JUTLP
Journal of University Teaching & Learning Practice

What and How: Investigating the Use of Student Evaluations in Scholarship of Teaching and Learning Research

Robyn Moore[a], Dr Jill M. McSweeney[a, b], and Dr Elizabeth A.L. Gillis[a, c]
[a] Dalhousie University, Canada; [b] Elon University, United States of America, [c] Saint Mary's University, Canada

Abstract

Student evaluations of teaching (SETs) have been a longstanding form of data intended to provide feedback on teaching from the student perspective. Little work has been done prior to this study to systematically explore how SET data is applied as part of research on teaching and learning and how/what data is being used in this form of scholarship. Through a scoping review of 223 studies and a survey with individuals engaging in research on teaching and learning (n = 47), this study offers insight into how SET data can and is being used in scholarship on teaching and learning, and common barriers to their use for research purposes. For example, concerns regarding data quality, inherent bias in the tool, access to data, and SET design limitations were identified as challenges to use. Despite such barriers, the results of this work indicate SETs are frequently used in a less formal capacity to improve teaching and learning practice and that the data is largely used in combination with other forms of evidence to provide empirical insight into teaching practices. Pathways for institutions to enhance access to SET data and develop processes for contextualizing it in ways that support faculty in using the information to improve their teaching are discussed.

Editors

Section: Curriculum and Assessment Design
Senior Editor: Dr Alison Purvis
Associate Editor: Professor Victoria Nicholas

Publication

Submission: 19 September 2024
Revised: 1 July 2025
Accepted: 21 October 2025
Published: 27 February 2026

Practitioner Notes

1. SETs are *one* evidence source in Scholarship of Teaching and Learning (SoTL) and can be triangulated with grades, observations, interviews, or other surveys before drawing conclusions on teaching practice.
2. Prioritize open-ended SET comments for actionable insight and pair them with specific quantitative data to check for trends over time.
3. Interpret SET results with caution and account for low response rates or known biases and avoid treating SETs as direct measures of teaching effectiveness.
4. Advocate for better departmental and institutional processes that clarify issues of ownership, access, and permissions to better support the ethical use of SET data for SoTL.
5. Improve the usefulness of SETs by adding course-specific questions aligned to your teaching context and tell students how their feedback will be used.

Keywords

SETs, teaching excellence, scholarship of teaching and learning, scoping review

Citation:

Moore, R., McSweeney, J.M., & Gillis, E.A.L. (2026). What and how: Investigating the use of student evaluations in scholarship of teaching and learning research. *Journal of University Teaching and Learning Practice, 23*(1). https://doi.org/10.53761/a6hmyb53

Introduction

Student evaluations of teaching (SET) have been a longstanding form of institutionally driven data collection for the evaluation of pedagogy, course delivery, the student experience, and more tenuously, hiring and promotion. For instructors and institutions, they are meant to be one source of formal data used to understand pedagogical effectiveness and impact; and for students, a place to reflect, evaluate, and express their experience of learning and course instruction. For decision makers, they have long been relied on as a measure of teaching effectiveness, though this practice is shifting to more holistic approaches that recognize the limitations of SET data in this regard. For most institutions, SETs are administered during the final weeks of a course, before final grades are provided to students. While originally administered via pencil and paper, most institutions have transitioned to online versions, providing students the flexibility to complete SETs inside and outside of class time. Most SETs are scored using a Likert scale ranging from "Strongly Disagree" to "Strongly Agree" (Uttl, 2024; Uttl et al., 2016), and ask students about their perceptions of instructional quality, course delivery, and instructor characteristics at the broader 'overall' course-level and may also include more specific instructor-written questions about their course experience. Anonymized data is typically reported back to instructors and administrators and includes a distribution of responses and mean ratings, as well as open-ended student comments. Institutions and administration often view SETs as a cost-effective and easily administered assessment of classroom learning (Klopfer et al., 2024; Uttl, 2024; Uttl et al., 2016), providing a pathway for students to share their experiences and their voices to be integrated into instructional and curricular decisions.

Recent work has explored the effectiveness of SET data as a measure of teaching effectiveness (Boring et al., 2016; Clayson, 2009; Kreitzer & Sweet-Cushman, 2022; Uttl, 2024) and found that they can misrepresent the connection between quality teaching and student experience. Within this context and the broader conversation of SET data, they have become commonly accepted as a measure of student satisfaction and often disregarded as useful data (Klopfer et al., 2024; Uttl, 2024). However, we believe that SETs are a valuable tool when used with appreciation of their limitations and in conjunction with other sources of data and teaching artefacts to better understand the impact of pedagogical interventions, instructional decisions, and course experiences from a student perspective. Thus, when considered as part of a landscape of data on teaching and learning, SET have the potential to lead to quality improvements in curriculum delivery, individual teaching practices, course development, and student learning (Linse, 2017; McDonald et al., 2024). SETs therefore offer an opportunity for scholarly inquiry and reflection on teaching and learning practices at course, department, and institutional levels, and given that they are used in nearly all universities and colleges within North America (Murray, 2005; Uttl et al., 2024) and throughout the international landscape, they are an abundant and accessible source of data for instructors and administration to use when exploring research on teaching and learning.

Literature

The scholarship of teaching and learning (SoTL) is an expanding area of research in post-secondary education. SoTL's systematic inquiry into pedagogical practices has the intended goal of enhancing the student experience through improving teaching and/or learning at the classroom, departmental, or disciplinary level (Godbold et al., 2021; McKinney, 2006; Potter & Kustra, 2011).

Boyer's definition of teaching as a scholarly activity (Boyer, 1990) has served to transform SoTL into a field that provides evidence-based change to teaching and learning for instructors across all disciplines (Felten & Chick, 2018; Hubball & Clarke, 2010), and has evolved into a mainstay of scholarly work at universities (Botnaru et al., 2022; Wuetherick et al., 2016). As academics look to engage more frequently in SoTL, the ubiquity of SET data and its potentially valuable insight into the student perspective makes it a convenient and readily available source of data that is often underutilized in our experience.

Research using SETs as a primary source of data for department and institutional level research exists. For example, there are several studies where SET data is used to explore the impact of instructional techniques (Mahendra, 2018; Middendorf & McNary, 2011), student experience and satisfaction (Calvo et al., 2010; Slocum-Schaffer & Bohrer, 2019; Wilson et al., 2011), and pilot studies on new teaching initiatives (Carbone, 2014; Harsy et al, 2020). Explicit examples of SETs utilized for SoTL can be seen by Bengtson et al. (2017) who utilized SET data to measure the impact of course development changes guided by student partners, and without this data, they would not have been unable to show the success of this partnership and the impact on student learning. Similarly, Bailey et al. (2020) used SETs as a data source to measure the impact of their implementation of team-based projects on student performance and perceptions of impact. Though SET data is widely collected, there are well-documented concerns regarding the quality and ownership of the data, leading to questions about how, when, and where it is suitable for use (see, for example: Boring et al., 2016; Maurer, 2006; Mitchell & Martin, 2018; Mowatt, 2019; Otu & Otu, 2023; Stoesz et al., 2022; Svanum & Aigner, 2011; Uttl, 2024; Zumbach & Funke, 2014). These concerns often lead to bureaucratic barriers for individual use (Schnurr & Taylor, 2019), limiting the ability for individual instructors to conduct classroom-based SoTL which can help to reflect, measure and inform their teaching practice and course decisions. Despite such concerns, SET data remains widely used as part of scholarly teaching and as evidence of teaching effectiveness in, for example, promotion and tenure processes and institutional/programmatic decision making. In an editorial, Ali et. al (2021) called for 'critical and reflexive' engagement with the use of SET data for SoTL, and acknowledged the complexities involved in using such data as evidence throughout the research process (e.g., impact of contextual factors, influence of student and instructor intersectionality on interpretation, and socio-cultural norms around providing evaluations.)

No work to date has explored the specific use of SET data for SoTL purposes and how instructors could potentially utilise this data to enhance their teaching and contribute to their scholarship practices. To provide support for SoTL researchers, and address Ali et. al's (2021) call for 'critical and reflexive' use of SET data, we must understand how researchers are currently integrating this data into their SoTL work, and if they are not, why. Here, we explore this line of inquiry with the aim of uncovering:

1. In what context is SET data being used for SoTL?
2. What data from SETs is most often used?
3. What barriers exist to utilising SET data for SoTL?

Method

This study used a concurrent mixed-methods design to explore past and current practices of SET data in SoTL (Creswell et al., 2003). These two approaches allowed us to explore the historical use of SET data for scholarly purposes in teaching and learning literature, as well as understand how current scholars perceive and utilise this type of data for scholarly purposes. We performed the study following the Declaration of Helsinki guidelines and the regulations of the Canadian Tri-Council Policy Statement (TCPS) on research ethics which governed the institutional body who approved this work.

Scoping Review

Using scoping review procedures by Arksey and O'Malley (2005), a variety of databases across disciplinary scholarship (i.e., ERIC, PsycINFO, Sociological Abstracts, Academic Search Premier, Taylor and Francis Online, and Scopus), as well as reference lists of relevant articles were used to identify articles. Articles were included if they were written in English and published within the twenty years prior to the year of our study (2003 - 2023). Search terms were constructed by our research question and were refined through reviewing the literature (Table 1).

Table 1

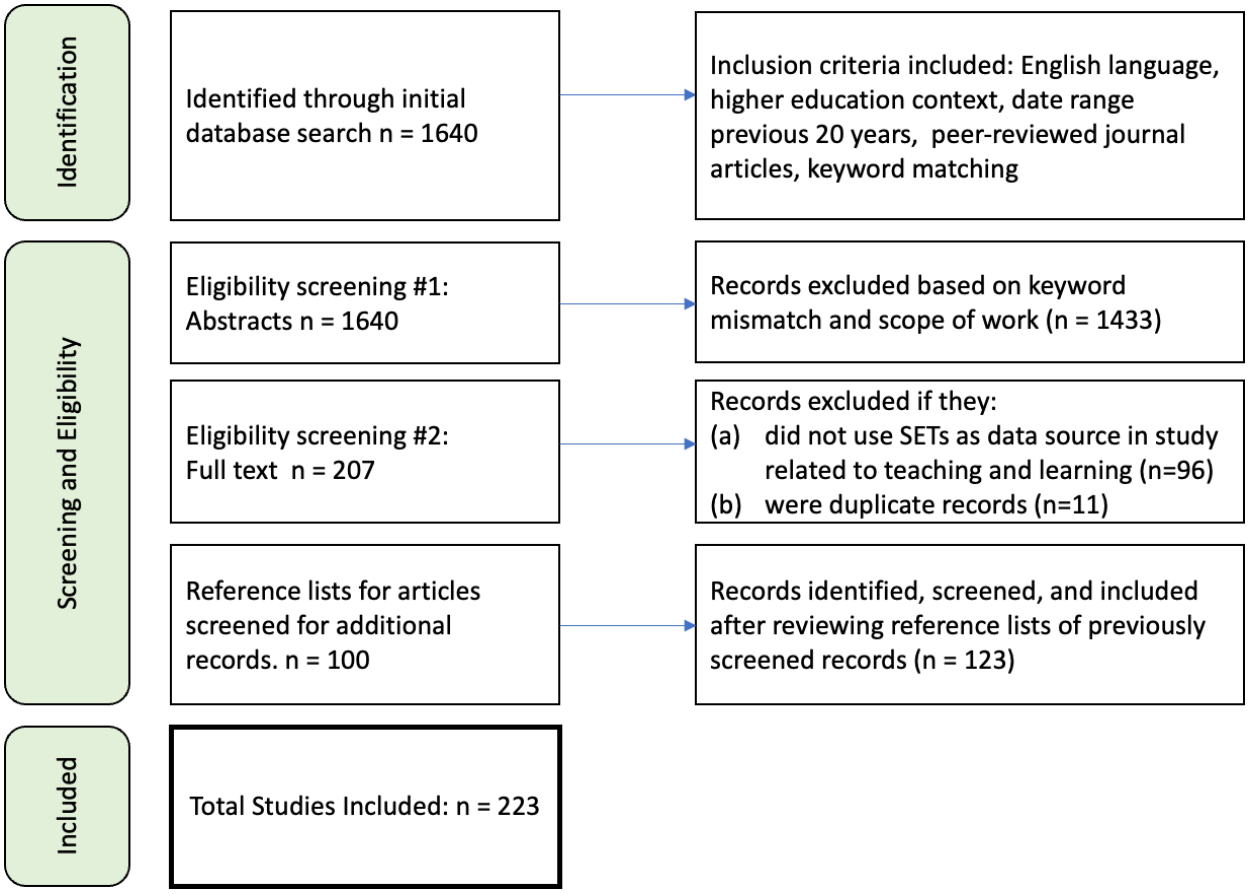
Search terms used for scoping study.

SET terms	Research terms	Education level terms
"Student Rating of Instruction" OR "Student Evaluation of Teach*" OR "Course Evaluation" OR "Teacher Evaluation" OR "Student Learning Experience" OR "Student Evaluations" OR "Student Experience"	"Scholarship of Teaching and Learning" OR "SOTL" OR "Disciplinary Based Education" OR "DBER" OR "teaching research" OR "education research"	Universit* OR "higher ed*" OR Tertiary ed* OR college OR "post secondary" OR "undergraduate"

The initial search (n = 1640) screened for inclusion according to predefined criteria (e.g. date range of publication, language, type of publication; Figure 1.) A total of 223 articles were examined.

Figure 1

Summary of scoping review process.



Survey

Participants were recruited through an open invitation on well-established North American teaching listservs (e.g. the Society of Teaching and Learning in Higher Education and the Professional and Organizational Development Network.) These listservs were chosen given the diversity of subscribers (e.g., faculty, educational developers, institutional researchers, higher education administrators) across disciplines and their use for recruitment in other teaching and learning scholarship (e.g. Kolomitro & Anstey, 2017; Sorcinelli & Austin, 2006). The survey was 23 questions long and included a mix of open- and closed-ended responses, such as "To your knowledge, are instructors at your institution able to personalise some questions on their student evaluation questionnaires?" and "Is there any data that is not provided that would make it more likely for you to use the data in research on teaching and learning?" Questions were developed and reviewed for readability in consultation with educational development colleagues and from our own experiences supporting instructors in using SET data for research purposes. An initial scan of the literature indicated that no studies to date had asked instructors about this topic, so we opted for a set of exploratory questions.

In total, 47 participants across Canada (57.4%, n = 27) and the United States (42.6%, n = 20) completed the survey. Most (68.1%, n = 32) indicated that they were currently in a

faculty/instructor position, followed by administrative (27.7%, n = 13) or staff (21.3%, n = 10). Participants could select that they held more than one role.

Analysis

Quantitative data from the survey and scoping review were analysed through descriptive statistics via Excel. Open-ended questions were independently coded by each author with a discussion on various codes to guarantee consistency and shared understanding of how we interpreted responses. This was especially important given the varied positions from which our respondents engage in data usage within institutions. From this, we held group discussions of potential themes emerging from the responses, giving time to review discrepancies until consensus was met. A single author compiled the final dataset based on the agreed-upon codes. Examples of emerging themes included 'data quality', which encompassed low response rates, limited data points, imbalances samples, ambiguous responses; 'bias related' including gender, racial, implicit bias in student responses and bias in the perception of data from SETs; 'relevance' and 'alignment of SET questions' and data collection format to SoTL questions and methodological approaches; and 'logistics', including limitations in timing and design of instrument as well as access to data.

Results

Summary of Publications

The 223 articles reviewed were published broadly. There were 208 unique first authors with publications in 122 different publications distributed across 2003 - 2023. The distribution of year of publications is provided in Figure 2 and Journals that had 3 or more contributions to our review are presented in Table 2.

Figure 2

Distribution of publications by year included in scoping review.

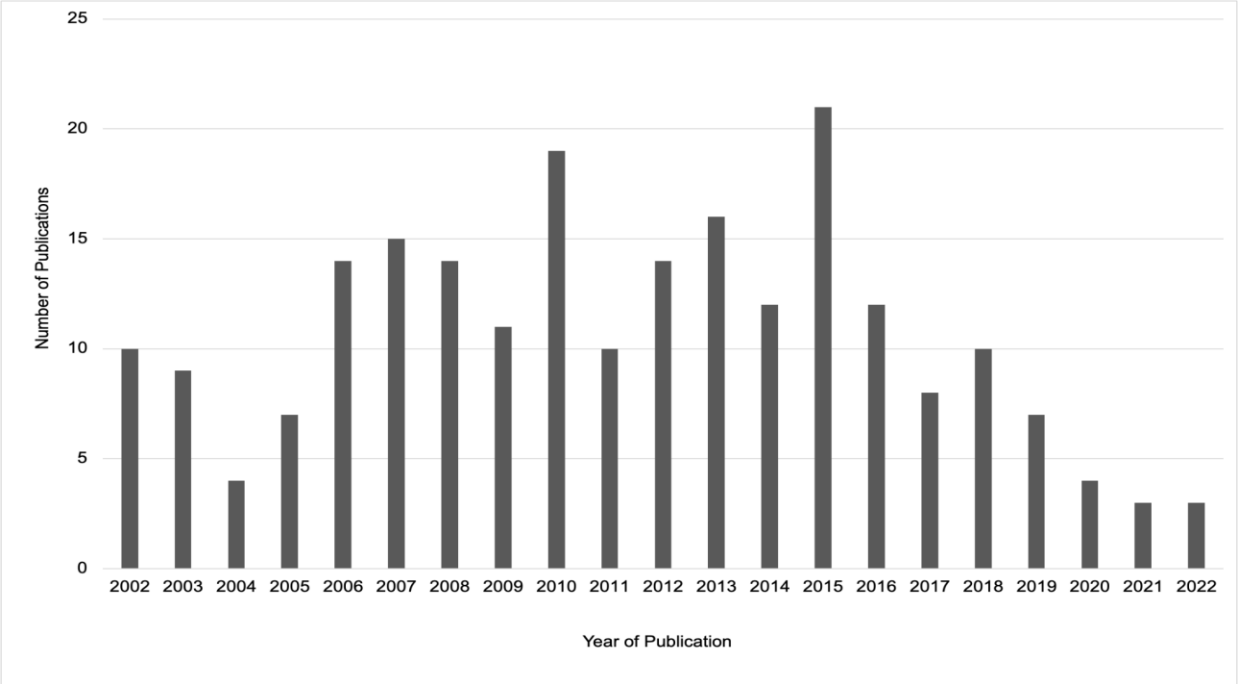


Table 2

Summary of publications in scoping review.

Journal Name	Number of publications included in review
Assessment & Evaluation in Higher Education	18
International Journal for the Scholarship of Teaching and Learning	11
Economics of Education Review	9
Teaching Sociology	9
Higher Education Research & Development	6
Nurse Educator	6
Research in Higher Education	6
Journal of Nursing Education	5
Journal of the Scholarship of Teaching and Learning	5
Decision Sciences Journal of Innovative Education	4
Journal of Chemical Education	4
Scholarship of Teaching and Learning in Psychology	4
International Journal of Educational Management	3
Journal of Education for Business	3
Statistics Education Research Journal	3
Teaching in Higher Education	3
Teaching of Psychology	3
Journal of Economic Education	3

Levels of Study

Using Simmon's (2020) work on the 4M framework's application of SoTL, we examined the levels at which the scholarship was explored at in the studies to better understand the scope and context of the research being conducted. The 4M framework has four levels: where *micro* refers to the level of the individual (in this sense, scholarship taking place in a single classroom), *meso* to department level (e.g., the same course across different instructors), *macro* to institutional level (e.g., different courses across different instructors), and *mega* to inter-institutional/disciplinary/interdisciplinary level. The scoping review revealed just over half (54%, n = 121) of the studies included after screening were focussed on investigations of SET tools and processes themselves, rather than using SETs as a form of data to explore questions related to teaching and learning practices. These studies offered insight into SoTL work with impact largely at the *macro* or *mega* level and were distinctly different from studies using SET data at the *micro* and *meso* levels. For example, studies at the course and department level that used SET data were often conducted by individual or groups of faculty who were using the data with the intention of enhancing or improving student learning through their teaching practice; however, studies at the broader institutional level largely focus on questions of validity, reliability, and bias within SETs

and were carried out by researchers that were not related to the courses in which the SET data was pulled from. We note, however, that the results of this second group of studies can directly impact the use and perception of SET data for SoTL at the micro and meso levels. To gain deeper insight into the broad use of SET data in SoTL across all levels, we felt it was necessary to include both use types in this review and to separate them to support our analysis of use at the different levels. Therefore, we present here the 223 articles as two distinct groups of comparison, Group A (micro, meso) and Group B (macro, mega) where we define Group A articles as those that use SET data to explore SoTL questions at the instructor/departmental level (e.g. how a course change impacted student experience or teaching approaches) and Group B articles as those that use SET data to investigate some aspect of SET tools and/or data, which may or may not include questions framed within a SoTL lens (e.g. the effect of incentives on response rate, relationship between instructor identity and SET responses). Examples of Group A and Group B articles are outlined in Table 3.

Table 3

Categories of studies from scoping review

Group	Expanded Definition	Examples
A	Works that use SETs as a source of data for investigations largely at the course or departmental level; they do not aim to answer questions about the SET instrument or process itself but rather use SET data to answer questions about student perceptions and/or experience of learning.	Impact of instructional technologies on student learning and evaluation of instruction (e.g., Carpenter et al., 2013) Examining student perceptions of learning experience after introducing active learning techniques (e.g., Flynn, 2015) Examining students' conceptions of topics related to teaching and learning (e.g. Mitchell, 2013) Comparison of student perceptions of learning and achievement of learning (van Sickle, 2016)
B	Works that use SETs to investigate the SET process or tool either for the purpose of continuous improvement of SETs, to better understand how students engage with the tool, or to measure and assess its appropriateness as a source of data.	Examining the mode of SET delivery (e.g. Ardalan et al., 2007) Identifying factors that impact response rates (e.g. Goodman et al., 2015) Examining factors that impact ratings of teaching effectiveness outside of the instructor's control such as race, gender, class size (e.g. Bedard & Kuhn 2008; Boring, 2017)

Note: Figure 2 provides a chronological overview of studies with the majority being published prior to 2016. Thus, the studies listed above as examples are somewhat limited by date.

Context of Use

Of the 223 studies, those falling into Group A (46%, n = 102) and Group B (54%, n = 121) were reviewed for two core areas of context: (1) what was the overall purpose of the work at hand, and (2) what was the scope of data included in the study (e.g., single/multiple courses and/or single/multiple iterations of a course). Our survey data was used to provide additional insight into these questions.

Studies at the course and department level (Group A) primarily used SETs to assess the impact of changes to course curriculum (34.3%, n = 35) or decisions around instructional approaches (34.3%, n = 35), while exploring/assessing teaching effectiveness (9.8%, n = 10), evaluating student experience including learning within a new course (8.8%, n = 9), and the assessment of a specific technology-based teaching tool (4.9%, n = 5) were identified less frequently. The studies showed a broad variation in the numbers and range of SET data being employed. They applied SET data to explore questions of instructional/pedagogical effectiveness over multiple iterations of a single course (39.2%, n = 40), multiple courses taught by different instructors (37.3%, n = 38), or a single iteration of a single course (16.7%, n = 17).

As previously noted, Group B studies largely examined factors that impact SET data (86%, n = 104) and examinations of SETs as a measure of effectiveness and/or productivity (e.g. for tenure and promotion (10.7%, n = 13)), rather than inquiry into teaching and learning directly. They predominantly used data from multiple courses over multiple instructors (96.7%, n = 117, Table 4) - as is expected at the macro level (Simmons, 2020).

Table 4

Application of data as determined via scoping review for both Group A and Group B studies.

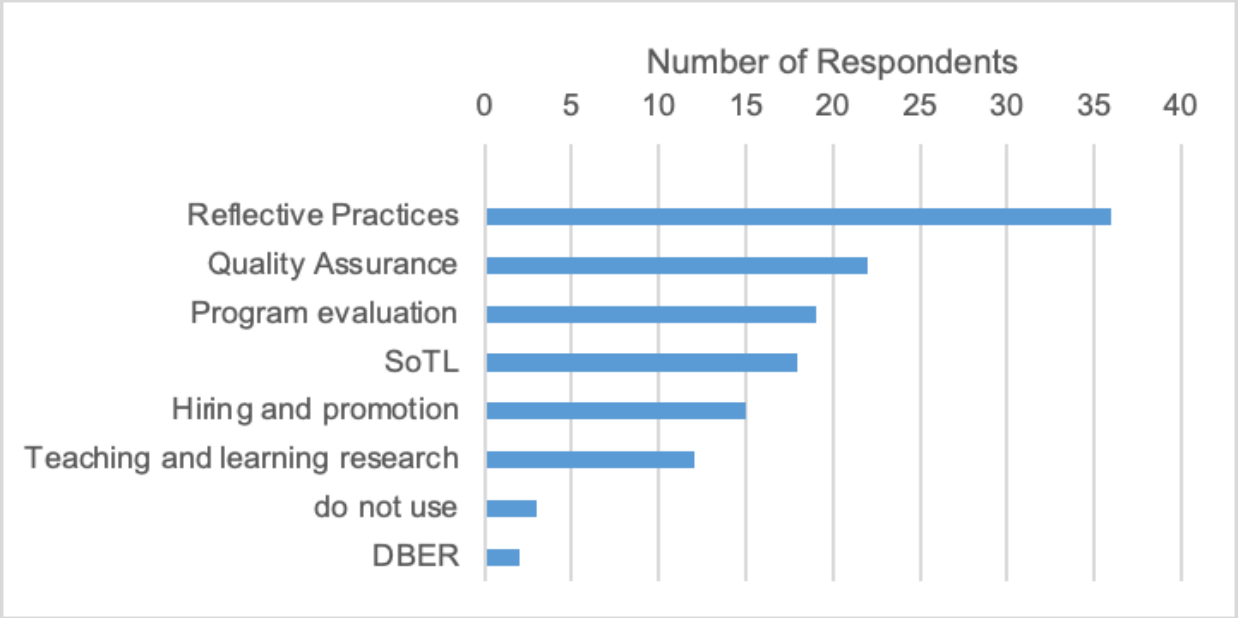
Context	Total Number of Studies	
	Group A (n = 102)	Group B (n = 121)
Single Course, Single Iteration	17	-
Single Course, Multiple Iterations	40	-
Multiple Courses, Multiple Instructors	38	117
Multiple Courses, Single Instructor	4	2
Other/unclear	3	2

Note. Multiple course studies may include some combination of single and multiple iterations.

Responses to our survey aligned with the results of the scoping review and provided further insight into why SET data was being used in scholarly practice more generally (often a precursor or parallel practice with SoTL). Participants indicated using SET data across a range of teaching activities (Figure 2), such as reflective practice (76.6%, n = 36), quality assurance (46.8%, n = 22), program evaluation (40.4%, n = 19), and SoTL (38.3%, n = 18). Participants shared that they use "...data to help guide faculty through program quality reviews (to identify program strengths, challenges, and recommendations" (Participant 023), that they use SETs to "reflect on my own teaching practice, demonstrate improvement of my teaching over time" (002) and "I like to 'read between the lines' to gain a sense of what students are struggling with in terms of course tools: which tools work well, and which need adjustment" (021). Other uses noted were exploring the learner experience (8.5%, n = 4) and institutional research and development (8.5%, n = 4).

Figure 2

Reported uses of SET data in teaching and learning by survey respondents (n=47). Respondents selected all that apply.



Over 42.6% of participants (n = 20) agreed that they use SET data as part of formal research on teaching (e.g. SoTL), and of those, the majority (75.0%, n = 15) indicated that they use it to explore course-level SoTL questions, compared to departmental (40.0%, n = 8), institutional (20.0%, n = 4), multi-departmental (10.0%, n = 2), or multi-institutional (5.0%, n = 1) level inquiries. Respondents of our survey who indicated they were using the data for research purposes were largely those holding either or both faculty (75%, n = 15) and administrator (35%, n = 7) roles.

What SET Data is Being Used?

Table 5 compares the types of SET data used in research (e.g. open-ended, closed-ended, aggregated departmental summaries, etc.) from both survey participants who use SET data for SoTL and our findings of Group A from the scoping review. Survey participants reported the use of both open- (90.0%, n = 18) and closed-ended (70%, n = 14) questions for SoTL use, while Group A studies revealed more studies included data from closed- (86.3%, n = 88) than open-ended (49.0%, n = 50) questions. No survey participants indicated using only closed-ended data from SETs, in contrast to the scoping review where many studies reported only data from closed-ended questions (47.1%, n = 48). About 39.2% (n = 40) of studies reported both data types. Interestingly, 90.0% (n = 18) of survey respondents who use SET data for SoTL reported that open-ended questions were the most useful form of SET data for SoTL "*It can be useful for exploratory or confirmatory use combined with other data*" (011).

Table 5

Comparison of types of data used from SET for research: Survey group vs published reports

SET Data Type	Survey (n = 20)	Scoping review	
		Group A (n = 102)	Group B (n = 121)
aggregated/departmental scores	3	n/a	n/a
open-ended	18	10	7
closed-ended	14	48	97
open + closed	n/a	40	6
self-created/personalised Qs	8	n/a	n/a
statistical summaries	10	-	11
unclear/other	n/a	4	-

Note. Survey participants also reported using both end-of-term (n = 16) and mid-course (n = 5) data.

The majority of participants who use SET data for research purposes (95.0%, n = 19) and the majority of Group A (92.2%, n = 94) and Group B (81.8%, n = 99) studies indicated that SET data was used in conjunction with other sources of data, such as student grades, student and faculty demographics (e.g., gender, major), course characteristics (e.g., online vs in-person, major requirement vs elective), surveys that included other measures, pedagogical or instructional information (e.g., active vs didactic), and faculty research productivity (Table 6).

Table 6

Types of data used in conjunction with SET data for research: Survey group vs published reports

Data Type	Survey (n = 19)	Scoping Review Group A (n = 102)	Scoping Review Group B (n = 121)
Only SETs	-	8	22
Student Grades	12	46	33
Pedagogical or instructional approach	5	32	7
Other survey measures	15	48	15
Classroom observations	4	9	0
Student or faculty focus groups / interviews	10	16	0
Course characteristics	7	16	48
Student demographics	-	18	29
Faculty demographics	-	10	43
Faculty research productivity	-	4	7

Discussion

What may be impacting the use of various forms of data?

Data from our survey, scoping review, and experience suggests that the use of open- and/or closed-ended questions may be influenced by several factors, including availability of and access to data, support and tools for data analysis, the appropriateness of data to answer SoTL questions, and shifts in institutional approaches for data collection. When asked to consider if there were any institutional supports that could be provided that would make it more likely to use SET data for SoTL, participants reported that having the ability to customise questions (14.9%, n = 7) and/or being able to connect SETs to demographic data (12.8%, n = 6) would be two supports for increasing their use of this data for research purposes, and we believe that this may also help with perceptions and issues of bias and validity with SET data. Triangulation (17.0%, n = 8) was one of the top three suggested best practices when using SET data in SoTL. One participant explained, "I only use it in combination with other data. It can be useful for exploratory or confirmatory use combined with other data" (009). Additional best practices highlighted by respondents included transparency to students in how the data is used (25.5%, n = 12) and following accepted practices in education and SoTL research (e.g. protecting student identity (14.9%, n = 7).

Our scoping review revealed that the SET instrument itself and the usefulness of its data is a significant area of interest in the literature (Group B studies), and lead to two points of further inquiry in our work:

> (1) Given the nature of these studies as typically larger scale, institutional or longitudinal studies examining multiple courses with multiple instructors looking to generalise the student and instructor experience, it is of little surprise to us that there would be a reliance on closed or quantitative data for this work (as confirmed in Table 5), and
>
> (2) The perceptions of the quality of SET data for SoTL was identified by survey participants as a barrier to use. In fact, data quality (57.4%, n = 27, e.g., low response rates, limited data points, imbalance in sample sizes) and perceptions of bias in student responses (27.7%, n = 13) were the two most identified barriers for the use of SET data in SoTL. For example, participants shared, "Student response rates are so low that I do not consider the data to be accurate or representative" (006), and "There is a common sentiment that they are biassed against female and racialized faculty. This sentiment causes some faculty to dismiss them completely." (002)

Additional barriers focused on the logistics (e.g. limitations in timing and design of questionnaire) and access/permission of using the data (34.0%, n = 16). For example, "Limitations around when/where I can use the data (internal use only versus use in publications/conference presentations." (017). Of the 223 studies examined in the scoping review (both Group A and B), only a small portion noted obtaining ethics approval (11.2%, n = 25) or indicated they were exempt (8.5%, n = 19). In contrast, over half of survey participants indicated that ethics approval was required by their institution to use SET data in SoTL (51.1%, n = 24), while only 36.2% (n = 17) said they did not know their institution's stance on this topic. Lastly, participants mentioned that

often SET questions did not provide relevant insight into their SoTL questions (14.9%, n = 7) and 44.7% (n = 21) said that, to their knowledge, instructors at their institution were not able to personalise questions on their instruments to enhance that relevance. Thus, for many, relevance of SET data for SoTL use was another major barrier to its integration into SoTL inquiry.

Implications for Practice

This work explored how SET data is, or can be, utilised in SoTL, and what existing barriers limit their use across academia. Informed by the results of our survey and scoping review, we have developed a list of considerations for how institutions might create pathways for instructors to use this data to explore meaningful questions around teaching practices, student engagement, learning processes, and curriculum changes. We offer six implications for practice.

First, provide leadership in addressing concerns regarding the validity and reliability of SETs by engaging in ongoing assessment of institutional SET instruments and procedures, as well as providing transparency and guidance for how and when the data can be used for SoTL purposes. Institutions might consider an on-going holistic approach to reviewing, revising, drafting, and validating SET questions to ensure the tool incorporates both student and faculty perspectives. Second, offer the opportunity for individual instructors to add personalised close- and open-ended questions to their individual SETs to allow for the collection of data that can be more relevant when exploring SoTL questions (e.g. related to course elements, instruction, program evaluation). In doing so, this provides instructors with the opportunity to strategically incorporate aspects of their pedagogy that they would like to explore through a SoTL lens, while also increasing instructor's ability to have greater autonomy of the narrative that their SETs provide. At an institutional and programmatic level, this functionality opens opportunities for collecting program-level student feedback as part of evidence-based program review and evaluation and offers the opportunity to explore SoTL at micro- to mega-systemic levels.

Third, streamline ethics approval processes for use of SET data for SoTL and clearly outline in policies of what SET data can/cannot be used for SoTL purposes. For many instructors, institutional bureaucratic barriers limit their desire to utilise this data beyond routine reflection of teaching practices and/or teaching evaluation processes like tenure and promotion. If institutions have clear and standard procedures for the use of this data, staff supporting SoTL at institutions can purposefully direct programming that targets ethical and appropriate use of this data. Fourth, if possible, provide options for the use of demographic data (both instructor and students) in relation to SET data, and offer opportunities for instructors to learn how to responsibly and ethically use demographic information and SET data in their SoTL. For many, the well-documented biases embedded in this data dissuades scholars in using it to explore SoTL questions. However, when appropriate, allowing instructors access to student demographics not only allows them to control for these biases in their SoTL work, it also opens the opportunity to explore this issue more widely, potentially contributing to scholarship that seeks to improve the use of SET tools. The ability to disaggregate teaching and learning data so we can best understand how (and whether) we offer equitable and inclusive education is a considerable area of growth in SoTL and higher education administration and leadership research in recent years.

Fifth, offer institutional-level systems that allow for enhanced data visualisation, reporting, and dissemination capabilities of both open- and closed-ended data at instructor, course,

departmental, and institutional levels. If instructors can easily view trends across their courses while reviewing SET data, even at aggregate levels, they may be prompted to ask and explore questions about their teaching and learning through a scholarly lens. In this way, providing greater capabilities to easily review their data offers more pathways for greater use of this data within their teaching and scholarly development. Finally, Provide resources for instructors, faculty, and staff as they engage in and develop capacity to analyse and accurately represent qualitative and quantitative data as part of research on teaching and learning. For the above implications to be successful, institutions must also explore additional resources for instructors/faculty and staff to utilise this data for scholarship purposes. Given that the administration and utilisation of SETs is largely across several institutional units, this offers the opportunity for cross-collaboration with research services, informational technology, and teaching and learning centres.

Limitations

Our ability to fully explore the application of this type of data in SoTL was limited by the number of respondents and by the nature of scoping reviews. While our work revealed trends in how SET data has been applied to SoTL questions and what barriers to use exist, the review was only exploratory in nature, thus limiting both generalizability and causation. Systematic studies would provide a finer analysis of the relationship of SoTL and SET data, including the use and effectiveness in exploring SoTL questions, and how barriers to use impact that effectiveness (e.g., reliability and validity of data, institutional policies to data access for research). Two areas of future exploration include (1) developing a better understanding of the impact of bias and data quality on the ability to make inferences and conclusions about pedagogy, course design, curriculum and the student experience, and (2) exploring the alignment between thoughtful and intentional research design and methods, and the use of SET data to answer questions related to teaching and learning.

We note also that the scoping review revealed many publications present in discipline-specific education journals (e.g., *Teaching Sociology, Journal of Nursing Education, Journal of Chemical Education,* and *Statistics Education Research Journal* (see Table 2)). Researchers often apply their own disciplinary norms and expectations when submitting to these journals with reviewers most often from the same general field rather than specialising solely in teaching and learning research. We would expect then that the context of SET use and how it is applied in use with other data is likewise affected by these disciplinary norms. While our survey outreach methods intended to target those working across groups and disciplines in teaching and learning, the catch-all type of listservs we used for survey promotion may not have captured the full breadth of discipline-based education research, and thus that lens of comparison between the scoping review and survey is difficult. An expanded survey net that targets discipline-specific listservs may provide further nuance to disciplinary norms of SET use in research.

Another limitation to this work is the time difference between those who completed the survey and the sample from which we reviewed data for the scoping review. A large portion of the literature originated prior to 2015. The tools available to researchers now including machine learning and question personalization have changed the way researchers view and use SET data. This shift in data use may not yet be reflected in our scoping review results.

Finally, we noticed a decline in the number of available publications in recent years (after 2016, Figure 2) that use SET data for research purposes. Though the reasons behind such a decline are beyond the scope of this work, a variety of factors are likely contributing to this trend including: increased attention towards the bias of SET data impacting both validity and reliability of the tool (see Group B studies), movement to digital collection of SET data rather than physical form resulting in lower response rates across institutions, and pivots in evaluation strategies in response to the COVID-19 pandemic may all be factors that influence the landscape of literature on this topic. Future studies may focus on the impact of ongoing changes to how SETs are administered (e.g. introducing the ability to add customised questions), the increased expectation on faculty to engage in SoTL as a regular part of teaching activities and their importance to tenure and promotion processes, and technological advancements that allow easier analysis of these types of data (e.g. text mining).

Conclusion

While there is a large body of work examining the structural bias and procedural processes of SETs (e.g., timing of administration, online vs paper), less is known about how SETs are used for SoTL purposes. This work provides insight into this gap by both exploring existing SoTL literature for SET use in teaching and learning research, as well as asking SoTL researchers about their current use of SETs in their work. While perspectives of SETs effectiveness as a measure of teaching quality remains debated, current SoTL work suggests that when used in different contexts and with additional sources of evidence on teaching and learning practices, SETs can offer a range of quantitative and qualitative data. Rather than focusing on a culture where SET data is excluded from research on teaching and learning, institutions can take a range of approaches to increase pathways of access and processes for contextualising SET data that support faculty use to improve their teaching.

Acknowledgements

The authors discloses that they have no actual or perceived conflicts of interest. The authors disclose that they received the financial support of Explorance for their BlueNotes Faculty Research Grant in the production of this work. The authors have not used artificial intelligence in the ideation, design, or write-up of this research as per Crawford et al. (2023). The authors confirm that they have met the ethical standards expected as per Purvis and Crawford (2024). The authors list the following CRediT contributions: **Moore**: Data curation, Formal analysis, Visualisation, Writing-Original Draft; **McSweeney:** Project administration, Funding acquisition, Conceptualisation, Methodology, Formal analysis, Supervision, Writing-Original Draft, Writing-Review & Editing; **Gillis**: Conceptualisation, Data curation, Methodology, Formal analysis, Visualisation, Writing-Original Draft, Writing-Review & Editing.

References

Ali, A., Crawford, J., Cejnar, L., Harman, K., & Sim, K. (2021). What student evaluations are not: Scholarship of Teaching and Learning using student evaluations. *Journal of University Teaching & Learning Practice*, *18*(8). https://doi.org/10.53761/1.18.8.1

Ardalan, A., Ardalan, R., Coppage, S., & Crouch, W. (2007). A comparison of student feedback obtained through paper-based and web-based surveys of faculty teaching. *British Journal of Educational Technology*, *38*(6), 1085–1101. https://doi.org/10.1111/j.1467-8535.2007.00694.x

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bailey, R. L., Kiesel, V. A., Lobene, A. J., & Zou, P. (2020). Redesigning an undergraduate nutrition course through active learning and team-based projects enhances student performance. *Current developments in nutrition*, *4*(4). https://doi.org/10.1093/cdn/nzaa039.

Bedard, K., & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review*, *27*(3), 253–265. http://dx.do]i.org/10.1016/j.econedurev.2006.08.007

Bengtson, C., Ahlkvist, M., Ekeroth, W., Nilsen-Moe, A., Proos Vedin, N., Rodiuchkina, K., Ye, S. & Lundberg, M. (2017). Working as partners: course development by a student–teacher team. *International Journal for the Scholarship of Teaching and Learning*, *11*(2), 6. https://doi.org/10.20429/ijsotl.2017.110206

Boring, A., Ottoboni, K., & Stark, P. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*.0(0), 1-11. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, *145*, 27–41. https://doi.org/10.1016/j.jpubeco.2016.11.006

Botnaru, D., Maurer, T. W., & Langdon, J. (2022). SoTL at Georgia Southern: Perceptions, engagement and impact. *International Journal for the Scholarship of Teaching and Learning*, *16*(1), 4. https://doi.org/10.20429/ijsotl.2022.160104

Boyer, E. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton University Press. https://www.umces.edu/sites/default/files/al/pdfs/BoyerScholarshipReconsidered.pdf

Calvo, R., Markauskaite, L., & Trigwell, K. (2010). Factors Affecting Students' Experiences and Satisfaction about Teaching Quality in Engineering. *Australasian Journal of Engineering Education*, *16*(2), 139-148. https://doi.org/10.1080/22054952.2010.11464049

Carbone, A. (2014). A peer-assisted teaching scheme to improve units with critically low student satisfaction: Opportunities and challenges. *Higher Education Research and Development*, *33*(3), 425-439.https://doi.org/10.1080/07294360.2013.841644

Carpenter, R., Theeke, L., & Smothers, A. (2013). Enhancing course grades and evaluations using distance education technologies. *Nurse Educator*, *38*(3), 114–117. https://doi.org/10.1097/NNE.0b013e31828dc2d7

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of marketing education*, *31*(1), 16-30. https://doi.org/10.1177/0273475308324086

Crawford, J., Cowling, M., Ashton-Hay, S., Kelder, J. A., Middleton, R., & Wilson, G. S. (2023). Artificial intelligence and authorship editor policy: ChatGPT, Bard Bing AI, and beyond.

Journal of University Teaching and Learning Practice, 20(5). https://doi.org/10.53761/1.20.5.01

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced Mixed Methods Research Designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 209-240). Thousand Oaks, CA: Sage

Felten, P., & Chick, N. (2018). Is SoTL a signature pedagogy of educational development? *Improve the Academy*, *37*(1), 4–16.https://doi.org/10.1002/tia2.20077

Flynn, A. B. (2015). Structure and evaluation of flipped chemistry courses: Organic & spectroscopy, large and small, first to third year, English and French. *Chemistry Education Research and Practice*, *16*(2), 198–211. https://doi.org/10.1039/C4RP00224E

Godbold, N., Irving-Bell, D., McSweeney-Flaherty, J. M., Prusko, P. T., Schlesselman, L. S., & Smith, H. (2021). The Courage to SoTL. *Teaching & learning inquiry*, *9*(1), 380-394. https://doi.org/10.20343/teachlearninqu.9.1.25

Goodman, J., Anson, R., & Belcheir, M. (2015). The Effect of Incentives and Other Instructor-Driven Strategies to Increase Online Student Evaluation Response Rates. *Assessment & Evaluation in Higher Education*, *40*(7), 958–970. https://doi.org/10.1080/02602938.2014.960364

Harsy, A., Carlson, C., & Klamerus, L. (2021). An Analysis of the Impact of Mastery-Based Testing in Mathematics Courses. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies,* *31*(10), 1071-1088.http://dx.doi.org/10.1080/10511970.2020.1809041

Hubball, H., & Clarke, A. (2010). Diverse methodological approaches & considerations for SoTL in higher education. *Canadian Journal for the Scholarship of Teaching and Learning*, *1*(1), 2. http://dx.doi.org/10.5206/cjsotl-rcacea.2010.1.2

Klopfer, M. D., Knight, D. B., Grohs, J. R., & Case, S. W. (2024). Course-level factors and undergraduate engineering students' ratings of instruction. *European Journal of Engineering Education*, *49*(4), 769-784. https://doi.org/10.1080/03043797.2023.2224245

Kolomitro, K., & Anstey, L. M. (2017). A survey on evaluation practices in teaching and learning centres. I*nternational Journal for Academic Development, 22*(3), 186-198.http://dx.doi.org/10.1080/1360144X.2017.1313162

Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. Journal of Academic Ethics, 1-12. https://doi.org/10.1007/s10805-021-09400-w

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106.https://doi.org/10.1016/j.stueduc.2016.12.004

Mahendra, N. (2018). Teaching about aphasia: Speech-language pathology students' perceptions of different instructional techniques. *Aphasiology*, 32, 136-137. https://doi.org/10.1080/02687038.2018.1489122

Maurer, T. (2006). Cognitive Dissonance or Revenge? Student Grades and Course Evaluations. *Teaching of Psychology*, *33*(3), 176–179. https://doi.org/10.1207/s15328023top3303_4

McDonald, M. A., Noble, M. A., Harris, B., Cortés, V., & Jeffery, K. (2024). Fostering a teaching and learning opportunity: Toward equity in student feedback of teaching. *Papers on Postsecondary Learning and Teaching*, *7*, 113-129. https://doi.org/10.55016/ojs/pplt.v7Y2024.77561

McKinney, K. (2006). Attitudinal and structural factors contributing to challenges in the work of the scholarship of teaching and learning. *New Directions for Institutional Research*, *129*, 37–50. https://doi.org/10.1002/ir.170

Middendorf, J., & McNary, E. (2011). Development of a Classroom Authority Observation Rubric. *College Teaching, 59*(4), 129-134. https://doi.org/10.1080/87567555.2011.580690

Mitchell, E. (2013). Reflective course construction: An analysis of student feedback and its role in curricular design. *Education for Information*, *30*(3–4), 149–166. https://doi.org/10.3233/EFI-130942

Mitchell, K. M. W., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, *51*(3), 648–652. https://doi.org/10.1017/S104909651800001X

Mowatt, R. A. (2019). Twelve years a servant: Race and the student evaluation of teaching. *SCHOLE: A Journal of Leisure Studies and Recreation Education*, *34*(2), 109–110. https://doi.org/10.1080/1937156X.2019.1622949

Murray, H. G. (2005). Student evaluation of teaching: Has it made a difference. In *Annual meeting of the society for teaching and learning in higher education* (pp. 1-15).

Otu, N., & Otu, N. E. (2023). Student Evaluations of Teaching Are Mostly Awfully Wrong. *Universal Journal of Educational Research*.

Potter, M. K., & Kustra, E. (2011). The relationship between scholarly teaching and SoTL: Models, distinctions, and clarifications. *International Journal for the Scholarship of Teaching and Learning*, *5*(1), 23. https://doi.org/10.20429/ijsotl.2011.050123

Purvis, A.J. & Crawford, J. (2024). Ethical Standards in Social Science Publications. *Journal of University Teaching and Learning Practice, 21*(9). https://doi.org/10.53761/hqnqr710

Schnurr, M. A., & Taylor, A. (2019). Bridging the Gap between the Research Ethics Board and the Scholarship of Teaching and Learning. *Canadian Journal for the Scholarship of Teaching and Learning*, *10*(1), n1. https://doi.org/10.5206/cjsotl-rcacea.2019.1.8003

Simmons, N. (2020). The 4M Framework as Analytic Lens for SoTL's Impact: A Study of Seven Scholars. *Teaching & Learning Inquiry*, *8*(1), 76-90. http://dx.doi.org/10.20343/teachlearninqu.8.1.6

Slocum-Schaffer, S., & Bohrer, R. (2019). Information Literacy for Everyone: Using Practical Strategies to Overcome 'Fear and Loathing' in the Undergraduate Research Methods Course. *Journal of Political Science Education*, 1-17. https://doi.org/10.1080/15512169.2019.1694935

Sorcinelli, M. D., & Austin, A. E. (2006). Developing faculty for new roles and changing expectations. *Effective practices for academic leaders, 1(*11), 1-16.

Stoesz, B. M., De Jaeger, A. E., Quesnel, M., Bhojwani, D., & Los, R. (2022). Bias in student ratings of instruction: a systematic review of research from 2012 to 2021. *Canadian Journal of Educational Administration and Policy*, (201), 39-62. https://journalhosting.ucalgary.ca/index.php/cjeap/article/view/73769

Svanum, S., & Aigner, C. (2011). The influences of course effort, mastery and performance goals, grade expectancies, and earned course grades on student ratings of course satisfaction. *British Journal of Educational Psychology*, *81*(4), 667–679. https://doi.org/10.1111/j.2044-8279.2010.02011.x

Uttl, B. (2024). Student Evaluation of Teaching (SET): Why the Emperor has no clothes and what we should do about it. *Human Arenas*, *7*(2), 403-437. https://psycnet.apa.org/doi/10.1007/s42087-023-00361-7

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

Van Sickle, Jenna R. (2016). Discrepancies between student perception and achievement of learning outcomes in a flipped classroom. *Journal of the Scholarship of Teaching and Learning*, *16*(2), 29–38. https://doi.org/10.14434/josotl.v16i2.19216

Wilson, R., Georgakis, S., & Hu, X. (2011). Meeting the Challenges of Electrical Engineering Service Courses. *Australasian Journal of Engineering Education*, 17(2), 91-100. https://doi.org/10.1080/22054952.2011.11464062

Wuetherick, B., Yu, S., & Greer, J. (2016). Exploring the SoTL landscape at the University of Saskatchewan. *New Directions for Teaching and Learning*, *2016*(146), 63–70. https://doi.org/10.1002/tl.20188

Zumbach, J., & Funke, J. (2014). Influences of mood on academic course evaluations. *Practical Assessment, Research & Evaluation*, *19*. http://files.eric.ed.gov/fulltext/EJ1031265.pdf