدنی- تحصیلی می‌شود (Golparvar, 2015)؛ چنانچه دریافت رفتارهای عادلانه با خودداری از خشونت و درگیر شدن بیشتر در فعالیت‌های فوق‌برنامه (Resh & Sabbagh, 2017)، خستگی هیجانی، بی‌علاقگی تحصیلی و ناکارآمدی تحصیلی مرتبط است (Marzooqi et al., 2013)؛ و حساسیت به عدالت با پذیرش و انجام رفتارهای فریبکارانه تحصیلی رابطه‌ی معنادار دارد؛ چنانچه دریافت رفتارهای ناعادلانه با تقلب (Lemons & Seaton, 2011) و مدرسه‌گریزی (Ishak & Fin, 2013) مرتبط است و روش‌های نمره‌دهی ناعادلانه منجر به خطر افزایش ترک تحصیل دانشجویان می‌شود (Burger & Gross, 2016).

با توجه به موارد مطروحه بالا، پیامدهای اجتماعی، هیجانی و تحصیلی سنجش و به‌ویژه اینکه پژوهش‌ها نشان داده‌اند که برداشت از عدالت می‌تواند تحت تأثیر باورها و ارزش‌های فرهنگی باشد (e.g. Tata, 2005; Liu et al., 2016) و با عنایت به این که هیچ ابزاری برای بررسی عدالت در سنجش کلاسی در ایران ساخته نشده است و در روبریک‌ها تفسیر معنای هر نقطه‌ی مقیاس روبریک در مقایسه با قالب‌های پاسخ کیفی (مانند مقیاس درجه‌بندی لیکرت) راحت‌تر است و برچسب‌های کلامی به‌عنوان لنگرهای نقطه‌ای مقیاس، انگیزه‌ی پاسخ‌دهندگان و کیفیت اندازه‌گیری را در مقایسه با سایر قالب‌های پاسخ‌دهی افزایش می‌دهد و ارزشیابی عادلانه‌تر و هماهنگ‌تر را با اجتناب از ذهنیت‌گرایی می‌توانند فراهم کنند (Pancorbo et al., 2020)؛ این پژوهش به دنبال آن است تا مبتنی بر پژوهش‌های Baniasadi et al. (2021) و Baniasadi et al. (2022) و نشانگرها و مضمون‌های به‌دست‌آمده در آن‌ها روبریکی معتبر و متناسب با بافت فرهنگی و بومی برای ارزشیابی عدالت در سنجش کلاسی تدوین کند؛ به-عبارت‌دیگر این پژوهش به دنبال پاسخگویی به این سؤالات است: ۱- ملاک‌های سنجش کلاسی عادلانه چیست؟ ۲- آیا روبریک سنجش کلاسی عادلانه از روایی برخوردار است؟ ۳- آیا روبریک طراحی شده برای سنجش کلاسی عادلانه از اعتبار برخوردار است؟

روش پژوهش

با توجه به اینکه پاسخ‌دهی به مسئله پژوهش حاضر در برگیرنده دو بخش الف) شناسایی ملاک‌های سنجش کلاسی عادلانه و ب) تعیین ویژگی‌های روان‌سنجی ابزار طراحی شده است،

پژوهش حاضر به روش پژوهش آمیخته اکتشافی^۱ از نوع ساخت ابزار انجام شد. بخش اول این پژوهش مطالعه‌ای کیفی و از نوع پدیدارشناسی است که به شناسایی ملاک‌های سنجش کلاسی عادلانه منجر گردید. بدین‌منظور دانشجویان با استفاده از نمونه‌گیری ملاکی انتخاب شدند و با آن‌ها مصاحبه شد. این دانشجویان به‌گونه‌ای انتخاب شدند که حداقل تجربه دو ترم مورد سنجش عملکرد واقع شدن در دانشگاه تهران داشته باشند و حداکثر دو ترم نیز از فراغت از تحصیل آن‌ها نگذشته باشد. همچنین اعضای هیأت علمی به روش نمونه‌گیری سهمیه‌ای غیرنسیبی^۲ انتخاب شدند؛ به این ترتیب که با در نظر گرفتن تجارب اعضای هیأت علمی در رشته‌های مختلف از هر کدام از حوزه‌های پنج‌گانه علوم انسانی، علوم پایه، فنی و مهندسی، کشاورزی و هنر دو نفر انتخاب شدند و با آن‌ها مصاحبه شد. اشباع مضمونی - استقرایی^۳ داده‌ها بعد از مصاحبه با ۱۰ عضو هیأت علمی، ۲۹ دانشجو و یک مصاحبه گروهی و یک مصاحبه گروه کانونی حاصل شد. در مطالعه دوم، از میان دانشجویان تهران تعداد ۵۱۱ نفر از طریق نمونه‌گیری خوشه‌ای سه‌مرحله‌ای انتخاب شدند. یافته‌های بخش کیفی، مبنایی برای شناسایی و رواسازی ملاک‌ها برای، ساخت روبریک گردید. بخش کمی با استفاده از روش پژوهش همبستگی انجام شد. جامعه آماری تمامی دانشجویان کارشناسی، کارشناسی ارشد و دکتری دانشگاه تهران در سال تحصیلی ۱۴۰۰-۱۳۹۹ بودند. نمونه‌گیری به‌صورت خوشه‌ای سه-مرحله‌ای انجام گرفت به این صورت که سه دانشکده از هر یک از دانشکده‌های فنی، علوم، کشاورزی، هنر، علوم انسانی و علوم اجتماعی و رفتاری به‌تصادف انتخاب شده و دانشجویانی که در این دانشکده‌ها مایل به همکاری با پژوهشگر بودند به روبریک پاسخ دادند. هنگام صحبت درباره‌ی روایی محتوایی روبریک، منابع اسناد محتوای روبریک و بررسی محتوای روبریک توسط متخصصان مورد استفاده قرار می‌گیرد و ارتباط بین ملاک‌های روبریک اغلب با استفاده از تحلیل عاملی نشان داده می‌شود (De la Rosa Gómez et al., 2019)؛ از این رو، در این پژوهش توافق بین متخصصان به‌عنوان شاخص روایی محتوایی محاسبه شده است. همچنین روایی سازه به روش تحلیل عاملی اکتشافی با استفاده از روش عامل‌یابی مؤلفه‌های اصلی^۴ با چرخش

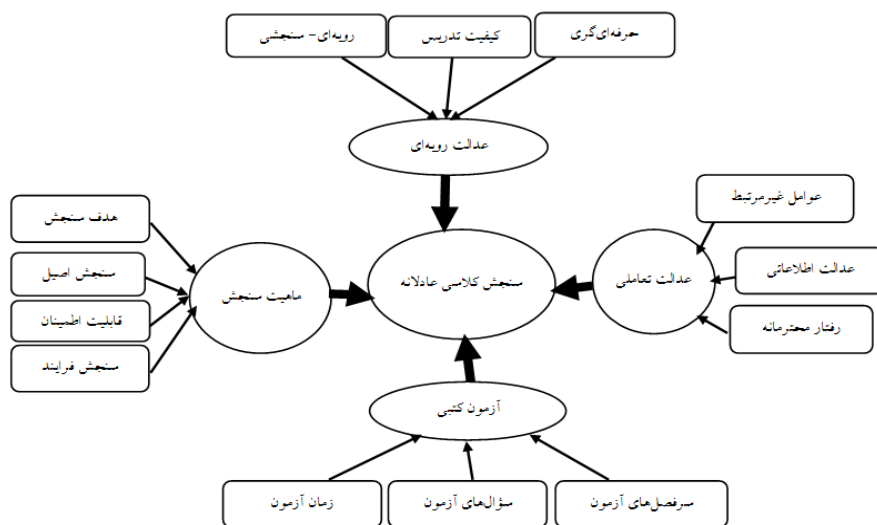
- 1- exploratory sequential mixed method
- 2- Non-proportional quota sampling
- 3- inductive thematic saturation
- 4- Principal components

واریماکس محاسبه شد. برای تعیین اعتبار روبریک با تأکید بر همسانی درونی از روش آماری ضریب آلفا و برای تعیین اعتبار با تأکید بر ثبات از روش بازآزمایی و محاسبه ضریب همبستگی پیرسون استفاده شد.

یافته‌های پژوهش

سؤال اول: ملاک‌های سنجش کلاسی عادلانه چیست؟

۱۱۰ نشانگر ارائه شده توسط پژوهشگران در (Baniasadi et al. (2022) به همراه نشانگرهای به دست آمده از مصاحبه با دانشجویان در مورد سنجش کلاسی عادلانه در Baniasadi et al. (2021) در قالب ۵۹ نشانگر بدون جهت یکپارچه شدند و در قالب مدلی شامل چهار مضمون اصلی «عدالت رویه‌ای»، «ماهیت سنجش»، «عدالت تعاملی» و «عدالت در آزمون کتبی» مبنای شکل‌گیری ۲۰ ملاکی قرار گرفتند که در تدوین روبریک مورد استفاده قرار گرفتند (شکل ۱ و جدول ۱ را ببینید). با توجه به این که روبریک تحلیلی است برای هر ملاک توصیف‌گرهایی در ۴ سطح کیفی عالی، خوب، متوسط و ضعیف تعریف کردیم^۱ (جدول ۲).



شکل ۱. مضمون‌ها و زیرمضمون‌های سنجش کلاسی عادلانه

Figure 1. Themes and sub-themes of fair classroom assessment

۱- در این مقاله، اشاره کامل به توصیف سطوح چهارگانه همه ملاک‌ها به علت گستردگی امکان‌پذیر نبوده است.

جدول ۱. ملاک‌ها و نشانگرهای سنجش کلاسی عادلانه^۱

Table 1. Criteria and indicators of fair classroom assessment

ملاک Criteria	نشانگرها Indicators
فرصت	سنجش مجدد، ارائه دیدگاه دانشجویان، اصالت‌دهی به بررسی مجدد و تقلب
بازخورد	به‌موقع بودن، وضوح و آینده‌نگر بودن
دسترسی	دسترسی حضوری به استاد، دسترسی در فضای مجازی (رایانامه)، دسترسی تلفنی به استاد
پژوهش‌های تخصیص یافته به دانشجویان	پژوهش برای یادگیری، راهنمایی استاد و فرصت انجام پژوهش
بارکاری	یکسان بودن بارکاری، تکالیف گسترده برای ارتقا یادگیری، تناسب با وزن درس و تناسب با نمره تعیین شده
تطابق	نوع سنجش با محتوای سنجش، محتوای سنجش با محتوای آموزش داده شده، سطح سنجش با کیفیت تدریس، سطح سنجش با دوره تحصیلی و منبع سنجش با موضوع سنجش
توجیه	ارائه شواهد قانع‌کننده، ارائه استدلال‌های قانع‌کننده
مقیاس نمره‌دهی	سنجش متأثر از شهرت کلاس، نمره‌دهی بر اساس عملکرد دیگران، وزن متناسب با تکلیف و نمره کار گروهی
حرفه‌ای‌گری در تدریس	مدیریت زمان، سنجش آغازین، طرح درس و بررسی تکالیف
کیفیت تدریس	تسلط و آمادگی تدریس، ارائه توسط دانشجو، آموزش برای سنجش عملی، به-روز بودن استاد، تناسب کیفیت و حجم مطالب ارائه شده با زمان
سنجش فرایند	سنجش مستمر، در نظر گرفتن فعالیت‌های طول ترم و توجه به تلاش دانشجو
قابلیت اطمینان	روایی محتوا، روایی منابع و اعتبار
سنجش اصیل	دانش کاربردی، سنجش کاربرد محتوا
عوامل غیرمرتبط	خطای هاله‌ای (ناشی از قومیت، جنسیت، مذهب، گرایش‌های سیاسی و نظایر آن)، دانشجو قربانی روابط اساتید، روابط استاد-دانشجو و رفتار مناسب با میل استاد
عدالت اطلاعاتی	تعیین انتظارات و اهداف دوره و عمل بر اساس ضوابط توافق شده
رفتار محترمانه	اعتراض به نمره، پذیرش اشتباه و نمره ابزار تهدید دانشجو
سرفصل‌های آزمون	مشخص بودن، به‌موقع بودن و به‌روز بودن
سؤال‌های آزمون	دشواری، ابهام و سطح شناختی
زمان آزمون	تناسب حجم زمان آزمون با سؤال‌ها، مناسبت زمان انجام سنجش، تصادف در تخصیص زمان سنجش به افراد
هدف سنجش	سنجش برای نمره‌دهی یا کمک به فرایند آموزش

۱- ملاک‌های اول تا هشتم در قالب مضمون «رویه‌ای-سنجشی» در شکل ۱، یکپارچه شده‌اند.

جدول ۲. مثالی از ملاک «حرفه‌ای‌گری در تدریس» و سطوح عملکردی آن

Table 2. An example of the criterion of "professionalism in teaching" and its functional levels

حرفه‌ای‌گری در تدریس Professionalism in teaching	
• استاد از زمان کلاس برای آموزش استفاده بهینه می‌کند و تقریباً تمام وقت کلاس، آموزش و بیان سخنان مرتبط با موضوع درسی بود.	۳
•• استاد با انجام یک سنجش ابتدایی به شناخت کلی از دانش پایه کلاس رسید و تدریس را بر آن مبنا شروع کرد.	
••• تقریباً در همه جلسات مشخص بود که استاد با برنامه‌ریزی قبلی و به صورت هدفمند تدریس می‌کند.	
•••• استاد تقریباً به طور کامل تکالیف (پروژه، کارگروهی، پژوهش و نظایر آن) انجام شده توسط ما را بررسی کرده و بر این مبنا میزان یادگیری ما را ارزیابی کرد. می‌کرد	
• استاد گاهی اوقات به بیان نظرات و خاطرات موردعلاقه شخصی خود می‌پرداخت با این حال از آموزش غافل نمی‌شد.	۴
•• استاد با پرسش و پاسخ از تعدادی از دانشجویان، سنجش ابتدایی را انجام داده و آن را مبنای آموزش خود قرار داد.	
••• در اغلب جلسات تدریس مشخص بود که استاد با برنامه‌ریزی قبلی و به صورت هدفمند تدریس می‌کند.	
•••• استاد تا حدود زیادی تکالیف انجام شده توسط ما را بررسی کرده و بر این مبنا میزان یادگیری ما را ارزیابی می‌کرد.	
• استاد بیشتر وقت کلاس را به بیان نظرات و خاطرات موردعلاقه شخصی خود می‌پرداخت و کمتر به آموزش مفید می‌پرداخت.	۵
•• استاد سنجش ابتدایی را انجام داد ولی آن را مبنای آموزش و تدریس خود قرار نداد.	
••• در اغلب جلسات، استاد تدریس منسجم و هدفمندی نداشت.	
•••• استاد نسبتاً تکالیف انجام شده توسط ما را بررسی کرده و بر این مبنا میزان یادگیری ما را ارزیابی می‌کرد.	
• استاد تقریباً تمام وقت کلاس را به بیان نظرات و خاطرات موردعلاقه شخصی خود می‌پرداخت و به آموزش اصلاً توجهی نداشت.	۶
•• استاد هیچ‌گونه سنجش ابتدایی برای آگاهی از سطح علمی کلاس انجام نداد.	
••• تقریباً در تمام طول ترم استاد تدریس منسجم و هدفمندی نداشت و پیدا بود که برنامه مشخصی برای آموزش ما ندارد.	
•••• استاد، تکالیف انجام شده توسط ما را اصلاً بررسی نکرده و در ارزیابی نهایی مدنظر قرار نداد.	

سؤال دوم: آیا روبریک سنجش کلاسی عادلانه از روایی برخوردار است؟

برای پاسخ به این سؤال دو دسته شواهد وابسته به محتوا (روایی محتوا^۱) و شواهد وابسته به سازه (روایی سازه^۲) بررسی شدند. روایی محتوایی شامل دو مرحله متمایز است؛ تلاش‌های

1- Content validity
2- Construct validity

پیشینی به وسیلهٔ تدوین‌کننده‌ی ابزار از طریق مفهوم‌سازی دقیق برای تولید ملاک‌ها و سلول‌های روبریک و تلاش‌های پسین برای ارزیابی محتوایی مقیاس توسط متخصصان را در برمی‌گیرد (Polit & Beck, 2006). در این مرحله بعد از تدوین ملاک‌ها، طبق نظر یکی از اساتید سنجش در بعضی از ملاک‌ها و نشانگرها اصلاحاتی صورت گرفت؛ اما تلاش‌های پسینی دو نوع روایی محتوایی را در برمی‌گیرد که شامل روایی محتوایی گویه‌ها^۱ و روایی محتوایی مقیاس کلی^۲ است. متداول‌ترین راهبردی که برای روایی محتوایی مورد استفاده قرار می‌گیرد، ارزیابی متخصصان است (De la Rosa Gómez et al., 2019). برای محاسبه‌ی روایی محتوایی از گروهی از متخصصان خواسته می‌شود تا هر گویه^۳ را از نظر ارتباط با سازه‌ی زیربنایی درجه‌بندی کنند. با توجه به اینکه روایی محتوایی شاخصی از توافق ارزیابان و متخصصان است و این توافق با پاسخ تصادفی و شانسی متورم می‌شود (Lynn 1986) حداقل سه متخصص را توصیه می‌کند اما بیش از ۱۰ نفر را ضروری نمی‌داند (Polit & Beck, 2006)؛ بر این اساس متخصصانی که ما برای رواسازی روبریک استفاده کردیم شامل ۹ استاد دانشگاه بودند. ملاک انتخاب این افراد، تخصص در حوزه سنجش و تجربه تدریس حداقل ۵ سال در دانشگاه بود. در این پژوهش از متخصصان خواسته شد تا میزان ارتباط هر گویه با سازه‌ی «سنجش کلاسی عادلانه» را در یک طیف لیکرت چهارنقطه‌ای، «کاملاً مرتبط»، «بسیار مرتبط»، «تا حدودی مرتبط» و «مرتبط نیست» درجه‌بندی کنند. انتخاب مقیاس رتبه‌ای چهار نقطه‌ای به جای مقیاس درجه‌بندی ۳ یا ۵ نقطه‌ای برای اجتناب از داشتن نقطهٔ میانی خنثی و دوسویه بود. برای هر گویه شاخص روایی محتوایی گویه با تقسیم تعداد متخصصانی که به گویه رتبه ۳ یا ۴ داده‌اند بر تعداد کل متخصصان محاسبه می‌شود. (Lynn 1986) با ۶ متخصص یا بیشتر، شاخص روایی محتوایی گویه کمتر از ۰,۷۸ را توصیه نمی‌کند (Polit & Beck, 2006).

در این پژوهش شاخص روایی محتوایی مقیاس کلی به توصیه‌ی Polit and Beck (2016) میانگین شاخص‌های روایی محتوایی گویه‌ها در نظر گرفته شد؛ چراکه به جای میانگین عملکرد متخصصان بر میانگین کیفیت گویه‌ها متمرکز است. بسیاری از پژوهشگران مقدار قابل قبول

1- content validity index for items (I-CVI)

2- content validity index for scale (S-CVI)

۳- در تمام متن هر گویه متناظر یک ملاک در روبریک است.

روایی محتوایی مقیاس کلی را ۰/۸۰ و بالاتر از آن می‌دانند؛ اما در مورد سلول‌های روبریک از متخصصان خواسته شد تا بعد از نظردهی در مورد میزان ارتباط ملاک مورد نظر با سازه سنجش کلاسی عادلانه، در مورد دقت و درستی توصیف سطوح عالی، خوب، متوسط و ضعیف هر ملاک داوری کنند و یکی از گزینه‌های «موافقم، تغییری نیاز ندارد»، «موافقم ولی نیازمند تغییراتی است» و «مخالفم» را انتخاب کرده و نظرات اصلاحی خود را در صورت انتخاب گزینه اول یا دوم مطرح کنند تا مبتنی بر این نظرات در مراحل بعدی تغییراتی که ضروری باشد روی محتوای سلول‌ها انجام گیرد.

جدول ۳. روایی محتوایی ملاک‌ها و سطوح مختلف آن‌ها

Table 3. Content validity of the criteria and their different levels

ملاک	روایی محتوایی ملاک	سطح عالی	سطح خوب	سطح متوسط	سطح ضعیف
بازخورد	1	1	1	0.89	1
فرصت	1	0.89	0.89	0.89	0.89
دسترسی	1	1	1	1	1
پژوهش	1	1	1	1	1
بارکاری	1	1	1	1	1
تطابق	1	1	1	1	1
توجیه	0.88	1	1	1	1
مقیاس نمره‌دهی	0.88	1	1	1	1
حرفه‌ای‌گری	0.88	1	1	0.89	0.89
کیفیت تدریس	1	1	1	1	1
سنجش فرایند	1	1	1	1	1
قابلیت اطمینان	1	1	1	1	1
سنجش اصیل	0.88	1	1	0.89	0.89
عوامل غیر مرتبط	1	1	0.89	0.89	0.89
عدالت اطلاعاتی	1	1	1	1	1
رفتار محترمانه	1	1	1	1	1
سرفصل‌های آزمون	1	1	1	1	1
سؤال‌های آزمون	1	1	1	1	1
زمان آزمون	1	1	1	1	1
هدف سنجش	0.88	1	1	1	1

با توجه به اینکه ارائه تمام نظرات به دلیل محدودیت مقاله امکان‌پذیر نیست به نمونه‌ای از این نظرات و اصلاحی که انجام شده اشاره می‌شود. در مورد ملاک «رفتار محترمانه» یکی از متخصصان معتقد بود که این ملاک صرفاً بر عمل و نحوه تعامل استاد تأکید می‌کند و نقش دانشجو و رفتار او مدنظر قرار نگرفته است: «... فریره از مفهوم انضباط متقابل صحبت می‌کند. شما کفه را دارید به نفع دانشجو و علیه استاد سنگین می‌کنید». بر همین مبنا تغییراتی در توصیف سطوح مختلف این ملاک اعمال شد و عبارت «با دانشجویانی که درباره نمره داده شده به آن‌ها توضیح می‌خواستند برخوردی محترمانه و شایسته داشت» به «با دانشجویانی که مؤدبانه و با رعایت جایگاه و منزلت استاد درباره نمره داده شده توضیح می‌خواستند برخوردی محترمانه و شایسته داشت» و عبارت «در صورتی که دانشجو نسبت به نادرستی سؤالی در آزمون اعتراض داشت محترمانه آن را پذیرفته و بررسی می‌کرد» به «در صورتی که دانشجو نسبت به نادرستی سؤالی در آزمون با لحنی مؤدبانه و رسمی نقد داشت، محترمانه آن را پذیرفته و بررسی می‌کرد» تغییر یافت.

روایی محتوایی هر یک از ملاک‌ها بالاتر از مقدار معیار ۰/۷۸ به دست آمد (جدول ۳). روایی محتوایی مقیاس کلی نیز بنا به تعریف برابر ۰/۹۴ بود که بالاتر از مقدار معیار ۰/۸۰ است.

روایی سازه

تحلیل به شیوه اکتشافی و با استفاده از روش عامل‌یابی مؤلفه‌های اصلی و با چرخش واریماکس انجام شد. مقدار KMO کایزر، مایر و اولکین (۰/۸۵۹) بیانگر کفایت داده‌ها برای انجام تحلیل عاملی است و آزمون کرویت بارتلت ($3,832e^3$, $P < 0,0001$) نیز نشان می‌دهد که ماتریس همبستگی داده‌ها در جامعه صفر نیست و بنابراین عامل‌یابی قابل توجیه است. ارزش ویژه و نمودار اسکری بیانگر استخراج ۳ عامل از داده‌ها بود؛ که این عوامل به ترتیب ۲۶/۴۰، ۲۰/۳۳ و ۱۰/۳۵ درصد واریانس متغیرها را تبیین می‌کنند و درصد پوشش واریانس مشترک بین متغیرها برای این سه عامل بر روی هم ۵۷/۰۸ درصد کل واریانس متغیرها را تبیین می‌کند. برای تفسیر نتایج از ملاک حداقل بار عاملی ۰/۳۰ استفاده شد که در تحلیل‌های عاملی معمول است (Foster et al., 2006, 82 cited in Baniasadi & Pourshafei, 2013).

با توجه به جدول ۴ تمام گویه‌های ۱، ۲، ۳، ۶، ۷، ۸، ۹، ۱۰ از مضمون «عدالت رویه‌ای» و تمامی گویه‌های مضمون «عدالت در آزمون کتبی» یعنی گویه‌های ۱۷، ۱۸ و ۱۹ و گویه ۲۰ از

مضمون «ماهیت سنجش» در عامل اول، بار بالایی گرفتند. تمام این گویه‌ها به‌جز گویه ۲۰ به شیوه‌های آموزشی و رویه‌های سنجشی که در تخصیص نمره‌های تأثیر گذاشته و از این طریق ادراک عدالت را تحت تأثیر قرار می‌دهد، اشاره می‌کنند؛ بنابراین گویه ۲۰ حذف می‌شود. گویه ۱۹ نیز به‌واسطه این‌که در دو عامل بار تقریباً یکسانی گرفته و با حذف آن از عامل اول مقدار ضریب آلفا افزایش می‌یابد و هم‌چنین ضریب تمیز آن پایین‌تر از مقدار ۰/۳۰ بوده (جدول ۵ را ببینید) و مناسب نیست گویه ضعیفی تشخیص داده می‌شود و از فرم نهایی روبریک حذف می‌شود. گویه ۱۸ علی‌رغم اینکه در دو عامل، بار بالایی گرفته است در عامل اول بار بیشتری گرفته و با حذف آن مقدار ضریب آلفا نیز کاهش پیدا می‌کند؛ بنابراین این گویه در تحلیل باقی می‌ماند.

جدول ۴. نتایج تحلیل عاملی اکتشافی با چرخش واریماکس

Table 4. Results of exploratory factor analysis with varimax rotation

Component مؤلفه			Item گویه
3	2	1	
-	-	0.58	فرصت
-	-	0.62	بازخورد
-	-	0.73	دسترسی
-	0.76	-	پژوهش
0.49	0.57	-	بارکاری
-	-	0.72	تطابق
-	-	0.73	توجه
-	-	0.75	مقیاس نمره‌دهی
-	-	0.71	حرفه‌ای‌گری در تدریس
-	-	0.56	کیفیت تدریس
-	0.76	-	سنجش فرایند
-	0.81	-	قابلیت اطمینان
-	0.79	-	سنجش اصیل
0.86	-	-	عوامل غیر مرتبط
0.78	-	-	عدالت اطلاعاتی
0.74	-	-	رفتار محترمانه
-	-	0.70	سرفصل‌های آزمون
-	0.39	0.56	سؤال‌های آزمون
-	0.32	0.33	زمان آزمون
-	-	0.72	هدف سنجش

عامل اول را تحت عنوان «عدالت رویه‌ای» نام‌گذاری می‌کنیم. تمام گویه‌های مضمون «ماهیت سنجش» یعنی گویه‌های ۱۱، ۱۲ و ۱۳ و گویه‌های ۴ و ۵ در عامل ۲ بار بالای ۰/۳۵ داشتند با توجه به اینکه گویه ۵ از مضمون عدالت آموزشی در دو عامل ۲ و ۳ بار بالایی گرفته است حذف می‌شود. گویه ۴ هرچند مرتبط با مضمون «عدالت آموزشی» است اما با توجه به این‌که ملاک «پژوهش» به یکی از شیوه‌های سنجش اشاره می‌کند و مقدار ضریب آلفای این عامل با حذف این گویه از ۰/۸۴ به ۰/۸۱ کاهش می‌یابد در تحلیل باقی می‌ماند. این عامل با توجه به محتوای گویه‌ها «ماهیت سنجش» نامیده شد. تمام گویه‌های مضمون «عدالت تعاملی» در عامل ۳ بار بالایی گرفتند و این عامل تحت همین عنوان نام‌گذاری شد؛ بنابراین با حذف سه گویه ۵، ۱۹ و ۲۰ روبریک نهایی شامل ۱۷ گویه شد.

سؤال سوم: آیا روبریک طراحی شده برای سنجش کلاسی عادلانه از اعتبار برخوردار است؟
به‌منظور پاسخ‌دهی به این سؤال پژوهش اقدامات ذیل انجام شد: ۱. بررسی اعتبار و ۲. تحلیل گویه‌ها بر اساس نظریه کلاسیک.

اعتبار روبریک

اعتبار روبریک از طریق هر دو دسته همسانی درونی و ثبات مورد بررسی قرار گرفت. از روش بازآزمایی به منظور برآورد ثبات و از ضریب آلفا کرونباخ به منظور بررسی همسانی درونی استفاده گردید؛ بررسی شد. محاسبه‌ی ضرایب اعتبار با استفاده از شیوه بازآزمایی به فاصله دو هفته در یک نمونه ۱۶ نفری انجام شد، ضریب بازآزمایی برای زیرمقیاس‌های «عدالت رویه‌ای» و «ماهیت سنجش» به ترتیب ۰/۶۵، ۰/۶۳ به دست آمدند که در سطح ۰/۰۰۱ معنی‌دار بودند. همچنین ضریب بازآزمایی برای «عدالت تعاملی» ۰/۵۴ به دست آمد که در سطح ۰/۰۵ معنی‌دار است. ضرایب آلفای کرونباخ برای زیرمقیاس عدالت رویه‌ای با ۱۰ گویه باقیمانده ۰/۸۷، برای زیرمقیاس ماهیت سنجش با ۴ گویه باقی‌مانده ۰/۸۴ و برای زیرمقیاس عدالت تعاملی با ۳ گویه ۰/۷۹ به دست آمد.

تحلیل گویه‌ها

از ضریب دشواری، تمیز و بررسی آلفای کرونباخ در صورت حذف گویه^۱ برای تحلیل

1- alpha-if-item-deleted (AID)

گویه‌ها استفاده شد. همبستگی بین گویه و نمره عامل مربوط به آن گویه به‌عنوان شاخص تمیز گویه در نظر گرفته شد؛ از این رو همبستگی تصحیح‌شده^۱ مورد استفاده قرار گرفت؛ چراکه تعداد گویه‌های هر عامل کم است و موفقیت در گویه‌ای که قرار است ضریب همبستگی آن با نمره کل آزمون محاسبه شود نمره آزمودنی‌ها را در کل آزمون افزایش می‌دهد و ضریب همبستگی محاسبه‌شده افزایش خواهد یافت (Sharifi, 2009). به‌علاوه این همبستگی با اعتبار ابزار نیز مرتبط است زیرا هر چه گویه و زیرمقیاس مربوط به آن همبستگی بیشتری داشته باشند احتمال همبستگی همه گویه‌های آن مقیاس با یکدیگر بیشتر است (Krishnan, 2013). همچنین شاخص دشواری هر گویه با محاسبه میانگین نمره‌های نسبی هر گویه و به کمک

جدول ۵. تحلیل گویه‌های روبریک

گویه	دشواری	تمیز	آلفا در صورت حذف سؤال
Item	Difficulty	Discrimination	Alpha if item deleted
1	0.735	0.502 (0.499)*	0.850
6	0.805	0.618 (0.639)	0.840
7	0.750	0.649 (0.645)	0.837
8	0.815	0.646 (0.659)	0.838
9	0.800	0.624 (0.612)	0.840
10	0.785	0.474 (0.479)	0.851
19	0.757	0.268	0.865
17	0.778	0.592 (0.613)	0.842
18	0.740	0.500 (0.468)	0.849
2	0.892	0.519 (0.521)	0.848
3	0.865	0.649 (0.661)	0.838
آلفای کرونباخ: 0.858			
4	0.895	0.645	0.815
11	0.890	0.680	0.799
12	0.912	0.717	0.783
13	0.905	0.671	0.804
آلفای کرونباخ: 0.843			
14	0.935	0.733	0.612
15	0.942	0.582	0.776
16	0.937	0.610	0.756
آلفای کرونباخ: 0.749			

1- Corrected Correlation

* ضرایب تمیز سوالات بعد از حذف سوال ضعیف ۱۹

نرم افزار R و پکیج sjPlot محاسبه شد. جدول ۵ نشان می‌دهد ضرایب تمام گویه‌ها در بازه ۰/۵۰ تا ۰/۷۳ قرار دارد که در حد قابل قبولی است. همچنین شاخص AID گویه‌های باقیمانده در هر زیرمقیاس از مقدار آلفای آن زیرمقیاس کمتر است که نشان می‌دهد حذف آن گویه همسانی گویه‌های باقیمانده را کاهش می‌دهد و تأثیر گویه در اعتبار روبریک را بیان می‌کند. در مورد ضریب دشواری گویه‌ها با توجه به اینکه گویه‌ها چندارزشی هستند، مقادیر بزرگ‌تر انتخاب سطوح بالاتر گویه را نشان می‌دهد. ضریب دشواری گویه‌ها بین ۰/۷۳ تا ۰/۹۴ تغییر می‌کند که نشان‌دهنده تمایل به انتخاب سطوح بالای گویه‌هاست.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف ساخت ابزاری برای ارزشیابی سنجش کلاسی عادلانه در دانشگاه انجام شد و با توجه به اینکه روبریک‌ها ارزشیابی عادلانه‌تر مهارت‌ها را ممکن می‌کنند و با عنایت به اینکه در پژوهش‌های انجام‌شده در این حوزه پرسشنامه‌هایی آن‌هم در قالب کلی‌تر و نه در حوزه خاص عدالت در سنجش کلاسی ساخته شده است؛ تدوین روبریک سنجش کلاسی عادلانه هدف اساسی این پژوهش بود. با توجه به بافتاری بودن موضوعاتی اخلاقی مانند عدالت، نشانگرها و مدل تدوین‌شده توسط نویسندگان این پژوهش در مطالعه کیفی و نشانگرهای به‌دست‌آمده از مطالعه مروری مبنای تدوین مدل نهایی سنجش کلاسی عادلانه با چهار مضمون اصلی عدالت رویه‌ای، عدالت تعاملی، ماهیت سنجش و عدالت در آزمون کتبی و ساخت روبریک بر اساس آن شد. عدالت آموزشی در این پژوهش بر خلاف پژوهش Bempechat et al. (2013) تحت عنوان عدالت رویه‌ای-آموزشی به همراه عدالت رویه‌ای-سنجشی تحت عنوان کلی عدالت رویه‌ای قرار گرفتند؛ چراکه به لحاظ نظری هر دو به رویه‌های سنجشی و آموزشی اثرگذار بر سنجش دانشجویان اشاره می‌کنند همچنین تحلیل عاملی نیز از بار گرفتن ملاک‌های این دو مضمون در مضمون عدالت رویه‌ای حمایت کردند. پس از تعیین ملاک‌ها و نوشتن توصیفگرهای هر ملاک در چهار سطح عالی، خوب، متوسط و ضعیف، روایی محتوایی آن‌ها توسط متخصصان حوزه سنجش برای اجرا روی نمونه در مرحله کمی تأیید شد. با انجام تحلیل عاملی در مرتبه اول برخی از ملاک‌ها حذف شد و چهار مضمون اولیه در قالب سه مضمون، ساختار جدیدی گرفتند. تنها تغییر مدل مفهومی اولیه بار گرفتن همه ملاک‌های مضمون «عدالت در آزمون کتبی» و ملاک‌های مضمون «عدالت رویه‌ای»

در یک عامل بود. با توجه به اینکه عدالت رویه‌ای به سیاست‌ها و رویه‌های مربوط به درجه‌بندی و نمره‌دهی اشاره دارد (Bempechat et al., 2013) آزمون‌های کتبی به‌عنوان یکی از روش‌های رایج دانشگاه برای تخصیص نمره و سنجش یادگیری دانشجویان و ملاک‌های مرتبط با آن می‌تواند زیر این مفهوم گسترده عدالت رویه‌ای قرار گیرد. در تحلیل عاملی مشخص شد که ملاک «بارکاری» در دو عامل بار بالایی گرفته است که هیچ‌گونه توجیه نظری برای آن وجود نداشت از تحلیل‌های بعدی کنار گذاشته شد. ملاک «هدف سنجش» نیز که در عامل ۱ بار بالایی آستانه مشخص شده گرفته بود که محتوای آن به لحاظ نظری همخوان با سایر گویه‌های آن عامل نبود از تحلیل کنار گذاشته شد. دو گویه «سؤال‌های آزمون» و «زمان آزمون» نیز شرایطی مشابه گویه «بارکاری» داشتند اما در یکی از عواملی که در آن بار بالایی گرفته بودند از دیدگاه نظری منطقی بود چراکه این دو گویه به‌عنوان بخشی از عدالت در آزمون کتبی به رویه‌های موجود در سنجش به‌وسیله این آزمون‌ها اشاره می‌کنند و می‌توانند زیر عنوان بزرگ‌تر عدالت رویه‌ای قرار گیرند؛ با این حال گویه «زمان آزمون» به دلیل اینکه ضریب تمیز آن پایین بود و با حذف آن میزان آلفا افزایش می‌یافت حذف شد. در نهایت روبریکی با ۱۷ ملاک که ویژگی‌های روان‌سنجی مطلوبی داشت برای کاربردهای آتی تدوین شد. این روبریک می‌تواند با شناسایی نقاط قوت و قابل بهبود (ضعف) اعضای هیات علمی در حوزه عدالت سنجش به افزایش سواد سنجشی اعضای هیات علمی، افزایش کیفیت تدریس و سنجش آن‌ها، افزایش بهزیستی دانشجویان و تقویت رفتارهای مدنی و تحصیلی آنان و تدوین دستورالعمل ملی سنجش کلاسی عادلانه کمک کند. همچنین می‌تواند مبنای مناسبی برای آموزش دوره‌های ضمن خدمت ارتقای سواد سنجشی مورد استفاده قرار گیرد. از سوی دیگر واکاوی عمیق مشکلات ساختار نظام آموزش عالی که تحقق ملاک‌ها و نشانگرهای سنجش کلاسی عادلانه را دچار چالش می‌کند، پیشنهاد می‌شود. در انتها، با توجه به اینکه در مقایسه با سایر ابزارها نظیر پرسشنامه‌ها و مقیاس‌های متداول مدادکاغذی، روبریک می‌تواند بخش قابل توجهی از خطاهای ذهنی در پاسخگویی را کاهش داده بر دقت و عینیت داده‌ها بیفزاید، پیشنهاد می‌شود در مطالعات آینده ابزارسازی به استفاده بهینه از این قابلیت و ساخت و میزان‌سازی سایر روبریک‌ها توجه ویژه‌ای شود. در انتها امید است متخصصان سنجش آموزش و روان‌سنجی^۱ در

کشورمان با فاصله گرفتن از استیلای سنجش‌های مداد کاغذی و بهره‌گیری از قابلیت‌های هوش مصنوعی و ظرفیت‌سازی برای استفاده بهینه از کلان‌داده‌ها، خود را با تحولات آینده سنجش عملکرد همراه نماید.

سهم مشارکت نویسندگان: در پژوهش حاضر نویسنده دوم، به‌عنوان استاد راهنما، نظارت و راهبردی روند کلی پژوهش و تدوین و نهایی‌سازی اصلاحات مقاله را بر عهده داشته‌اند. نویسنده اول در تدوین طرح تحقیق، فرآیند گردآوری، تحلیل و تفسیر یافته‌ها و نگارش متن مقاله را برعهده داشته و در مجموع نتیجه‌گیری از یافته‌ها و بسط و تفسیر به صورت مشترک و با بحث و تبادل نظر کلیه همکاران و با همراهی نویسنده سوم، چهارم و پنجم به‌عنوان اساتید مشاور، انجام شد.

تضاد منافع: نویسندگان اذعان دارند که در این مقاله هیچگونه تعارض منافی وجود ندارد.

منابع مالی: پژوهش حاضر از هیچ موسسه و نهادی حمایت مالی دریافت نکرده و کلیه هزینه‌ها در طول فرآیند اجرای پژوهش بر عهده پژوهشگران بوده است.

تشکر و قدردانی: پژوهش حاضر بدون همکاری مشارکت‌کنندگان امکان‌پذیر نبود؛ بدینوسیله از کلیه مشارکت‌کنندگان تقدیر و تشکر به عمل می‌آید.

References

- Allal, L. (2013). Teachers' Professional Judgement in Assessment: A Cognitive Act and a Socially Situated Practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20–34.
- Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education*, 5(3), 197–203.
- Baniasadi, A., & Pourshafei, H. (2013). Validation of Studying Approaches Test (validity and reliability of revised short form). *Journal of Educational Sciences*, 20(1), 221-242. [Persian]
- Bani Asadi, A., Salehi, K., Khodaie, E., Bagheri Noanarast, K., Izanloo, B. (2021). Students' perceptions of fair classroom assessment: A qualitative study. *Educational Measurement and Evaluation Studies*, 11(35), 68-88. doi: 10.22034/emes.2021.249163 [Persian]
- Baniasadi, A., Salehi, K., Khodaie, E., Bagheri, Kh., & Izanloo, B. (2022). Fairness in classroom assessment: A systematic review. *The asia pacific education researcher*. 1-19. <https://doi.org/10.1007/s40299-021-00636-z>
- Bempechat, J., Ronfard, S., Li, J., Mirny, A., & Holloway, S. D. (2013). She always gives grades lower than one deserves: A qualitative study of Russian adolescent perceptions of fairness in the classroom. *Journal of Ethnographic & Qualitative Research*, 7(4), 169- 187.
- Brame, C. J. (2019). Rubrics: Tools to Make Grading More Fair and Efficient. *Teach. Essentials*, 175–184.

- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Burger, R., & Gross, M. (2016). Fairness and university dropout: The role of grading procedures in the development of dropout intentions. *Zeitschrift für Erziehungswissenschaft*, 19(3), 625-647.
- Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly*, 50(1), 58–77.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386.
- Daas, R., Dijkstra, A. B., & Karsten, S. (2019). Assessing young people's citizenship attitudes using rubrics. *Studies in Educational Evaluation*, 62, 118–128
- De la Rosa Gómez, A., Meza Cano, J. M., & Miranda Díaz, G. A. (2019). Validation of a Rubric to Evaluate Open Educational Resources for Learning. *Behavioral Sciences*, 9(12), 126-134.
- DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, 34(5–6), 576–591.
- Duplaga, E. A., & Astani, M. (2010). An exploratory study of student perceptions of which classroom policies are fairest. *Decision Sciences Journal of Innovative Education*, 8(1), 9–33.
- Ghanbari, S., Ardalan, M. R., & Karimi, I. (2015). Effect of the Challenges of Student Learnings Evaluation on Deliberate Practice Study Approach. *Educ Strategy Med Sci*, 8(2), 105-113. [Persian]
- Ghorbankhani, M., Salehi, K., & Moshaddam Zadeh, A. (2020). Construction of a Standardized Questionnaire to Detect the Pseudo Evaluation in Elementary Schools. *Journal of Educational Sciences*, 27(2), 91-116. doi: 10.22055/edus.2020.35053.3114 [Persian]
- Golparvar, M. (2015). The Role of Educational Ethics in Relation between Educational Justice and Iniustice with Educational Cheating: Structural Equation Modeling. *Educational Psychology*, 11(37), 51-67. [Persian]
- González-Chordá, V. M., Mena-Tudela, D., Salas-Medina, P. S., Cervera-Gasch, A., Orts-Cortés, I., & Maciá-Soler, L. (2016). Assessment of bachelor's theses in a nursing degree with a rubrics system: Development and validation study, *Nurse Education Today*, 37, 103–107.
- Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching*, 58(3), 93–98.
- Green, S. K., Johnson, R. L., Kim, D.H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999–1011.

- Ishak, Z., & Fin, L. S. (2013). Truants' and Teachers' Behaviors in the Classroom. *Procedia - Social and Behavioral Sciences*, 103.
- Johnson, R. L., Green, S. K., Kim, D.H., & Pope, N. S. (2008). Educational leaders' perceptions about ethical practices in student evaluation. *American Journal of Evaluation*, 29(4), 520-530.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Krishnan, V. (2013). The early child development instrument (EDI): An item analysis using classical test theory (CTT) on Alberta's data. *Early Child Development Mapping (ECMap) Project Alberta, Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta*.
- Lemons, M., & Seaton, J. (2011). Justice in the classroom: Does fairness determine student cheating behaviors? *Journal of Academic Administration in Higher Education*, 7(1).
- Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation*, 48, 57-66.
- Marzooqi, R., Heidari, M., & Heidari, A. (2013). Investigating the relationship between educational justice and academic burnout of students of the University of Social Welfare and Rehabilitation Sciences. *Strides In Development Of Medical Education*, 10(3), 210-216. [Persian]
- Pancorbo, G., Primi, R., John, O. P., Santos, D., Abrahams, L., & De Fruyt, F. (2020). Development and psychometric properties of rubrics for assessing social-emotional skills in youth. *Studies in Educational Evaluation*, 67, 100938.
- Pepper, M. B., & Pathak, S. (2008). Classroom contribution: What do students perceive as fair assessment? *Journal of Education for Business*, 83(6), 360-368.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497.
- Rodabaugh, R. C. (1994). College students' perceptions of unfairness in the classroom. *To Improve the Academy*, 13(1), 269-282.
- Schmidt, T. A., Houston, M. B., Bettencourt, L. A., & Boughton, P. D. (2003). The impact of voice and justification on students' perceptions of professors' fairness. *Journal of Marketing Education*, 25(2), 177-186.
- Seraji, F., Maroofi, Y., & Razeqi, T. (2013). Identifying the challenges of evaluating what students have learned in the Iranian higher education system. *Educational Measurement and Evaluation Studies*, 4(5), 54-33. [Persian]
- Sharifi, H. (2009). *Principles of psychometrics and psychoanalysis*. Rishd. [Persian]

- Usman-Stillman, A. S., Englund, M., Webb, C., & Grenell, A. (2018). Reliability and validity of a measure of preschool children's theatre arts skills: The Preschool Theatre Arts Rubric. *Early Childhood Research Quarterly*, 45, 249-262.
- Tata, J. (2005). The influence of national culture on the perceived fairness of grading procedures: A comparison of the United States and China. *The Journal of Psychology*, 139(5), 401-412.
- Tierney, R. D. (2013). *Fairness in classroom assessment*. In J. H. McMillan (Ed.). SAGE Handbook of Research on Classroom Assessment. Thousand Oaks, CA: Sage Publications.
- Tierney, R. D. (2014). Fairness as a Multifaceted Quality in Classroom Assessment. *Studies in Educational Evaluation*, 43, 55-69.
- Vaghoor Kashani. M., Haii Hossein Neiad. G., Moosapour. N., & Ebrahim Zadeh. I. (2019). Designing and Validating a Conceptual Model for the Evaluation of Teachers' Professional Competencies in Iran. *Journal of Educational Scinces*, 26(2), 27-50. doi: 10.22055/edus.2019.29678.2818 [Persian]

