

Use of artificial intelligence algorithms and mixed effects models to analyze educators' and students' perceptions and attitudes towards objective structured clinical examinations

Asma Ben Amor¹, Hassan Farhat², Guillaume Alinier³, Amina Aounallah⁴, and Olfa Bouallegue⁵

¹MRes, Lecturer of Allied Health Sciences, School of Health Sciences and Techniques; Faculty of Medicine, Ibn El Jazzar, University of Sousse, Sousse, Tunisia

²PhD, Adjunct Faculty, Epidemiology of Mental Illnesses, Screening and Early Management, Faculty of Medicine, Ibn El Jazzar, University of Sousse, Sousse, Tunisia; Major Incident Preparedness and Resilience & Ambulance Service Group, Hamad Medical Corporation, Doha, Qatar

³PhD, Visiting Professor of Simulation in Healthcare Education, University of Hertfordshire & Northumbria University, Newcastle upon Tyne, United Kingdom; Adjunct Professor of Education in Medicine, Weill Cornell Medicine-Qatar, Doha, Qatar; Major Incident Preparedness and Resilience & Ambulance Service Group, Hamad Medical Corporation, Doha, Qatar

⁴MD, Professor, Department of Dermatology, Academic Hospital, Farhat Hached; Faculty of Medicine, Ibn El Jazzar, University of Sousse, Sousse, Tunisia

⁵MD, Professor, Microbiology Laboratory, Hygiene and Critical Care Departments, Faculty of Medicine, Ibn El Jazzar; Academic Hospital of Sahloul, University of Sousse, Sousse, Tunisia

Abstract

Background: During the COVID-19 pandemic, Tunisia implemented Objective Structured Clinical Examinations (OSCEs) for healthcare students at the School of Health Sciences and Technologies in Sousse. This study analyzed satisfaction levels using advanced analytical techniques. **Methods:** A cross-sectional study was conducted involving 128 final-year students across all healthcare programs and 31 educators. Two anonymous satisfaction surveys using five-point Likert scales were administered. The analysis employed supervised machine learning (SML), unsupervised machine learning (UML), and linear mixed-effects models (LMEM). The SML utilized Random Forest classification, whilst UML implemented K-means clustering and principal component analysis. **Results:** The SML demonstrated robust predictive accuracy for satisfaction categories (students: 78%; educators: 100%, though the latter may reflect overfitting due to the small sample size) with strong reliability (kappa values 0.65 and one). UML identified three distinct satisfaction clusters among

students and seven among educators, with the highest satisfaction cluster showing a mean score of 3.82 ± 0.34 . LMEM revealed that student satisfaction increased with age ($\beta=0.041$), whilst educator satisfaction correlated negatively with age ($\beta=-0.017$) but positively with teaching experience ($\beta=0.016$). Those from the Emergency Medical Care program consistently demonstrated higher student satisfaction levels. **Conclusions:** Integrating advanced analytical methods provided a more profound understanding of OSCE satisfaction patterns than traditional statistical approaches. The findings suggest multiple demographic factors influence OSCE satisfaction, necessitating context-specific approaches to the educational specialty rather than universal solutions. These results have important implications for healthcare education practice, suggesting the need for differentiated approaches to OSCE implementation based on student specialty and educators' experience levels.

Key Words: Machine learning, Mixed Effect Models, OSCE, Medical Education, Simulation

Date submitted: 18-April-2025

Email: Hassan Farhat (hassen.farhat@gmail.com)

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

Citation: Ben Amor A, Farhat H, Alinier G, Aounallah A, and Bouallegue O. Use of artificial intelligence algorithms and mixed effects models to analyze educators' and students' perceptions and attitudes towards objective structured clinical examinations. *Educ Health* 2025;38:244-257

Online access: www.educationforhealthjournal.org

DOI: 10.62694/efh.2025.330

Published by The Network: Towards Unity for Health

1. Introduction

Objective Structured Clinical Examinations (OSCEs) have been used in medical education since 1975 and have since been introduced in many countries in most healthcare disciplines.^{1,2} Implementing the OSCE assessment method for healthcare students during the COVID-19 pandemic in Tunisia represented a significant advancement over the previously used skills and knowledge examination approaches. Our initial evaluations have demonstrated high satisfaction levels among students and educators, prompting deeper analytical exploration to improve and sustain them effectively in our institution.^{3,4}

Further, OSCEs have emerged as a fundamental global aspect of healthcare education, providing a standardized assessment approach of clinical competencies. International consensus supports their effectiveness in enhancing students' clinical capabilities, communication skills, and professional development.^{5,6} Bill Gates said, "We all need people who will give us feedback. That's how we improve," emphasizing the importance of feedback in improving processes, including in the medical field.⁷

Undeniably, artificial intelligence (AI) analysis methods have shown remarkable potential in medical education and simulation, particularly in identifying complex patterns and predicting outcomes. The artificial intelligence-based method has demonstrated its effectiveness in medical education. The AI transforms medical education by personalizing learning through adaptive content and targeted feedback while enhancing teaching methods via improved assessment accuracy and realistic clinical simulations.⁸⁻¹⁰ It strengthens skills development in diagnosis, problem-solving, and clinical decision-making through AI-assisted analysis.⁸ Through virtual reality training and real-time progress tracking, AI is revolutionizing traditional teaching approaches and modernizing healthcare delivery.

However, no studies have employed AI-based analytical methods to explore students' and educators' satisfaction with healthcare education, specifically OSCEs.

While numerous studies have investigated OSCE satisfaction worldwide, the analytical approaches have predominantly relied on traditional statistical methods.^{3,6,11-13} No related research has used advanced analytical techniques, such as artificial

intelligence and mixed effects modelling methods, to examine satisfaction patterns with educational methods in healthcare education, considering the particularities of low-resource, low-income countries of the Middle East and North Africa.

Building on AI's transformative potential in medical education, we hypothesized that supervised machine learning models could accurately classify satisfaction categories. We also hypothesized that distinct satisfaction profiles exist among students and educators in medical education regarding OSCEs, which could be identified through unsupervised machine learning techniques. Furthermore, we posited that demographic and educational variables would significantly predict satisfaction levels, quantifiable through mixed-effects modelling.

This study aims to investigate the satisfaction levels of students and educators with OSCEs at the School of Health Sciences and Technologies in Sousse (SHSTS), Tunisia, and identify predictors and clusters within satisfaction data, including supervised and unsupervised machine learning and mixed-effects models.

2. Methods

2.1. Study design and setting

This cross-sectional satisfaction survey study was conducted in June 2022. It included final-year students from four specialties of health sciences education at the University of Sousse in Tunisia: Emergency Medical Care (EMC), Surgical Technology (ST), Pediatric Care (PC), and Podology. The students of these healthcare programs completed both OSCE and traditional practical examinations in May–June 2022, noting that they participated in at least three other OSCE sessions in the SHSTS.

Two distinct anonymous satisfaction surveys were developed in French for students and educators, employing a five-point Likert scale (from "Strongly Disagree" to "Strongly Agree"). Both surveys were included in the appendices in a previous publication.³ The survey structure and validation process have been previously described in detail.³ Both instruments comprised 41 items across four domains: OSCE characteristics (17 items), structural aspects (10 items), organizational elements (9 items), and assessment efficiency (5 items), along with six demographic questions. Participation was voluntary, with written informed consent obtained from all participants. Anonymity

was maintained throughout the data collection process to ensure unbiased responses.

2.2. Ethics and Reporting

The article's structure adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. The research protocol received ethical approval from the Doctoral School Review Board of the Faculty of Medicine "Ibn Eljazzar", University of Sousse (Approval date: 27/03/2021)

2.3. Population and sampling

The target population consisted of all final-year students (N=133) in their three-year program and their educators (N=33) across all specialties at the SHSTS. Using Slovin's formula with a 5% margin of error, the calculated minimum sample sizes for students and educators were 98 and 31, respectively.¹⁴

2.4. Data Management and Cleaning

Two datasets containing OSCE satisfaction survey responses from students and educators were imported. First, to ensure data quality and consistency, column names were standardized. Second, data transformation was performed, and Likert-scaled responses from a categorical format to a numeric scale ("Strongly Disagree", "Disagree", "Undecided", "Agree", and "Strongly Agree" to respectively, 1, 2, 3, 4, and 5). Third, individual satisfaction scores were calculated as the mean of all responses and subsequently categorized into five levels: Very Low (0–1.5), Low (1.5–2.5), Moderate (2.5–3.5), High (3.5–4.5), and Very High (4.5–5). The initial proportion of missing data was 2.1% for students and 1.6% for educators. Missing values were addressed using Multiple Imputation by Chained Equations.¹⁵ Classification and Regression Trees were employed for categorical variables to predict missing values based on existing data patterns.¹⁵ Predictive Mean Matching was utilized for numerical variables, maintaining the natural distribution of the data by imputing observed values from similar cases.¹⁵ Fifteen imputations were performed to ensure stable estimates. Continuous variables used predictive mean matching. Categorical variables employed classification and regression trees. Convergence was verified through trace plots of imputed variable means and variances across 15 iterations, which showed stable chains.

2.5. Statistical Analysis

In this study, we applied Supervised Machine Learning (SML), Unsupervised Machine Learning

(UML), and Linear Mixed-Effects Models (LMEM). For the SML analysis, students' satisfaction scores were categorized into five levels: Very Low (0–1.5), Low (1.5–2.5), Moderate (2.5–3.5), High (3.5–4.5), and Very High (4.5–5). The intervals are based on equal-width splits of the scale, a common practice in survey research, to balance the distribution of responses across categories, aligning with established conventions in educational and psychometric literature, and were further adjusted, allowing for more precise comparison and interpretation of satisfaction patterns.¹⁶

The dataset was partitioned into training (70%) and testing (30%) sets using stratified sampling to maintain class distributions. The model incorporated students' specialties and OSCE-specific questions as predictive variables, employing 10-fold cross-validation, within the training set, to minimize overfitting.¹⁵ The Random Forest (RF) algorithm was utilized for the SML. The RF was selected as the primary classification algorithm due to its versatility, strong performance with categorical predictors, and reduced risk of overfitting compared to single decision trees. The RF was implemented using 500 decision trees, with the number of variables considered at each split set to the default value (the square root of the number of predictors). No explicit maximum tree depth was imposed, so each tree grew until nodes were pure.

Given the limited number of predictors (specialty for students and educators' teaching classes), grid search for hyperparameter optimization was not performed. Instead, the model stability was verified through repeated random splits, and consistent accuracy was observed across runs. Model performance was evaluated using three metrics: classification accuracy, kappa statistics for inter-rater reliability, and confusion matrices visualized through heatmaps. Distribution plots were generated to illustrate satisfaction patterns across student specialties and educators teaching classes.

The UML utilized K-means clustering to identify natural response patterns, complemented by Principal Component Analysis (PCA) for dimensionality reduction while preserving essential data structure and visualizing the identified clusters.

LMEM were employed to examine OSCE satisfaction determinants, simultaneously analyzing fixed effects (showing consistent influence across participants) and random effects (group-specific

variations). LMEM are statistical tools that analyze data with consistent effects across all groups (fixed effects, like age affecting everyone similarly) and group-specific variations (random effects, like different educators having varying impacts).¹⁷ Age was considered a fixed effect for students, reflecting its anticipated uniform implications for satisfaction, while specialty was a random effect to account for unmeasured differences between training programs. For educators, age and years of teaching experience were analyzed as fixed effects, as these are likely to influence satisfaction consistently across individuals, with background and teaching classes as random effects, to accommodate variability attributable to institutional or teaching context. Individual survey questions were incorporated as random effects to account for participant interpretation variations.

3. Results

128 students and 31 educators participated in the OSCE satisfaction surveys, resulting in a 96.2% and 93.4% participation rate. The students' mean age was 21.8 ± 0.9 years, ranging from 21 to 28 years. The educators had a mean age of 41.0 ± 9.9 years, ranging from 25 to 62 years, and reported an average number of years of teaching experience of 13.0 ± 10.9 years, ranging from 4 to 39 years.

Figures 1, Figure 2 and Table 2 display the performance metrics for the RF in students and educators and their confusion matrix. For students, the model was accurate at 0.78, a Kappa score of 0.65, Precision of 0.54, Recall of 0.61, and F1 of 0.52, which suggests good agreement between predicted and actual satisfaction levels. For educators, the model achieved perfect test-set metrics, reflecting the dataset's extreme class imbalance (94% "High" satisfaction) and small sample size ($n = 31$), limiting the predictive capability. A larger and more balanced sample of educators could improve the performance of the predictive modelling. The confusion matrix heat maps showed how well the RF predicted satisfaction levels. For students, the model performed well in predicting "High" satisfaction (13 correct predictions) and "Moderate" satisfaction (12 correct predictions). The model correctly identified 3 "High" and 4 "Very High" satisfaction cases for educators. The darker blue colors in the confusion matrices in Figures 1 and 2 show where the model made more correct predictions.

Figure 3, which displays the satisfaction patterns across different specialties, shows that students from the Emergency Medical Care program showed the highest "High" satisfaction ratings, while Pediatric Care had more balanced ratings across categories. The Emergency Medical Care program had the most responses from educators with "High" and "Very High" satisfaction levels. All specialties showed generally positive satisfaction levels. The RF effectively predicted satisfaction levels, especially for educators, showing that most participants were generally satisfied with the OSCE. UML was applied to the OSCE satisfaction survey data from 128 students and 31 educators. The analysis used PCA and cluster analysis to explore patterns in satisfaction responses.

The optimal number of clusters (k) in the UML was determined using the silhouette method. Clustering was performed with a range of k values, and the average silhouette width was calculated for each solution. The value of k that maximized the average silhouette score was selected, as this metric reflects both the cohesion within and the separation between clusters. The best clustering value k was determined to be $k=3$ for students and $k=7$ for educators. Figure 4 presents the principal component plots for both groups. For students, clustering with $k=3$ revealed three distinct response groups with clear separation along the first two principal components. For educators, clustering with $k=7$ generated more dispersed and less clearly separated clusters, with some overlap evident in the principal component space.

The silhouette plot in Figure 5 provides a quantitative assessment of clustering quality. The average silhouette score for students' clusters was 0.23, indicating weak cluster structure. Cluster 3 (blue) showed the highest within-cluster cohesion, while Cluster 1 (red) included several points with low or negative silhouette widths, suggesting possible misclassification or overlap with other clusters. For educators ($k=7$), the average silhouette score was even lower at 0.18, reflecting poor separation between clusters and substantial overlap. These low silhouette scores suggest that, although clusters can be mathematically defined, the natural grouping within the data is limited for educators.

Figure 6 summarizes the distribution of student satisfaction scores by cluster. Cluster 3 ($n=60$) exhibited the highest mean satisfaction (3.82 ± 0.34), representing students with

predominantly positive OSCE experiences. Cluster 1 (n=54) had moderate satisfaction (mean 2.83 ± 0.31), while Cluster 2 (n=14) reported the lowest satisfaction (mean 1.59 ± 0.35). The density and box plots show minimal overlap between clusters, each displaying a relatively compact distribution of satisfaction scores.

Figure 7 shows the satisfaction score distributions for educators across seven clusters. Despite the mathematical partitioning, the clusters are small and unevenly distributed (cluster sizes ranging from 2 to 10), and the satisfaction scores are generally high and tightly clustered (means ranging from 3.67 to 4.35). The density and box plots confirm the lack of clear separation, consistent with the low silhouette score.

The LMEM in Figure 8 and Table 1 revealed variant patterns in OSCE satisfaction between students and educators. In the student model, age was demonstrated to have a significant positive effect on satisfaction ($\beta=0.041$, standard error (SE)=0.012, $t=3.358$), with a model intercept of 1.907 (SE=0.519, $t=3.671$). The random effects analysis showed that educators' academic qualification contributed the highest variance ($\sigma^2=0.293$, standard deviation (SD)=0.541), followed by students' specialty ($\sigma^2=0.193$, SD=0.439), while question-level variance was negligible. For educators, the model revealed contrasting age-related effects, with age slightly negatively associated with satisfaction ($\beta=-0.017$, SE=0.009, $t=-1.91$). However, educators' teaching experience demonstrated a positive effect ($\beta=0.016$, SE=0.008, $t=2.01$), with a model intercept of 4.656 (SE=0.297, $t=15.66$). The random effects structure for educators showed that question-level variations contributed the highest variance ($\sigma^2=0.035$, SD=0.186), followed by teaching classes ($\sigma^2=0.027$, SD=0.164), while academic background showed minimal variance ($\sigma^2=0.0003$, SD=0.017). These findings suggest that older students generally reported higher satisfaction levels. The opposite was true for educators, though positive associations with years of teaching experience offset this negative age effect. Furthermore, the variance components indicate that educational specialty was the primary grouping factor affecting student satisfaction, whereas question-specific variations played a more crucial role in educators' satisfaction.

4. Discussion

OSCEs are generally positively perceived by learners and educators as they can be used to practice or assess a broad range of skills and knowledge through multiple stations, and help identify weaknesses in the learners' abilities or the curriculum.¹³ They, however, rely on careful planning and preparation on the part of the educators with learning outcomes well mapped against the program's curriculum and adequate orientation of the learners.¹⁸ Learners' early exposure and experience of OSCEs will influence their level of anxiety and acceptability of this educational or assessment approach.¹⁹ This study's analytical approach demonstrated that integrating supervised and unsupervised ML revealed specific relationships and patterns that traditional statistical methods overlook. This integration enables an in-depth understanding of how students and educators perceive educational methods, facilitating robust planning for improving or sustaining these approaches. Previous studies have relied on traditional statistical methods to assess satisfaction levels among students and educators and reported generally high satisfaction with OSCEs.³

Furthermore, through RF classification, the SML demonstrated robust predictive accuracy (students: 78%, educators: 100%) with strong reliability (kappa values 0.65 for students and 0.1 for educators) for satisfaction categories across both groups (students and educators). The analysis assessed variables, including students' specialty, educators' educational level, year of teaching experience, instructional classes, and both groups' ages, thereby identifying key satisfaction indicators within these populations. Further, students from the Emergency Medical Care program consistently demonstrated higher satisfaction levels than their peers in other specialties, whilst age demonstrated divergent effects, a positive correlation for students but a negative one for educators. This finding may reflect the alignment between OSCE scenarios and high-acuity care environments, such as prehospital and emergency care settings.

The structured nature of OSCEs, with clear time limits, standardized guidelines, and simulated urgency, likely resonates more strongly with emergency care pedagogy, prioritizing rapid decision-making under pressure. Conversely, specialties such as pediatric care and longitudinal patient relationships may perceive OSCEs as less authentic, contributing to lower satisfaction. These

findings align with our previous traditional statistical analyses of the same dataset, providing additional validation through advanced analytical methods.³

The principal advantage of ML in this context lies in its capacity to process multiple complex variables simultaneously, offering more robust predictions by considering the effect of all potential factors influencing satisfaction levels. The positive correlation between age and students' satisfaction was likely due to the students' clearer understanding of the educational learning objectives of the designed OSCE programs and enhanced adaptability to educational environments. Older students typically demonstrate stronger academic engagement and performance due to their broader developed study strategies.²¹ Conversely, the negative correlation of satisfaction with age for educators is counterbalanced by the positive effect of teaching experience, suggesting that pedagogical competency development, rather than age alone, is a crucial determinant of satisfaction levels.

This pattern aligns with research showing that educators' satisfaction generally increases with age until their late career before declining and that experienced educators value different aspects of their professional roles compared to their younger colleagues, as explained by similar studies.^{22,23} These findings highlight the importance of considering age and years of teaching experience in understanding satisfaction with educational assessment methods. In the same context of using ML, a recent study using ML methods in medical education also demonstrated its advantage in predicting clinical performance, where key performance indicators were identified in clinical rotations that remained undetected through conventional statistical approaches.²⁴

Additionally, the UML identified three main satisfaction profiles among students: highly satisfied (Cluster 3), moderately satisfied (Cluster 1), and least satisfied (Cluster 2). The silhouette analysis indicates these clusters are statistically distinct regarding satisfaction scores, though the overall cluster structure is weak. For educators, the attempt to partition responses into seven clusters did not identify important separation, with most clusters being small and overlapping, and the silhouette score confirming poor cluster quality. This suggests that, while students' satisfaction responses show some natural grouping, educators' responses are

more homogeneous and do not cluster well. Overall, three clusters with distinct student satisfaction profiles, but a weak overall cluster structure (Average silhouette=0.23). For educators, seven clusters with poor separation and low interpretability (Average silhouette=0.18).

USML clustering analyses reveal significant heterogeneity among students' satisfaction but slightly among educators. Understanding the characteristics of each cluster facilitates the development of specific interventions for groups showing lower satisfaction levels. This understanding can help to ease the path of educational, strategic review decisions about OSCE resource allocation and training design, enabling evidence-based improvements in OSCE delivery within the SHSTS. Besides that, the data-driven clustering approaches allow healthcare educators to move beyond "one-size-fits-all" solutions and instead develop cluster-specific strategies to enhance the OSCE across different student groups. A recent study found comparable clustering patterns in its analysis of clinical skills assessments and identified specific skill combinations that consistently predicted high satisfaction levels.²⁵

The LMEM revealed, in addition, age-related patterns as students' satisfaction increased with age ($\beta=0.041$), reflecting their enhanced metacognitive abilities, demonstrating better self-regulation and academic engagement. Educators' satisfaction with OSCEs correlated negatively with age ($\beta=-0.017$), which might reflect the increasing awareness of assessment limitations with accumulated experience or a reluctance to change in their ingrained teaching and assessment practices. Nevertheless, the positive correlation of educators' satisfaction level with their teaching experience ($\beta=0.016$) suggests that pedagogical expertise is an important driver of satisfaction, not age alone. This controversial correlation of educators' satisfaction indicates that while older educators might be more critical of assessment methods, their teaching experience provides them with better coping strategies and a deeper understanding of the value of educational assessment. Research reported similar findings in a multicenter study, where it was identified that experienced educators achieved higher student satisfaction rates, despite age-related variations, leading to the conclusion of the importance of experience over age in teaching effectiveness.²⁶

Overall, advanced analytical methods enable specific, targeted educational interventions and allow for proactively predicting satisfaction levels through demographic information using predictive modelling capabilities, which allows for proactive improvement interventions.¹⁹ They also help identify natural student and educator groupings, enabling the stratification of proactive improvement interventions. Another study implemented similar complex analytical frameworks in various medical schools, reporting a 25% improvement in OSCE delivery efficiency and reduced assessment bias by 30% through automated pattern detection.^{27,28} These advanced analytical methods contribute to the evolution of healthcare education assessment, enabling evidence-based improvements in OSCE design and implementation. The ability to predict and understand satisfaction patterns leads to more proactive, effective assessment strategies and improved educational outcomes.

5. Limitation

The study has several methodological limitations. First, educators' relatively small sample size ($n=31$) might contribute to decreased accuracy and computational complexity when employing machine learning methods and raising the risk of misclassification, although patterns can still be identified when adequately validated. To mitigate these issues in future research, larger and more balanced samples should be collected, mainly for the educator group, and alternative clustering methods less sensitive to group size and distribution, such as hierarchical clustering or model-based approaches, should be considered.

Second, the satisfaction surveys simultaneously captured students' and educators' perspectives, missing temporal variations in satisfaction patterns and the influence of external factors. Retrospective data collection may also be subject to recall bias regarding specific aspects of the OSCE experience. Another limitation is that it is a single-site study, and there was no guarantee that the students' OSCE experience was comparable across the various healthcare programs. Future research would benefit from prospective, longitudinal designs to better understand how satisfaction patterns evolve over time and validate these analytical methods' predictive capabilities.

Additionally, RF was utilized in this study due to its robustness in handling categorical predictors, resilience to overfitting, and ability to quantify

feature importance—critical attributes for interpreting satisfaction patterns in educational datasets with heterogeneity in previous research. This is because the primary aim was to contrast ML methods with conventional statistical approaches, rather than optimize algorithmic performance. Future studies should consider other algorithms such as support vector machine and neural network.

Lastly, our analysis focused on demographic and educational variables as predictors of OSCE satisfaction and did not consider the potential influence of unmeasured confounding factors not captured in our dataset, such as environmental variables (timing of surveys relative to exams), prior clinical experience, or institutional policies affecting OSCE implementation. Future studies should validate these findings through multicenter collaborations with larger, balanced cohorts.

6. Conclusion

This study demonstrates the effectiveness of advanced analytical methods in understanding OSCE satisfaction patterns among health sciences students and educators. The combination of supervised machine learning, unsupervised machine learning, and linear mixed-effects modelling provided an in-depth exploration of the satisfaction determinants and patterns.

These analytical approaches were advantageous over traditional statistical methods. They enabled the simultaneous analysis of multiple complex variables, provided predictive capabilities, and identified groupings within the studied populations, facilitating targeted interventions and evidence-based improvements.

Furthermore, the findings suggest that multiple factors, demographics specifically, can influence OSCE satisfaction, requiring context-specific approaches rather than universal solutions to improve OSCE delivery. These results have important implications for medical education practice. They suggest the need for differentiated approaches to OSCE implementation based on students' healthcare specialty and educators' experience levels. They also emphasize the importance of considering age and years of teaching experience in faculty development programs.

Future research should focus on validating these findings with larger samples and exploring the implementation of targeted interventions based on

Table 1: Linear Mixed Regression Model summary of students' and educators' satisfaction categories

	Fixed Effects						Random Effects		
	Term	Estimate	SE	p-value	CI		Group	Variance	Std. Dev.
					Lower	Upper			
Students	(Intercept)	1.91	0.52	0.03	0.29	3.52	Question	0.00	0.00
	age	0.04	0.01	0.00	0.02	0.07	Specialty	0.19	0.44
							Educational level	0.29	0.54
							Residual	0.58	0.76
Educators	(Intercept)	4.66	0.30	0.00	4.06	5.25	Group	Variance	Std. Dev.
	age	-0.02	0.01	0.06	-0.04	0.00	question	0.04	0.19
	experience	0.02	0.01	0.05	0.00	0.03	Highest academic degree	0.01	0.11
							Classes teaching	0.03	0.16
							Background	0.00	0.02
							Residual	0.77	0.88

Table 2: Random Forest Model Performance Metric in Classifying the Satisfaction for Educators and Students

Group	Model	Accuracy	Kappa	Precision	Recall	F1
Students	Random Forest	0.78	0.65	0.54	0.61	0.52
Educators	Random Forest	01	01	01	01	01

Figure 1: Supervised Machine Learning Random Forest model performance metrics and confusion metrics in predicting the students' satisfaction categories

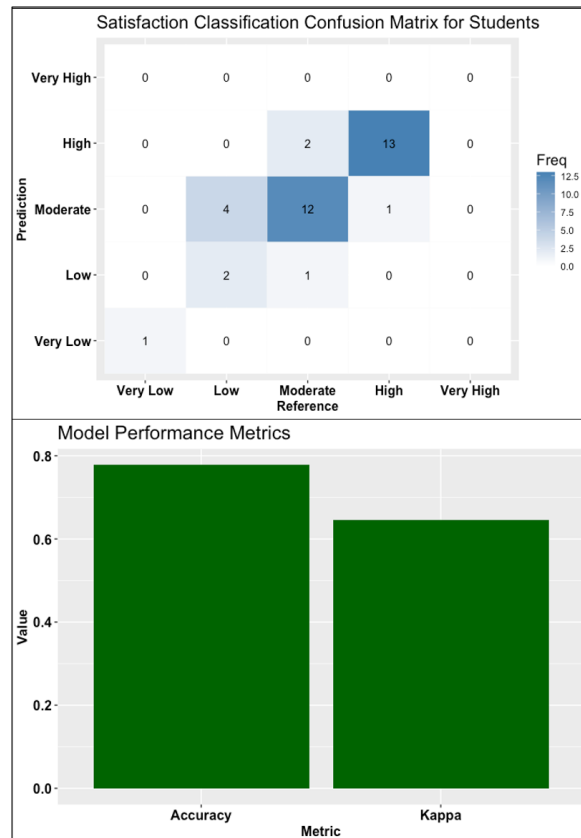


Figure 2: Supervised Machine Learning Random Forest model performance metrics and confusion metrics in predicting the educators' satisfaction categories

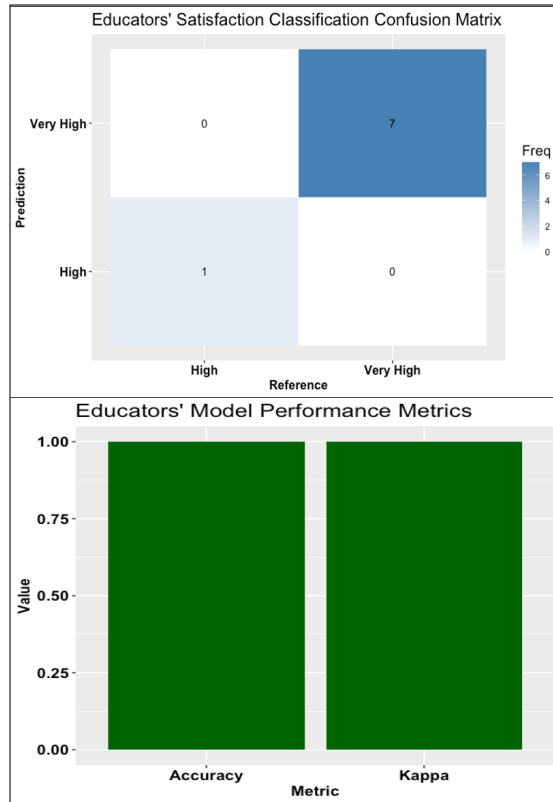


Figure 3: Supervised Machine Learning Random Forest Model enabling the prediction of the educators' and students' satisfaction categories across students' specialties and educators' teaching classes

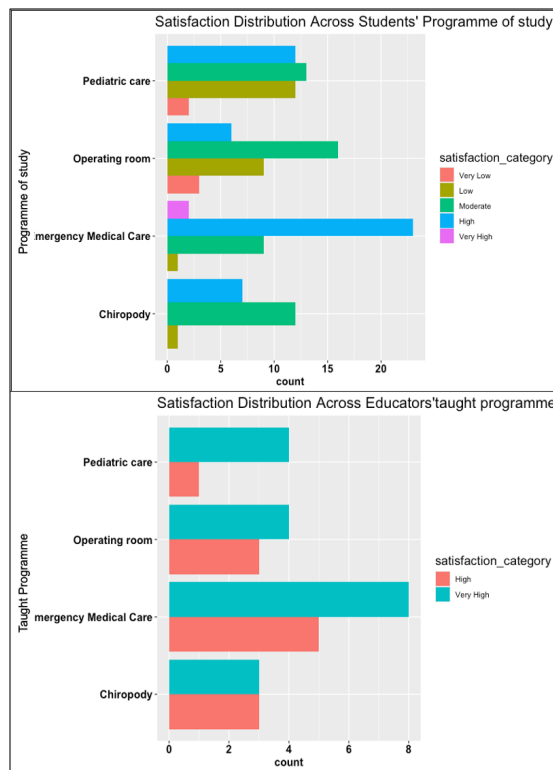


Figure 4: Unsupervised Machine Learning Clusters, principal components for students and educators

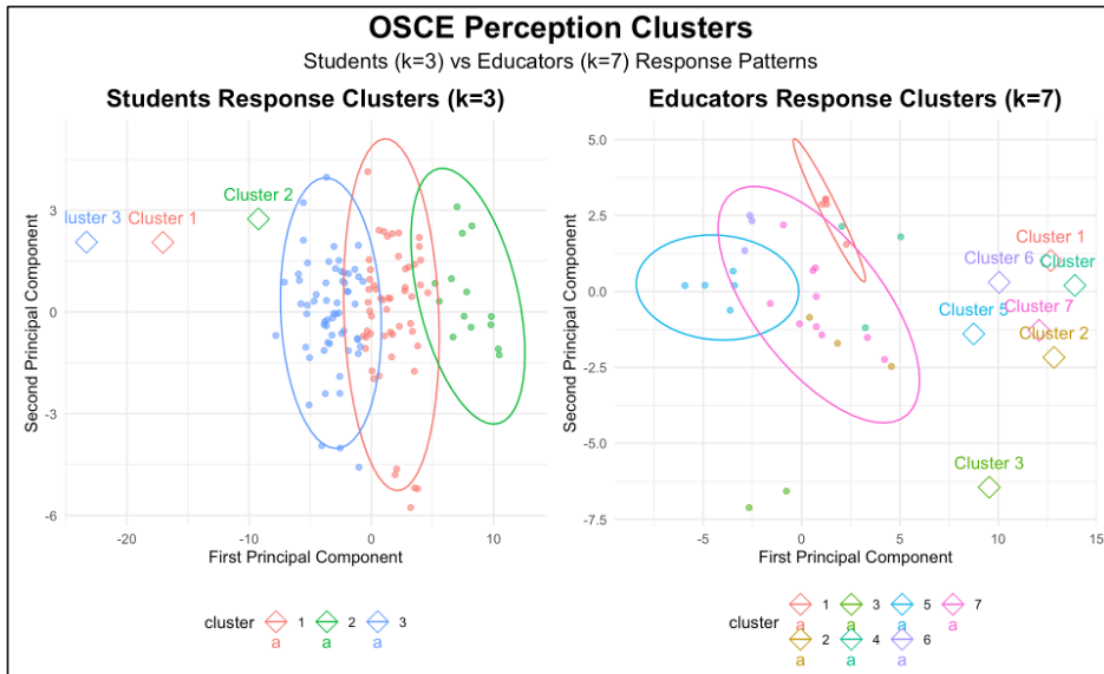


Figure 5: Silhouette analysis results for the clustering analysis quality

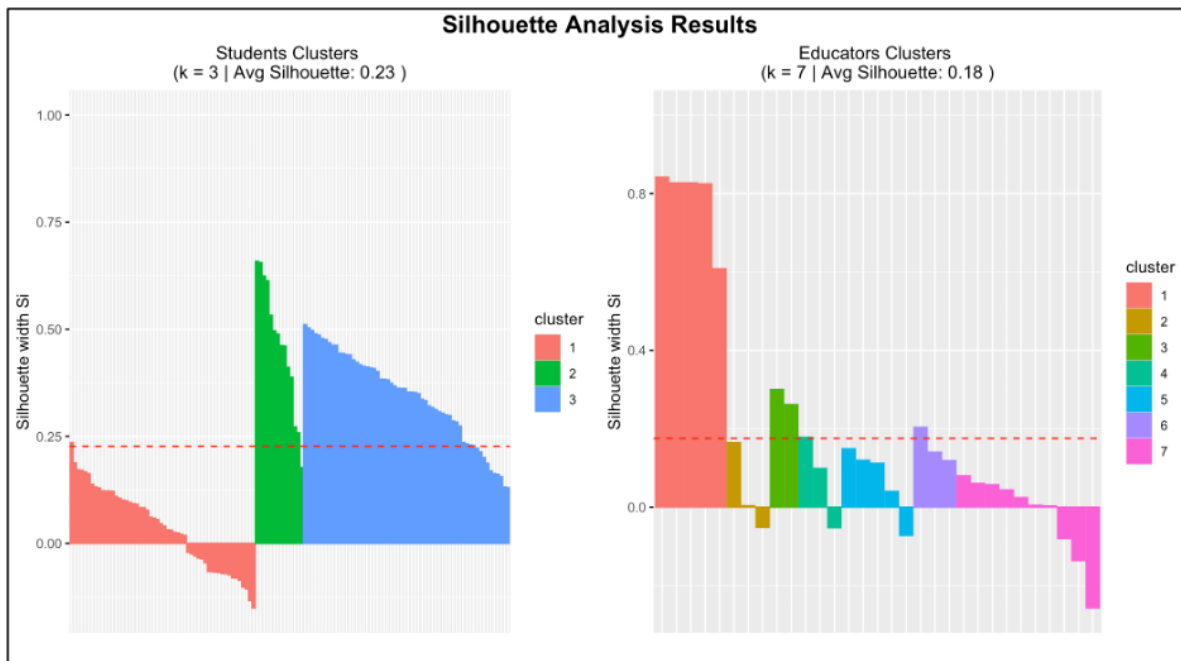


Figure 6. Students' satisfaction scores distribution, by cluster

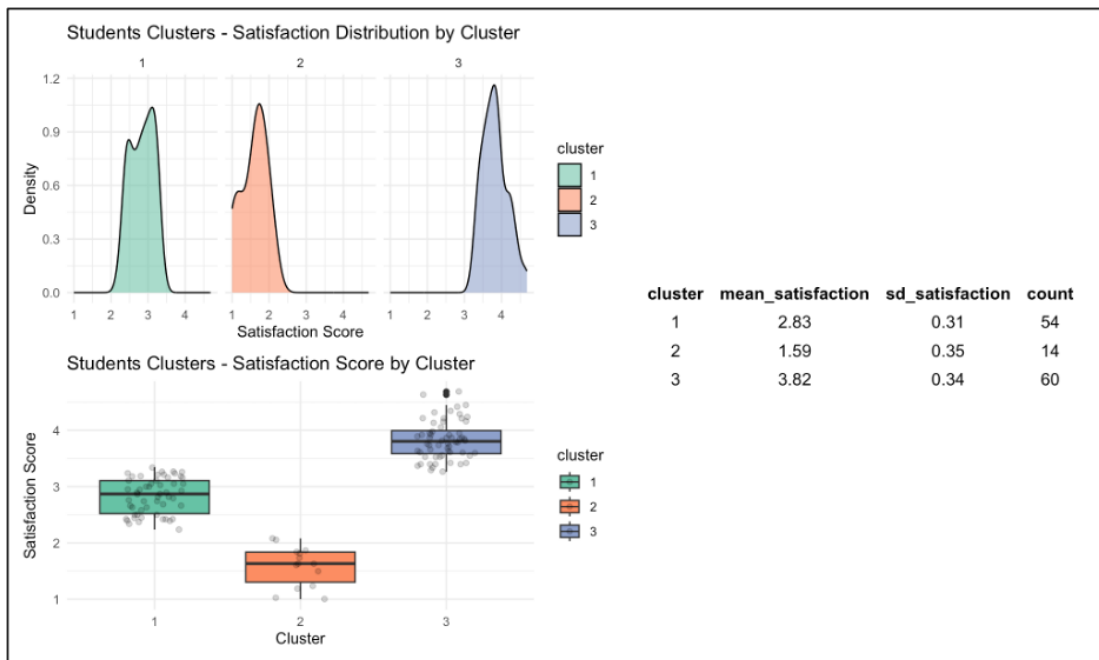


Figure 7: Educators' satisfaction scores distribution, by cluster

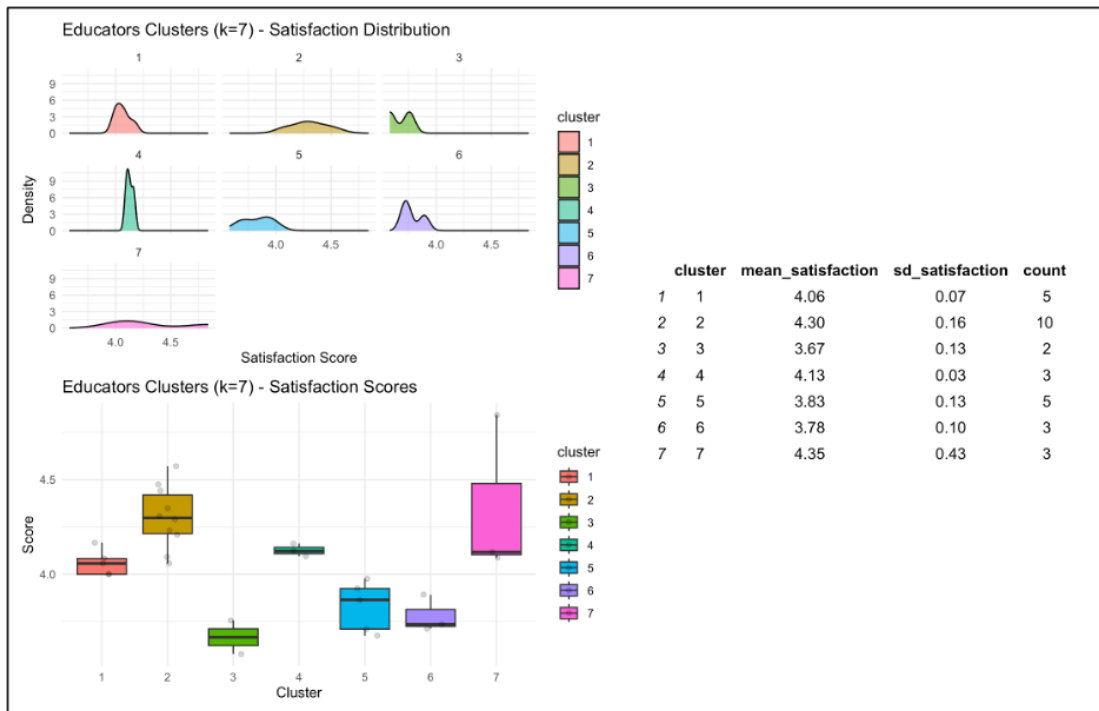
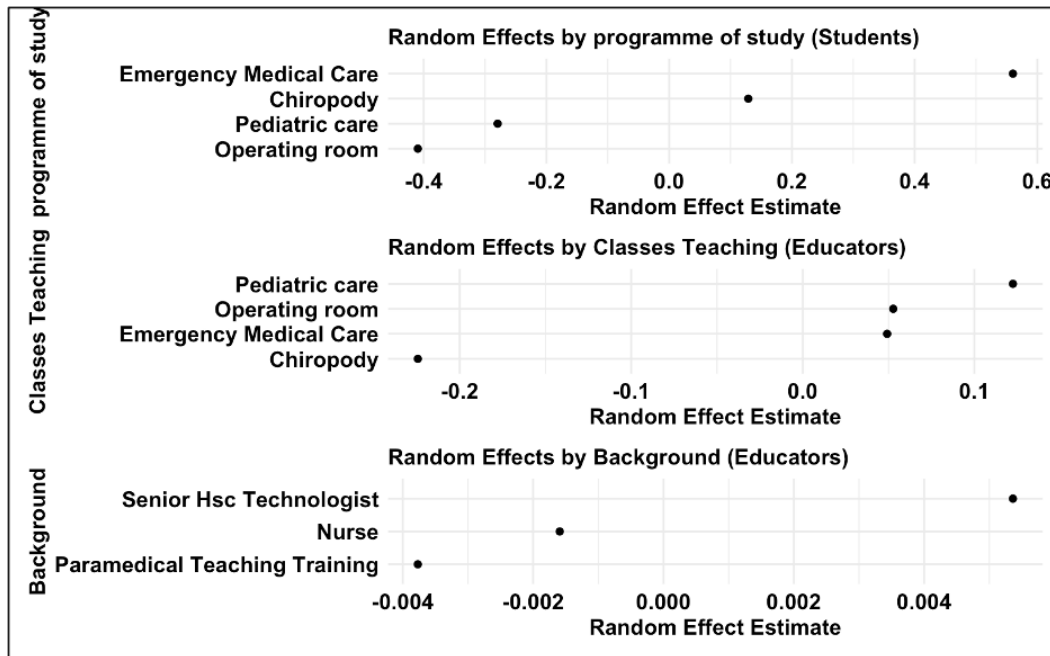


Figure 8: Linear Mixed-Effects Models Random Effects for students' and educators' satisfaction scores



the identified satisfaction patterns. The success of these advanced analytical methods in understanding OSCE satisfaction suggests their potential application in other areas of medical education assessment and healthcare quality improvement.

7. List of Tables

Table 1: Linear Mixed Regression Model summary of students' and educators' satisfaction categories
 Table 2: Random Forest Model Performance Metric in classifying the satisfaction for educators and students

8. List of Figures

Figure 1: Supervised Machine Learning Random Forest model performance metrics and confusion metrics in predicting the students' satisfaction categories
 Figure 2: Supervised Machine Learning Random Forest model performance metrics and confusion

metrics in predicting the educators' satisfaction categories

Figure 3: Supervised Machine Learning Random Forest model feature importance enabling in predicting the educators' and students' satisfaction categories across students' specialties and educators' teaching classes

Figure 4: Unsupervised Machine Learning Clustering and Principal Component Analysis of students' and educators' satisfaction scores

Figure 5: Linear Mixed-Effects Models Random Effects for students' and educators' satisfaction scores

9. Funding

No funds, grants, or other support were received.

10. Declaration

No funds, grants, or other support were received.

References

1. Evans BW, Alinier G, Kostrzewski A, Lefteri KA, Dhillon S. Development and design of objective structured clinical examinations (OSCE) in undergraduate pharmacy education in a new School of Pharmacy in England. *Currents in Pharmacy Teaching and Learning*. 2011 Jul;3(3):216–223. <https://doi.org/10.1016/j.cptl.2011.01.003>
2. Onwudiegwu U. OSCE: Design, development and deployment. *Journal of the West African College of Surgeons*. 2018;8(1):1–22. https://doi.org/10.4103/jwas.jwas_1_18

3. Ben Amor A, Farhat H, Alinier G, Ounallah A, Bouallegue O. Evaluation of the implementation of the objective structured clinical examination in health sciences education from a low-income context in Tunisia: A cross-sectional study. *Health Science Reports*. 2024 May;7(5):e2116. <https://doi.org/10.1002/hsr2.2116>
4. Boukhris H, Zidani H, Bouslema G, Hajjaji S, Mghirbi N, Hajjemi H, et al. Dental students' perception of the objective structured clinical examination (OSCE) in Tunisia: A cross-sectional study. *Clinical and Medical Engineering Live*. 2024 Aug 10;2(2):1–6. <https://doi.org/10.5281/zenodo.1234567>
5. Hsu YH, Wang YCL, Chou P-N, Wang CC, Liu SC, Chen TC, et al. Pharmacists' perceptions of objective structured clinical examination (OSCE) educational training: A survey research. *Journal of Medical Education*. 2023 Jun;27(2):97–104. <https://doi.org/10.6145/jme.2023.27.2.97>
6. Montgomery A, Chang HC (Rita), Ho MH, Smerdely P, Traynor V. The use and effect of OSCEs in post-registration nurses: An integrative review. *Nurse Education Today*. 2021 May;100:104845. <https://doi.org/10.1016/j.nedt.2021.104845>
7. Prayson RA, Rowe JJ. Effective feedback in the workplace. *Critical Values*. 2017 Jul;10(3):24–27. <https://doi.org/10.1093/crival/vax017>
8. Hussain S, Bhatti DE. Artificial intelligence and medical education. *Annals of King Edward Medical University*. 2022 Apr 30;28(1):3–6. <https://doi.org/10.21649/akemu.v28i1.4990>
9. Franco D'Souza R, Mathew M, Mishra V, Surapaneni KM. Twelve tips for addressing ethical concerns in the implementation of artificial intelligence in medical education. *Medical Education Online*. 2024 Dec 31;29(1):2330250. <https://doi.org/10.1080/10872981.2024.2330250>
10. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. *Medical Teacher*. 2023 Jun 3;45(6):574–584. <https://doi.org/10.1080/0142159X.2023.2186203>
11. Latjatih NHF, Roslan NS, Jahn Kassim PS, Adam SK. Medical students' perception and satisfaction on peer-assisted learning in formative OSCE and its effectiveness in improving clinical competencies. *Journal of Applied Research in Higher Education*. 2021 Jan;14(1):171–179. <https://doi.org/10.1108/JARHE-07-2020-0212>
12. Jenkins D, Nashed JY, Touma NJ. Virtual OSCE examinations during COVID-19. *Canadian Urological Association Journal*. 2023 Oct;17(10):E315–E318. <https://doi.org/10.5489/cuaj.8265>
13. Alinier G. Nursing students' and lecturers' perspectives of objective structured clinical examination incorporating simulation. *Nurse Education Today*. 2003 Aug;23(6):419–426. [https://doi.org/10.1016/S0260-6917\(03\)00044-3](https://doi.org/10.1016/S0260-6917(03)00044-3)
14. Farhat H, Alinier G, Gangaram P, El Aifa K, Khenissi MC, Bounouh S, et al. Exploring pre-hospital healthcare workers' readiness for chemical, biological, radiological, and nuclear threats in the State of Qatar: A cross-sectional study. *Health Science Reports*. 2022;5(5):e803. <https://doi.org/10.1002/hsr2.803>
15. Farhat H, Makhoulouf A, Gangaram P, El Aifa K, Howland I, Babay Ep Rekik F, et al. Predictive modelling of transport decisions and resources optimisation in pre-hospital setting using machine learning techniques. *PLOS ONE*. 2024 May 3;19(5):e0301472. <https://doi.org/10.1371/journal.pone.0301472>
16. López-Guerra VM, Pucha-Loarte TI, Angelucci LT, Torres-Carrión PV. Psychometric properties and factor structure of the satisfaction with life scale in Ecuadorian university students. *Frontiers in Psychology*. 2025 Mar 20;16:1536973. <https://doi.org/10.3389/fpsyg.2025.1536973>

17. Brown VA. An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*. 2021 Jan;4(1):2515245920960351. <https://doi.org/10.1177/2515245920960351>
18. Majumder MAA, Kumar A, Krishnamurthy K, Ojeh N, Adams OP, Sa B. An evaluative study of objective structured clinical examination (OSCE): students and examiners perspectives. *Advances in Medical Education and Practice*. 2019 Dec 31;10:387–397. <https://doi.org/10.2147/AMEP.S197275>
19. Al-Hashimi K, Said UN, Khan TN. Formative objective structured clinical examinations (OSCEs) as an assessment tool in UK undergraduate medical education: A review of its utility. *Cureus*. 2023 May;15(5):e38519. <https://doi.org/10.7759/cureus.38519>
20. Covas F, Veiga FH. Student engagement in higher education, age and parental education level. *Estudos de Psicologia (Campinas)*. 2021 May 19;38:e200020. <https://doi.org/10.1590/1982-0275202138e200020>
21. Bhatnagar K, Srivastava K, Singh A, Jadav SL. A preliminary study to measure and develop job satisfaction scale for medical teachers. *Industrial Psychiatry Journal*. 2011 Dec;20(2):91. <https://doi.org/10.4103/0972-6748.102480>
22. Narimani M, Zamani BE, Asemi A. Qualified instructors, students' satisfaction and electronic education. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*. 2015 Sep 19;6(3):31–39. <https://doi.org/10.30476/ijvlms.2015.46151>
23. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: Exploring key domains and essential functions. *Computer Methods and Programs in Biomedicine Update*. 2024 Jan 1;5:100148. <https://doi.org/10.1016/j.cmpbup.2024.100148>
24. Brentnall J, Thackray D, Judd B. Evaluating the clinical reasoning of student health professionals in placement and simulation settings: A systematic review. *International Journal of Environmental Research and Public Health*. 2022 Jan;19(2):936. <https://doi.org/10.3390/ijerph19020936>
25. Martínez GMÁ, Oscullo G, Gomez-Olivas JD, Gozal D. Measuring severity in OSA: The arguments for collaboratively developing a multidimensional score. *Journal of Clinical Sleep Medicine*. 2023 Oct;19(10):1705–1707. <https://doi.org/10.5664/jcsm.10722>
26. Johnson MW, Gheihman G, Thomas H, Schiff G, Olson APJ, Begin AS. The impact of clinical uncertainty in the graduate medical education (GME) learning environment: A mixed-methods study. *Medical Teacher*. 2022 Oct 3;44(10):1100–1108. <https://doi.org/10.1080/0142159X.2022.2058383>
27. Park J, Jang KJ, Alasaly B, Mopidevi S, Zolensky A, Eaton E, et al. Assessing modality bias in video question answering benchmarks with multimodal large language models. *arXiv*. 2024. <http://arxiv.org/abs/2408.12763>