

Technical Note

Open Access

A Fair Scoring System with Score Calibration to Reduce Evaluator Bias and Question Difficulty in the Student Scientific Olympiad of Health Management

Abtin Heidarzadeh^{1,2} , Ideh Dadgaran^{1*} 

1. Medical Education Research Center, Education Development Center, Guilan University of Medical Sciences, Rasht, Iran
2. Department of Community Medicine, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran

Received: 2024/09/20

Accepted: 2024/11/21

Keywords:

Calibration
Scientific Olympiad
Students
SoTL

***Corresponding author:**

Ideh Dadgaran, Medical
Education Research Center,
Education Development Center,
Guilan University of Medical
Sciences, Rasht, Iran
i.dadgaran@gmail.com

**ABSTRACT**

Today, educational systems are vital for social and economic development. In this regard, making fair and accurate assessments of student competencies, particularly in medical sciences, is essential. The purpose of this study is to explain the challenges and necessities of designing a new scoring system in the student Olympiad of health system management and to seek to provide practical and scientific solutions to improve the scoring process and achieve educational justice in this field.

This developmental process utilized the Scholarship of Teaching & Learning (SoTL) approach within an innovative Multi-Dimensional Scores Calibration Model (MDSCM). The model comprises seven steps: needs and objectives analysis, scoring system design, anchor point determination, evaluator training, system implementation, analysis and evaluation, and continuous improvement.

The process successfully met the defined educational objectives. The novel scoring system, based on score calibration and anchor point adjustments, minimized evaluator discrepancies and question difficulty errors, enhancing scoring accuracy and fairness. Additionally, it empowered evaluators and led to the development of score calculation software.

The findings indicated that innovative approaches and calibration techniques can significantly mitigate scoring biases and promote educational equity. This system enables students to showcase their true abilities and fosters a fairer, more efficient educational environment. The achievements of this process can be considered a model for other educational systems and scientific assessments, contributing to the improvement of quality in education and evaluation across various fields.

How to Cite This Article: Heidarzadeh A, Dadgaran I. A Fair Scoring System with Score Calibration to Reduce Evaluator Bias and Question Difficulty in the Student Scientific Olympiad of Health Management. *Res Med Edu.* 2025;16 (4):68-74.

 [10.32592/rmegums.16.4.68](https://doi.org/10.32592/rmegums.16.4.68)

Copyright © 2025 Research In Medical Education, and Guilan University of Medical Sciences.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
License (<https://creativecommons.org/mission/by-nc/4.0/>)

نظام نمره‌دهی عادلانه با ترازسازی نمرات برای کاهش خطای داوران و سختی سوالات در المپیاد علمی دانشجویی مدیریت نظام سلامت

آبتین حیدرزاده^{۱*}، ایده دادگران^۲

۱. مرکز تحقیقات آموزش پزشکی، مرکز مطالعات و توسعه آموزش پزشکی، دانشگاه علوم پزشکی گیلان، رشت، ایران

۲. گروه پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی گیلان، رشت، ایران

اطلاعات مقاله

نوع مقاله:

یادداشت فنی

تاریخچه:

تاریخ دریافت: ۱۴۰۳/۰۶/۳۰

تاریخ پذیرش: ۱۴۰۳/۰۹/۰۱

کلیدواژه‌ها:

المپیاد علمی

ترازسازی

دانشجویان

رویکرد SoTL

*نویسنده مسئول:

ایده دادگران، مرکز تحقیقات

آموزش پزشکی، مرکز مطالعات و

توسعه آموزش پزشکی، دانشگاه

علوم پزشکی گیلان، رشت، ایران

i.dadgaran@gmail.com

مقدمه

در دنیای امروز، نظام‌های آموزشی به‌عنوان یکی از ارکان اصلی توسعه اجتماعی و اقتصادی کشورها شناخته می‌شوند. در این راستا، ارزیابی عادلانه و دقیق از توانمندی‌های دانشجویان، به‌ویژه در حوزه‌های علوم پزشکی، از اهمیت ویژه‌ای برخوردار است. المپیادهای علمی دانشجویی، به‌عنوان یک بستر رقابتی و چالشی، فرصتی مناسب برای شناسایی و پرورش استعدادها و جوانان در این حوزه به‌شمار می‌آیند. از آغاز برگزاری این المپیادها در کشور، هدف اصلی این المپیادها شناسایی، هدایت و شکوفایی توانمندی‌های دانشجویان بوده

در دنیای امروز، نظام‌های آموزشی به‌عنوان یکی از ارکان اصلی توسعه اجتماعی و اقتصادی کشورها شناخته می‌شوند. در این راستا، ارزیابی عادلانه و دقیق از توانمندی‌های دانشجویان، به‌ویژه در حوزه‌های علوم پزشکی، از اهمیت ویژه‌ای برخوردار است. هدف از این مطالعه تبیین چالش‌ها و ضرورت‌های طراحی نظام نوین نمره‌دهی در المپیاد دانشجویی مدیریت نظام سلامت است و به دنبال ارائه راهکارهایی عملی و علمی برای بهبود فرایند نمره‌دهی و تحقق عدالت آموزشی در این زمینه است. این فرایند توسعه‌ای بوده و با رویکرد SoTL (Scholarship of Teaching & Learning) انجام شده است. این فرایند در قالب مدل نوآورانه ترازسازی نمرات، تحت‌عنوان مدل نوآورانه ترازسازی چندوجهی نمرات MDSCM (Multi-Dimensional Scores Calibration Model) مشتمل بر ۷ گام تحلیل نیازها و اهداف، طراحی سیستم نمره‌دهی، تعیین نقطه لنگر، آموزش ارزیابان و داوران، اجرای سیستم نمره‌دهی، تحلیل و ارزیابی و بهبود مستمر طراحی و اجرا شد. نتایج حاصل از اجرای این فرایند بیانگر دستیابی به اهداف آموزشی تعیین‌شده بود و طراحی نظام نوین نمره‌دهی براساس ترازسازی نمرات و جابه‌جایی نقطه لنگر، نتایج مثبتی را به همراه داشت و به‌طور مؤثری به کاهش خطای ناشی تفاوت در نمره‌دهی داوران و سختی سؤالات و به بهبود دقت و عدالت در نمره‌دهی منجر شد. از دیگر نتایج مثبت این فرایند می‌توان به توانمندسازی داوران در نمره‌دهی و توسعه نرم‌افزاری برای محاسبه نمرات اشاره کرد. یافته‌ها نشان می‌دهد که با اتخاذ رویکردهای نوین و استفاده از تکنیک‌های ترازسازی می‌توان به کاهش سوگیری‌های نمره‌دهی و تحقق عدالت آموزشی نزدیک‌تر شد. این نظام نه تنها می‌تواند به دانشجویان کمک کند تا توانایی‌های واقعی خود را نشان دهند، بلکه می‌تواند به ایجاد یک نظام آموزشی عادلانه‌تر و کارآمدتر کمک کند. دستاوردهای این فرایند می‌تواند به‌عنوان یک مدل برای سایر نظام‌های آموزشی و آزمون‌های علمی در نظر گرفته شود و به بهبود کیفیت آموزش و ارزیابی در حوزه‌های مختلف کمک کند.

است. با این حال، چالش‌های موجود در فرایند نمره‌دهی به آزمون‌های تشریحی، ضرورت طراحی یک نظام نوین نمره‌دهی در این آزمون‌ها را به‌وضوح نمایان می‌کند (۱، ۲). آزمون‌های تشریحی به دلیل نوع سؤالات و آزادی عمل بیشتری که برای دانشجویان فراهم می‌کنند، به‌عنوان ابزاری مؤثر در ارزیابی توانایی‌های تحلیلی و انتقادی دانشجویان شناخته می‌شوند. اما نمره‌دهی به این نوع آزمون‌ها با چالش‌های متعددی روبه‌روست. از جمله این چالش‌ها می‌توان به عدم وجود پاسخ‌های قطعی، تعداد محدود مصححین و زمان بر بودن

داوران یا آزمون‌ها، نادرست یا ناعادلانه به نظر می‌رسند. با جابه‌جایی نقطه لنگر، دقت نمرات افزایش می‌یابد و نمرات به واقعیت نزدیک‌تر می‌شوند. این روش به کاهش نابرابری‌های نمره‌دهی کمک می‌کند و موجب می‌شود که همه داوران به یک استاندارد مشترک نزدیک شوند. ترازسازی نمرات با جابه‌جایی نقطه لنگر یک ابزار مؤثر برای بهبود دقت و عدالت در نمره‌دهی است (۸، ۹).

این مقاله به بررسی چالش‌ها و ضرورت‌های طراحی نظام نوین نمره‌دهی در المپیاد دانشجویی مدیریت نظام سلامت می‌پردازد و به دنبال ارائه راهکارهایی عملی و علمی برای بهبود فرایند نمره‌دهی و تحقق عدالت آموزشی در این زمینه است. با توجه به اهمیت موضوع و تأثیر آن بر آینده تحصیلی و شغلی دانشجویان، امید است که این تلاش‌ها به ایجاد یک نظام آموزشی عادلانه‌تر و کارآمدتر منجر شود.

روش‌ها

این فرایند توسعه‌ای بوده و با رویکرد (Scholarship of Teaching & Learning) انجام شده است. از ابتدای آبان‌ماه ۱۴۰۲ که فرایند المپیاد شانزدهم شروع شد، با توجه به شکایات مکرر دانشجویان شرکت‌کننده در المپیادهای سال‌های قبل، اولین اقدامی که انجام شد، برگزاری جلسه ویناری با نمایندگان دانشجویان مدال‌آور سال‌های گذشته بود. این دانشجویان لیستی بالغ بر ۱۵ مشکل فرایندی و ساختاری به دبیر حیطه المپیاد مدیریت سلامت ارائه کردند و سپس از طریق سامانه پرس‌لاین از دانشجویان متقاضی شرکت در المپیاد شانزدهم سؤال شد که به نظر آنان بیشترین اولویت مربوط به کدام یک از مشکلات است. عیناً همین سؤالات در قالب پرس‌لاین جهت مسئولان حیطه مدیریت المپیاد شانزدهم در دانشگاه‌های علوم پزشکی مختلف کشور اجرا شد. درنهایت، مهم‌ترین و پُرچالش‌ترین مشکل، عدم رعایت عدالت در تصحیح سؤالات تشریحی و پروپوزال‌های گروه‌های شرکت‌کننده در المپیاد و داوری ناعادلانه درمورد ارائه نهایی تیم‌ها در سنوات گذشته تعیین شد.

در ادامه با برگزاری نشست‌های مشورتی با اساتید و کارشنان خبره سنجش، ارزشیابی و آموزش پزشکی دانشگاه‌های (علوم پزشکی و غیرعلوم پزشکی) کشور و مرکز سنجش آموزش پزشکی کشور و دو تن از متخصصان سنجش و ارزشیابی از دانشگاه علوم و تکنولوژی نروژ (Norwegian University of Science and Technology) و دانشگاه (UBC)

فرایند تصحیح اشاره کرد. علاوه‌براین، سوگیری‌های فردی و سلیقه‌های شخصی مصححان می‌تواند به نمرات غیرعادلانه منجر شود. به‌همین دلیل، بررسی و طراحی یک نظام نمره‌دهی عادلانه و دقیق به‌عنوان یک ضرورت در نظام آموزشی مطرح می‌شود (۳، ۴، ۵).

در حال حاضر چالش‌های مختلف موجود در امر ارزیابی دانشجو توسط آزمون‌های تشریحی وجود دارد و انتخاب روش مناسب نمره‌دهی و اثرات متفاوت ناشی از روش نمره‌دهی، ضرورت پرداختن به این موضوع را به‌عنوان یکی از اولویت‌های حیطه ارزیابی، بیش‌ازپیش مورد تأکید قرار می‌دهد (۶). از طرف دیگر، نتایجی که از تحقیقات مختلف در این زمینه به دست آمده، نشان داده است که معلمان مدارس، استادان دانشگاه و کلیه کسانی که درباره نوشته‌های دیگران قضاوت می‌کنند، تفاوت‌های فردی بسیار زیادی را نشان می‌دهند. این تفاوت‌ها در حدی است که می‌تواند بر روی سرنوشت صاحبان نوشته‌ها تأثیرگذار باشد (۷). اهمیت این چالش‌ها به‌حدی است که استفاده از فناوری‌های روز دنیا را برای نمره‌گذاری آزمون‌های تشریحی به بحث جدیدی در این باره تبدیل کرده است (۶).

روش‌های نوین نمره‌دهی، از جمله ترازسازی نمرات با جابه‌جایی نقطه لنگر، به‌عنوان راهکارهایی برای کاهش خطاهای موجود در نمره‌دهی و افزایش دقت ارزیابی‌ها، اهمیت ویژه‌ای دارند. این روش‌ها به ما این امکان را می‌دهند که نمرات را براساس استانداردهای مشترک تنظیم و از نابرابری‌های موجود در نمره‌دهی جلوگیری کنیم. در این راستا، طراحی یک نظام نوین نمره‌دهی که به عدالت آموزشی احترام بگذارد و توانایی‌های واقعی دانشجویان را به‌درستی نمایان کند، بیش‌ازپیش احساس می‌شود.

ترازسازی نمرات با جابه‌جایی نقطه لنگر (Anchor Point Shifting) یک روش متداول در ارزیابی و نمره‌دهی است که به‌منظور بهبود دقت و عدالت در نمره‌دهی استفاده می‌شود. نقطه لنگر به یک استاندارد یا مرجع مشخص اشاره دارد که برای تراز کردن نمرات استفاده می‌شود. این نقطه می‌تواند یک نمره خاص، یک سطح عملکرد یا یک معیار دیگر باشد که به‌عنوان مبنای مقایسه عمل می‌کند. جابه‌جایی نقطه لنگر به معنای تغییر یا تنظیم این مرجع به‌گونه‌ای است که نمرات به‌صورت بهتری با یکدیگر مقایسه شوند. این کار معمولاً در مواردی انجام می‌شود که نمرات اولیه به دلیل تفاوت‌های ساختاری یا ناهم‌هنگی‌های بین

University of British Columbia مقرر گردید یک روش جدید برای تصحیح، استانداردسازی و نمره‌دهی به سوالات تشریحی، بررسی پروپوزال‌ها و نمره‌دهی به ارائه‌های گروهی طراحی شود، به گونه‌ای که ابتدا این روش به صورت آزمایشی اجرا شود و پس از برطرف شدن مشکلات، نهایی‌سازی و رسماً لحاظ گردد. بدین منظور مجریان این فرایند جهت طراحی نظام نوین نمره‌دهی، مدل نوآورانه ترازسازی چندوجهی نمرات MDSCM (Multi-Dimensional Scores Calibration Model) را معرفی کردند که شامل ۷ گام و در مطالعه حاضر به شرح زیر انجام شده است:

۱- تحلیل نیازها و اهداف

شناسایی مشکلات موجود: مشکلات فعلی در نظام نمره‌دهی، مانند تفاوت در نمرات داوران و سختی سوالات، بررسی و تحلیل شد.

تعیین اهداف: اهداف سیستم جدید، از جمله ترازسازی نمرات، افزایش دقت نمره‌دهی و کاهش نابرابری‌ها مشخص شد.

۲- طراحی سیستم نمره‌دهی

توسعه معیارهای نمره‌دهی: معیارهای مشخص برای نمره‌دهی سوالات تشریحی، شامل استانداردهای ارزیابی و نقاط قوت و ضعف تعیین شد. قبل از تصحیح سوالات، ارزیابان با یکدیگر برای تصحیح سوالات براساس استانداردهای تعیین شده به توافق رسیدند.

ایجاد فرم‌های نمره‌دهی: فرم‌های نمره‌دهی که شامل بخش‌های مختلف برای ارزیابی جنبه‌های مختلف پاسخ‌ها بودند، طراحی شدند.

۳- تعیین نقطه لنگر

شناسایی نقاط لنگر: سوالات و پاسخ‌های مرجع انتخاب شدند و از آن‌ها به عنوان نقاط لنگر برای نمره‌دهی استفاده

شد.

تدوین و ارزیابی روش جابه‌جایی نقاط لنگر: به منظور تراز کردن نمرات و کاهش خطاها، روشی برای جابه‌جایی نقطه لنگر طراحی شد و با استفاده از نرم‌افزار Stata تحلیل انجام گرفت. در این فرایند ۶ ارزیاب (داور) شرکت داشتند و تعداد سوالات تشریحی انفرادی ۱۶ سؤال بود. در ابتدای کار ترازسازی نمرات، با توجه به گستردگی دامنه نمرات، ابتدا باید نمرات در مبنای ۱۰۰۰ تراز می‌شد. بدین منظور در گام اول، ۳ گروه دونفره تشکیل شدند و به گروه‌های دونفره گفته شد که هر کدام جداگانه به یک سؤال نمره دهند؛ یعنی به ازای هر سؤال، ۶ نمره به دست آمد. حال باید نمرات داوران مختلف نسبت به همدیگر در یک مسیر قرار می‌گرفت که بدین منظور از ضریب K_1 برای تراز کردن نمرات ۶ داور استفاده شد. در گام بعد مرکز ثقل عوض شد و روی فرد ارزیاب متمرکز گردید و نمرات بقیه به نسبت داور محاسبه شد که همان ضریب K_2 بود. سپس در فاز سوم ضریب K_3 محاسبه گردید که شامل میانگین نمرات گروه داوران (هر ۶ نفر بود). برای هم‌ترازی نمرات در گام چهارم، هر سه ضریب K_1 ، K_2 و K_3 در یک معادله Kmeans Cluster Analysis تحت عنوان K_4 گذاشته شدند که توسط آن تمامی نمرات برای هر سؤال تشریحی برای هر فرد محاسبه شد. در نهایت، برای ۱۶ سؤال تشریحی ۱۶ بار این روش محاسبه گردید و یک معادله نمره تراز نهایی نوشته شد. دقیقاً همین مراحل برای تصحیح پاسخ‌های پروژه‌های مرحله گروهی اجرا شد. در جدول ۱ نمونه‌ای از نمرات تراز شده بر مبنای ۱۰۰۰ حول ۳ لنگر مربوط به پروژه‌های مرحله گروهی آمده است (جدول ۱).

جدول ۱. نمونه‌ای از نمرات تراز شده بر مبنای ۱۰۰۰ حول ۳ لنگر

ردیف	پروژه	نمره تراز شده گروهی لنگر ۱	نمره تراز شده گروهی لنگر ۲	نمره تراز شده گروهی لنگر ۳	نمره تراز شده گروهی نهایی در مبنای ۱۰۰۰ هر سه لنگر
۱	A	۴۷۳	۲۳۶	۱۶۸۹	۳۸۴/۶۱۵۴
۲	B	۹۴۰	۴۷۱	۱۷۶۴	۹۳۴/۰۶۵۹
۳	C	۴۶۱	۵۷۷	۴۰۶	۳۹۵/۶۰۴۴
۴	D	۹۸۵	۷۰۵	۹۵۱	۸۳۵/۱۶۴۸
۵	E	۸۷۷	۲۹۲	۱۷۵۱	۸۶۸/۱۳۱۹
۶	F	۱۰۱۶	۹۰۴	۱۱۵۴	۹۶۷/۰۳۳
۷	G	۱۰۴۵	۴۰۵	۱۷۴۷	۸۴۶/۱۵۳۸
۸	H	۴۱۳	۸۷۶	۴۰۶	۳۹۵/۶۰۴۴

۴- آموزش ارزیابان و داوران

برگزاری کارگاه‌های آموزشی: کارگاه‌های آموزشی برای ارزیابان و داوران به‌منظور آشنا کردن آن‌ها با معیارهای نمره‌دهی و روش‌های جدید برگزار شدند.

تدوین راهنماهای نمره‌دهی: تهیه و توزیع راهنماهای نمره‌دهی که شامل مثال‌ها و توضیحات دقیق بود.

۵- اجرای سیستم نمره‌دهی

اجرای آزمایشی: یک دوره آزمایشی از سیستم جدید نمره‌دهی برای شناسایی مشکلات و نقاط قوت اجرا شد.

جمع‌آوری داده‌ها: داده‌های نمرات و بازخورد از داوران و دانشجویان جمع‌آوری گردید.

۶- تحلیل و ارزیابی

تحلیل داده‌ها: داده‌های جمع‌آوری‌شده تحلیل و الگوها و نقاط ضعف در سیستم شناسایی شدند.

ترازسازی نمرات: روش‌های ترازسازی بر روی نمرات به‌منظور کاهش تفاوت‌ها و نابرابری‌ها اعمال شدند.

۷- بهبود مستمر

بازخورد و اصلاح: بازخورد داوران و دانشجویان جمع‌آوری گردید و برای اصلاح سیستم راهکارهایی ارائه شد.

توسعه و به‌روزرسانی: معیارها و روش‌های نمره‌دهی براساس تجربیات به‌دست‌آمده و نیازهای جدید به‌روزرسانی شدند.

یافته‌ها

در رابطه با هدف ویژه اول- طراحی نظام نمره‌دهی مبتنی بر ترازسازی نمرات: براساس ترازسازی نمرات و جابه‌جایی نقطه لنگر، نظام نمره‌دهی جدید طراحی و نمرات دانشجویان با رویکرد عدالت‌محور محاسبه شد.

در رابطه با هدف ویژه دوم- کاهش خطای تفاوت داوران: نتایج حاصل از این فرایند بیانگر این بود که با استفاده از رویکرد نوین ترازسازی نمرات، خطای تفاوت نمرات داوران به‌طور معناداری کاهش یافت.

در رابطه با هدف ویژه سوم- کاهش خطای سختی سوالات تشریحی: نتایج نشان داد که روش‌های ترازسازی به کاهش خطای سختی سوالات تشریحی کمک کردند و نمرات به واقعیت نزدیک‌تر شدند.

در رابطه با هدف ویژه چهارم- ارزیابی و ارزشیابی نظام نمره‌دهی: نظام نمره‌دهی جدید به‌طور کامل ارزیابی شد و

کارایی و اثربخشی آن با توجه تحلیل نتایج نمرات ارزیابان تأیید گردید.

در رابطه با هدف ویژه پنجم- توسعه نرم‌افزار برای محاسبه نمرات: از نرم‌افزار Stata و روش آماری Cluster: Analysis Kmeans استفاده شد و نمرات ترازسازی شدند.

در رابطه با هدف ویژه ششم- توانمندسازی داوران: جلسات توجیهی و توانمندسازی قبل و حین اجرای فرایند برای ارزیابان تشکیل شدند. داوران با استفاده از کارگاه‌های آموزشی، با اطلاعات لازم در ارتباط با نظام نمره‌دهی جدید که مشتمل بر ارزیابی چندمرحله‌ای و نحوه نمره‌دهی بود، آشنا شدند.

پایش و ارزیابی نظام نمره‌دهی: نظام نمره‌دهی جدید به‌طور مستمر پایش و ارزیابی شد تا اطمینان حاصل شود که کارایی و اثربخشی آن در طول زمان حفظ می‌شود.

نقاط قوت و ضعف

نقاط قوت

- **عدالت آموزشی:** طراحی نظام نمره‌دهی جدید می‌تواند از ایجاد بی‌عدالتی‌های موجود در فرایند ارزیابی جلوگیری کند و شرایط بهتری برای تمامی شرکت‌کنندگان فراهم نماید.

- **کاهش سوگیری‌های نمره‌دهی:** با جابه‌جایی نقطه لنگر و استفاده از ترازسازی نمرات، می‌توان خطای ناشی از تفاوت در ارزیابی نمره‌دهندگان را کاهش داد که به بهبود دقت نمره‌دهی کمک می‌کند.

- **تقویت اعتبار آزمون:** ایجاد سیستم نمره‌دهی عادلانه و استاندارد موجب اعتماد بیشتر به نتایج آزمون‌ها می‌شود و این می‌تواند به افزایش اعتبار و وجهه المپیادها کمک کند.

- **افزایش انگیزه و اعتماد به نفس:** با تضمین عدالت در ارزیابی، شرکت‌کنندگان احساس می‌کنند که عملکردشان به‌درستی شناسایی می‌شود و این امر می‌تواند انگیزه و اعتماد به نفس آن‌ها را افزایش دهد.

- **تحلیل دقیق تر نمرات:** با استفاده از روش‌های آماری پیشرفته و ترازسازی نمرات می‌توان تحلیل‌های دقیق‌تری انجام داد که به درک بهتری از سختی سوالات و عملکرد شرکت‌کنندگان منجر می‌شود.

نقاط ضعف

- **پیچیدگی اجرا:** اجرای یک نظام نوین نمره‌دهی با

جابه‌جایی نقطه لنگر ممکن است پیچیده باشد و به زمان و منابع انسانی و مالی قابل توجهی نیاز داشته باشد.

- مقاومت در برابر تغییر: نمره‌دهندگان و آموزگاران ممکن است نسبت به تغییر در روش‌های نمره‌دهی و ارزیابی مقاوم باشند و این می‌تواند به چالش‌هایی در اجرای سیستم جدید منجر شود.

- نیاز به آموزش: برای اجرای موفق این سیستم، نمره‌دهندگان و متصدیان به آموزش‌های لازم برای درک و بهره‌برداری مؤثر از نظام جدید نیاز خواهند داشت.

- محدودیت‌های فناوری: وابستگی به تکنولوژی برای ترازسازی نمرات ممکن است در مناطق کم‌برخوردار با مشکلاتی مواجه شود، به‌ویژه در تأمین زیرساخت‌های لازم.

بحث و نتیجه‌گیری

در فرایند ارزیابی دانشجویان، یکی از چالش‌های اصلی، تفاوت‌های قابل توجه در نمرات داوران و سختی سوالات تشریحی است. این تفاوت‌ها می‌تواند ناشی از سوگیری‌های شخصی، خستگی یا سلیقه‌های متفاوت داوران باشد که در نهایت به نمرات ناعادلانه و غیرواقعی منجر می‌شود. بررسی‌های انجام‌شده نشان می‌دهند که نمرات دانشجویان در یک گروه با نمرات دانشجویان گروه دیگر در آزمون‌های مشابه به‌شدت متفاوت است و این امر تأحد زیادی به تفکر و سلیقه فردی داوران بستگی دارد (۱۰، ۱۱، ۱۲).

استفاده از یک فرد برای ارزیابی پاسخ‌ها به‌دلیل زمان‌بر بودن و عدم وجود پاسخ‌های قطعی، مشکلاتی را به همراه دارد. به‌همین دلیل، ایجاد یک نظام نمره‌دهی که بتواند این چالش‌ها را برطرف کند، ضروری است. مدل نوآورانه ترازسازی عادلانه نمرات (MDSCM) که در این تحقیق معرفی می‌شود، به‌دنبال طراحی یک نظام نمره‌دهی دقیق و عادلانه است که با استفاده از روش‌های ترازسازی و جابه‌جایی نقطه لنگر، به کاهش خطای تفاوت داوران و سختی سوالات کمک می‌کند. این نظام نمره‌دهی شامل مراحل مختلفی از جمله تحلیل نیازها و اهداف، طراحی سیستم نمره‌دهی، تعیین نقطه لنگر، آموزش داوران، اجرای سیستم نمره‌دهی، تحلیل و ارزیابی داده‌ها و بهبود مستمر است. هریک از این مراحل به‌دقت طراحی شده‌اند تا اطمینان حاصل شود که نمرات به‌صورت عادلانه و دقیق ارزیابی می‌شوند.

همچنین با توجه به اهمیت آموزش و توانمندسازی داوران، برگزاری کارگاه‌های آموزشی و توسعه راهنماهای نمره‌دهی نیز

در این نظام در نظر گرفته شده است. این اقدامات نه تنها به افزایش دقت نمره‌دهی کمک می‌کنند، بلکه به ایجاد یک فرهنگ عادلانه و شفاف در ارزیابی دانشجویان نیز می‌انجامد. یافته‌ها نشان می‌دهند که با اتخاذ رویکردهای نوین و استفاده از تکنیک‌های ترازسازی می‌توان به کاهش سوگیری‌های نمره‌دهی و تحقق عدالت آموزشی نزدیک‌تر شد. این نظام نه تنها به دانشجویان کمک می‌کند تا توانایی‌های واقعی خود را نشان دهند، بلکه می‌تواند به ایجاد یک نظام آموزشی عادلانه‌تر و کارآمدتر کمک کند. با توجه به چالش‌های موجود در نظام‌های نمره‌دهی سنتی و نابرابری‌های سیستمیک، طراحی یک نظام نمره‌دهی نوین که به عدالت آموزشی احترام بگذارد، بیش‌ازپیش احساس می‌شود (۱۳).

یافته‌ها بیانگر این هستند که طراحی نظام نوین نمره‌دهی براساس ترازسازی نمرات و جابه‌جایی نقطه لنگر به‌طور مؤثری توانسته است خطای تفاوت داوران و سختی سوالات را کاهش دهد (۱۴). این نظام به بهبود دقت و عدالت در نمره‌دهی کمک می‌کند و به تحقق اهداف آموزشی نزدیک‌تر می‌شود. از دیگر نتایج مثبت این مطالعه می‌توان به توانمندسازی داوران و توسعه نرم‌افزار برای محاسبه نمرات اشاره کرد. آموزش داوران برای استفاده از نظام نمره‌دهی جدید و پایش و ارزیابی مستمر این نظام از جمله اقداماتی است که برای اطمینان از کارایی و اثربخشی آن در طول زمان ضروری است.

به‌طور کلی، این فرایند نشان می‌دهد که با اتخاذ رویکردهای نوین در نمره‌دهی و توجه به چالش‌های موجود می‌توان به کاهش سوگیری‌های تصحیح و نمره‌دهی آزمون‌ها و در نهایت تحقق عدالت آموزشی نزدیک‌تر شد. با این حال، محدودیت‌هایی مانند تمرکز بر یک حوزه خاص، عدم بررسی تأثیرات بلندمدت، وابستگی به منابع انسانی و مالی و عدم استفاده کامل از فناوری‌های پیشرفته مانند هوش مصنوعی از جمله چالش‌های اصلی آن هستند. همچنین، مطالعه حاضر به تأثیرات روان‌شناختی نظام جدید بر داوران و دانشجویان نپرداخته و مقایسه سیستماتیک با روش‌های سنتی نمره‌دهی انجام نداده است. برای افزایش اعتبار و کارایی این نظام، پیشنهاد می‌شود مطالعات گسترده‌تر در سایر حوزه‌ها انجام گیرند، از فناوری‌های پیشرفته بهره گرفته شود و تأثیرات بلندمدت و روان‌شناختی آن بررسی گردد.

در نهایت، دستاوردهای این فرایند می‌تواند به‌عنوان یک مدل برای سایر نظام‌های آموزشی و آزمون‌های علمی در نظر گرفته شود و به بهبود کیفیت آموزش و ارزیابی در حوزه‌های مختلف

کمک کند.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

در این تحقیق، تمامی مراحل طراحی و اجرای نظام نمره‌دهی با رعایت اصول اخلاقی و توجه به عدالت آموزشی انجام شد. همچنین، نظرات و بازخوردهای داوران و دانشجویان به‌عنوان منبعی برای بهبود مستمر نظام نمره‌دهی در نظر گرفته شدند. جهت تدوین و خلاصه‌سازی این مقاله از هوش مصنوعی ۳،۱ Llama کمک گرفته شده است.

حمایت مالی

این پژوهش هیچگونه کمک مالی از سازمان‌های دولتی، خصوصی یا غیرانتفاعی دریافت نکرده است.

مشارکت نویسندگان

مفهوم‌سازی و طراحی مطالعه؛ جمع‌آوری، تحلیل و تفسیر

داده‌ها و تهیه پیش‌نویس دست‌نوشته؛ بازبینی نقادانه دست‌نوشته برای محتوای فکری مهم؛ آبتین حیدرزاده ایده‌پردازان

تعارض منافع

نویسندگان اظهار می‌دارند که هیچ‌گونه تعارض منافع در این مقاله وجود ندارد.

استفاده از هوش مصنوعی در فرآیند نگارش

نویسندگان اعلام می‌دارند که در فرآیند نگارش این مقاله از هیچ‌گونه فناوری‌های هوش مصنوعی برای تولید محتوا، ایده، تجزیه و تحلیل و سایر موارد استفاده نشده است.

قدردانی

نویسندگان بر خود لازم می‌دانند از زحمات مسئولین و برگزارکنندگان آزمون المپیاد علمی دانشجویی و دانشجویان شرکت‌کننده در این فرایند تشکر و قدردانی می‌گردد.

References

1. Education Development Center. Objective of the Student Scientific Olympiad; 2025.[Persian] [Link]
2. Ministry of Health and Medical Education. Regulations of the medical sciences students' scientific olympiad; 2017. [Link]
3. Sadr H, Nazari Soleimandarabi M, Khodaverdian Z. Automatic assessment of short answers based on computational and data mining approaches. J Decisions and Operations Res 2021;6(2):242-255. [DOI:10.22105/dmor.2021.251713.1230]
4. Magirans. Essay questions 2025. [Link]
5. Barkaoui K. Variability in ESL essay rating processes: the role of the rating scale and rater experience. Lang Assess Quarter 2010;7(1):54-74. [DOI:10.1080/15434300903464418]
6. Moradi E, Didehban H. Scoring in the essay tests questions: methods, challenges and strategies. Nurs Midwifery J 2015;13(8):692-698. [Link]
7. Alimohammadi m, Farakhi N. Effective factors on composition tests' correction by teachers. Quart Educa Measure 2015;6(20):211-225. [DOI:10.22054/jem.2015.1678]
8. Kolen M, Brennan R. Test equating, scaling, and linking: methods and practices. Springer; 2004. [DOI:10.1007/978-1-4757-4310-4]
9. Asadi V, Moghadamzadeh A, Salehi K. The Effect of the anchor to total test correlation on equating results: a systematic review. Educa Measure Evalua Studies 2023;13(43):7-38. [DOI:10.22034/emes.2023.1971260.2430]
10. Huggins-Manley AC, Qiu Y, Penfield RD. Exploring a source of uneven score equity across the test score range. Int J Test 2018;18(1):50-70. [DOI:10.1080/15305058.2017.1396463]
11. Nabbout M. A study of discrepancies in the assessment of probabilistic tasks: why might teachers grade and evaluate inconsistently a given answer?. Assessing Student Learning In Statistics IASE Satellite Conference: 2007. [Link]
12. Nikmard F, Tavassoli K. The Impact of Test Length on Raters' Mental Processes During Scoring Test-Takers' Writing Performance. J Lang Horizons 2023;7(1):159-182. [Link]
13. Effatpanah F, Baghaei P. Exploring rater quality in rater-mediated assessment using the non-parametric item characteristic curve estimation. Psychol Test Asses Model 2022;64(3):216-252. [Link]
14. Zolfagharnasab S, Moghadamzadeh A, Sharifi Yeganeh N. Equating and linking scores in national exams. Educa Measure Evalua Studies 2023;13(44):111-132. [DOI: 10.22034/emes.2024.561619.2412]