



OPEN ACCESS

EDITED BY

Gladys Sunzuma,
Bindura University of Science
Education, Zimbabwe

REVIEWED BY

Valentine Ukachukwu Okwara,
University of the Free State, South Africa
Svetoslav Atanasov,
Trakia University, Bulgaria

*CORRESPONDENCE

Máté Pusker

✉ pusker.mate@kre.hu

RECEIVED 13 March 2025

ACCEPTED 02 September 2025

PUBLISHED 18 September 2025

CITATION

Pusker M, Takács S and Takács R (2025)
Student evaluation of teaching prior, during
and after the COVID-19 pandemic.
Front. Educ. 10:1593000.
doi: 10.3389/feduc.2025.1593000

COPYRIGHT

© 2025 Pusker, Takács and Takács. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Student evaluation of teaching prior, during and after the COVID-19 pandemic

Máté Pusker^{1*}, Szabolcs Takács² and Rita Takács³

¹Doctoral School of Psychology, Károli Gáspár University of the Reformed Church in Hungary, Budapest, Hungary, ²Institute of Psychology, Department of General Psychology and Methodology, Károli Gáspár University of the Reformed Church in Hungary, Budapest, Hungary, ³Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary

Introduction: The primary aim of this study is to determine whether the quality of education declined at a Hungarian university during the COVID-19 lockdown. Contrary to studies with smaller sample sizes, we analyzed SET results from the university's IT faculty (both undergraduate and postgraduate programs) to provide empirical evidence for the academic discourse on the quality of education before, during, and after the COVID-19 pandemic. Although controversy surrounds SET questionnaires regarding their validity, reliability, and application, valuable information can be extracted from student feedback.

Method: This research retrospectively investigates the quality of education during the COVID-19 pandemic by analyzing student evaluation of teaching (SET) questionnaires at a Hungarian university collected from 2015 to 2022. This timeframe allows for a unique comparison between pre-COVID, COVID, and post-COVID educational periods, providing essential context for the pandemic's impact.

Results: Our analysis of SET questions yielded statistically significant results; however, the effect sizes were negligible ($\epsilon^2 < 0.01$), indicating a lack of meaningful differences between the COVID and non-COVID periods.

Discussion: Results suggests that despite the widespread disruption across all areas of life during the pandemic, the quality of education, with all its inherent challenges, did not substantially decrease in this Hungarian university.

KEYWORDS

SET, COVID-19, student feedback, higher education, STEM-field

Introduction

The COVID-19 pandemic required an immediate response from higher education institutions, leading to the suspension of on-campus attendance. While traditional in-person education could not be maintained, online education presented new challenges for institutions, teachers, and students alike (Adedoyin and Soykan, 2020; Daniel, 2020). Given the scarcity of examples of such drastic global educational reorganization, ensuring the quality of education became a central concern, as the transition to online education had to occur with extreme haste (Means et al., 2020). This adaptation was facilitated by advancements in teaching methodologies and support systems during the pandemic (Abu Talib et al., 2021; Lin et al., 2022; Zhang et al., 2022; Boyle and Cook, 2023). This research aims to contribute to the academic literature by evaluating feedback on the quality of education during the pandemic through the analysis of large-sample quantitative data from Student Evaluation of Teaching (SET) surveys.

Concerns regarding the negative effects of the pandemic on higher education were widely expressed (Boysen, 2023), encouraging researchers to investigate the issue. Abu Talib et al. (2021), in their review of research on student and faculty feedback, found

that approximately 22% of articles addressed negative aspects of online education (e.g., anxiety, doubts, overwhelming effects), while the remaining research proactively gathered information concerning the exceptional situation caused by the pandemic.

Jin et al. (2021) observed that students recognized the convenience introduced by online learning during COVID-19 and found the quality of online lectures equally satisfactory (Angelova, 2020). Surprisingly, Alam and Singh (2021) found that IT students performed better with online teaching and learning compared to in-person education. Not only did student performance remain stable, but teacher performance also appeared unaffected by COVID-19 (Reis et al., 2021), irrespective of staff gender, academic title, or field of teaching (Berniak-Wozny et al., 2021). From an institutional perspective, no significant difference in course feedback was detected during the switch to online learning (Rodrigo et al., 2022). Furthermore, Charytanowicz et al. (2022) found no clear evidence that pandemic disruption and online learning led to knowledge deficiencies.

However, one study found that 80% of students would choose a traditional educational setting (Kazainé, 2021), indicating dissatisfaction with online teaching and a preference for on-site learning (Guo, 2020). There is only a slight predominance of students who perceive the intensity of online education as less compared to traditional lectures (Angelova, 2020). Moreover, the transition from traditional to online education was evaluated negatively, specifically noting that courses became less enjoyable, less interesting, and decreased in learning value (Garris and Fleck, 2022). When comparing the experiences of engineering students, Warfvinge et al. (2022) found that during the pandemic, student satisfaction with courses decreased, they received less feedback, and understanding course completion requirements became more difficult.

Ahmed et al. (2022) examined feedback from IT students on teaching in a fully online environment during the COVID period to identify key factors for effective education. Analyzing data from 27,622 individuals, two factors characterized successful education:

1. Student-faculty relationship, encompassing:

- Respectful relationships between teachers and students
- Ensuring a supportive educational climate
- Care and support for students
- Fair, objective, and consistent treatment of students

2. Routine activities, such as:

- Tracking student attendance
- Providing accurate class status
- Availability of online teaching materials
- Clear structure and communication of course framework and expectations

Sepahi et al. (2021), in their small-sample study, found that the most important teaching characteristics during the COVID period, from the students' perspective, were effective presentation, observable professional and ethical attitudes, and consultation and support in problem-solving.

LeBlanc (2021) investigated SET results during pandemic-affected semesters and found that SET scores deteriorated, despite still receiving relatively high average scores. Boysen (2023) also investigated whether SET results would negatively change, aligning with assumptions that the pandemic adversely affected online education. Boysen's results showed a significant difference between non-COVID and COVID periods (the latter receiving significantly higher ratings, but with a low degree of impact), concluding that average SET results were not affected by the magnitude of external events. Campos et al. (2022) followed up on an education reform implemented during the COVID-19 period and found that alongside a more positive perception of the reform, there were no negative changes in SET assessments. A more recent study by Fullerton et al. (2024), comparing SET scores across pre-COVID and COVID periods, did not reveal meaningful changes.

Despite the ongoing debate surrounding the reliability of SET as a source of information, it remains a widely applied instrument at universities. Pineda and Steinhardt (2023), in their comprehensive analysis, noted that the focus of SET is divided between viewing the methodology as a management tool (American continent) or as a tool to improve the quality of education. The aim of SET surveys is, among other things, to support the quality assessment of teaching (Zabaleta, 2007), aid decisions regarding promotions (Clayson, 2009; Berezvai et al., 2020; Goos and Salomons, 2017) and provide students an opportunity to express opinions about a subject or instructor (Baxter, 1991; Safavi et al., 2013). Professional judgment on the methodology is heterogeneous, with sharp criticism (Cook et al., 2022; Lakeman et al., 2023; Spooren et al., 2013) alongside views of the method as a useful approach for gathering information (Ali et al., 2021; Murray, 1997; Ulker, 2021).

Factors unrelated to teaching quality, such as age (Murray et al., 2020; Stonebraker and Stone, 2015), likability (Clayson, 2022), physical appearance (Reinsch et al., 2020), and cultural characteristics (Arnold and Versluis, 2019), can all impact SET ratings. Some research has examined, with varying results, whether gender also affects student assessment (Boring and Ottoboni, 2016; Kreitzer and Sweet-Cushman, 2022; Tangalakakis et al., 2022). Feistauer and Richter (2018) also investigated the effect of sympathetic teachers on SET outcomes. Their results indicated that teacher sympathy positively affects SET scores; however, they noted the low number of participants in their research due to voluntary participation. Students who positively assess the SET questionnaire and its usefulness tend to give higher scores, making the reliability of the measurement questionable (Spooren and Christiaens, 2017).

SET also appears to be an unreliable measure of teaching effectiveness, as only a weak relationship was found between current academic performance and teaching effectiveness (Sánchez et al., 2020). Esarey and Valdes (2020) argue that unbiased, reliable SETs may contain a reporting error that should not be ignored when interpreting results. According to Uttl (2021, 2023), it is a mistake to assess teachers' educational effectiveness based solely on SET results. Additionally, Zabaleta (2007) found a moderate correlation between students with low grades and low SET scores in a sample of 18,175 students. This could lead to teachers awarding better grades in hopes of receiving higher SET scores, resulting in grade inflation (Nowell, 2007; Berezvai et al., 2020), when examining factors influencing SET evaluation, found that results are not independent of the grade received for

the given subject; a better grade correlates with a higher SET rating for the instructor. Berezvai et al. (2019) also found that financial incentives for educators significantly increased grade averages in courses where students had previously received low average grades.

When reviewing SET literature, one encounters various experimental settings and sometimes conflicting results; thus, appropriate context is required to support the future use of identified relationships and effects in the research field (Ali et al., 2021). Active student participation in SET surveys is essential for ensuring quality education (Salvo-Garrido et al., 2022). However, SET data collection itself proves challenging, which forms the cornerstone of validity concerns (Ali et al., 2021). In many cases, the completion rate of SET is considered low (Nulty, 2008), increasing the chance of sampling error (Goos and Salomons, 2017; Wolbring and Treischl, 2016). Meta-analyses on the relationships between student studies and SET were reviewed by Uttl et al. (2017), who found that previous conclusions were often based on small sample sizes. For this reason, they re-ran the analysis adjusting for sample size and found no correlation between the variables (Uttl et al., 2017).

The widespread use of SET results raises additional expectations regarding the reliability and validity of the measurement (Zhao et al., 2022); however, it is increasingly accepted that SET can play an active role in improving the quality of education (Ali et al., 2021; Penny, 2003). We observe examples of SET use in improving the quality of educational work (Golding and Adam, 2016) and in research supporting administrative activity (Barrow and Grant, 2016). SET scales focusing on fundamental aspects of learning (e.g., teacher helpfulness) highly predict high teacher ratings (Park and Dooris, 2020). It is also important to highlight that students associate value with SET surveys (Stein et al., 2021), supporting the use of this key measurement for both administration and educators (Linse, 2017).

Following prior research building on SET results (LeBlanc, 2021; Boysen, 2023), our aim was to determine if differences could be detected by comparing not only the COVID-19 period SET scores to pre-COVID results but also by including post-COVID SET results in the analysis. The scale and type of this analysis remain scarce in the literature to date (Fullerton et al., 2024). Based on Warfvinge et al. (2022), we sought to detect the potential negative effects of “distance learning” utilizing the available data. Based on these considerations, our research question is:

RQ: What effect did the COVID-19 period have on SET results that reflect on the quality of education?

Based on Ahmed et al. (2022), we examined key SET questions related to the course framework and requirements, which were found to be indispensable during COVID-19 education.

H1: The perception of course structure received lower ratings during the COVID-19 period compared to pre-COVID and post-COVID periods.

H2: The assessment of the feasibility of course requirements is worse in the COVID-19 period compared to pre-COVID and post-COVID periods.

We assume that due to the COVID-19 pandemic, teachers were required to provide more support and motivation, which would manifest in higher scores on the corresponding SET questions (Sepahi et al., 2021).

H3: Students gave higher scores on the “teacher’s professional, motivating and supportive attitude” question during the COVID-19 period compared to pre-COVID and post-COVID periods.

Based on Means et al. (2020), we assume that during the COVID-19 pandemic, due to the transition to online education, the interactivity of practice courses could not be achieved as effectively as in the previous period. However, after the COVID-19 experiences and practices, students consider available teaching methodologies successful.

H4: When judging the interactivity of practice courses, students gave lower ratings during the COVID-19 period compared to pre-COVID and post-COVID periods.

Materials and methods

Hungarian universities conduct SET surveys every semester to monitor educational quality. The surveys and evaluations are available at all faculties, serving multiple purposes. Students provide separate feedback on the course quality and the teacher’s performance. This research is based on SET data from one Hungarian university’s IT Faculty, including both undergraduate and postgraduate students feedback. The completion period for the surveys begins on the first day of the exam period and concludes on the penultimate day. The Quality Office is responsible for gathering information across all faculties. Students provide anonymous, voluntary feedback on courses they have signed up for each semester. This anonymity aims to promote full disclosure and authenticity of information. The university’s Senate stipulates that students who complete all questionnaires receive 8 extra points as a reward, usable for course admission in the subsequent semester. This practice aims to remedy low completion rates by introducing incentives, such as the possibility of early subject admission for survey respondents (Lukáts et al., 2023). Naturally, with an increased number of evaluable fill-ins, the number of non-evaluable fills may also increase, and response bias must be considered; however, an overall increase in evaluable data can be achieved (Lukáts et al., 2023).

TABLE 1 Labels of individual SET questions.

Label	SET question
Course structure	How well was the course structured?
Feasibility of course requirements	To what extent were the course requirements achievable?
Teacher’s professional, motivating and supportive attitude	To what extent was the instructor professionally motivating and inspiring?
Interactivity	To what extent did the instructor conduct the classes in an interactive manner?

The research builds on individual SET questions to test the hypotheses. All individual SET questions were assigned a label to make the analysis and its interpretation easier (Table 1).

The university collects student feedback using a four-point Likert scale, where 1 is the minimum and 4 is the maximum score. While this raises questions regarding potential validity and reliability errors, Chang (1994) suggests that the choice of the number of points in a Likert scale may depend on the empirical setting. In our research, students are prevented from choosing a neutral stance, which serves institutional purposes.

This study is based on a secondary analysis of existing data obtained from the university. The SET results were extracted from the NEPTUN educational system in an anonymized format. Student identities and feedback were anonymized to the extent that only the number of survey responses is available for each course and the corresponding teacher evaluations. Consequently, no information is available regarding response rates, gender, age, or other sample characteristics. While this represents a limitation of the study, the volume of available data provides a sufficient basis for interpreting SET scores across the key time periods.

We defined a “Period” variable to categorize information into pre-COVID (spring 2015 to autumn 2019), COVID (spring 2020 to spring 2021), and post-COVID (autumn 2021 to spring 2022) periods. Our second definition was the “Course Type” variable, enabling the analysis to differentiate between Lectures, Practice courses, and a combination of the two labeled as Lecture and practice. All other courses (labs, seminars) were treated uniformly, as the research did not aim to investigate these types of courses.

Results

IBM SPSS 25.0 for Windows (IBM Corp., Armonk, NY, USA) software was used to conduct the analysis. The sample size (Table 2) is sufficient for the analysis (Nulty, 2008; Uttl et al., 2017).

A Kruskal-Wallis H test was conducted to examine differences between ordinal data (Field, 2017), which is suitable for this dataset as the assumption of equality of variances was violated (Table A1). Mean values of the selected SET questions based on Period can be found in Table 3. Since the individual SET questions assessed distinct constructs rather than a single underlying factor, scale reliability analysis was not conducted.

The first analysis involved the first three hypotheses, investigating differences among the three Periods across all course types concerning the selected SET questions. Bonferroni correction was applied to control for multiple comparisons (adjusted $p = 0.01$ for three tests, rounded down) to control for Type I error. The analysis revealed statistically significant ($p < 0.001$) differences among the periods (Table 4). However, these differences demonstrated negligible effects ($\epsilon^2 < 0.01$), indicating minimal practical implications (Tomczak and Tomczak, 2014). For this reason, *post hoc* comparisons were not conducted, in line with recommendations to avoid overinterpreting trivial effects (Funder and Ozer, 2019).

The second analysis focused on the Interactivity of Practice courses across the three periods. Similarly, a Kruskal-Wallis H test was conducted with Bonferroni correction. The test revealed

TABLE 4 Kruskal-Wallis test.

SET question	N	df.	KW	p	ϵ^2
Course structure	145,740	2	1083.033	<0.001	0.0074
Feasibility of course requirements	58,412	2	255.901	<0.001	0.0043
Teacher's professional, motivating and supportive attitude	143,051	2	855.578	<0.001	0.0059

N, number of cases; df, degree of freedom; KW, Kruskal-Wallis H test; p, significance level; ϵ^2 , Epsilon² effect size.

TABLE 2 Frequency of feedback based on periods.

SET question	Course type	Period			
		Pre-COVID	COVID	Post-COVID	SUM
Course structure	All courses	6,289	25,946	26,006	58,241
Feasibility of course requirements	All courses	6,289	25,946	26,006	58,241
Teacher's professional, motivating and supportive attitude	All courses	84,876	45,986	12,189	143,051
Interactivity	Practice courses	27,543	12,094	3,559	43,196

TABLE 3 Mean value of SET questions based on the period variable.

SET question	Period					
	Pre-COVID		COVID		Post-COVID	
	Mean	SD	Mean	SD	Mean	SD
Course structure	3.35	0.796	3.45	0.736	3.49	0.724
Feasibility of course requirements	3.34	0.782	3.48	0.717	3.49	0.730
Teacher's professional, motivating and supportive attitude	3.25	0.914	3.38	0.854	3.41	0.849
Interactivity	3.26	0.886	3.37	0.849	3.42	0.848

SD, Standard Deviation.

TABLE 5 Kruskal-Wallis test.

SET question	N	df.	KW	p	ϵ^2
Interactivity of practice courses	43,196	2	71.999	<0.001	0.0016

N, number of cases; df, degree of freedom; KW, Kruskal-Wallis H test; p, significance level; ϵ^2 , Epsilon² effect size.

significant differences ($p < 0.001$) among the three periods regarding the Interactivity of Practice courses (Table 5); however, we found a negligible effect (Tomczak and Tomczak, 2014). Since the effect size was negligible ($\epsilon^2 < 0.01$), *post-hoc* testing was discarded.

Summary of the results can be found in Table 6.

Discussion

Following Warfvinge et al. (2022), we initially suspected a decrease in SET scores influenced by the COVID-19 pandemic. Based on Ahmed et al. (2022), we examined SET questions related to the framework and requirements of the courses. Our findings indicate that the perception of course structure and feasibility of course requirements during the COVID-19 period did not differ meaningfully, despite statistical significance. We suggest this is attributable to the inherent stability of well-established course structures and requirements. Furthermore, considering the unique circumstances of the pandemic, educators likely tailored certain requirements to online education, ensuring students had a fair and manageable workload throughout the semester.

The third hypothesis investigated the level of support teachers provided to students. We assumed that students' heightened need for support and motivation (Sepahi et al., 2021) would manifest in higher scores on the corresponding SET question. This assumption was not supported by the data; despite statistically significant differences in "teacher's professional, motivating and supportive attitude," the practical difference was negligible. The data suggest that students' evaluation of teacher support remained similarly high across the pre-COVID, COVID, and post-COVID periods. While it is possible that students' need for support and motivation did not change considerably during the COVID-19 period, it seems more probable that teachers were able to provide the required amount of support just as effectively as before and after the pandemic.

The final hypothesis focused on the Interactivity of Practice courses, as the transition of these courses may have been particularly challenging (Means et al., 2020), potentially leading to decreased SET evaluations. The analysis established a statistically significant, yet negligible, difference regarding the Interactivity of Practice courses. It is important to note that interactivity does not solely rely on in-person education, as online platforms offer a variety of tools to ensure collaboration and interactivity during lessons. This novelty, brought about by the pandemic, could have factored into the evaluation of interactivity.

Contrary to LeBlanc (2021), we did not find a deterioration in SET scores; however, our results align with the relatively high overall evaluation of the selected SET questions observed in his study. Boysen (2023) and Fullerton et al. (2024) reported significant

TABLE 6 Hypothesis and results.

Hypothesis	Corresponding SET question	Predicted direction	Outcome
H1	Course structure	Decreases in COVID-19 period	Not supported
H2	Feasibility of course requirements	Decreases in COVID-19 period	Not supported
H3	Teacher's professional, motivating and supportive attitude	Increases in COVID-19 period	Not supported
H4	Interactivity	Decreases in COVID-19 period	Not supported

differences, with the COVID-19 period receiving higher SET scores, but these effects were also marginal. A key novelty of this research is the inclusion of the post-COVID period in the analysis, providing a more comprehensive context for comparison. Our results support the conclusion that no meaningful negative change can be detected in SET scores during the COVID-19 pandemic.

We acknowledge limitations in the data collection process, as this secondary analysis relied on an anonymized dataset provided by the university. To comply with data protection guidelines, student identities and feedback were anonymized to the extent that only the number of survey responses was available. Consequently, no information was accessible regarding response rates, gender, age, or other sample characteristics. Incentives offered to students for completing SET surveys may distort responses; however, they can also increase the overall volume of evaluable data (Lukáts et al., 2023). The use of a four-point Likert scale may polarize student responses, potentially affecting the results. Further limitation is that outcomes are single-item measures, so internal-consistency reliability is not applicable. Although the validity of using the mean value as an indicator has been questioned (Fullerton et al., 2024), our analysis appears to support contemporary findings on the subject. While limitations exist, the volume of available data provides a sufficient basis for interpreting SET scores across the key time periods.

Our research aimed to provide factual data, based on a large sample size, regarding the quality of education before, during, and after the COVID-19 pandemic. Our objective was to understand how students experienced education during these periods and to contribute to the academic literature concerning their feedback. By building on results from a sufficient sample size and providing contextual information, we aimed to offer useful insights despite the controversy surrounding the value of SET questionnaires.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institute Research Ethics Committee of the Institute of Psychology at Károli Gáspár University of the Reformed Church in Hungary. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

MP: Writing – review & editing, Writing – original draft, Formal analysis. ST: Writing – original draft, Writing – review & editing, Formal analysis, Data curation, Methodology. RT: Data curation, Formal analysis, Methodology, Writing – review & editing, Conceptualization, Supervision, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

References

- Abu Talib, M., Bettayeb, A. M., and Omer, R. I. (2021). Analytical study on the impact of technology in higher education during the age of COVID-19: systematic literature review. *Educ. Inf. Technol.* 26, 6719–6746. doi: 10.1007/s10639-021-10507-1
- Adedoyin, O. B., and Soykan, E. (2020). Covid-19 pandemic and online learning: the challenges and opportunities. *Interact. Learn. Environ.* 31, 863–875. doi: 10.1080/10494820.2020.1813180
- Ahmed, N., Nandi, D., and Zaman, A. G. M. (2022). Analyzing student evaluations of teaching in a completely online environment. *Int. J. Modern Educ. Comput. Sci.* 14:13. doi: 10.5815/ijmecs.2022.06.02
- Alam, M., and Singh, P. (2021). Performance feedback interviews as affective events: an exploration of the impact of emotion regulation of negative performance feedback on supervisor–employee dyads. *Hum. Resour. Manag. Rev.* 31:100740. doi: 10.1016/j.hrmr.2019.100740
- Ali, A., Crawford, J., Cejnar, L., Harman, K., and Sim, K. N. (2021). What student evaluations are not: scholarship of teaching and learning using student evaluations. *J. Univ. Teach. Learn. Pract.* 18. doi: 10.53761/1.18.8.1
- Angelova, M. (2020). Students' attitudes to the online university course of management in the context of COVID-19. *Int. J. Technol. Educ. Sci.* 4, 283–292. doi: 10.46328/ijtes.v4i4.111
- Arnold, I. J. M., and Versluis, I. (2019). The influence of cultural values and nationality on student evaluation of teaching. *Int. J. Educ. Res.* 98, 13–24. doi: 10.1016/j.ijer.2019.08.009
- Barrow, M., and Grant, B. M. (2016). Changing mechanisms of governmentality? Academic development in New Zealand and student evaluations of teaching. higher education. *Int. J. High. Educ. Res.* 72, 589–601. doi: 10.1007/s10734-015-9965-8
- Baxter, E. P. (1991). The teval experience, 1983–88: the impact of a student evaluation of teaching scheme on university teachers. *Stud. High. Educ.* 16, 151–178. doi: 10.1080/03075079112331382954
- Berezvai, Z., Lukáts, G., and Molontay, R. (2020). Can professors buy better evaluation with lenient grading? The effect of grade inflation on student evaluation of teaching. *Assess. Eval. High. Educ.* 46, 1–16. doi: 10.1080/02602938.2020.1821866
- Berezvai, Z., Lukáts, G. D., and Molontay, R. (2019). A pénzügyi ösztönzők hatása az egyetemi oktatók osztályozási gyakorlatára. *Közgazdasági Szemle* 66, 733–750. doi: 10.18414/KSZ.2019.7-8.733
- Berniak-Wozny, J., Rataj, M., and Plebańska, M. (2021). The impact of learning mode on student satisfaction with teaching quality: evaluation of academic staff teaching before and during COVID-19. *Educ. Sci.* 11:643. doi: 10.35808/ersj/2497
- Boring, A., and Ottoboni, K. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Sci. Open Res.* doi: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Boyle, F., and Cook, E. J. (2023). Developmental evaluation of teaching quality: evidencing practice. *J. Univ. Teach. Learn. Pract.* 20. doi: 10.53761/1.20.01.11
- Boysen, G. A. (2023). Student evaluations of teaching during the COVID-19 pandemic. *Scholarsh. Teach. Learn. Psychol.* 9, 254–263. doi: 10.1037/stl0000222
- Campos, E., Daruich, S. D. N., de la O, J. F. E., Castaño, R., Escamilla, J., and Hosseini, S. (2022). Educational model transition: Student evaluation of teaching amid the COVID-19 pandemic. *Front. Educ.* 7. doi: 10.3389/feduc.2022.991654
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Appl. Psychol. Meas.* 18, 205–215. doi: 10.1177/014662169401800302
- Charytanowicz, M., Zola, M., and Suszyński, W. (2022). The impact of the COVID-19 pandemic on higher education: assessment of student performance in computer science. *PLOS ONE* 17:e0273135. doi: 10.1371/journal.pone.0305763
- Clayson, D. (2009). Student evaluations of teaching: are they related to what students learn? *J. Market. Educ.* 31, 16–30. doi: 10.1177/0273475308324086
- Clayson, D. (2022). The student evaluation of teaching and likability: what the evaluations actually measure. *Assess. Eval. High. Educ.* 47, 313–326. doi: 10.1080/02602938.2021.1909702
- Cook, C., Jones, J., and Al-Twal, A. (2022). Validity and fairness of utilising student evaluation of teaching (SET) as a primary performance measure. *J. Furth. High. Educ.* 46, 172–184. doi: 10.1080/0309877X.2021.1895093
- Daniel, S. J. (2020). Education and the COVID-19 pandemic. *Prospects* 49, 91–96. doi: 10.1007/s11125-020-09464-3
- Esarey, J., and Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assess. Eval. High. Educ.* 45, 1106–1120. doi: 10.1080/02602938.2020.1724875

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Feistauer, D., and Richter, T. (2018). Validity of students' evaluations of teaching: biasing effects of likability and prior subject interest. *Stud. Educ. Eval.* 59, 168–178. doi: 10.1016/j.stueduc.2018.07.009
- Field, A. (2017). *Discovering Statistics using IBM SPSS Statistics (North American Ed.)*. SAGE Publications.
- Fullerton, J. A., Kendrick, A., and Schoeneman Jr., J. P. (2024). A “COVID bump” in communication course evaluations: implications for future assessment. *J. Mark. Commun.* 30, 221–241. doi: 10.1080/13527266.2023.2273537
- Funder, D. C., and Ozer, D. J. (2019). Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* 2, 156–168. doi: 10.1177/2515245919847202
- Garris, C. P., and Fleck, B. (2022). Student evaluations of transitioned-online courses during the COVID-19 pandemic. *Scholarsh. Teach. Learn. Psychol.* 8, 119–139. doi: 10.1037/stl0000229
- Golding, C., and Adam, L. (2016). Evaluate to improve: useful approaches to student evaluation. *Assess. Eval. High. Educ.* 41, 1–14. doi: 10.1080/02602938.2014.976810
- Goos, M., and Salomons, A. (2017). Measuring teaching quality in higher education: assessing selection bias in course evaluations. *Res. High. Educ.* 58, 341–364. doi: 10.1007/s11162-016-9429-8
- Guo, S. (2020). Synchronous vs asynchronous online teaching of physics during the COVID-19 pandemic. *Phys. Educ.* 55:065007. doi: 10.1088/1361-6552/ab1c5
- Jin, Y. Q., Lin, C. L., Zhao, Q., Yu, S. W., and Su, Y. S. (2021). A study on traditional teaching method transferring to e-learning under the Covid-19 pandemic: from Chinese students' perspectives. *Front. Psychol.* 12. doi: 10.3389/fpsyg.2021.632787
- Kaziné, A. (2021). Online vagy hagyományos tantermi oktatás? Hallgatói elégedettség kérdőíves felmérése egy rendkívüli oktatási félévről. *Education* 30, 508–514. doi: 10.1556/2063.30.2021.3.10
- Kreitzer, R. J., and Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: a review of measurement and equity bias in sets and recommendations for ethical reform. *J. Acad. Ethics* 20, 73–84. doi: 10.1007/s10805-021-09400-w
- Lakeman, R., Coutts, R., Hutchinson, M., Massey, D., Nasrawi, D., Fielden, J., et al. (2023). Playing the SET game: how teachers view the impact of student evaluation on the experience of teaching and learning. *Assess. Eval. High. Educ.* 48, 749–759. doi: 10.1080/02602938.2022.2126430
- LeBlanc, H. P. (2021). COVID-19 effects on communication course and faculty evaluations. *J. Mass Commun. Educ.* 76, 469–476. doi: 10.1177/10776958211034116
- Lin, P.-H., Huang, L.-R., and Lin, S.-H. (2022). Why teaching innovation matters: evidence from a pre- vs peri-COVID-19 pandemic comparison of student evaluation data. *Front. Psychol.* 13. doi: 10.3389/fpsyg.2022.963953
- Linse, A. R. (2017). Interpreting and using student ratings data: guidance for faculty serving as administrators and on evaluation committees. *Stud. Educ. Eval.* 54, 94–106. doi: 10.1016/j.stueduc.2016.12.004
- Lukáts, G. D., Berezvai, Z., and Molontay, R. (2023). Assessing the effects of a reformed system of student evaluation of teaching. *Periodica Polytechnica Soc. Manag. Sci.* 31, 164–177. doi: 10.3311/PPSo.18817
- Means, B., Peters, V., Neisler, J., and Griffiths, R. (2020). *STEM Courses during the COVID Pandemic: Lessons from Spring 2020*. Digital Promise Global.
- Murray, D., Boothby, C., Zhao, H., Minik, V., Bérubé, N., Larivière, V., et al. (2020). Exploring the personal and professional factors associated with student evaluations of tenure-track faculty. *PLOS ONE* 15:e0233515. doi: 10.1371/journal.pone.0233515
- Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *Int. J. Acad. Dev.* 2, 8–23. doi: 10.1080/1360144970020102
- Nowell, C. (2007). The impact of relative grade expectations on student evaluation of teaching. *Int. Rev. Econ. Educ.* 6, 42–56. doi: 10.1016/S1477-3880(15)30104-3
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assess. Eval. High. Educ.* 33, 301–314. doi: 10.1080/02602930701293231
- Park, E., and Dooris, J. (2020). Predicting student evaluations of teaching using decision tree analysis. *Assess. Eval. High. Educ.* 45, 776–793. doi: 10.1080/02602938.2019.1697798
- Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: four shortcomings of srt research. *Teach. High. Educ.* 8, 399–411. doi: 10.1080/13562510309396
- Pineda, P., and Steinhardt, I. (2023). The debate on student evaluations of teaching: global convergence confronts higher education traditions. *Teach. High. Educ.* 28, 859–879. doi: 10.1080/13562517.2020.1863351
- Reinsch, R. W., Goltz, S. M., and Hietapelto, A. B. (2020). Student evaluations and the problem of implicit bias. *J. Coll. Univ. Law* 45, 114–139. Available online at: <https://digitalcommons.mtu.edu/michigantech-p/1712>
- Reis, C. F. S., Simões, M. F. S., and Flores-Tena, M. J. (2021). Students' pre and during COVID-19 perception of higher education switch to online: an exploratory study in Portugal. *Cypriot J. Educ. Sci.* 16, 2368–2377. doi: 10.18844/cjes.v16i5.6352
- Rodrigo, C., Herbert, C., Saunders, D., Thomas, S., and Polly, P. (2022). Influence of COVID-19 restrictions on student satisfaction with undergraduate pathology teaching in an Australian university. *Front. Educ.* 7. doi: 10.3389/feduc.2022.1014906
- Safavi, S. A., Bakar, K. A., Tarmizi, R. A., and Alwi, N. H. (2013). Faculty perception of improvements to instructional practices in response to student ratings. *Educ. Assess. Eval. Account.* 25, 143–153. doi: 10.1007/s11092-013-9160-3
- Salvo-Garrido, S., Sagner-Tapia, J., Bravo-Sanzana, M., and Torralbo, C. (2022). Profiles of good teaching practices in STEM disciplines: an analysis of mixed methods of academic and assessment variables of teaching in the first cycle of civil engineering. *Front. Educ.* 7. doi: 10.3389/feduc.2022.849849
- Sánchez, T., Gilar-Corbi, R., Castejón, J.-L., Vidal, J., and León, J. (2020). Students' evaluation of teaching and their academic achievement in a higher education institution of ecuador. *Front. Psychol.* 11. doi: 10.3389/fpsyg.2020.00233
- Sepahi, V., Salari, F., Khoshay, A., and Rezaei, M. (2021). Evaluation of professors toward e-learning during COVID-19 and its associated factors from the perspective of the students of kermanshah university of medical sciences (2020). *Educ. Res. Med. Sci.* 10. doi: 10.5812/erms.111994
- Spooren, P., Brockx, B., and Mortelmans, D. (2013). On the validity of student evaluation of teaching: the state of the art. *Rev. Educ. Res.* 83, 598–642. doi: 10.3102/0034654313496870
- Spooren, P., and Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET): students' perceptions of a teaching evaluation process and their relationships with SET scores. *Stud. Educ. Eval.* 54, 43–49. doi: 10.1016/j.stueduc.2016.12.003
- Stein, S. J., Goodchild, A., Moskal, A., Terry, S., and McDonald, J. (2021). Student perceptions of student evaluations: enabling student voice and meaningful engagement. *Assess. Eval. High. Educ.* 46, 837–851. doi: 10.1080/02602938.2020.1824266
- Stonebraker, R. J., and Stone, G. S. (2015). Too old to teach? The effect of age on college and university professors. *Res. High. Educ.* 56, 793–812. doi: 10.1007/s11162-015-9374-y
- Tangelakis, K., Kelly, K., KonYu, N., and Hall, D. (2022). The impact of teaching from home during the covid-19 pandemic on the student evaluations of female academics. *J. Univ. Teach. Learn. Pract.* 19, 160–175. doi: 10.53761/1.19.1.10
- Tomczak, M., and Tomczak, E. (2014). The need to report effect size estimates revisited: an overview of some recommended measures of effect size. *Trends Sport Sci.* 1, 19–25.
- Ulker, N. (2021). How can student evaluations lead to improvement of teaching quality? A cross-national analysis. *Res. Post Comput. Educ.* 26, 19–37. doi: 10.1080/13596748.2021.1873406
- Uttl, B. (2021). “Lessons Learned from Research on Student Evaluation of Teaching in Higher Education,” in *Student Feedback on Teaching in Schools: Using Student Perceptions for the Development of Teaching and Teachers* eds. W. Rollett, H. Bijlsma, and S. Röhl (Springer Nature Switzerland AG), 237–256. doi: 10.1007/978-3-030-75150-0_15
- Uttl, B. (2023). Student Evaluation of Teaching (SET): why the emperor has no clothes and what we should do about it. *Hum. Arenas* 7. doi: 10.1007/s42087-023-00361-7
- Uttl, B., White, C. A., and Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* 54, 22–42. doi: 10.1016/j.stueduc.2016.08.007
- Warfvinge, P., Löfgreen, J., Andersson, K., Roxå, T., and Åkerman, C. (2022). The rapid transition from campus to online teaching—how are students' perception of learning experiences affected? *Eur. J. Eng. Educ.* 47, 211–229. doi: 10.1080/03043797.2021.1942794
- Wolbring, T., and Treischl, E. (2016). Selection bias in students' evaluation of teaching. *Res. High. Educ.* 57, 51–71. doi: 10.1007/s11162-015-9378-7
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teach. High. Educ.* 12, 55–76. doi: 10.1080/13562510601102131
- Zhang, L., Carter, J. R., Qian, X., Yang, S., Rujimora, J., and Wen, S. (2022). Academia's responses to crisis: a bibliometric analysis of literature on online learning in higher education during COVID-19. *Br. J. Educ. Technol.* 53, 620–646. doi: 10.1111/bjet.13191
- Zhao, L., Xu, P., Chen, Y., and Yan, S. (2022). A literature review of the research on students' evaluation of teaching in higher education. *Front. Psychol.* 13. doi: 10.3389/fpsyg.2022.1004487

Appendix

TABLE A1 Levene’s test of equality of error variances.

SET question		Levene statistic	df	<i>p</i>
Course structure	Based on mean	48,316	11	<0.001
	Based on median	85,537	11	<0.001
	Based on median and with adjusted df	85,537	11	<0.001
	Based on trimmed mean	72,205	11	<0.001
Feasibility of course requirements	Based on mean	31,361	11	<0.001
	Based on median	28,012	11	<0.001
	Based on median and with adjusted df	28,012	11	<0.001
	Based on trimmed mean	37,191	11	<0.001
Teacher’s professional, motivating and supportive attitude	Based on mean	42,712	11	<0.001
	Based on median	82,987	11	<0.001
	Based on median and with adjusted df	82,987	11	<0.001
	Based on trimmed mean	63,832	11	<0.001
Interactivity	Based on mean	32,218	2	<0.001
	Based on median	103,469	2	<0.001
	Based on median and with adjusted df	103,469	2	<0.001
	Based on trimmed mean	64,607	2	<0.001

df, degree of freedom; p, significance level.