AI DECISIONS — Author Templates (DOCX) & Manuscript Guide

Purpose. This pack provides the exact text and structure to create two DOCX files for submission under **double-blind** review, plus formatting, declarations and examples.

---

1) Submission file (required)

- **Manuscript (DOCX)** — a single file for review. Include **title, authors and affiliations** in the document; the editorial office will **create the anonymised copy** for double-blind review.
- **Optional ZIP** — figures/supplementary.

Author identities and full metadata are captured in the submission form (no separate Title Page file). You **do not** need to anonymise the manuscript yourself; we will do this internally.

Required sections (single file): - Title - Authors and affiliations (ORCID may be provided in the form) - Abstract (150–250 words); Keywords (3–6) - 1. Introduction and decision problem - 2. Related work and design rationale - 3. Methods: data, models, interfaces, human–AI protocols - 4. Evaluation design (e.g., RCT/A-B/DiD/ITS); outcomes and power - 5. Results (decision-level metrics) - 6. Operations & governance (monitoring, escalation, audit) - 7. External validity and generalisation - 8. Limitations and risks - References (author–year; include DOIs where available)

Brief **Data/Code availability** notes at the end are optional at submission and will be finalised after acceptance.

---

2) Formatting & layout

- **A4**, margins **2.5 cm**; **Times New Roman 12 pt**; line spacing **1.15**.
- Use Word styles **Heading 1–3**; avoid numbering beyond **H2**.
- **Tables:** create as Word tables (not images). Caption **above** the table.
- **Figures:** 300 dpi (vector preferred). Caption **below** the figure.
- **Footnotes:** keep to a minimum; endnotes only if necessary.
- **Headers/footers:** optional — the editorial office will remove identifiers during anonymisation.

---

3) Citations & references (author–year)

- Use **author–year** in-text citations (Harvard/APA style): e.g., *Smith (2023)* or *(Smith & Lee, 2023)*.
- Include **DOI links** for references where available.
- Examples (APA-like):
  - Article: *Smith, J. A., & Jones, M. B. (2022).* Title. *Journal Name, 15*(3), 45–58. https://doi.org/10.1234/jn.2022.015
  - Conference: *Green, P. R. (2021).* Title. In *Proceedings of …* (pp. 123–130). https://doi.org/10.1007/978-3-030-12345-6_12
  - Dataset/Software: *Doe, J. (2024).* Title (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1234567

---

4) Required declarations (collected in the submission form)

You will fill these items **in the submission form** — do **not** attach separate documents: - **Keywords (3–6)** and **Submission domain(s)** - **Data/Code availability plan** (links or justified restriction) - **Ethics** (IRB/IEC if applicable; GDPR/UK GDPR de-identification for any human/clinical data) - **Disclosures** (funding; competing interests; AI-tool use; dual-use/misuse risks if relevant) - **High-stakes declaration** (healthcare/finance/government etc.; short checklist via form) - **Preprint DOI/URL** (optional) - **ORCID** for the corresponding author (required); **ROR** for affiliations (optional)

> If you prefer, you may also include brief **Data/Code availability** notes at the end of your manuscript — this is **optional at submission** and will be finalised after acceptance.

---

5) Figures, tables, equations

- Provide original editable tables; avoid screenshots.
- Prefer vector graphics (PDF/SVG/EPS) for line art; 300 dpi for raster images.
- Ensure symbols and units are consistent; define abbreviations at first use.
- Use equation editor (not images); number displayed equations if referenced.

---

6) Article types & length limits

- **Research Articles:** 7,000–9,000 words; up to 8 figures/tables.
- **Methods & Tools:** 4,000–6,000 words; up to 6 figures/tables.
- **Case Studies / Deployed Systems:** 5,000–8,000 words; up to 8 figures/tables.
- **Governance & Policy:** 3,000–5,000 words; up to 4 figures/tables.
- **Systematic Reviews:** 6,000–10,000 words; up to 6 figures/tables.
- **Registered Reports:** Stage 1: 3,000–5,000; Stage 2: 5,000–7,000; up to 6 figures/tables per stage.
- **Negative/Null Results:** 3,000–6,000 words; up to 4 figures/tables.
- **Perspectives / Viewpoints:** 1,500–2,500 words; up to 2 figures/tables.

Lengths exclude references, tables/figures and appendices. Supplementary materials are encouraged.

---

7) Submission package checklist (quick)

- **DOCX manuscript (anonymised)** — single file
- **Figures/Supplement:** ZIP (optional)
- **Everything else is entered in the submission form:** keywords; submission domain(s); Data/Code availability; Ethics (IRB/IEC; de-ID if applicable); Disclosures (funding, COI, AI-tool use, dual-use if relevant); High-stakes declaration; optional preprint link; ORCID/ROR.

---

8) After acceptance — final files & publication

- Camera-ready DOCX + final figures (hi-res); permissions for third-party materials.
- Completed **CRediT** with author names; **licence** (CC BY 4.0) agreement.
- Model/Data Cards (final); any repository links (datasets/code) updated.
- Production → author proofs (48–72 h) → publication **Open Access (CC BY 4.0)** with **Crossref DOI** and **Crossmark** status badge.

Questions? Contact the editorial office: **journal@aidecisions.ai**

AI DECISIONS — Submission Domains (Guide)

What we publish (in one paragraph)

AI DECISIONS publishes rigorous, practice-oriented research on **human–AI decision-making** in real settings. We treat methods, interfaces, operations and policy as a **single socio-technical system**, focusing on how information is collected, modelled, presented, acted upon, monitored and governed. We welcome work that **improves real decisions**—making them more accurate, fair, explainable, auditable, robust and affordable—with particular attention to **high-stakes domains**.

How to use this guide

- **Pick a primary domain** (add a secondary if helpful). Cross-domain submissions are welcome.
- Domains clarify **fit**; they don't determine acceptance. Fit depends on the contribution to decision quality.
- If your work spans prototypes and deployments, choose the domain of the **decision context** you study.
- If in doubt, choose the closest domain and explain the context briefly in your cover note.

Quick fit checklist (self-check)

Your manuscript is likely in scope if it clearly reports:

- **Decision context** (who decides, stakes/constraints, workflow).
- **Decision-level metrics** (e.g., calibration, subgroup performance, error costs, time-to-decision, workload, incidents).
- **Human factors & interfaces** (trust, escalation, handoffs).
- **Operations & governance** (monitoring, drift, audit logs, rollback).
- **Transparency & reproducibility** (code/data plans, **Model/Data Cards**).
- **Ethics** (IRB/IEC if applicable; **GDPR/UK GDPR** de-identification for human data).

Article types we accept

Research Articles · Case Studies / Deployed Systems · Methods & Tools · Governance & Policy · Systematic Reviews · Registered Reports (Stage 1/2) · Negative/Null Results · Perspectives/Viewpoints.

Submission Domains

1. **Healthcare & Clinical Decisions** — triage, diagnosis support, care pathways, escalation in real clinics.

   Examples: risk scores with calibrated triage, clinician-in-the-loop tools, post-deployment incident learning.

2. **Finance, Banking & Insurance** — credit/underwriting, fraud/AML, risk alerts with human oversight.

   Examples: affordability/cost-of-error analyses, reviewer workload effects, auditability and controls.

3. **Government & Public Services (incl. Emergency & Public Safety)** — eligibility, prioritisation, inspections, emergency dispatch/escalation.

   Examples: case routing with contestability, inspection targeting, call-centre decision support.

4. **Law, Justice & Compliance** — case triage, discovery, risk flags, contestable decisions.

   Examples: explainability for legal users, safeguards against misuse, measurable fairness/harms.

5. **Critical Infrastructure, Energy & Environment** — grids, water, buildings: operations, outages, resilience, environmental impact.

   Examples: demand response decisions, maintenance triage, air/water monitoring with escalation.

6. **Transportation & Mobility** — dispatch, routing, safety monitoring, traffic and logistics.

   Examples: operator-in-the-loop routing, incident response playbooks, reliability under drift.

7. **Education & Assessment** — placement, tutoring support, fair assessment workflows.

   Examples: calibrated feedback, workload/time savings, subgroup reliability over terms/semesters.

8. **Manufacturing & Supply Chains** — quality control, predictive maintenance, scheduling, exceptions.

   Examples: fault detection with human verification, cost/time impacts, rollback procedures.

9. **Customer Operations & Service** — contact centres/service desks: routing, priorities, quality control.

   Examples: triage cueing, agent assist UX, escalation rules and performance guarantees.

10. **Physical AI & Robotics (Embodied Systems)** — robots and embodied AI that perceive, decide and act in the physical world.

    Examples: manipulation and mobile robots, human–robot collaboration, deployment safety cases.

Notes for high-stakes submissions

If your work concerns **healthcare, finance or government/public services**, attach the short **High-stakes Checklist** (safety case, escalation, monitoring, rollback, costs). Where applicable, reference domain standards/reporting (e.g., CONSORT-AI/SPIRIT-AI, PROBAST) in your methods or appendix.

Submission at a glance

Submit (DOCX template) →

Editorial screening (Days 1–7) →

Peer review (Weeks 1–4) →

First decision (by Week 4) →

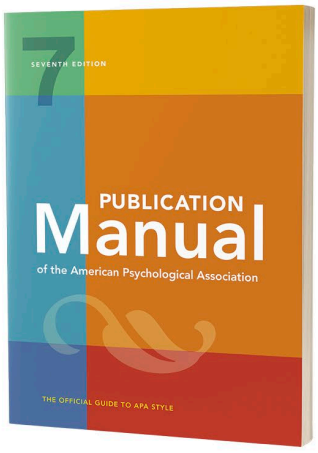
Revision (if requested) →

Acceptance & APC (per Fees & Waivers) →

Final files & publication (**CC BY 4.0**, Crossref DOI, Crossmark).

Questions? Contact: **journal@aidecisions.ai**

American Psychological Association | APA Style
750 First Street, NE, Washington, DC 20002



https://apastyle.org

Hi, APA Styler!

Thank you for using the APA Style annotated sample professional paper for guidance when writing your paper or assignment.

This sample paper PDF contains annotations that draw attention to key APA Style content and formatting such as the title page, headings, in-text citations, references, and more. Relevant sections of the seventh edition of the *Publication Manual* are also provided for your reference.

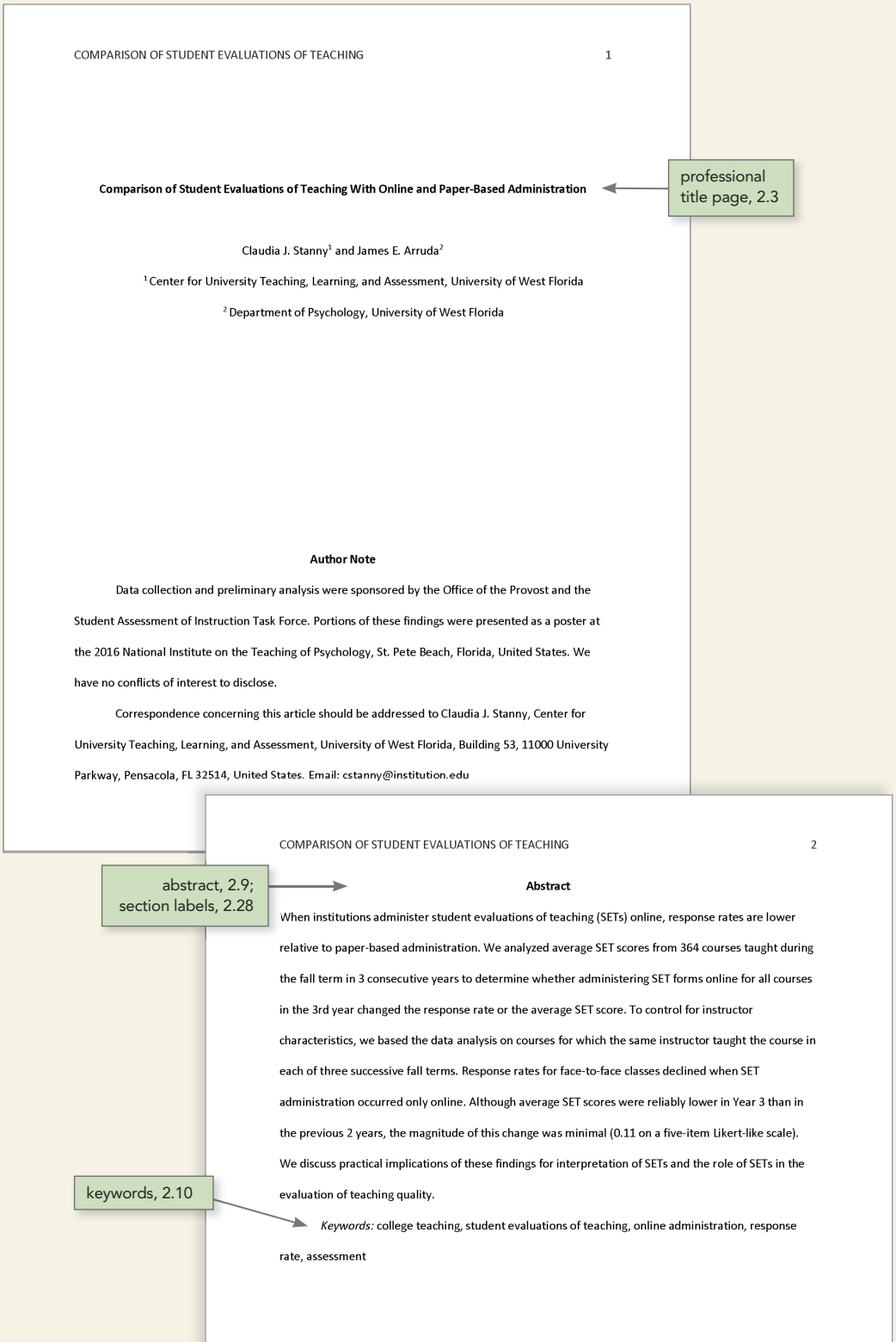
You can find this sample paper and many other resources in the seventh editions of the *Publication Manual* and *Concise Guide to APA Style*.

Please use discount code STYLEPAPER15 for 15% off **APA Style print products** with free shipping in the United States.

Happy writing!

Sample Papers

Sample Professional Paper



COMPARISON OF STUDENT EVALUATIONS OF TEACHING 1

Comparison of Student Evaluations of Teaching With Online and Paper-Based Administration

professional title page, 2.3

Claudia J. Stanny[1] and James E. Arruda[2]

[1] Center for University Teaching, Learning, and Assessment, University of West Florida

[2] Department of Psychology, University of West Florida

Author Note

Data collection and preliminary analysis were sponsored by the Office of the Provost and the Student Assessment of Instruction Task Force. Portions of these findings were presented as a poster at the 2016 National Institute on the Teaching of Psychology, St. Pete Beach, Florida, United States. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Claudia J. Stanny, Center for University Teaching, Learning, and Assessment, University of West Florida, Building 53, 11000 University Parkway, Pensacola, FL 32514, United States. Email: cstanny@institution.edu

COMPARISON OF STUDENT EVALUATIONS OF TEACHING 2

abstract, 2.9; section labels, 2.28

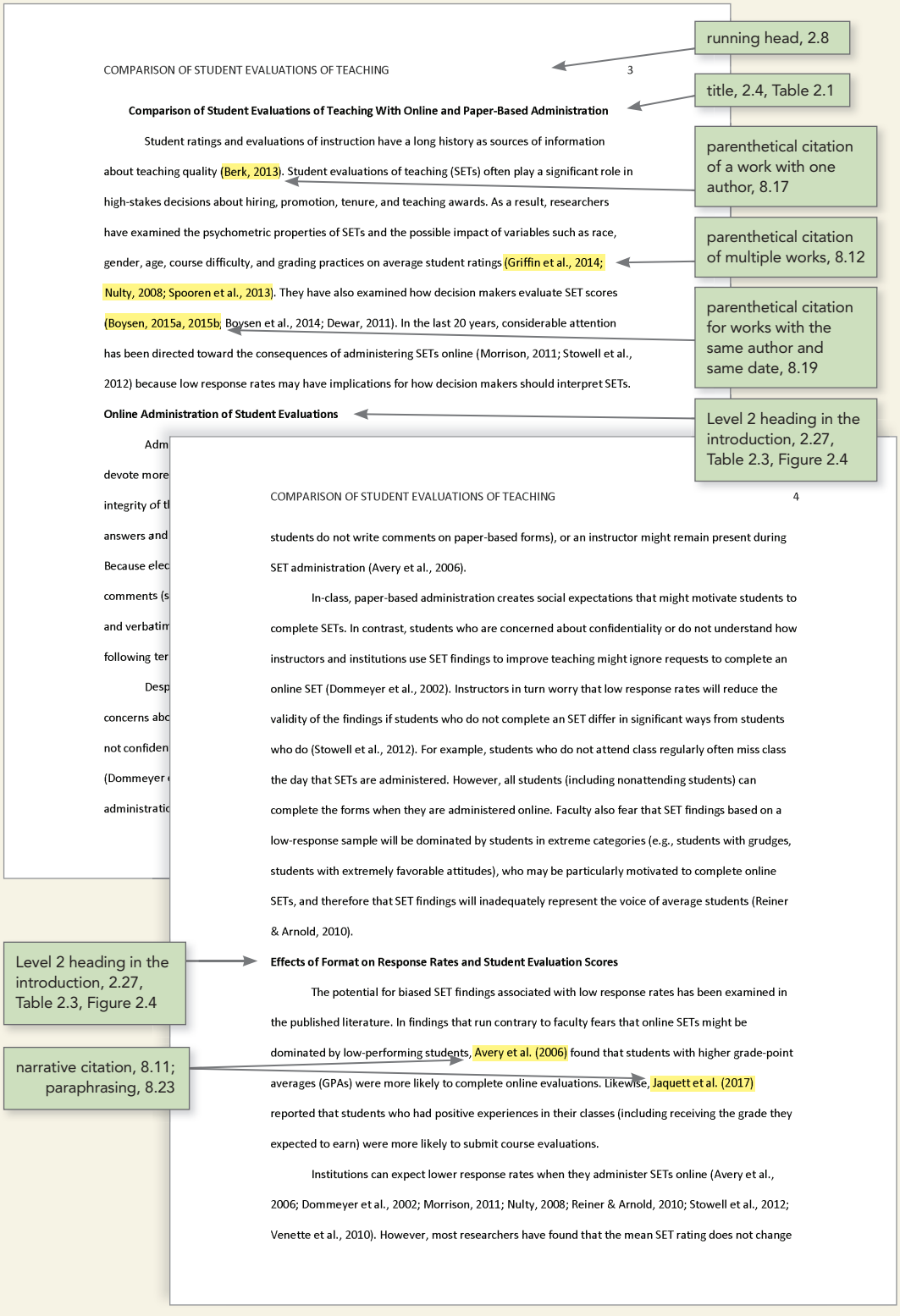
Abstract

When institutions administer student evaluations of teaching (SETs) online, response rates are lower relative to paper-based administration. We analyzed average SET scores from 364 courses taught during the fall term in 3 consecutive years to determine whether administering SET forms online for all courses in the 3rd year changed the response rate or the average SET score. To control for instructor characteristics, we based the data analysis on courses for which the same instructor taught the course in each of three successive fall terms. Response rates for face-to-face classes declined when SET administration occurred only online. Although average SET scores were reliably lower in Year 3 than in the previous 2 years, the magnitude of this change was minimal (0.11 on a five-item Likert-like scale). We discuss practical implications of these findings for interpretation of SETs and the role of SETs in the evaluation of teaching quality.

keywords, 2.10

Keywords: college teaching, student evaluations of teaching, online administration, response rate, assessment

Sample Professional Paper *(continued)*



COMPARISON OF STUDENT EVALUATIONS OF TEACHING 3

Comparison of Student Evaluations of Teaching With Online and Paper-Based Administration

Student ratings and evaluations of instruction have a long history as sources of information about teaching quality (Berk, 2013). Student evaluations of teaching (SETs) often play a significant role in high-stakes decisions about hiring, promotion, tenure, and teaching awards. As a result, researchers have examined the psychometric properties of SETs and the possible impact of variables such as race, gender, age, course difficulty, and grading practices on average student ratings (Griffin et al., 2014; Nulty, 2008; Spooren et al., 2013). They have also examined how decision makers evaluate SET scores (Boysen, 2015a, 2015b; Boysen et al., 2014; Dewar, 2011). In the last 20 years, considerable attention has been directed toward the consequences of administering SETs online (Morrison, 2011; Stowell et al., 2012) because low response rates may have implications for how decision makers should interpret SETs.

Online Administration of Student Evaluations

Adm
devote more
integrity of t
answers and
Because elec
comments (s
and verbatim
following ter

Desp
concerns abo
not confiden
(Dommeyer
administratio

COMPARISON OF STUDENT EVALUATIONS OF TEACHING 4

students do not write comments on paper-based forms), or an instructor might remain present during SET administration (Avery et al., 2006).

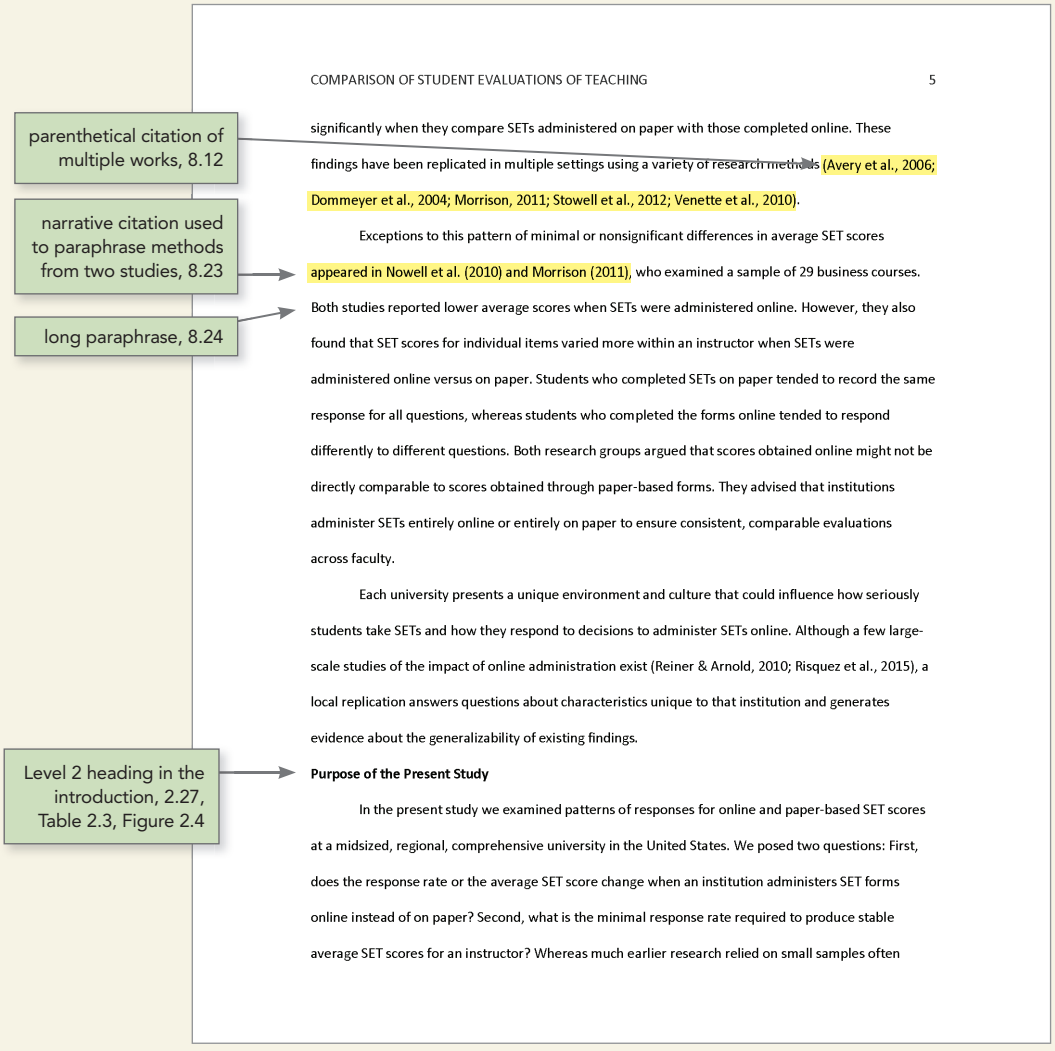
In-class, paper-based administration creates social expectations that might motivate students to complete SETs. In contrast, students who are concerned about confidentiality or do not understand how instructors and institutions use SET findings to improve teaching might ignore requests to complete an online SET (Dommeyer et al., 2002). Instructors in turn worry that low response rates will reduce the validity of the findings if students who do not complete an SET differ in significant ways from students who do (Stowell et al., 2012). For example, students who do not attend class regularly often miss class the day that SETs are administered. However, all students (including nonattending students) can complete the forms when they are administered online. Faculty also fear that SET findings based on a low-response sample will be dominated by students in extreme categories (e.g., students with grudges, students with extremely favorable attitudes), who may be particularly motivated to complete online SETs, and therefore that SET findings will inadequately represent the voice of average students (Reiner & Arnold, 2010).

Effects of Format on Response Rates and Student Evaluation Scores

The potential for biased SET findings associated with low response rates has been examined in the published literature. In findings that run contrary to faculty fears that online SETs might be dominated by low-performing students, Avery et al. (2006) found that students with higher grade-point averages (GPAs) were more likely to complete online evaluations. Likewise, Jaquett et al. (2017) reported that students who had positive experiences in their classes (including receiving the grade they expected to earn) were more likely to submit course evaluations.

Institutions can expect lower response rates when they administer SETs online (Avery et al., 2006; Dommeyer et al., 2002; Morrison, 2011; Nulty, 2008; Reiner & Arnold, 2010; Stowell et al., 2012; Venette et al., 2010). However, most researchers have found that the mean SET rating does not change

Sample Professional Paper *(continued)*



COMPARISON OF STUDENT EVALUATIONS OF TEACHING 5

significantly when they compare SETs administered on paper with those completed online. These findings have been replicated in multiple settings using a variety of research methods (Avery et al., 2006; Dommeyer et al., 2004; Morrison, 2011; Stowell et al., 2012; Venette et al., 2010).

Exceptions to this pattern of minimal or nonsignificant differences in average SET scores appeared in Nowell et al. (2010) and Morrison (2011), who examined a sample of 29 business courses. Both studies reported lower average scores when SETs were administered online. However, they also found that SET scores for individual items varied more within an instructor when SETs were administered online versus on paper. Students who completed SETs on paper tended to record the same response for all questions, whereas students who completed the forms online tended to respond differently to different questions. Both research groups argued that scores obtained online might not be directly comparable to scores obtained through paper-based forms. They advised that institutions administer SETs entirely online or entirely on paper to ensure consistent, comparable evaluations across faculty.

Each university presents a unique environment and culture that could influence how seriously students take SETs and how they respond to decisions to administer SETs online. Although a few large-scale studies of the impact of online administration exist (Reiner & Arnold, 2010; Risquez et al., 2015), a local replication answers questions about characteristics unique to that institution and generates evidence about the generalizability of existing findings.

Purpose of the Present Study

In the present study we examined patterns of responses for online and paper-based SET scores at a midsized, regional, comprehensive university in the United States. We posed two questions: First, does the response rate or the average SET score change when an institution administers SET forms online instead of on paper? Second, what is the minimal response rate required to produce stable average SET scores for an instructor? Whereas much earlier research relied on small samples often

Sample Professional Paper *(continued)*

limited to a single academic department, we gathered SET data on a large sample of courses (N = 364) that included instructors from all colleges and all course levels over 3 years. We controlled for individual differences in instructors by limiting the sample to courses taught by the same instructor in all 3 years. The university offers nearly 30% of course sections online in any given term, and these courses have always administered online SETs. This allowed us to examine the combined effects of changing the method of delivery for SETs (paper-based to online) for traditional classes and changing from a mixed method of administering SETs (paper for traditional classes and online for online classes in the first 2 years of data gathered) to uniform use of online forms for all classes in the final year of data collection.

Method

Sample

Response rates and evaluation ratings were retrieved from archived course evaluation data. The archive of SET data did not include information about personal characteristics of the instructor (gender, age, or years of teaching experience), and students were not provided with any systematic incentive to complete the paper or online versions of the SET. We extracted data on response rates and evaluation ratings for 36

(2012, 2013,

The s

instructors (3

provided dat

courses, and

face-to-face

Level 1 heading after the introduction, 2.27, Table 2.3, Figure 2.5

Level 2 heading, 2.27, Table 2.3, Figure 2.5

Level 2 heading, 2.27, Table 2.3, Figure 2.5

beginning undergraduate level (1st- and 2nd-year students), 205 courses (56%) at the advanced undergraduate level (3rd- and 4th-year students), and 52 courses (14%) at the graduate level.

Instrument

The course evaluation instrument was a set of 18 items developed by the state university system. The first eight items were designed to measure the quality of the instructor, concluding with a global rating of instructor quality (Item 8: "Overall assessment of instructor"). The remaining items asked students to evaluate components of the course, concluding with a global rating of course organization (Item 18: "Overall, I would rate the course organization"). No formal data on the psychometric properties of the items are available, although all items have obvious face validity.

Students were asked to rate each instructo as *poor* (0), *fair* (1), *good* (2), *very good* (3), or *excellent* (4) in response to each item. Evaluation ratings were subsequently calculated for each course and instructor. A median rating was computed when an instructor taught more than one section of a course during a term.

The institution limited our access to SET data for the 3 years of data requested. We obtained scores for Item 8 ("Overall assessment of instructor") for all 3 years but could obtain scores for Item 18 ("Overall, I would rate the course organization") only for Year 3. We computed the correlation between scores on Item 8 and Item 18 (from course data recorded in the 3rd year only) to estimate the internal consistency of the evaluation instrument. These two items, which serve as composite summaries of preceding items (Item 8 for Items 1–7 and Item 18 for Items 9–17), were strongly related, $r(362) = .92$. Feistauer and Richter (2016) also reported strong correlations between global items in a large analysis of SET responses.

Design

This study took advantage of a natural experiment created when the university decided to administer all course evaluations online. We requested SET data for the fall semesters for 2 years

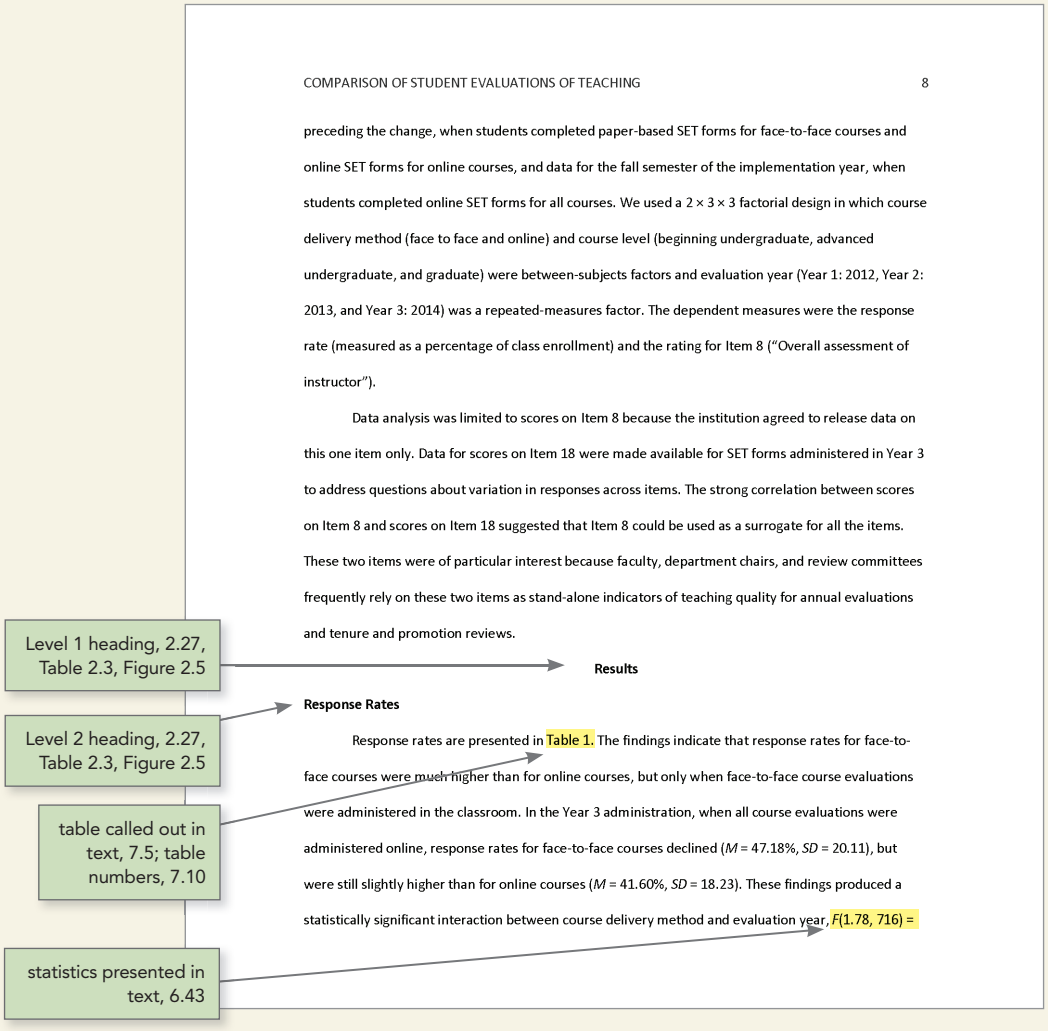
italics used for anchors of a scale, 6.22

en dash used in a numerical range, 6.6

statistics presented in text, 6.43

Level 2 heading, 2.27, Table 2.3, Figure 2.5

Sample Professional Paper *(continued)*



COMPARISON OF STUDENT EVALUATIONS OF TEACHING 8

preceding the change, when students completed paper-based SET forms for face-to-face courses and online SET forms for online courses, and data for the fall semester of the implementation year, when students completed online SET forms for all courses. We used a 2 × 3 × 3 factorial design in which course delivery method (face to face and online) and course level (beginning undergraduate, advanced undergraduate, and graduate) were between-subjects factors and evaluation year (Year 1: 2012, Year 2: 2013, and Year 3: 2014) was a repeated-measures factor. The dependent measures were the response rate (measured as a percentage of class enrollment) and the rating for Item 8 ("Overall assessment of instructor").

Data analysis was limited to scores on Item 8 because the institution agreed to release data on this one item only. Data for scores on Item 18 were made available for SET forms administered in Year 3 to address questions about variation in responses across items. The strong correlation between scores on Item 8 and scores on Item 18 suggested that Item 8 could be used as a surrogate for all the items. These two items were of particular interest because faculty, department chairs, and review committees frequently rely on these two items as stand-alone indicators of teaching quality for annual evaluations and tenure and promotion reviews.

Level 1 heading, 2.27, Table 2.3, Figure 2.5

Results

Level 2 heading, 2.27, Table 2.3, Figure 2.5

Response Rates

table called out in text, 7.5; table numbers, 7.10

Response rates are presented in Table 1. The findings indicate that response rates for face-to-face courses were much higher than for online courses, but only when face-to-face course evaluations were administered in the classroom. In the Year 3 administration, when all course evaluations were administered online, response rates for face-to-face courses declined (M = 47.18%, SD = 20.11), but were still slightly higher than for online courses (M = 41.60%, SD = 18.23). These findings produced a statistically significant interaction between course delivery method and evaluation year, $F(1.78, 716)$ =

statistics presented in text, 6.43

Sample Professional Paper *(continued)*

101.34, *MSE* = 210.61, $p < .001$.[1] The strength of the overall interaction effect was .22 (η_p^2). Simple main-effects tests revealed statistically significant differences in the response rates for face-to-face courses and online courses for each of the 3 observation years.[2] The greatest differences occurred during Year 1 ($p < .001$) and Year 2 ($p < .001$), when evaluations were administered on paper in the classroom for all face-to-face courses and online for all online courses. Although the difference in response rate between face-to-face and online courses during the Year 3 administration was statistically reliable (when both face-to-to-face and online courses were evaluated with online surveys), the effect was small ($\eta_p^2 = .02$). Thus, there was minimal difference in response rate between face-to-face and online courses when evaluations were administered online for all courses. No other factors or interactions included in the analysis were statistically reliable.

Evaluation Ratings

The same 2 × 3 × 3 analysis of variance model was used to evaluate mean SET ratings. This analysis produced two statistically significant main effects. The first main effect involved evaluation year, $F(1.86, 716) = 3.44$, *MSE* = 0.18, $p = .03$ ($\eta_p^2 = .01$; see Footnote 1). Evaluation ratings associated with the Year 3 administration (*M* = 3.26, *SD* = 0.60) were significantly lower than the evaluation ratings associated with both the Year 1 (*M* = 3.35, *SD* = 0.53) and Year 2 (*M* = 3.38, *SD* = 0.54) administrations. Thus, all courses received lower SET scores in Year 3, regardless of course delivery method and course level. However, the size of this effect was small (the largest difference in mean rating was 0.11 on a five-item scale).

[1] A Greenhouse–Geisser adjustment of the degrees of freedom was performed in anticipation of a sphericity assumption violation.
[2] A test of the homogeneity of variance assumption revealed no statistically significant difference in response rate variance between the two delivery modes for the 1st, 2nd, and 3rd years.

Stability of Ratings

The scatterplot presented in Figure 1 illustrates the relation between SET scores and response rate. Although the correlation between SET scores and response rate was small and not statistically significant, $r(362) = .07$, visual inspection of the plot of SET scores suggests that SET ratings became less variable as response rate increased. We conducted Levene's test to evaluate the variability of SET scores above and below the 60% response rate, which several researchers have recommended as an acceptable threshold for response rates (Berk, 2012, 2013; Nulty, 2008). The variability of scores above and below the 60% threshold was not statistically reliable, $F(1, 362) = 1.53$, $p = .22$.

Discussion

Online administration of SETs in this study was associated with lower response rates, yet it is curious that online courses experienced a 10% increase in response rate when all courses were evaluated with online forms in Year 3. Online courses had suffered from chronically low response rates in previous years, when face-to-face classes continued to use paper-based forms. The benefit to response rates observed for online courses when all SET forms were administered online might be attributed to increased communications that encouraged students to complete the online course evaluations. Despite this improvement, response rates for online courses continued to lag behind those for face-to-face courses. Differences in response rates for face-to-face and online courses might be

footnote callout, 2.13

Level 2 heading, 2.27, Table 2.3, Figure 2.5

referring to a previous footnote, 2.13

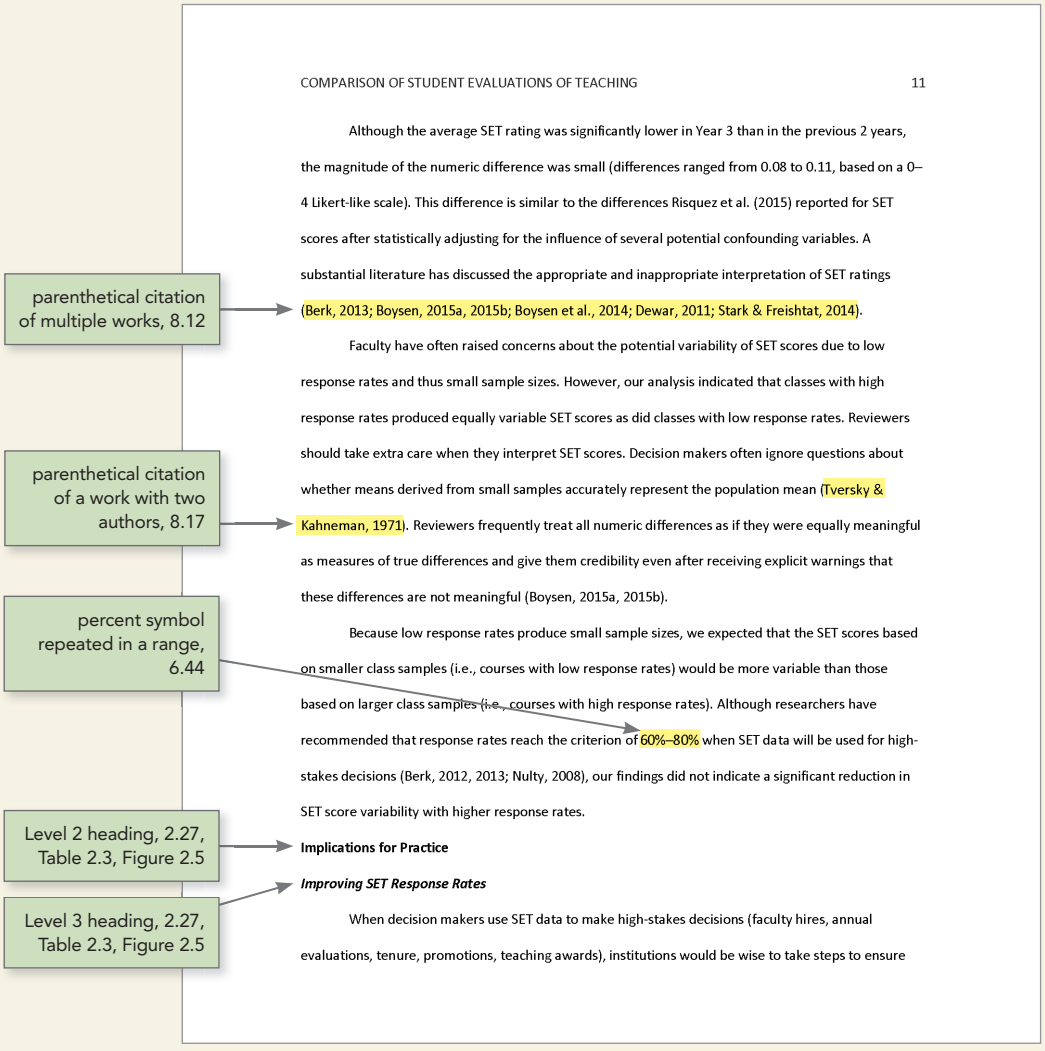
footnote in page footer, 2.13

figure called out in text, 7.5; figure numbers, 7.24

parenthetical citation of multiple papers by the same author, 8.12

Level 1 heading, 2.27, Table 2.3, Figure 2.5

Sample Professional Paper *(continued)*



Sample Professional Paper *(continued)*

that SETs have acceptable response rates. Researchers have discussed effective strategies to improve response rates for SETs (Nulty, 2008; see also Berk, 2013; Dommeyer et al., 2004; Jaquett et al., 2016). These strategies include offering empirically validated incentives, creating high-quality technical systems with good human factors characteristics, and promoting an institutional culture that clearly supports the use of SET data and other information to improve the quality of teaching and learning. Programs and instructors must discuss why information from SETs is important for decision-making and provide students with tangible evidence of how SET information guides decisions about curriculum improvement. The institution should provide students with compelling evidence that the administration system protects the confidentiality of their responses.

"see also" citation, 8.12

Evaluating SET Scores

Level 3 heading, 2.27, Table 2.3, Figure 2.5

In addition to ensuring adequate response rates on SETs, decision makers should demand multiple sources of evidence about teaching quality (Buller, 2012). High-stakes decisions should never rely exclusively on numeric data from SETs. Reviewers often treat SET ratings as a surrogate for a measure of the impact an instructor has on student learning. However, a recent meta-analysis (Uttl et al., 2017) questioned whether SET scores have any relation to student learning. Reviewers need evidence in addition to SET ratings to evaluate teaching, such as evidence of the instructor's disciplinary content expertise, skill with classroom management, ability to engage learners with lectures or other activities, impact on student learning, or success with efforts to modify and improve courses and teaching strategies (Berk, 2013; Stark & Freishtat, 2014). As with other forms of assessment, any one measure may be limited in terms of the quality of information it provides. Therefore, multiple measures are more informative than any single measure.

parenthetical citation of a work with one author, 8.17

parenthetical citation of two works, 8.12

A

include su

assignmer

samples of student work. Course syllabi can identify intended learning outcomes; describe instructional strategies that reflect the rigor of the course (required assignments and grading practices); and provide other information about course content, design, instructional strategies, and instructor interactions with students (Palmer et al., 2014; Stanny et al., 2015).

Level 2 heading, 2.27, Table 2.3, Figure 2.5

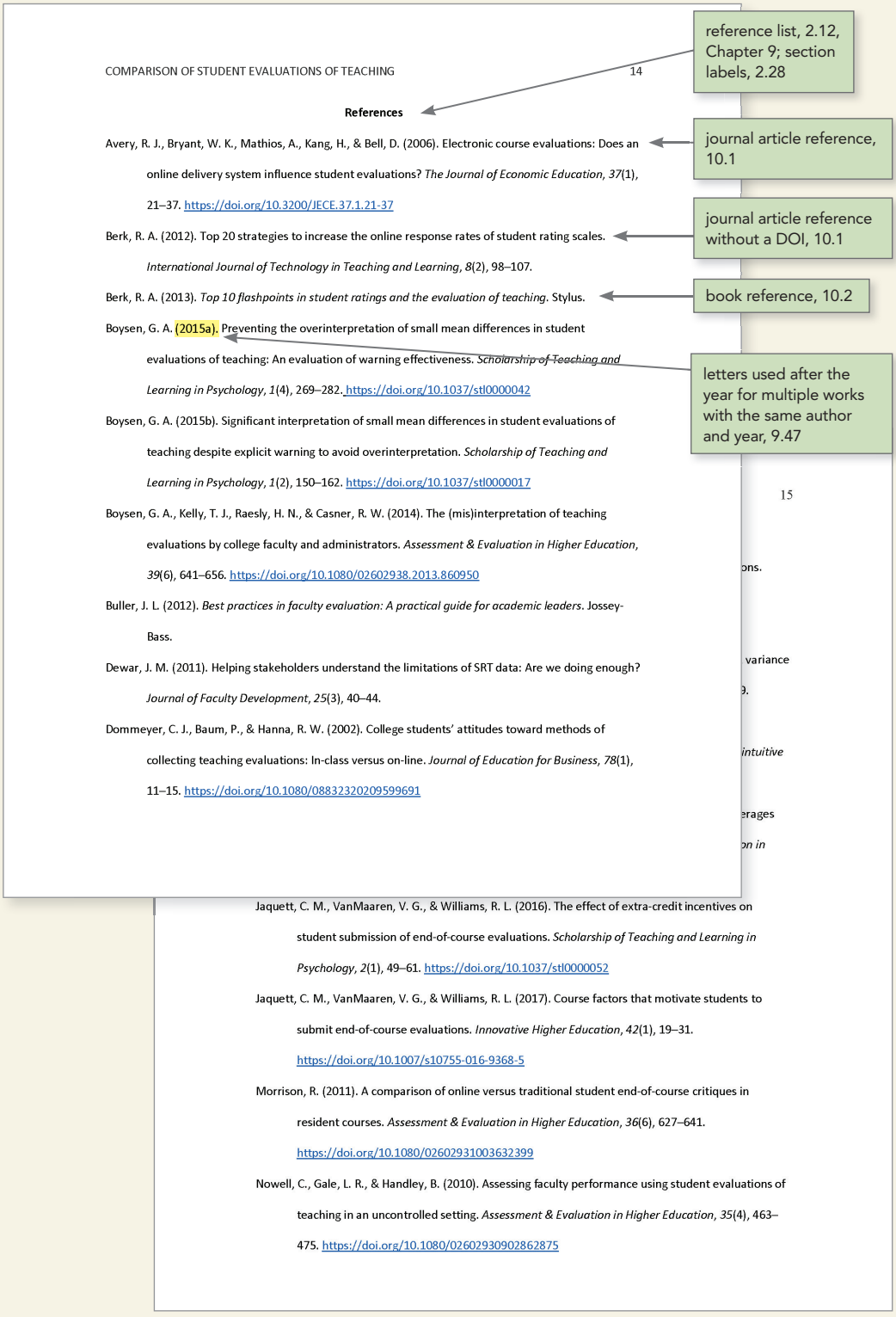
Conclusion

Psychology has a long history of devising creative strategies to measure the "unmeasurable," whether the targeted variable is a mental process, an attitude, or the quality of teaching (e.g., Webb et al., 1966). In addition, psychologists have documented various heuristics and biases that contribute to the misinterpretation of quantitative data (Gilovich et al., 2002), including SET scores (Boysen, 2015a, 2015b; Boysen et al., 2014). These skills enable psychologists to offer multiple solutions to the challenge posed by the need to objectively evaluate the quality of teaching and the impact of teaching on student learning.

quotation marks used to indicate an ironic comment, 6.7

Online administration of SET forms presents multiple desirable features, including rapid feedback to instructors, economy, and support for environmental sustainability. However, institutions should adopt implementation procedures that do not undermine the usefulness of the data gathered. Moreover, institutions should be wary of emphasizing procedures that produce high response rates only to lull faculty into believing that SET data can be the primary (or only) metric used for high-stakes decisions about the quality of faculty teaching. Instead, decision makers should expect to use multiple measures to evaluate the quality of faculty teaching.

Sample Professional Paper *(continued)*



COMPARISON OF STUDENT EVALUATIONS OF TEACHING 14

References

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education*, *37*(1), 21–37. https://doi.org/10.3200/JECE.37.1.21-37

Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, *8*(2), 98–107.

Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching*. Stylus.

Boysen, G. A. (2015a). Preventing the overinterpretation of small mean differences in student evaluations of teaching: An evaluation of warning effectiveness. *Scholarship of Teaching and Learning in Psychology*, *1*(4), 269–282. https://doi.org/10.1037/stl0000042

Boysen, G. A. (2015b). Significant interpretation of small mean differences in student evaluations of teaching despite explicit warning to avoid overinterpretation. *Scholarship of Teaching and Learning in Psychology*, *1*(2), 150–162. https://doi.org/10.1037/stl0000017

Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis)interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in Higher Education*, *39*(6), 641–656. https://doi.org/10.1080/02602938.2013.860950

Buller, J. L. (2012). *Best practices in faculty evaluation: A practical guide for academic leaders*. Jossey-Bass.

Dewar, J. M. (2011). Helping stakeholders understand the limitations of SRT data: Are we doing enough? *Journal of Faculty Development*, *25*(3), 40–44.

Dommeyer, C. J., Baum, P., & Hanna, R. W. (2002). College students' attitudes toward methods of collecting teaching evaluations: In-class versus on-line. *Journal of Education for Business*, *78*(1), 11–15. https://doi.org/10.1080/08832320209599691

reference list, 2.12, Chapter 9; section labels, 2.28

journal article reference, 10.1

journal article reference without a DOI, 10.1

book reference, 10.2

letters used after the year for multiple works with the same author and year, 9.47

15

Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2016). The effect of extra-credit incentives on student submission of end-of-course evaluations. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 49–61. https://doi.org/10.1037/stl0000052

Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2017). Course factors that motivate students to submit end-of-course evaluations. *Innovative Higher Education*, *42*(1), 19–31. https://doi.org/10.1007/s10755-016-9368-5

Morrison, R. (2011). A comparison of online versus traditional student end-of-course critiques in resident courses. *Assessment & Evaluation in Higher Education*, *36*(6), 627–641. https://doi.org/10.1080/02602931003632399

Nowell, C., Gale, L. R., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, *35*(4), 463–475. https://doi.org/10.1080/02602930902862875

Sample Professional Paper *(continued)*

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

Palmer, M. S., Bach, D. J., & Streifer, A. C. (2014). Measuring the promise: A learning-focused syllabus rubric. *To Improve the Academy: A Journal of Educational Development*, *33*(1), 14–36. https://doi.org/10.1002/tia2.20004

Reiner, C. M., & Arnold, K. E. (2010). Online course evaluation: Student and instructor perspectives and assessment potential. *Assessment Update*, *22*(2), 8–10. https://doi.org/10.1002/au.222

Risquez, A., Vaughan, E., & Murphy, M. (2015). Online student evaluations of teaching: What are we sacrificing for the affordances of technology? *Assessment & Evaluation in Higher Education*, *40*(1), 210–234. https://doi.org/10.1080/02602938.2014.890695

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Stanny, C. J., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education*, *40*(7), 898–913. https://doi.org/10.1080/02602938.2014.956684

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stowell, J. R
eva
htt

Tversky, A.,
105

title ending with a question mark, 9.19

journal article reference with missing issue number, 9.26

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/10.1016/j.stueduc.2016.08.007

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, *35*(1), 101–115. https://doi.org/10.1080/02602930802618336

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally.

Sample Professional Paper *(continued)*

table number, 7.10

table title, 7.11

table note, 7.14

Table 1

Means and Standard Deviations for Response Rates (Course Delivery Method by Evaluation Year)

Administration year	Face-to-face course		Online course	
	M	*SD*	*M*	*SD*
Year 1: 2012	71.72	16.42	32.93	15.73
Year 2: 2013	72.31	14.93	32.55	15.96
Year 3: 2014	47.18	20.11	41.60	18.23

Note. Student evaluations of teaching (SETs) were administered in two modalities in Years 1 and 2: paper based for face-to-face courses and online for online courses. SETs were administered online for all courses in Year 3.

figure number, 7.24

figure title, 7.25

Figure 1

Scatterplot Depicting the Correlation Between Response Rates and Evaluation Ratings

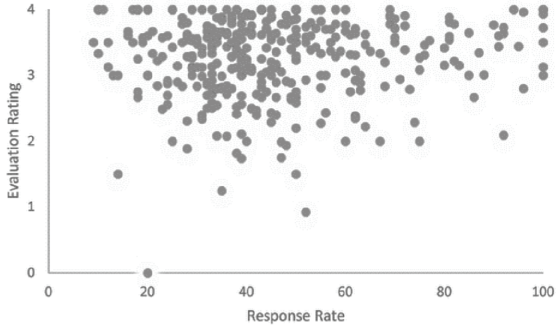


figure note, 7.28

Note. Evaluation ratings were made during the 2014 fall academic term.