

Reimagining Teaching Evaluations

Nichole D. Barta, Ed.D.

Director, Center for Teaching & Advising

*Associate Professor, Kinesiology & Sport
Management Department*

Gonzaga University

Physical Fitness--Components


Fitness Component	Definition	Example Assessments
Body Composition	The proportion of fat, muscle, bone, and other tissues in the body.	Body fat percentage measurement (e.g., skinfold calipers, hydrostatic weighing).
Aerobic Fitness	The ability of the cardiovascular and respiratory systems to supply oxygen during sustained physical activity.	VO2 max test, step test, or a 1.5-mile run test.
Muscular Strength	The maximum force that a muscle or muscle group can generate.	One-repetition max (1RM) test (e.g., bench press, squat).
Muscular Endurance	The ability of a muscle or muscle group to perform repeated contractions over time.	Push-up test or sit-up test for repetitions.
Flexibility	The range of motion around a joint or group of joints.	Sit-and-reach test or Apley scratch test



Session Objectives

By the end of this session, participants will be able to:

- Discuss why evaluating teaching effectiveness requires drawing on multiple sources of evidence to reflect its multi-dimensional nature.
- Examine the strengths and limitations of commonly used methods for assessing teaching effectiveness, including student evaluations, peer review, and self-reflection.
- Envision how a holistic evaluation framework for teaching could be implemented within their unit.



Examine the strengths and limitations of commonly used methods for assessing teaching effectiveness including student evaluations, peer review, and self-reflection.

Evaluating Teaching Effectiveness Scenario:

Faculty A and Faculty B are in the same department, both applying for reappointment. As the department chair, you must evaluate their teaching effectiveness. **Performance Scale: 1.0 to 5.0:** Scale of performance for each category. **3.0:** Minimum satisfactory performance

Data Snapshot of 1 Class Across 3 Semesters:

Faculty Member	Semester	Total # of Responses	Class Size	Response Rate (%)	Q1: Instructor Support Mean	Q2: Presentation of Concepts Mean	Q3: Overall Teaching Effectiveness Mean
A	F23	7	25	28%	4.8	4.7	4.9
A	S24	6	20	30%	4.7	4.6	4.8
A	F24	7	18	39%	4.6	4.5	4.9
B	F23	14	16	87.5%	3.4	3.8	3.6
B	S24	19	22	86%	3.9	3.7	4.0
B	F24	23	25	92%	4.4	4.3	4.4

Turn and Talk:

- **Question:** Based on this data, what rating would you give Faculty Member A and Faculty Member B for **Teaching Effectiveness on their Faculty Performance Assessment**? Who would you say is more effective? Faculty Member A or B?
- **Question:** What additional questions would you need answers to assign your ratings?



My Responses Would Be

Faculty Member A: Inconclusive

Due to the very low response rates (28–39%), the data is unreliable and unrepresentative of those classes. Based on this data alone (SET scores that are in these ranges), it is impossible to evaluate teaching effectiveness confidently. Additional sources of evidence, such as peer evaluations, classroom observations, or a teaching portfolio, would be required to make an accurate assessment.

Faculty Member B: Additional Analysis Needed

While the response rates for Faculty Member B are consistently high (85–90%), making the data generally representative, additional statistics such as the mode, median, and standard deviation are necessary to make a more accurate evaluation. Understanding the distribution of scores and variability would help clarify whether the data reflects a broad consensus or is influenced by outliers. Additionally, analyzing qualitative comments could provide insight into whether students are identifying specific issues contributing to the lower scores. There is a positive growth trajectory across the questions.

Impact of Outliers on Mean Scores

Faculty Member B

Fall 2023 Class Data:

- 14 students scored: 4/5
- 2 students scored: 1/5

Impact Analysis of Q3:
With outliers: **3.6/5.0**
Without outliers: **4.0/5.0**

Note: Two low scores (1/5) (12.5% of students) reduced the class average by 0.375 points, 87.5% of students gave high scores (4/5).

Additional Statistics: Mode= 4, Median = 4, Standard Deviation= 0.992

This relatively high standard deviation (nearly 1 point on a 5-point scale) indicates a significant spread in the scores despite most scores being 4. This is due to the impact of the two outlier scores of 1

Influence of Response Rates on SET Reliability

Low response rates increase the margin of error and reduce the reliability of SETs, especially in small classes (James et al., 2015)

Recommended response rates based on class size:

- **<30 students: ≥80%**
- **31–50 students: 66%–75%**
- **>50 students: ≥50%** (minimally acceptable, but higher rates reduce sampling error)

(Al Kuwaiti et al., 2016; James et al., 2015)

Selection Bias Effects

- **Upward Bias:** Students who are satisfied may be more likely to respond (Goos & Salomons, 2016).
- **Downward Bias:** Dissatisfied students may be more likely to respond (Luo, 2020).

Implication: Decisions based on SETs with low response rates risk misinterpretation and unreliable conclusions about teaching effectiveness. Institutions must consider response rates and class sizes to ensure appropriate SET data use.



Influence of Response Rates on SET Reliability

- **<20 students:** Requires a **58% response rate or higher** for reliable results.
- **>30 students:** Response rates as low as **47%** are **only just sufficient** to ensure meaningful interpretation (Nulty, 2008).

If response rates do not meet these thresholds, the data is unreliable—not only for guiding course improvements but especially for making personnel decisions, as it risks being biased and unrepresentative.

As academics, we place great importance on peer review and reliable data to ensure the credibility of our research and scholarship. Yet, we often fail to uphold the same standards when using data (SETs) to evaluate teaching effectiveness. Using unreliable data for high-stakes decisions directly contradicts the principles we strive to uphold in scholarly work.



SETs Are Influenced by Factors Unrelated to Teaching Quality

Bias occurs when factors unrelated to actual teaching quality affect evaluations (Centra & Gaubatz, 2000).

Female and marginalized groups are disproportionately affected by bias, leading to lower ratings (Adams et al., 2022; Kreitzer & Sweet-Cushman, 2022; MacNeil et al., 2015; Mitchell & Martin, 2018; Heffernan, 2022; Wallace et al., 2019).

Bias examples:

- **Professional Titles and Approachability:** Female professors addressed by their titles are perceived as less approachable, reflecting a gendered double bind between status and approachability (Takiff et al., 2001.) Similarly, Black male instructors received lower evaluations when they preferred to be addressed by their professional title rather than their first name (Foster, 2023).
- **Attire and Likability:** Academics dressed in casual attire are rated as less likable than those in business casual or professional attire (Chatelain, 2015).

SET Scores: Unintended Consequences

“Teachers who are more effective in promoting future performance receive worse evaluations from their students” (Braga, et al, 2014, p. 81).

Active learning and "desirable difficulties" enhance long-term knowledge but often result in lower SET scores as students misinterpret the cognitive effort required as poor teaching (Deslauriers et al., 2011; Kornell & Hausman, 2016).

Courses with lenient grading receive higher evaluations, yet students invest less time studying in these courses (Greenwald & Gillmore, 1997b; Babcock, 2010).

“Teachers should serve their students broccoli, but they tend to get higher ratings when they serve chocolate, and this is not just an analogy—one study showed that ratings increased when teachers literally served their students chocolate” (Kornell & Hausman, 2016)

Implication: Heavy reliance on SET scores may inadvertently encourage instructors to adapt their teaching to align with student preferences, sometimes at the expense of evidence-based practices. This can unintentionally lead to less demanding coursework, grade inflation, and a focus on student satisfaction rather than long-term learning outcomes (Strobe, 2020).



Mismatch Between Numeric Scores and Comments

Gonzaga University Study: In a study evaluating our Course Evaluation Survey, we examined responses to the scaled question: *“Approachability of the Instructor Outside of Class.”*

While analyzing a random sample of 100 responses, we observed a small inconsistency: 5% of students gave high numeric ratings on a scale of 1-7 but included negative comments about the instructor’s approachability in the open-ended sections.

Possible Causes:

This minor discrepancy may highlight areas for improving alignment between numeric and written feedback:

- 1. Student Inattention:** Some students may quickly select a high score but then express criticisms in their written responses.
- 2. Differing Interpretations of the Rating Scale:** With limited anchors, one student’s “5” could represent another’s “3” or “6,” leading to inconsistent interpretations.
- 3. Social Desirability Bias:** Students may hesitate to provide “punitive” low ratings but feel more comfortable voicing critical feedback in the written sections. *(Note: some students said in our study that their professors asked them not to rate them low)*



Mismatch Between Numeric Scores and Comments

Implications

- **Validity Concerns:** Numeric ratings alone may not fully capture students' perspectives, leading to potential misinterpretations of what works well and what needs improvement.
- **Qualitative Alignment:** Open-ended comments should confirm, not contradict, numeric ratings. Discrepancies between the two may signal issues with how students interpret the rating scale or the questions.
- **Training for Accurate Feedback:** Educating students on how to provide specific and actionable feedback can enhance the reliability and usefulness of both numeric and written responses.

Key Takeaway: Use a mixed-methods approach to reviewing student feedback. Combining coding of qualitative comments with appropriate statistical analysis of quantitative data provides a more accurate understanding of student perceptions.



Reflection and Discussion

1. Reflect on the response rates of your own evaluations. How do you think response rates have impacted the reliability of your SET results for decision-making purposes?

2. Discuss-Turn and Talk to Your Neighbor:

- How might this information about SETs shape how we think about teaching evaluations in our unit?
- What questions or concerns does this information raise for you regarding the use of SETs in evaluating teaching?

Student Feedback: A Valuable Perspective

While SETs have limitations, they are a critical source for understanding and improving teaching effectiveness. Students provide unique insights as the ones directly experiencing the teaching and learning process. Moreover, students are our “**why**” in higher education. Helping them learn is central to our mission, and understanding how our actions influence their learning is essential for becoming more effective educators.

Student Feedback Can Highlight:

- **Clarity of Instruction:** How well course objectives, expectations, and materials are communicated.
- **Engagement and Motivation:** Whether teaching strategies inspire interest, active participation, and curiosity.
- **Inclusivity and Classroom Dynamics:** perceptions of whether the learning environment feels equitable, supportive, and respectful.
- **Support and Responsiveness:** How accessible and approachable the instructor is, and how effectively they address questions or offer additional help.
- **Relevance of Content:** How well students feel the material connects to real-world applications and their academic and personal goals.



Peer Review: Expert Insights

Peer review offers expert insights into the intersection of disciplinary knowledge, pedagogy, and the learning environment.

Peer Review Highlights:

- **Subject Matter Expertise:** Observations on how effectively the instructor translates complex disciplinary knowledge into understandable and engaging content.
- **Pedagogical Alignment:** Feedback on whether teaching methods align with best practices in the discipline and evidence-based strategies.
- **Learning Environment:** Observation of equitable participation strategies, active classroom interactions, and whether students engage in higher-order thinking and ask meaningful questions.
- **Objective Observation:** An external lens that identifies elements of instructional delivery that support student learning (e.g., pacing, time management, adaptive responses, facilitation of discussion).



Limitations of Peer Review

Key Limitations

- **Snapshot Nature:** Observations provide only a brief view of teaching, capturing isolated sessions rather than long-term practices or instructional trends.
- **Bias Risks:** Evaluations may be shaped by subjective standards, personal relationships, or power dynamics. Clear criteria and standardized protocols are essential to minimize bias. (e.g., pre and post-observation discussions, common templates)
- **Collegiality Challenges:** Personal relationships can make reviewers hesitant to provide honest, constructive feedback, fearing it may harm professional relationships. This hesitancy may result in overly positive evaluations or avoidance of critical issues.

(Zeng, 2020)



Value of Self-Reflection

Self-reflection provides instructor insights into teaching practices, challenges, and the rationale behind instructional choices. It enables educators to take ownership of their professional development by fostering intentional and evidence-based approaches to teaching.

Key Values of Self-Reflection:

- **Intentionality in Instructional Choices:** Encourages educators to reflect on their teaching rationale, drawing on evidence from the literature and applying scholarly practices to improve learning outcomes.
- **Context-Specific Successes:** Provides a better understanding of teaching challenges and achievements within the unique dynamics of the discipline and class formats—teaching effectiveness is contextualized.
- **Commitment to Continuous Improvement:** Encourages educators to analyze their practices critically, refine strategies, and describe changes implemented to demonstrate teaching effectiveness over time.



Limitations of Self-Reflection

Key Limitations:

- **Inconsistent Quality:** The effectiveness of self-reflection depends heavily on the instructor's ability to critically evaluate their practices. Without proper training or structured guidance, reflections may lack depth, overlook critical evidence, and miss opportunities to draw meaningful connections or identify impactful strategies for growth.
- **Time-Consuming:** Meaningful self-reflection requires time and focus. Competing priorities may lead instructors to rush the process or perform it superficially, reducing its value.
- **Fear of Honesty:** Instructors may hesitate to fully acknowledge challenges or areas for growth, fearing self-criticism or potential implications for their professional identity.
- **Confirmation Bias:** Tendency to seek out evidence that supports existing beliefs or practices while disregarding contradictory evidence. This can lead instructors to justify ineffective practices instead of critically evaluating them.



From Limitations to Solutions

Recognizing the Flaws: Overreliance on a single metric or perspective can yield inaccurate or biased results.

A Potential Solution: Adopt a holistic evaluation approach to capture the multidimensional nature of teaching effectiveness.

Discuss why evaluating teaching effectiveness requires drawing on multiple sources of evidence to reflect its multi-dimensional nature.



Key Elements of Holistic Evaluation

Multiple Perspectives: Incorporates feedback from students, peers, and self as evidence.

Diverse Data Sources: Uses a mix of quantitative and qualitative data.

Contextual Relevance: Allows the evaluation process to reflect the unique discipline practices, course format, and learning goals within the teaching environment.

Benefits

- Promotes equity by reducing bias from over-reliance on a single perspective.
- Encourages a more comprehensive understanding of teaching effectiveness.



Triangulation: A Foundation for Holistic Teaching Evaluation

What is Triangulation? A process that combines multiple methods, perspectives, or sources of data to develop a comprehensive understanding of a phenomenon. Initially used in navigation, triangulation involves using multiple reference points to locate or assess something more accurately (Arias Valencia, 2022). It ensures complex phenomena, like teaching effectiveness, are approached from various angles to account for different dimensions and contexts.

How It Relates to Holistic Evaluation

- Holistic evaluation of teaching applies **data source triangulation** by incorporating:
- **Student Feedback:** Perspectives on how course elements and instructor support influenced their learning.
- **Peer Observations:** Insights into how the delivery of instruction influences student learning and engagement, as well as reviews of course design and the use of subject matter expertise in lessons.
- **Self-Reflection:** Rationale for instructional approaches, a critical examination of outcomes, and plans for improvement.
- This approach uncovers **patterns** and **contradictions** across sources, offering a more accurate picture of teaching effectiveness.



Analyzing the Faculty Handbook

What does Bradley's Faculty Handbook have to say about a holistic evaluation of teaching?

Objective: Envision how a holistic evaluation framework for teaching could be implemented within their unit.



Analyzing the Faculty Handbook

Among the three areas, the highest priority is excellent teaching. Toward this end, faculty shall engage in the study and preparation necessary to demonstrate successful teaching. Teaching is complemented and enhanced by research and/or creative production.

The **criteria** to evaluate professorial faculty are the following:

1a) Teaching effectiveness: a) Successful teaching performance in the classroom, lab, studio, or other instructional sites; b) Regular and substantial investment of time in study to enhance one's knowledge of the field(s) taught; c) Conscientious preparation for instruction; d) Currency and innovation in pedagogy, course development, and course revision; e) Helpfulness and accessibility to students, both in and outside the classroom, such as in reviewing, counseling, or advising; f) Adequacy of instructional materials and their use; g) Critical self-evaluation;



Analyzing the Faculty Handbook

Bradley's Faculty Handbook implies that faculty will be evaluated using a holistic approach, as the criteria for teaching effectiveness are clearly multi-dimensional.

For instance, while SETs can provide relevant evidence for 'helpfulness and accessibility to students, both in and outside the classroom,' but they are not appropriate evidence for other criteria, such as 'critical self-evaluation' or the 'regular and substantial investment of time in study to enhance one's knowledge of the field(s) taught.' This reinforces the need for diverse sources of evidence to evaluate teaching effectiveness as defined in the handbook.

Dimensions of the Handbook

Dimension	FHB Criteria Count	FHB Criteria	My Interpretation
Reflective Practices/Continuous Improvement	3	<ul style="list-style-type: none"> - b) Regular and substantial investment of time in study to enhance one's knowledge of the field(s) taught. - d) Currency and innovation in pedagogy, course development, and course revision. - g) Critical self-evaluation. 	Faculty engage in ongoing professional development, revise practices based on new trends and research, and critically reflect on their teaching to improve student learning outcomes.
Instructional Design/Planning	2	<ul style="list-style-type: none"> - c) Conscientious preparation for instruction. - f) Adequacy of instructional materials and their use. 	Faculty design lessons and courses that align learning outcomes, activities, and assessments. They integrate evidence-based practices and ensure materials are up-to-date, accessible, and support learning outcomes.
Instructional Delivery	1	<ul style="list-style-type: none"> - a) Successful teaching performance in the classroom, lab, studio, or other instructional sites. 	Faculty use learner-centered strategies that actively engage students and guide them toward achieving the intended learning outcomes.
Student Support	1	<ul style="list-style-type: none"> - e) Helpfulness and accessibility to students, both in and outside the classroom, such as in reviewing, counseling, or advising. 	Faculty are accessible to students, provide timely and constructive feedback, and foster academic and personal growth in a supportive environment.

Dimension: Reflective Practices/Continuous Improvement

FHB Criterion:

- Regular and substantial investment of time in study to enhance one's knowledge of the field(s) taught. (b)
- Currency and innovation in pedagogy, course development, and course revision.(d)
- Critical self-evaluation.(g)

Measurable Indicators:

- Faculty demonstrate professional growth by engaging in ongoing development and applying new knowledge to teaching.
- Faculty regularly revise courses based on current trends, research, and student feedback to enhance teaching and learning.
- Faculty critically analyze teaching practices and implement targeted strategies for improvement.

Sources of Evidence:

- Professional development records (e.g., workshops, certifications).
- Reflective statements detailing changes in teaching practices.
- Course materials are revised to incorporate new knowledge or methodologies.



Dimension: Reflective Practices/Continuous Improvement

Performance Descriptors:

- **Unsatisfactory:** Limited or no engagement in professional development activities. Reflective statements are absent or fail to identify areas for growth. No evidence of updated course materials or applied new knowledge.
- **Satisfactory:** Regular participation in relevant professional development. Reflective statements demonstrate some analysis of teaching practices and identification of areas for improvement. Evidence of updated course materials incorporating new knowledge or practices.
- **Outstanding:** Proactive and frequent engagement in professional development activities clearly aligned with teaching goals. Reflective statements demonstrate deep analysis of practices, implementation of significant changes, and clear benefits for student learning. Course materials are frequently updated and demonstrate innovative application of current knowledge.

Faculty Performance Assessment

Teaching Effectiveness Evaluation: Summative Scoring

Use the table to assign a score for each dimension of teaching effectiveness. Provide comments that support your scoring, including strengths, areas of opportunity, and patterns observed in the evidence. Calculate the average score across all dimensions.

Dimension	Score (1–5)	Comments on Strengths	Comments on Areas of Opportunity
Instructional Delivery			
Reflective Practices/Continuous Improvement			
Instructional Design/Planning			
Student Support			
Average Score Across All Dimensions			



A Holistic Evaluation of Teaching Approach in Your Unit

Objective: Envision how a holistic evaluation framework for teaching could be implemented within their unit.

Reflection and Discussion:

- **Current Practices:** What dimensions of teaching are already evaluated in your unit, and how well do they align with a holistic framework?
- **Identifying Gaps:** Are there any gaps in the current evaluation process that need to be addressed?
- **Anticipating Challenges:** What challenges might arise in implementing a holistic evaluation framework, and how could they be addressed?



Q&A

Key References

- Adams, M. J. D., & Umbach, P. D. (2012). Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments. *Research in Higher Education*, 53(5), 576–591. <https://doi.org/10.1007/s11162-011-9240-5>
- Al Kuwaiti, A., AlQuraan, M., & Subbarayalu, A. V. (2016). Understanding the effect of response rate and class size interaction on students evaluation of teaching in a higher education. *Cogent Education*, 3(1), 1204082. <https://doi.org/10.1080/2331186X.2016.1204082>
- Arias Valencia MM. (n.d.). *Principles, Scope, and Limitations of the Methodological Triangulation*.
- Benton, S. L. (n.d.). *Best Practices in the Evaluation of Teaching*.
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020a). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, 9(2), 137–151. <https://doi.org/10.1016/j.jarmac.2019.12.009>
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020b). On students' (mis)judgments of learning and teaching effectiveness: Where we stand and how to move forward. *Journal of Applied Research in Memory and Cognition*, 9(2), 181–185. <https://doi.org/10.1016/j.jarmac.2020.04.003>
- Chatelain, A. M. (2015). The effect of academics' dress and gender on student perceptions of instructor approachability and likeability. *Journal of Higher Education Policy and Management*, 37(4), 413–423. <https://doi.org/10.1080/1360080X.2015.1056598>
- Chen, Y. (2023). Does students' evaluation of teaching improve teaching quality? Improvement versus the reversal effect. *Assessment & Evaluation in Higher Education*, 48(8), 1195–1207. <https://doi.org/10.1080/02602938.2023.2177252>
- Foster, M. M. (2023). Instructor Name Preference and Student Evaluations of Instruction. *PS: Political Science & Politics*, 56(1), 143–149. <https://doi.org/10.1017/S1049096522001068>
- Goos, M., & Salomons, A. (2017). Measuring teaching quality in higher education: Assessing selection bias in course evaluations. *Research in Higher Education*, 58(4), 341–364. <https://doi.org/10.1007/s11162-016-9429-8>
- Hunukumbure, A. D., Horner, P. J., Fox, J., & Thakerar, V. (2021). An online discussion between students and teachers: A way forward for meaningful teacher feedback? *BMC Medical Education*, 21(1), 289. <https://doi.org/10.1186/s12909-021-02730-8>
- James, D. E., Schraw, G., & Kuch, F. (2015). Using the sampling margin of error to assess the interpretative validity of student evaluations of teaching. *Assessment & Evaluation in Higher Education*, 40(8), 1123–1141. <https://doi.org/10.1080/02602938.2014.972338>

Key References

- Kornell, N., & Hausman, H. (2016). Do the Best Teachers Get the Best Ratings? *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00570>
- Lilly, J. D., Wipawayangkool, K., & Pass, M. (2022). Teaching Evaluations and Student Grades: That's Not Fair! *Journal of Management Education*, 46(6), 994–1023. <https://doi.org/10.1177/10525629221084338>
- Liu, S., Keeley, J. W., & Buskist, W. (2020). What Constitutes Poor Teaching? Perspectives From Chinese College Students. *Teaching of Psychology*, 47(1), 58–62. <https://doi.org/10.1177/0098628319888091>
- Luo, M. N. (2020a). Student Response Rate and Its Impact on Quantitative Evaluation of Faculty Teaching. *The Advocate*, 25(2). <https://doi.org/10.4148/2637-4552.1137>
- Luo, M. N. (2020b). Student Response Rate and Its Impact on Quantitative Evaluation of Faculty Teaching. *The Advocate*, 25(2). <https://doi.org/10.4148/2637-4552.1137>
- Miguel, C. V., Moreira, C., Alves, M. A., Campos, J. B. L. M., Glassey, J., Schaer, E., Kockmann, N., Porjazoska Kujundziski, A., Polakovic, M., & Madeira, L. M. (2019). Developing a framework for assessing teaching effectiveness in higher education. *Education for Chemical Engineers*, 29, 21–28. <https://doi.org/10.1016/j.ece.2019.06.001>
- Myllykoski-Laine, S., Postareff, L., Murtonen, M., & Vilppu, H. (2023). Building a framework of a supportive pedagogical culture for teaching and pedagogical development in higher education. *Higher Education*, 85(4), 937–955. <https://doi.org/10.1007/s10734-022-00873-1>
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, 33(3), 301–314. <https://doi.org/10.1080/02602930701293231>
- Reverter, A., Martinez, C., Currey, P., Van Bommel, S., & Hudson, N. J. (2020). Unravelling student evaluations of courses and teachers. *Cogent Education*, 7(1), 1771830. <https://doi.org/10.1080/2331186X.2020.1771830>
- 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(4), 276–294. <https://doi.org/10.1080/01973533.2020.1756817>
- Takiff, H. A., Sanchez, D. T., & Stewart, T. L. (2001). What's in a Name? The Status Implications of Students' Terms of Address for Male and Female Professors. *Psychology of Women Quarterly*, 25(2), 134–144. <https://doi.org/10.1111/1471-6402.00015>
- Zeng, L. M. (2020). Peer review of teaching in higher education: A systematic review of its impact on the professional development of university teachers from the teaching expertise perspective. *Educational Research Review*, 31, 100333. <https://doi.org/10.1016/j.edurev.2020.100333>