Acta Pedagogia Asiana

Short Communication
Volume 5(SI), 2026, 64-73
https://doi.org/10.53623/apga.v5iSI.1069

Leveraging Smart Campus Data to Improve Teaching Quality: Insights on Teaching Evaluations

Ao Zhang[1]*, Zhizhen Chen[2], Ruizhi Liao[3]

[1]School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, China
[2]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China
[3]Guangdong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence, School of Humanities & Social Science, The Chinese University of Hong Kong, Shenzhen, China

*Correspondence: aozhang@link.cuhk.edu.cn

SUBMITTED: 24 February 2026; REVISED: 26 March 2026; ACCEPTED: 30 March 2026

ABSTRACT: In higher education, student evaluations play a crucial role in assessing teaching quality. However, these evaluations areofteninfluenced byextraneous factors, e.g., false high-grade expectations indicated by course instructors. While previous research has extensively examined the long-term implications of grade inflation, the immediate impact of students' expectations for higher grades on their teaching evaluations has been less explored. This paper leverages smart campus data from The Chinese University of Hong Kong, Shenzhen, coveringthe periodfrom 2018 to 2020, to addressthis gap. By selecting four representative indicators, we investigate their potential to enhance teaching quality through student evaluations. Our analysis reveals that integrating additional data on student life and academic performance from Smart Campus systems can help identify key factors influencing students' expected grades. This, in turn, allows for more precise adjustments to teaching evaluation results, pave the way to develop AI models aimed at enhancing the accuracy and reducing the incredibility of student evaluation of teaching.

KEYWORDS: Smart campus; big data analysis; student evaluations of teaching

1. Introduction

Student evaluations of teaching are a crucial metric for assessing teaching quality in higher education. However, their validity is often compromised by subjective elements, such as the influence of students' anticipated grades on their evaluations. Studies show a notable correlation between expected grades and evaluation ratings, suggesting that educators might adopt lenient grading or simplify course content to boost students' anticipated grades in exchange for undeservedly high evaluation scores. This phenomenon distorts evaluation outcomes, inevitably undermining overall teaching quality [1, 2]. Within the evaluation score and grade-level exchange dynamic, students may use higher teaching evaluation scores as leverage to obtain more lenient grading, while instructors may exploit past impressions of generous grading or suggest lenient grading practices to elevate students' expected grades in exchange for higher evaluation scores. The critical question arises: does this seemingly mutually beneficial exchange truly exist? Can higher teaching evaluation scores genuinely

translate into more lenient grading practices? This inquiry forms the basis of our investigation into the reciprocal relationship between teaching evaluations and grading leniency. Leveraging extensive student life and academic performance data within a smart campus's big data system, and employing data mining techniques, this study statistically analyzes the correlation between students' expected grades and teaching evaluations. By adjusting evaluation results based on these insights, the process aims to make teaching evaluations more authentic and fair, ultimately enhancing the quality of higher education teaching.

2. Literature Review

Teaching evaluation is an essential metric for assessing instructional quality in higher education, providing a measurable method to evaluate educators' proficiency and commitment. Worldwide, student feedback is the primary method for these evaluations, favored because students are direct recipients of instruction and actively participate in the teaching process. Additionally, collecting student opinions is cost-effective and facilitates the analysis of large datasets [3, 4]. In contrast, internal peer reviews may be biased due to vested interests, and evaluations by administrators may lack persuasiveness since administrators are not directly involved in teaching activities.

Over the past few decades, widespread adoption of student evaluations has revealed notable limitations. Grade inflation has emerged as a key concern, influenced by students' anticipated grades, which depend not only on their understanding of course material but also on grading leniency and effort. Kanagaetnam's utility function analysis indicates that over-reliance on student evaluations for assessing instructional quality can incentivize instructors to lower course requirements and grading standards, leading to inflated grades and disproportionately high evaluation scores [5]. Furedi and Baker further note that relationships among administrators, teachers, and students have evolved into a business-like structure, with teachers motivated by pay and promotion, and administrators relying heavily on student evaluations due to the inability to directly observe teaching effectiveness [6, 7].

This dynamic fosters a cycle where students seek higher grades, teachers cater to students to enhance evaluations, and institutions conform to the trend of grade inflation to meet short-term employment and educational demands. The lack of uniform grading standards exacerbates the problem, creating a "prisoner's dilemma" where rational individual choices lead to unfair evaluations and deteriorating educational outcomes.

Research on grade inflation in international contexts began after the Vietnam War, with Juola pioneering this field. Analyzing the average GPA of 134 American universities from 1965 to 1973, he observed an increase of 0.404 over nine years [8]. Building on this, Rojstaczer analyzed nearly 70 years of GPA data from multiple American universities, finding that the proportion of students receiving "A" grades nearly tripled between 1940 and 2012, with some institutions awarding over 50% "A" grades while "C" grades decreased significantly [9, 10]. Studies in other Western countries and China similarly observed grade inflation, suggesting that global higher education institutions using student evaluations have encountered varying degrees of this phenomenon [11].

The negative impact of grade inflation on talent development is evident. Teachers influencing students' anticipated grades by reducing course difficulty or subtly encouraging high scores undermines evaluation fairness. Consequently, students often choose courses with

perceived grade inflation, avoiding more challenging ones. This grade-based rather than knowledge-based approach erodes academic integrity [12].

3. Data Selection

Student data within campuses are diverse, as summarized in Table 1. For this study, four representative indicators were selected: actual grade, expected grade, course content evaluation, and teaching evaluation. Data for actual grades were obtained from student grade records, while expected grades, course content evaluation, and teaching evaluation data were derived from Course and Teaching Evaluation (CTE) records. All data were obtained from The Chinese University of Hong Kong, Shenzhen, covering the period from 2018 to 2020.

Table 1. Selection of data and indicators,

Indicator	Data Source
Actual Grade	Student grade records
Expected Grade	Course and Teaching Evaluation (CTE)
Course Content Evaluation	Course and Teaching Evaluation (CTE)

4. Data Preprocessing

Prior to conducting data analysis, the raw datasets were carefully preprocessed to ensure accuracy, consistency, and suitability for statistical evaluation. The following steps were undertaken:

4.1. Clean datasets.

All datasets were examined for missing or blank values, which were subsequently removed to ensure the integrity of the data and prevent potential biases in the analysis.

4.2. Define low-rated courses.

Courses with evaluation scores lower than 5.5 (on a scale of 0 to 6) in the Course and Teaching Evaluation (CTE) records were classified as low-rated courses. This threshold allowed for the identification of courses requiring further investigation or potential improvement.

4.3. Convert student grade levels.

Student grade levels were transformed into numerical values ranging from 1 to 5. In this scale, lower numbers corresponded to better performance (higher grades), while higher numbers indicated poorer performance (lower grades). This conversion facilitated quantitative analysis and comparison across different courses and cohorts.

4.4. Re-label course codes.

Course codes were standardized for consistency. Courses offered by School A were labeled as "Axxxx," where “xxxx” indicated course-specific information. The first two digits represented the academic year (e.g., “18xx” denoted the 2018 academic year), and the last two digits corresponded to the semester within that year (10 for Fall, 20 for Spring). This labeling enabled clear tracking of courses across multiple years and semesters for longitudinal analysis.

5. Experimentation

5.1. Study on the effectiveness of the evaluation system.

Experiment 1: Investigating whether the CTE system can effectively enhance the quality of higher education teaching. To assess the impact of Course and Teaching Evaluation (CTE) on enhancing teaching quality in higher education, the study focused on a sample of 45 courses from three different schools (School A, B, and C), selecting 15 courses from each school. These courses were chosen based on the criterion that the same instructor taught the same course in both the Fall semester of 2018 (1810) and the Fall semester of 2019 (1910). This consistency allows for a more accurate comparison of the course content evaluation scores and teacher teaching evaluation scores across these periods. The hypothesis suggests that a significant improvement in CTE scores for courses with initially poor evaluations in the Fall semester of 2018 indicates that low CTE scores can motivate instructors to improve their course content and teaching methods. These enhancements are likely to result in higher CTE scores in the following year. Despite concerns about the authenticity of CTE scores, they can still play a crucial role in promoting better teaching quality. The effect demonstrates that CTE, despite its potential limitations, serves as a valuable tool for improving educational standards and teacher performance. Following a comparative analysis of the statistical data, Fig. 1 and Table 1 clearly illustrates that courses with lower CTE scores in the Fall of 2018 generally witnessed a significant increase in their CTE scores in 2019.

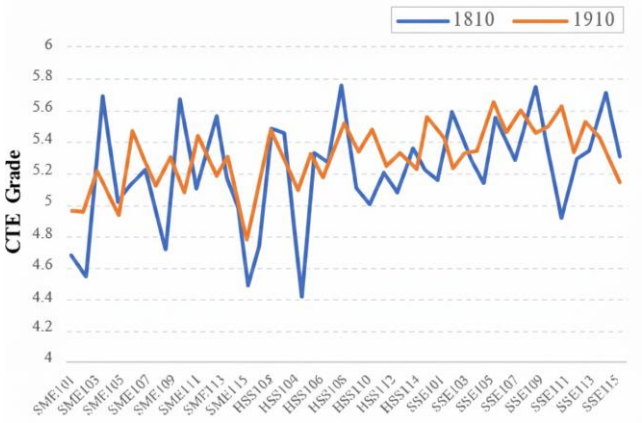

Figure 1. Comparison of course content evaluation scores between the Fall semester of 2018 and the Fall semester of 2019.

Table 2. CTE score improvements.

School	Course CTE Improvement Ratio	Teaching CTE Improvement Ratio	Average Improvement in Course CTE	Average Improvement in Teaching CTE
School A	93.3%	73.3%	0.179	0.115
School B	67.7%	80%	0.296	0.295
School C	67.7%	53.3%	0.201	0.111

Similarly, courses instructed by teachers with lower CTE scores in the Fall of 2018 also exhibited a substantial rise in these teachers' CTE scores in 2019 (Fig. 2). As shown in Table II, more than half of the courses in each school had an increase in CTE scores, with School A having the highest proportion of courses showing improvement in CTE scores at 93%. School B experienced the greatest average increase in CTE scores, nearly 0.3, and the absolute values of the increases in course CTE scores were generally higher than those of teacher CTE scores.

In response to the aforementioned statistical data, we employed the Shapiro-Wilk test to assess the normality of the data, and the samples showed significance at the 5% confidence level, indicating they did not follow a normal distribution. Given that the CTE score distribution does not adhere to a normal distribution and the data samples are from repeated observations of the same course over two time periods, we opted to use the paired sample Wilcoxon signed-rank test to evaluate the differences in CTE data between 1810 and 1910. As indicated in Table 3, the Wilcoxon test yielded a p-value of 0, and the effect size of Cohen's d-value was 1.27, suggesting a very large difference between the two semesters. The results of Experiment 1 suggest that the current methods of student evaluations of teaching hold some degree of effectiveness across all schools, improving the quality of courses and teaching. When teachers receive lower CTE scores for teaching or courses, they are more motivated to make improvements in course content or their teaching methods, with the aim of achieving higher CTE scores subsequently.

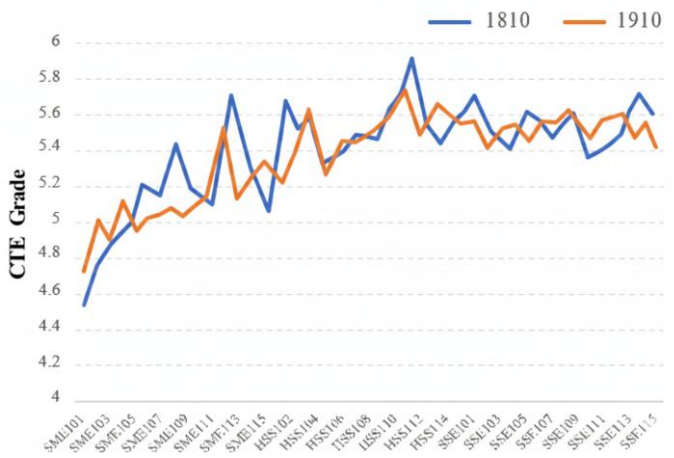

Figure 2. Comparison of teaching evaluation scores between the Fall semester of 2018 and the Fall semester of 2019.

Table 3. Wilcoxon signed-rank test results for course cte scores.

Paired Variables	z-value	df	P	Cohen's d
1810 Course CTE Scores vs 1910 Course CTE Scores	4.271	28	0.000***	1.27

5.2. Statistical relationships Between teaching evaluations and grade levels.

Given the effectiveness of CTE in enhancing teaching quality in higher education, considering the impact of the prisoner's dilemma involving teachers, school administrators, and students on the accuracy of CTE, students want to improve their grades to secure better employment and higher education opportunities, often leading them to exchange higher evaluation scores for inflated grades in teachers' courses. To explore the dynamics of this exchange, we analyze the relationship between course evaluations, professors' evaluations, and the actual grade versus expected grade scores. By examining these relationships, we seek to understand the extent to which the desire for higher grades influences student evaluations and whether students indeed receive the inflated grades they desire in exchange for higher CTE scores.

Experiment 2: The relationship between teaching evaluations and expected grade levels. The CTE data from The Chinese University of Hong Kong, Shenzhen, encompass students' expected grade levels, course content evaluations, and teaching evaluations. We employed the

Spearman correlation coefficient to analyze these data. The experimental results, as presented in Table 4, reveal a highly negative correlation of -0.829 between the expected grade levels and both course content evaluations and teacher teaching evaluations (a lower grade level signifies a better grade), confirming that as students' estimated grades improve, the scores they assign to courses and teachers increase correspondingly.

Table 4. Correlation between expected grade, course CTE score, and teaching CTE score.

Variable	Expected Grade	Course CTE Score	Teaching CTE Score
Expected Grade	1.000 (***0.000)	-0.829 (***0.042)	-0.829 (***0.042)
Course CTE Score	-0.829 (***0.042)	1.000 (***0.000)	1.000 (***0.000)
Teaching CTE Score	-0.829 (***0.042)	1.000 (***0.000)	1.000 (***0.000)

Experiment 3: Relationship Between Teaching Evaluations and Actual Grades. In Experiment 3, we continued to use the Spearman correlation coefficient but replaced expected grade levels with actual grade levels, representing the final grades students achieved in the course. The results, shown in Table 5, indicate that the correlation between course CTE scores and actual grades is -0.014, while the correlation between teaching CTE scores and actual grades is -0.019. These findings suggest no meaningful relationship between students' final grades and CTE scores.

Table 5. The relationship between course content evaluations, teaching evaluations,and actual grade.

Variable	Actual Grade	Course CTE Score	Teaching CTE Score
Actual Grade	1.000 (0.000***)	-0.014 (0.000***)	-0.019 (0.000***)
Course CTE Score	-0.014 (0.000***)	1.000 (0.000***)	0.774 (0.000***)
Teaching CTE Score	-0.019 (0.000***)	0.774 (0.000***)	1.000 (0.000***)

Across all three experiments, we examined correlations among four key indicators: actual grade, expected grade, course content evaluation, and teaching evaluation. A summary of these correlations is presented in Table 6.

Table 6. Correlation relationships among the four categories of datasets.

Variable	Actual Grade	Expected Grade	Course Content Evaluation	Teaching Evaluation
Actual Grade	—	Not Related	Not Related	Not Related
Expected Grade	Not Related	—	Positively Correlated	Positively Correlated
Course Content Evaluation	Not Related	Positively Correlated	—	Positively Correlated
Teaching Evaluation	Not Related	Positively Correlated	Positively Correlated	—

6. Discussion of the Experimental Results

Observations from Experiments 2 and 3 reveal a clear pattern: teaching evaluation scores are positively correlated with students' expected grade levels. However, for a specific course, no significant relationship was found among the evaluations of course content, instructor teaching performance, and students' actual final grades. Three plausible explanations for this phenomenon are outlined below. First, instructors may subtly imply lenient grading standards to elevate students' grade expectations, thereby strategically influencing teaching evaluations. Since final grades must be approved by the registrar's office, overt grade inflation is risky and can be detected. However, if instructors make vague or exaggerated promises about grading

leniency before the end of the term—essentially "promising the moon"—they may secure higher evaluation scores without altering actual grades. Such hints are typically indirect, making them difficult for the administration to detect or substantiate. Moreover, students often lack transparency into the grading process and may not recognize the insincerity of such promises. This asymmetry of information allows instructors to benefit from inflated evaluations with minimal risk, effectively creating an arbitrage opportunity that may shape their grading behavior. Thus, while final grades may not be significantly inflated, students' expectations are raised just enough to yield more favorable evaluations.

Second, instructors may reduce course difficulty during teaching, leading students to perceive a stronger grasp of the material than they actually possess. Although the overall difficulty is lowered for everyone, students tend to feel more confident and comfortable. However, because students are more attuned to their own progress than to that of their peers, they may overestimate their relative performance. This overestimation raises their expected grades, which in turn inflates their teaching evaluations, even if their actual grade percentiles remain unchanged [13].

Third, many universities worldwide schedule student evaluations of teaching before final exams, with exam scores typically being the main determinant of course grades. This timing is intended to prevent students from basing evaluations on received grades. Nevertheless, students remain sensitive to instructors' grading standards and often try to anticipate their own grade levels when completing evaluations. The experiments highlight that while course and teaching evaluations (CTE) remain useful tools for assessing faculty performance, their limitations are becoming increasingly evident. Some instructors, in pursuit of higher evaluation scores, cater to student desires for high grades, exacerbating grade inflation [14]. However, in the short term, there is no observable link between more favorable evaluations and higher class grades. This suggests that students' hopes of raising average class scores through inflated evaluations are unfounded. As the disadvantaged party in an information-asymmetric relationship, students' most rational approach is to evaluate teaching honestly and resist being swayed by implicit promises of grade leniency.

Although the strategy of leveraging inflated evaluations to boost class grades is unlikely to succeed, instilling this understanding in every student is difficult given current limitations in monitoring and guiding individual behavior. Therefore, it becomes essential for institutions to rely on structured communication and educational initiatives to gradually cultivate this awareness among students. That said, cultivating such awareness will take considerable time. In the interim, it may be necessary to adjust teaching evaluation scores to account for non-instructional influences. Factors beyond expected grades—such as course difficulty, instructor gender, and appearance—have also been shown to affect evaluation outcomes [15]. Moreover, students' expected grades are shaped not only by perceived grading leniency and teaching quality but also by their own mastery of the material, effort, and interest in the course. In fact, higher engagement and sustained interest can themselves be indicators of effective teaching.

Historically, capturing and analyzing non-academic behavioral data has been challenging. However, the rise of smart campuses and increased digitization in higher education have opened new avenues for applying big data technologies. Institutional information systems now store extensive behavioral data on students' academic and personal activities, providing a rich foundation for large-scale analysis. Combined with advances in data mining and machine learning, these data offer unprecedented opportunities to develop more

holistic and fairer evaluation systems. By examining correlations between behavioral patterns and expected grades, it becomes possible to assess students' real-time learning states, predict future behaviors, and adjust teaching evaluations accordingly. This approach can enhance the fairness and accuracy of evaluations, ultimately contributing to improved instructional quality.

Research by Aziz and Okoye [16, 17] has demonstrated the potential of data mining and machine learning in analyzing teaching-related datasets. These techniques enable robust contextual analysis and real-time prediction of student performance based on behavioral indicators. By identifying key factors influencing student outcomes, such tools can clarify the relationships between student behavior logged in smart campus systems and their expected grades. They also help quantify the impact of non-teaching elements on evaluation scores, allowing for more precise recalibration of results to better reflect instructors' true teaching competence.

7. Conclusion

This study has shown that representative indicators can effectively reveal the relationship between expected grades and teaching evaluations in higher education. Using smart campus data from The Chinese University of Hong Kong, Shenzhen, we identified key factors affecting students' grade expectations, which can inform the development of AI models aimed at enhancing the accuracy and credibility of student evaluations of teaching. While grade inflation may pose long-term challenges to evaluation systems, short-term improvements in class average grades will require adjustments to evaluation scores. Students should be encouraged to resist grade inflation and provide honest feedback to support genuine improvements in teaching quality. Future research will expand the scope of data sources within the smart campus environment. By integrating diverse dimensions of student life and academic performance—beyond traditional metrics—we hope to uncover additional influences on expected grades and evaluation scores. Ultimately, we aim to conduct comprehensive correlation analyses among various data types, including course engagement, student interests, and digital footprints in learning management systems, to further enhance teaching quality in higher education through AI-driven insights.

Acknowledgment

This work was supported in part by the Shenzhen Education Project (JYPJ22001), the National Natural Science Foundation of China (NSFC) under Grant 61902332, the Shenzhen Stability Science Program (20231130170021001), and the Guangdong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence (2023B1212010001).

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

All authors contributed equally to the conception, design, analysis, and writing of this manuscript. All authors have read and approved the final version of the manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Reference

[1] Nowell, C.; Alston, R.M. (2007). I thought I got an A! Overconfidence across the economics curriculum. *The Journal of Economic Education*, *38*, 131–142. https://doi.org/10.3200/JECE.38.2.131-142.

[2] Tiron-Tudor, A.; Nistor, C.S.; Fekete, S.; Alexandru, A. (2022). Factors Influencing Public Higher Education Institutions' Performance Reporting in the Romanian Context. *Administrative Sciences*, *12*, 163. https://doi.org/10.3390/admsci12040163.

[3] Ciuchi, O.M.; Șerbănescu, L.E.; Dobre, C.M.; Georgescu, B.G.; Țigănoaia, B.D.; Țucă, P.L. (2024). The Impact of Student Evaluation of Teaching Staff on Enhancing the Quality of Teaching in Higher Education in Romania. *Sustainability*, *16*, 10196. https://doi.org/10.3390/su162310196.

[4] Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*, 276–294. https://doi.org/10.1080/01973533.2020.1756817.

[5] Kanagaretnam, K.; Mathieu, R.; Thevaranjan, A. (2003). An economic analysis of the use of student evaluations: Implications for universities. *Managerial and Decision Economics*, *24*, 1–13. https://doi.org/10.1002/mde.1099.

[6] Díaz-Méndez, M.; Paredes, M.R.; Saren, M. (2019). Improving Society by Improving Education through Service-Dominant Logic: Reframing the Role of Students in Higher Education. *Sustainability*, *11*, 5292. https://doi.org/10.3390/su11195292.

[7] Baker, G.P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy*, *100*, 598–614.

[8] Salas-Velasco, M. (2023). Panel Data Models for School Evaluation: The Case of High Schools' Results in University Entrance Examinations. *Stats*, *6*, 312-321. https://doi.org/10.3390/stats6010019.

[9] Zhang, Y.; Hu, Z. (2025). Research on Grade Point Innovation and Grade Point Average Based on Deep Learning Networks and Evolutionary Algorithms for College Innovation Education. *Sustainability*, *17*, 2171. https://doi.org/10.3390/su17052171.

[10] Rojstaczer, S.; Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940–2009. *Teachers College Record*, *114*, 1–23. https://psycnet.apa.org/doi/10.1177/016146811211400707.

[11] Bojorque, R.; Moscoso, F.; Pesántez, F.; Flores, Á. (2025). Stress Factors in Higher Education: A Data Analysis Case. *Data*, *10*, 22. https://doi.org/10.3390/data10020022.

[12] Liao, R.; Ou, Q.; Zheng, W.; Shi, Z.; Li, S. (2022). Credibility based teaching assessment in smart campus. *2022 IEEE Integrated STEM Education Conference,* pp. 301–304. https://doi.org/10.1109/ISEC54952.2022.10025348.

[13] Boring, A.; Ottoboni, K. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*, *1*, 1–15. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.

[14] Nordin, M.; Heckley, G.; Gerdtham, U. (2019). The impact of grade inflation on higher education enrolment and earnings. *Economics of Education Review*, *73*, 101936. https://doi.org/10.1016/j.econedurev.2019.101936.

[15] Yilmaz, Y.; Sarikaya, O.; Senol, Y.; Baykan, Z.; Karaca, O.; Demiral Yilmaz, N.; Sayek, İ. (2021). RE-AIMing COVID-19 online learning for medical students: a massive open online course evaluation. *BMC Medical Education*, *21*, 1–14. https://doi.org/10.33899/berj.2020.166006.

[16] Aziz, S. F. A. (2020). Students' performance evaluation using machine learning algorithms. *College Of Basic Education Research Journal*, *16*, 977–986. https://doi.org/10.33899/berj.2020.166006.

[17] Okoye, K.; Arrona-Palacios, A.; Camacho-Zuñiga, C.; Achem, J. A. G.; Escamilla, J.; Hosseini, S. (2022). Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. *Education and Information Technologies*, *27*, 1–43. https://doi.org/10.1007/s10639-021-10751-5.

© 2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).