

Protocol paper: The impact of open-ended student evaluations of teaching on university professors' psychological health and well-being. A quick review

Authors: Poe, Edgard R.

Introduction

Student satisfaction surveys regarding teaching quality have been employed for decades as a tool to measure the quality of university teaching services and to propose improvement actions. In most cases, these surveys consist of closed-ended questions with multiple levels ranging from "strongly agree" to "strongly disagree" or with different verbal anchors. These questions are typically grouped into dimensions, ultimately yielding a numerical score that is sometimes rescaled to take values between 0 and 10.

Less frequently observed is the inclusion of open-text questions where students can express in their own words their perception of teaching quality, positive aspects, or suggestions for improvement.

This research is framed within the context of Action Research in the vice-rectorate of a public university. During the 2023-24 academic year, an improvement action was implemented that consisted of introducing two open-ended questions in the Student Evaluation of Teaching (SET). One question was designed to justify why, in cases where students indicated that the professor should be congratulated for outstanding teaching performance, they should provide their reasoning. The other question asked students to indicate aspects they would improve about the course or professor. Due to complaints from university unions, this initiative was halted in the same year of implementation to avoid potential psychosocial risks for faculty.

Consequently, the need arose to conduct a quick review that would allow the identification of SET studies that focus on examining the free-text responses that students provide in evaluations, beyond numerical ratings, to analyze the emotional and psychological impact on faculty. The review aims to examine how student comments, especially negative, abusive, or non-constructive ones, affect the mental health, wellbeing, and professional confidence of academics. In case sufficient meta-analyses on the topic are identified, we will complement this work with an umbrella review.

Previous research

Student Evaluation of Teaching (SET) refers to the systematic process of collecting feedback from students about their educational experiences, instructor effectiveness, and course quality. These evaluations typically occur at the end of a course or academic term and involve students rating various aspects of teaching performance and course design through questionnaires or surveys. The main synonyms or related terms are:

- Course evaluations
- Teaching evaluations
- Student ratings of instruction
- Student feedback questionnaires
- Student assessment of teaching
- Student evaluations of educational quality
- Student ratings of teaching effectiveness
- Student course experience questionnaires
- Teaching effectiveness measures

Open-ended questions are a form of interrogative statement that cannot be answered with a simple "yes" or "no" response or with a specific piece of information, and instead require the respondent

to elaborate on their thoughts. These questions are designed to encourage a full, meaningful answer using the respondent's own knowledge, feelings, and/or opinions. They do not constrain respondents to predefined options. The main synonyms or related terms are:

- Unstructured questions
- Free-response questions
- Qualitative questions
- Unconstrained student feedback

Higher education refers to post-secondary academic instruction conducted at universities, colleges, and other tertiary institutions. It encompasses undergraduate, graduate, and professional degree programs that build upon secondary education and typically requires completion of high school or equivalent qualifications for entry. Higher education is distinguished by its focus on advanced knowledge creation, transmission, and application through teaching, research, and service. It serves multiple functions including workforce preparation, intellectual development, cultural preservation and advancement, and societal problem-solving. The main synonyms or related terms are:

- Tertiary education
- Post-secondary education
- University
- College education
- Further education
- Third-level education
- Advanced education

Research on the use of open-ended questions in SET is scarce (Aznar-Mas et al., 2023). Most SET research has focused on using results from closed-ended questions to improve teaching service quality (Barnard, 1999) (Langbein, 2008) (Marimon et al., 2019), various biases (racial, gender, or other) that condition the value of scores received by a professor or course (Dee, 2007) (Griffin, 2004) (Centra, 2003), the debate over the reliability of quantitative SETs (Balam & Shannon, 2010; Kalender, 2015; Martínez-Gómez et al., 2011), perceptions of their utility (Suárez Monzón et al., 2022) (Marsh & Roche, 1992), and the relationship between SET scores and other variables of interest such as student learning (Stark-Wroblewski et al., 2007; Weinberg et al., 2009) (Clayson et al., 2006) (Uttl et al., 2017) or research productivity (Berbegal-Mirabent et al., 2018; Stapleton & Murkison, 2001).

Specifically regarding SETs with open-ended responses, there are studies in specific contexts that demonstrate they provide useful and relevant information for management decisions and continuous improvement of teaching to enhance student learning (Aznar-Mas et al., 2021; Aznar-Mas et al., 2023).

Nevertheless, some studies have documented stress, anxiety, emotional distress, and effects on professional self-esteem, showing tension between information utility and potential harm (Cunningham et al., 2023; Lakeman et al., 2021; Lakeman et al., 2022).

Contribution

Our research question aims to analyze how student comments, especially negative, abusive, or non-constructive ones, affect the mental health, wellbeing, and professional confidence of academics, what actions can be taken to prevent these problems, particularly how to automatically detect abusive comments and how to process and filter them.

This review will explore the paradox that while open-ended comments can provide valuable feedback to improve teaching, they can also generate significant negative consequences for educators, especially when they are anonymous and unfiltered.

Methods

We propose a systematic narrative literature review (Aguinis et al., 2020; Medina-López, 2010) complemented with an umbrella review if there are previously published meta-analyses addressing the same research question as ours.

Inclusion criteria:

- Scientific articles, theses, and conference communications indexed in WOS or Scopus databases
- In any language and any year
- Whose contribution focuses on how to analyze open-ended SET questions or the effect these open-ended questions have on faculty mood
- In the context of higher education

Exclusion criteria:

- The effect on faculty due to being evaluated or the value of closed-ended question scores
- Other outcomes not relevant to this work such as scientific productivity, student learning outcomes, implemented improvements, management decisions regarding hiring, promotion, or dismissal, etc.
- Do not analyze the effect of open-ended questions
- Not in the context of university education
- Evaluation of other quality areas unrelated to faculty (facilities, staff, etc.)

The automatic search strategy will be conducted by blocks in both databases, subsequently duplicates will be merged and results exported for title and abstract screening. A human coder will be used with artificial intelligence assistance (Marin-Garcia et al., 2024). Subsequently, there will be a coding task for all screened PDFs, performed by the two authors of this article, following an extraction protocol where open coding (Frieze, 2017) will be conducted in ZOTERO and focused and selective coding will be performed with the help of Atlas-ti or Excel templates. These tools will be used to reorganize information and provide support during the scientific report writing phase of the review. The entire process will be documented following the PRISMA framework (Page et al., 2021).

Pilot

A pilot test has been conducted in WOS with this search strategy: "student* evaluat* of teach*" (Topic) and Preprint Citation Index (Exclude -- Database), generating 1,325 results. The first 563 were screened with the help of a script using the gpt-4o-mini-2024-07-18 model via API (see prompt in appendix). This classification served to rank the 593 articles, and the first author manually screened the most relevant half according to artificial intelligence results, leaving 39 selected articles. This would guarantee the feasibility of the review (pending screening of the remaining articles and adding non-duplicates that might emerge from Scopus).

To verify the feasibility of the umbrella review, we proceeded with this search strategy: "student* evaluat* of teach*" (Topic) and Preprint Citation Index (Exclude -- Database) and Review Article (Document Types), yielding 35 articles that, at the time of writing this protocol, have not yet been screened by title and abstract directly on the WOS platform.

Since the number of false positives is high, an optimal search filter will be created to allow for a more specific search by incorporating the presence of more elements from the inclusion criteria.

References

- Aguinis, H., Ramani, R. S., & Alabduljader, N. (2020). Best-Practice Recommendations for Producers, Evaluators, and Users of Methodological Literature Reviews. *Organizational Research Methods*. <https://doi.org/10.1177/1094428120943281>
- Aznar-Mas, L. E., Atarés Huerta, L., & Marin-Garcia, J. A. (2021). Students have their say: factors involved in students' perception on their engineering degree. *European Journal of Engineering Education*, 46(6), 1007-1025. <https://doi.org/10.1080/03043797.2021.1977244>
- Aznar-Mas, L. E., Atarés Huerta, L., & Marin-Garcia, J. A. (2023). Effectiveness of the use of open-ended questions in student evaluation of teaching in an engineering degree [Teaching evaluation, higher education, student satisfaction, teaching improvement, open-ended questions]. *Journal of Industrial Engineering and Management*, 16(3), 14. <https://doi.org/10.3926/jiem.5620>
- Balam, E. M., & Shannon, D. M. (2010). Student ratings of college teaching: a comparison of faculty and their students. *Assessment & Evaluation in Higher Education*, 35(2), 209-221. <https://doi.org/10.1080/02602930902795901>
- Barnard, J. (1999). Using total quality principles in business courses: The effect on student evaluations. *Business Communication Quarterly*, 62(2), 61.
- Berbegal-Mirabent, J., Mas-Machuca, M., & Marimon, F. (2018). Is research mediating the relationship between teaching experience and student satisfaction? *Studies in Higher Education*, 43(6), 973-988. <https://doi.org/10.1080/03075079.2016.1201808>
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518. <Go to ISI>://000185429900001
- Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). Grades and the Student Evaluation of Instruction: A Test of the Reciprocity Effect. *Academy of Management Learning & Education*, 5(1), 52-65. <https://doi.org/Article>
- Cunningham, S., Cathcart, A., & Graham, T. (2023). Allegations, Abuse and Discrimination: Using Student Evaluation of Teaching Surveys to Support Student and Educator Wellbeing. *Student Success*, 14(3), 92-103. <https://doi.org/10.5204/ssj.2756>
- Dee, K. C. (2007). Student perceptions of high course workloads are not associated with poor student evaluations of instructor performance. *Journal of Engineering Education*, 96(1), 69-78. <Go to ISI>://000251314000007
- Friese, S. (2017). Translating the Process of Open / Initial Coding In Grounded Theory. <http://atlasti.com/2017/07/01/gt/>,
- Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, 29(4), 410-425. <Go to ISI>://000224050800003
- Kalender, I. (2015). Reliability-Related Issues in the Context of Student Evaluations of Teaching in Higher Education. *International Journal of Higher Education*, 4(3), 44-56. <https://doi.org/http://dx.doi.org/10.5430/ijhe.v4n3p44>
- Lakeman, R., Coutts, R., Hutchinson, M., Lee, M., Massey, D., Nasrawi, D., & Fielden, J. (2021). Appearance, insults, allegations, blame and threats: an analysis of anonymous non-constructive student evaluation of teaching in Australia. *Assessment & Evaluation in Higher Education*, 47(8), 1245-1258. <https://doi.org/10.1080/02602938.2021.2012643>
- Lakeman, R., Coutts, R., Hutchinson, M., Massey, D., Nasrawi, D., Fielden, J., & Lee, M. (2022). Playing the SET game: how teachers view the impact of student evaluation on the experience of teaching and learning. *Assessment & Evaluation in Higher Education*, 48(6), 749-759. <https://doi.org/10.1080/02602938.2022.2126430>
- Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review*, 27(4), 417-428. <https://doi.org/10.1016/j.econedurev.2006.12.003>
- Marimon, F., Mas-Machuca, M., Berbegal-Mirabent, J., & Llach, J. (2019). UnivQual: a holistic scale to assess student perceptions of service quality at universities. *Total Quality Management & Business Excellence*, 30(1-2), 184-200. <https://doi.org/10.1080/14783363.2017.1302795>
- Marin-Garcia, J. A., Martinez-Tomas, J., Juarez-Tarraga, A., & Santandreu-Mascarell, C. (2024). Protocol paper: From Chaos to Order. Augmenting Manual Article Screening with Sentence Transformers in Management Systematic Reviews. *WPOM-Working Papers on Operations Management*, 15, 172-208. <https://doi.org/10.4995/wpom.22282>
- Marsh, H. W., & Roche, L. A. (1992). *The Use of Student Evaluations of University Teaching in Different Settings: The Applicability Paradigm In File*.

- Martínez-Gómez, M., Carot Sierra, J. M., Jabaloyes, J., & Zarzo, M. (2011). A multivariate method for analyzing and improving the use of student evaluation of teaching questionnaires: a case study. *Quality & Quantity*, 45(6), 1415-1427.
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <https://doi.org/10.1136/bmj.n160>
- Stapleton, R. J., & Murkison, G. (2001). Optimizing the Fairness of Student Evaluations: A Study of Correlations Between Instructor Excellence, Study Production, Learning Production, and Expected Grades. *Journal of Management Education*, 25(3), 269-291. <http://jme.sagepub.com>
- Stark-Wroblewski, K., Ahlering, R., & Brill, F. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: supplementing student evaluations of teaching with pre-post learning measures [Journal]. *Assessment & Evaluation in Higher Education*, 32(4), 403-416.
- Suárez Monzón, N., Gómez Suárez, V., & Lara Paredes, D. G. (2022). Is my opinion important in evaluating lecturers? Students' perceptions of student evaluations of teaching (SET) and their relationship to SET scores. *Educational Research and Evaluation*, 27(1-2), 117-140. <https://doi.org/10.1080/13803611.2021.2022318>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. <https://doi.org/https://doi.org/10.1016/j.stueduc.2016.08.007>
- Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating Teaching in Higher Education. *Journal of Economic Education*, 40(3), 227-261.