Received: 8 April 2024 | Revised: 11 May 2024 | Accepted: 31 May 2024 | Published online: 6 June 2024

RESEARCH ARTICLE

International Journal of Changes in Education
2024, Vol. 1(4) 169–176
DOI: 10.47852/bonviewIJCE42023045

BON VIEW PUBLISHING

Unintended Consequences: When Innovation in Pedagogy Impacts Student Evaluations

Rosemary Fisher[1,*], Chamila Perera[1] and Richard Laferriere[1]

[1]*School of Business, Law and Entrepreneurship, Swinburne University of Technology, Australia*

Abstract: Student evaluation of teaching surveys (SETS) is one of the most controversial tools used in higher education, globally. Recently the impact that SETS had on innovation in pedagogy has been raised as a potential problem. The teaching team experienced a gap between the SETS results generated for a final-year undergraduate class that utilized flipped learning as the innovation and an empirical research project using the same cohort of students that investigated the effectiveness of that pedagogy. Accordingly, from 2014, we established this cross-sectional sequential study over seven semesters to understand how students used the SETS after experiencing this innovation. We conducted a thematic analysis on the 588 SETS results from a final-year undergraduate class studying at an Australian university and found resisting students used the SETS as a weapon with adoption of the innovation of the prime casualty. We recommend SETS be tailored for use where students experience innovation in pedagogy as the use of SETS by students may undermine the adoption of otherwise effective pedagogy and impact the willingness of faculty to pursue innovation in pedagogy.

Keywords: flipped learning, student evaluations, pedagogy, innovation

1. Introduction

Due to the rapidly changing competitive landscape and other challenges in the higher education sector, universities are actively exploring innovative teaching models in an effort to attract, engage, and retain students and ensure their learning outcomes competitively position them for the workforce as job-ready graduates of the digital age. The result of these efforts has seen a wide variety of new pedagogies emerge, of which the flipped learning approach is one.

Since its emergence into the higher education sector, flipped learning has enjoyed a rapid uptake as an approach that responds to the need for innovations in the sector, particularly those that leverage digital assets and improve student outcomes including satisfaction with the learning experience. Flipped learning is a philosophical teaching approach where the learning activities, independently achievable, are specified as homework and class time is reserved for seeking deeper richer learning through problem exploration with the lecturer and peers [1, 2]. Flipped learning has been credited with improving student outcomes including performance, engagement, and overall satisfaction with the learning experience, although results are mixed and student satisfaction can be negatively impacted [2–5]. Its contribution to the continuous improvement of critical thinking, which in turn leads to the capacity to make sound judgments, makes flipped learning a compelling pedagogy suited to any discipline [6].

Problematizing this innovate-on-pedagogy strategy is the rise in the importance of the student as consumer, where student satisfaction with their learning experiences is increasingly used as a proxy for evidence the university has delivered on its value proposition [7–10]. Typically collected via the end of semester, student evaluations of teaching surveys (SETS) are recognized as a blunt but cost-effective mechanism seen by some to have little or questionable value [11, 12] in part because their interpretation is not well understood by those who utilize them [13]. Nonetheless, they remain an influential source used to inform curriculum, teaching-related decisions, academic career progression [8, 9, 13], and university policy [10]. This is despite these surveys being well recognized as surveys that deal with student satisfaction with their learning experience and not satisfaction with the primary purpose of a university education – the achievement the of learning outcomes [14]. Developed as a source of insight for potential pedagogical improvements, the metamorphosis of SETS into a quality assurance mechanism creates fear and resentment among academics [9]. That many institutions discontinued or modified their use of SETS during the coronavirus disease 2019 pandemic-impacted semesters [15] points to knowledge of their potential use as a weapon in the hands of anxious students.

The role of the SETS in the context of innovation in pedagogy is particularly problematic. The trough of disillusionment that characterizes the lag between the implementation and adoption of an innovation [16] is a salient reminder of the time it can take for the normalization of any innovation. In the context of pedagogical innovations, the use of SETS has been questioned because such innovations challenge student expectations about their learning experience, experiment with habituated learning practices, and are potentially disruptive and difficult to master [9]. Thus, there is a recognized but potentially under-researched tension between pedagogical innovations and how students are motivated to use ratings in the face of novel situations.

*Corresponding author: Rosemary Fisher, School of Business, Law and Entrepreneurship, Swinburne University of Technology, Australia. Email: rlfisher@swin.edu.au

There is little research into the perceptions of flipped classrooms expressed through SETS [17] and none that we could find that examines the institution's own SETS in the context of flipped learning. Thus, we situate our research in an open space where research on the impact of novel pedagogy within a SETS context is largely silent [18]. We explore the research question: how do students use the SETS when experiencing pedagogical innovation? In this paper, we present a cross-sectional sequential case study in which the results of the institution's SETS, generated for a flipped learning third-year undergraduate subject, are explored. Emerging from the data analysis are key theoretical perspectives (diffusion of innovations [19], resistance to change [20], and theory of planned behavior [21] that help interpret the findings and understand the broader implications of SETS).

Our research uniquely contributes to the understanding of how student evaluations can inadvertently become tools of resistance against pedagogical innovation. The significance of this research lies in its potential to inform improvements in evaluation methods to better support and reflect the impacts of pedagogical innovations. By aligning evaluation tools with the objectives of modern educational practices, higher education institutions can more accurately gauge the effectiveness of innovative teaching methods and encourage their adoption, ultimately enhancing student learning experiences in preparation for the digital age. We offer our experience as a cautionary tale and invite classroom innovators to consider the impact on SETS when implementing innovation.

2. Literature Review

2.1. Student evaluation of teaching surveys

SETS are distributed to a cohort of enrolled students to gather their perceptions, feedback, opinions, and views on a range of aspects relating to their experience of a particular subject and its teacher. The surveys frequently use a form of scale scoring and often include open-ended questions and/or room for unprompted student comment. SETS are also known as evaluation forms, feedback surveys, satisfaction surveys, ratings, and course evaluations. Said to be the "most researched topic in higher education" [13], SETS are a contentious topic generating an impressive quantity of papers examining them in a diverse array of contexts [22]. The importance and value of SETS are underscored by the quantity of reviews and summaries of the SETS literature that exists [12, 22–28]. SETS have been examined from the point of view of discipline [27], specific influencing factors [29, 30], course delivery platforms [31], and deliberate interventions [32] to list but a few.

Acknowledged as a simple mechanism [33], SETS are used to source information about a teacher's in-class behavior [34] and for career management purposes [13], as well as gaining insights for the purpose of improving pedagogy [9] used to deliver course content. Implemented by most tertiary institutions and often a regulatory requirement of funding bodies, SETS may be intended as best practice but are variously misused, misinterpreted, or misunderstood by faculty, administrators, and the students responsible for completing them [9, 12, 13, 35].

Not surprisingly, reviews of the SETS literature suggest caution with their use and interpretation and have generated both definitive and equivocal findings regarding the value, reliability, and validity of SETS [13, 22, 23, 25, 28]. A re-analysis of three meta-analytic reviews of the SETS research literature found each meta-analytic review suffered from methodological flaws and raised questions about their findings [12]. SETS are recognized as surveys that provide insights into student satisfaction [12, 36] or perceptions of the completed course [13] but not into what drives that satisfaction or those perceptions [36]. Student satisfaction can variously include "teaching performance and teacher characteristics, career-related issues, program/course innovativeness or appropriateness, and classroom facilities among others" [36] as well as generous grading [37].

The challenges identified in using SETS include that students understand these surveys differently [7, 25] and may not be competent to rate a course design, curriculum development, or goals of the course [22]. Furthermore, students use SETS variously to aggress against the teaching staff including the following: to punish, bully, or harass lecturers [37, 38] and vent their frustrations with matters outside the teacher's scope of influence [24].

Recommendations to replace or improve SETS abound. Calls for mechanisms better suited to the complexity of teaching and learning in today's higher education institutions have been made [9], as have calls to use statistical data (means, confidence intervals, and test of significance) to avoid over interpreting small differences [35]. Ultimately, SETS are a source of perceptual data about the collective experience of a subject and need to be thoughtfully and knowledgably examined by users educated in their interpretation if they are to be effective in building a "high quality teaching ecosystem within an institution" [13]. The problems with SETS lead many teachers to see them as an unreliable tool and a threat to their career progression [39].

2.2. SETS as a weapon

An emergent literature reflects an increasing concern that SETS are not just unsuitable for the purposes for which they are used but that they have become a weapon in the hands of both students and institutions [40, 41]. At the root of our concern about the weaponization of SETS is how students use them to express incivility toward their teachers and the consequent career and personal implications of this for faculty. Incivility refers to those disruptive or discourteous verbal and non-verbal actions directed toward others that interfere with optimal learning atmospheres and includes aggression, bullying, and intimidation [38, 42]. SETS have been used to smear or intimidate faculty [38], take revenge for low grades or perceived insults [37], and threaten faculty with negative feedback [42]. Furthermore, this potential use for SETS is known to students [43].

Feldman [42] explains the incivility of students as rooted in one or more of "(a) a need to express power over another, (b) a need for verbal release due to frustration over an apparently unsolvable situation, or (c) a need to obtain something of value" and describes how some students will threaten to use political or social pressure on instructors via the SETS. May and Tenzek [44] explored student classroom bullying of faculty, identifying triggers for such behavior as concern over grades, students-as-consumers beliefs, diversity experiences, and learning expectations on and of students. Burke et al. [38] found understanding of incivility was nascent, and its causes could not be explained solely by situational or personal characteristics of instructor or student. Nordstrom et al. [45] found that students with a consumerist orientation toward their education or narcissistic tendencies were more likely to act in uncivil ways. Vaillancourt [37] demonstrates large effect sizes in experimental studies that demonstrate students use SETS to reward or punish teachers for grades awarded. Complicating the issue of incivility

by students is faculty reluctance to report or address it because of the impact any perception of classroom incompetence or unsuitability may have on career progression [38]. Since institutions use SETS in promotion and tenure applications, they are a primed weapon that can be used against faculty by both institutions [40, 41] and students [42].

3. Method

This study examined the SETS collected over seven semesters from 2014 to 2017 in a third-year on-campus undergraduate class that utilized flipped learning (research approved by the Swinburne University of Technology Human Research Ethics Committee).

The SETS comprise quantitative and qualitative data. In this research, we examine questions asked consistently in each of the seven semesters, "Overall, I am satisfied with this unit" captured on a 10-point scale, and two open-ended questions, "In my opinion, aspects of this unit that could be improved were . . . ." and "In my opinion, the best aspects of this unit were . . . ."

Using NVivo software, textual comments were thematically analyzed by identifying and describing both implicit and explicit ideas within the data set [46, 47]. Two experienced qualitative researchers separated their roles into primary and secondary researcher, and agreement was achieved through iterative discussion of the findings.

3.1. Case background

This new subject was designed such that students were required (marks attached) to complete preparatory content, which was then further explored and expanded upon in the subsequent face-to-face class using active and peer learning strategies. Learning time allocation was estimated based on the hours prescribed and recommended by the curriculum outline. The design objectives were to achieve good learning outcomes of engagement, performance achievement, and overall satisfaction with the learning experience.

In its first four semesters, this subject was the basis of a separate research initiative (Project A) (names withheld), thus allowing for a comparison of the SETS results with this alternative research approach. Using Structural Equation Modeling, Project A concluded the subject was effective in achieving good student outcomes of engagement, performance, and overall satisfaction (name withheld). Qualitative research components did surface areas of dissatisfaction (see name withheld 2014 and 2019).

By contrast with Project A's positive findings, the SETS results received in the same four semesters lead to the subject being identified as in need of quality assurance review. Given the important uses to which SETS are put, we sought to understand how students used them when experiencing innovative pedagogy by investigating the SETS results themselves.

3.2. SETS data gathering procedure

The institution distributed its SETS online via the student portal in the final 4 weeks of each semester and kept it open for a total of 8 weeks; SETS were collected before the institution released final academic results for the subject. The SETS was promoted to students through direct email, popups reminding students to complete the survey, reminders on the institution's social media outlets, lecturer announcements in class, and via the subject's learning management system. From 2015 onward, random draw prize incentives were offered by the institution to encourage completion of the SETS, for example, a $500 travel voucher was offered in one semester. The institution promoted scores over 8/10 as the desired benchmark, with scores below 7 considered indicative of a problematic subject. Over the seven semesters, SETS overall satisfaction rating for this subject averaged 6.5/10.

In comparison, Project A (see name withheld, early cite) invited the same cohort of students to complete an ethics-approved hardcopy survey distributed on the last day of classes; 714 students enrolled in the first 4 semesters the subject was taught completed this survey (response rate = 49%). Project A used a single item indicator measured on a 5-point scale to capture student perceptions of satisfaction with their learning experience ("Overall, class *X* was a satisfying learning experience for me") ($M = 3.55$, $SD = 1.46$). Path model analysis demonstrated the other three variables explored explained 67% of student overall satisfaction with the subject: benefits of this class, experience of engagement with flipped learning, and student perceptions of their performance in the subject.

The SETS data collected over 7 semesters comprised 588 individual responses (response rate = 39%, i.e., $n = 229$) and provided 29 pages of textual comments for analysis. Demographic data were not collected via the SETS. However, collection of demographic data for both Project A and informal classroom use indicates the majority of students are male (53.5%); under 23 years of age (72.3%), in their third year of studies (66%), and domestic students (60–70%).

4. Findings and Discussion

Three final thematic categories emerged from the SETS data analysis: (1) transferring from the known into the unknown, (2) empowering resistance-prone student feedback, and (3) fostering low-effort returns (The summary of open, axial, and selective codes and the final thematic categories is available from the lead author).

The sheer number of complaints compared to compliments in the SETS was confronting. The emotional tone of some of the negative feedback (e.g., "stupid," "horrifically boring," "so draining," "overwhelmed") gave the SETS results the inescapable flavor of pedagogical failure despite much of the negative commentary being focused on the novel structural features of the pedagogy: timetabling (Friday workshops commencing at 5.00 pm), workshop length (3 h), and perceived volume of work (assessed preparatory tasks). While the pedagogy did have its adherents and SETS scores crept up each semester, the students used the written comments predominantly for complaints, concerns, or suggestions for change.

4.1. Transferring from the known to the unknown

Emerging from the data analysis is evidence of a clear reluctance to embrace or adapt to innovative pedagogy and an accompanying tendency to reaffirm the value of established pedagogy. Thus, the SETS became an instrumental mechanism for students to express counter actions and negative comments in response to the novelty of the pedagogy.

> *'We had to learn the modules in the articles before going into a lecture theatre and going over what we found in the articles was backwards to what we have done in the past, meaning that everyone was in a challenging situation of re-learning how to learn.'*

> *'The class began with the mention of a study that said students preferred to study from home and have less contact hours. Personally this is quite the opposite and find that going into class is the best way for me to learn.'*

This confirms previous literature that found students' favorable and unfavorable evaluations of new pedagogy were partly influenced by their established expectations of what pedagogy should be like [4, 48, 49].

The SETS comments rarely reflected on the learning content or the novelty of the teaching approach itself; rather, they focused on comparison with what students expected to see and found absent from the subject's structure. For example, confirming previous research [50, 51], our analysis shows the students had a serious level of uncertainty with respect to the self-driven preparatory work and expected teachers to follow a traditional pedagogy. This manifested as a desire or demand by students to return to a teacher-dominated or teacher-directed form of learning, supporting previous observations that students are dependent on teacher centered pedagogy [2]. *"Go back to a traditional format including of test, assignment and exam. The 'contemporary style of teaching' is not effective . . . stick to what works . . . ."*

Noticeably, some comments suggested that it was positive experiences of group work that lead some students to adapt to the flipped pedagogy. *"In my opinion, the best aspects of this unit were that it was heavily based on group work, and everyone contributed in the classwork which made it more interesting."* This reinforces the importance of collaborative learning through social-constructive pedagogy for effective peer experiences in the learning environment [52], suggesting it is particularly important in the context of implementing innovative pedagogies.

Innovation in pedagogy is inherently designed to change the boundaries of the existing teaching and learning culture, and institutions do understand they should collect student insights using various mechanisms and not rely on "one-size-fits-all" SETS [13]. However, such recommendations are considered impractical to implement with resourcing an oft-cited concern [53], and as our experience demonstrates in the context of student as consumer, the SETS become a weapon that can be used to demand learning revert to expectations. In the context of innovation in pedagogy, SETS are flawed in that they can be used by students to undermine techniques designed to drive self-learning, develop critical thinking skills, and facilitate group engagement. Thus, SETS can act as a hurdle to the design, adoption, or development of innovative pedagogies [9].

4.2. Empowering resistance-prone students

Monitoring is characterized by noticing, correcting, or negatively gossiping about someone [54] and can involve direct reporting to organizational authorities; and when monitoring is a role responsibility, group members are more likely to engage in it [55]. The SETS that is the focus of this research is repeatedly promoted to students over an 8-week period with the phrase "your unit, your say" clearly positioning the students as customers of a service experience. In so doing, the institution could be signaling to students that part of their role as a student is to monitor the subject in which they are enrolled and that their existing expectations are the appropriate basis on which to do this. Monitoring is a form of social control used within organizations to achieve desired behaviors [56] and often manifests as promotion criteria. Our data demonstrate that student feedback ranged from constructive to emotive and sometimes vengeful, providing additional insight into the tension in using SETS as a tool for assurance of quality learning and promotion purposes.

> *'If you will take one bit of advice from me its this, REMOVE THIS SUBJECT FROM THE CORE BUSINESS UNITS. Thanks for your time. If you didn't read that's your loss.'*

In these data, there is clear evidence that students do notice, correct, and negatively gossip about the subject and to a lesser extent the teachers. "I asked other students if they gained anything out of it and they did not." Our data suggest some students believe part of their student role is to ensure the teacher constructs the learning environment in a manner that conforms to their understanding of teaching and learning norms and report teachers via the SETS when they do not. Therefore, we suggest the SETS may empower students to believe their resistance to embracing or adapting to the new norms inherent in the adoption of innovative pedagogy is not only acceptable but part of their role as a student.

The students use terms such as "fairness" and "equality" when negotiating for their preferred learning expectations. In using the SETS to negotiate for their preferences, students see the SETS as representing "the other party" with whom they should bargain for their goal attainment. Some students even saw the SETS as the vehicle by which they could force the subject out of existence, suggesting an almost unionized approach to learning can emerge when students are faced with innovation in pedagogy.

The SETS analysis also shows that the students looked to negotiate and build resistance toward considerations outside the scope of the SETS including teacher performance or characteristics, university requirements to engage in pedagogical innovation, and university administrative decisions (such as timetabling or room allocation). These student-identified flaws served as the ammunition needed to stop or prevent pedagogical innovation which lies at the root of the complaint.

> *'I also wanted to point out that I will take this matter further and make a formal complaint because I felt scammed. The unit cost quite a lot of money, and it was not worth the value I paid for.'*

A boundary condition of this finding is that the negotiations may result from a lack of holistic understanding of the purpose of, and requirements for, successful flipped learning. This is not surprising given students are largely not skilled or experienced enough to evaluate course design, curriculum development, subject or course goals, or their own performance within a course of study [22, 57], and the flipped learning environment was novel to them. Student attempts to negotiate for how they wished to learn stimulated energetic evaluation of the subject and emerged as empowered resistance to the pedagogy.

4.3. Fostering low-effort returns

We found the SETS did not necessarily provide students with an opportunity to reflect on their learning styles, achievement of learning outcomes, or connection to course outcomes. Repeatedly, at the end of each semester, some students used the SETS to foster a teaching and learning culture that made acceptable a lower level of effort than expected by the pedagogy or curriculum. By low effort, we refer to a preference for a subject that demands no changes to ingrained if not routine patterns of engagement for knowledge gain. Primarily in their final year of university, these students had strong views about the type and amount of effort they should be expected to expend for their return.

That return on student effort was specifically the marks available for effort expended and not their intellectual development or acquisition of subject knowledge. Beatty [58] notes that because grades focus students on the achievement of performance, they are diverted from the need to master what has been learned. They enact a direct quid pro quo of time on task for expected grade and willingly forgo the effort needed to achieve increased competency, understanding, and appreciation of the

learning materials [58] and reward easy courses with higher SETS [59]. This effort is an essential feature of flipped learning, and this type of complaint is noticeable in flipped learning feedback [60, 61].

> *'I also think that the amount of work given is far more than what would be done if you had the class each week and have had a number of people tell me they agree.'*

By attaching marks to the preparation, the objective of incentivizing engagement with the required work backfired. Suddenly student learning behaviors were exposed to the teaching team and to the students themselves, leaving no room to fly under the preparation radar and focusing students' attention to the trade-off between preparatory work and marks available for any given task. "Each article requires citation and a 400-word response for 2% grade." This contributed to some students positioning the pedagogy as unfair or unjust.

Given the discrepancy between how engaged students were during the face-to-face classes (as observed by the instructors in each semester) and the SETS results, we speculate that the reputation of the subject may also have contributed to student expectations and perception of the subject. McNatt [62] found that negative information passed to students formed a reputation for the ratee that persisted, even in the face of disconfirming information, and biased performance and ancillary ratings. That is, the reputation of the ratee acted as a "preconceived notion that disrupts accurate rating processes" [62]. In the second iteration of this subject, one student took it upon himself to (as he told the subject convenor) "shut your subject down" and was seen to openly agitate and recruit students to his cause during classes. This contributed further to general talk about the subject, and in subsequent semesters, there were notable instances where others from the same student's discipline volunteered, they had "heard all about this subject." Thus, we suggest a form of contagion occurs for subjects or innovative pedagogies as much as it does for individuals. That is, a negative reputation can persist despite disconfirming evidence adding to the problematic nature of using SETS in the context of innovation in pedagogy.

> *'How innovative the workshop was, with incorporating a new style of learning, less contact time, use of technology and more student to student interaction.'*

When students assess their learning experience based upon their preference for a return on low effort and are oblivious to, or unwilling to accept, they have a special role in their own learning, innovative pedagogical approaches such as flipped learning are undermined by conventional SETS. Students perceive SETS as a means to maintain their predilection for a particular level of effort status quo. Comments directed toward the subject cutting into the student's own time, bemoaning the requirement of preparatory work, or castigating the incentivized approach to encourage learning, all suggest that the flaw is not necessarily the pedagogy but rather the student perceptions of the amount of effort that should be sufficient for the volume of learning expected for a subject. The flipped class, which relies on the completion of all preparatory work, could be particularly dissatisfying for such students. These findings are consistent with the literature that recognizes the resistance of students to flipped learning lies in part to their perceptions regarding acceptable workloads [2, 49].

5. Conclusions and Recommendations

We developed a flipped class in an effort to improve student engagement and generate good student learning outcomes including overall satisfaction and made it the subject of a four-semester empirical research project (Project A) in which the model tested unequivocally supported the benefits of flipped learning. Separately, concurrent with and subsequent to Project A, SETS comments from the same cohort of students were noticeably critical and ratings below the institution's target benchmark. This raised the specter of a credibility gap for the teaching staff that defended the flipped learning pedagogy as successful. Thus, this research analyzed the SETS comments over seven semesters in an effort to understand the gap between the two sources of insight into student perceptions about their learning experience. Our analysis highlights a three-pronged attack strategy used by students – effectively weaponizing the SETS. The direct casualty was the innovation (flipped learning) and collateral damage coalesced around instructor career implications and willingness to innovate and student learning outcomes. To drive SETS up to the benchmark required, we capitulated to student and institution pressure and removed the marks for completing preparatory work and achieved the desired lift in SETS results in subsequent semesters. However, this also resulted in a dramatic reduction in the number of students prepared for the class, arguably undermining an essential feature of flipped learning.

Given the importance of pedagogical innovation to business schools, an open and frank discussion is needed to address the use of SETS within the context of innovation in pedagogy. In this context, a SETS is needed that captures the student voice with respect to their learning experience while bridging the gap between existing theoretic models of innovation in pedagogy and the imperatives that drive the design of the institutional SETS.

The completion of SETS at the end of the semester and before final marks are known positions them as a forum for discontent. As identified in this research, it is the pedagogy that can be viewed as broken or wrong by students reacting to the pressures that coalesce around end of semester. The sticky nature of a negative reputation [62], constraining nature of marks on students learning behaviors [58], and deliberate incivility of students [42] can all be seen emerging via the SETS results in this subject. There was clear evidence that the SETS are knowingly used by some students to resist the pedagogy and by others to voice discontent or anxiety often unrelated to matters of pedagogical quality.

We suggest where students are taught using innovative pedagogy the SETS applied by the institution be constructed explicitly for such situations, much like an intervention. Using an open, fluid, and dialogic/conversation design, could be instrumental in helping both the student and the institution understand the student experience in this context.

How students are treated in the SETS process is important. The paradigm of the student role in SETS could be changed from one where students are anonymous reporters of teacher behavior and customers to be satisfied by their learning experience, to one where students are invited to reflect on the learning they acquired and consider themselves as co-collaborators in the creation of pedagogy. In so doing, this can extend theoretic models that articulate how to develop, implement, and evaluate innovations [63] by articulating how the institution's SETS process can and should fit into such models.

The ongoing decision to ignore the SETS conundrum at the institutional level has consequences. At the micro level, faculty are asked to inductively design and develop pedagogical innovations. However, a discontinuity exists when faculty are then ultimately evaluated using a macro-level tool that in effect serves to maintain the status quo in part because it is designed as an accountability tool rather than one that aims to promote learning [64]. Demoralization of the faculty as a result of SETS should be

a major concern for higher education institutions as it is the faculty that observes problems and develops solutions via the curriculum they design and implement. Feedback that asks for a return to the old; is bullying, aggressive, or intimidating; or not considered against the level of engagement defensibly expected of students does not help faculty demonstrate to the institution how their course is preparing students for the realities of the workforce. However, such feedback does reinforce the critical importance of scaffolding change and novelty within teaching practice. Although this study emphasizes the divergence between student feedback in evaluations and their observable behaviors and interactions within flipped learning environments, the findings could be applicable to any pedagogical innovation where there is a discrepancy between student-reported experiences in evaluations and their actual engagement during class activities or objective learning outcomes.

The theoretical perspectives that inform this study emerged organically during the data analysis phase, rather than guiding the initial research design. This emergent approach to theory underscores the exploratory and inductive nature of our qualitative methodology, which seeks to uncover and understand phenomena through the lens of the collected data. While this method provides deep, context-specific insights, it also introduces certain limitations. For instance, the interpretative nature of qualitative analysis can introduce subjective biases, even as we adhere to rigorous thematic analysis procedures. Additionally, the findings are inherently contextual and may not be generalizable across all educational settings or disciplines.

Acknowledging these limitations, future research could benefit from a more diverse set of educational contexts to test the hypotheses and applicability of these findings more broadly. Comparative studies across different institutions, disciplines, or cultural settings could surface how contextual factors influence the reception and effectiveness of pedagogical innovations and their evaluation through SETS. Furthermore, incorporating other data collection methods and longitudinal design could provide insights into how perceptions and evaluations evolve over time as students and faculty become more accustomed to innovative pedagogies.

By explicitly identifying and discussing these emergent theoretical frameworks and methodological limitations, this study aims to contribute to a nuanced understanding of the complex dynamics between student evaluations and pedagogical innovation. This reflection not only enhances the transparency of the research process but also provides a foundation for ongoing scholarly discourse and further investigation into effective educational practices and evaluation methodologies.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

For information about the data, please contact the first author.

Author Contribution Statement

Rosemary Fisher: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Chamila Perera:** Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Richard Laferriere:** Writing – original draft, Writing – review & editing, Visualization.

References

[1] McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., . . . , & Mumper, R. J. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, *89*(2), 236–243. https://doi.org/10.1097/ACM.0000000000000086

[2] O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, *25*, 85–95. https://doi.org/10.1016/j.iheduc.2015.02.002

[3] Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology*, *49*(3), 398–411. https://doi.org/10.1111/bjet.12548

[4] Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *Journal of Nursing Education*, *52*(10), 597–599. https://doi.org/10.3928/01484834-20130919-03

[5] Wasserman, N. H., Quint, C., Norris, S. A., & Carr, T. (2017). Exploring flipped classroom instruction in calculus III. *International Journal of Science and Mathematics Education*, *15*, 545–568. https://doi.org/10.1007/s10763-015-9704-8

[6] Priem, R. L. (2018). Toward becoming a complete teacher of strategic management. *Academy of Management Learning & Education*, *17*(3), 374–388. https://doi.org/10.5465/amle.2017.0237

[7] Canning, J. (2017). Conceptualising student voice in UK higher education: Four theoretical lenses. *Teaching in Higher Education*, *22*(5), 519–531. https://doi.org/10.1080/13562517.2016.1273207

[8] Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education*, *5*(1), 52–65. https://doi.org/10.5465/amle.2006.20388384

[9] Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation*, *54*, 13–21. https://doi.org/10.1016/j.stueduc.2016.08.002

[10] Freeman, R. (2016). Is student voice necessarily empowering? Problematising student voice as a form of higher education governance. *Higher Education Research & Development*, *35*(4), 859–862. https://doi.org/10.1080/07294360.2016.1172764

[11] 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

[12] 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

[13] 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

[14] Chong, Y. S., & Ahmed, P. K. (2015). Student motivation and the 'feel good' factor: An empirical examination of motivational predictors of university service quality evaluation. *Studies in Higher Education*, *40*(1), 158–177. https://doi.org/10.1080/03075079.2013.818643

[15] Lederman, D. (2020). *Evaluating teaching during the pandemic*. Retrieved from: https://www.insidehighered.com/digital-learning/article/2020/04/08/many-colleges-are-abandoning-or-downgrading-student-evaluations

[16] McAndrew, P. (2015). Innovating for learning: Designing for the future of education. In *Proceedings of the 14th European Conference on E-Learning*, 356–363.

[17] Samuel, M. L. (2021). Flipped pedagogy and student evaluations of teaching. *Active Learning in Higher Education*, *22*(2), 159–168. https://doi.org/10.1177/1469787419855188

[18] Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods*, *11*(3), 481–509. https://doi.org/10.1177/1094428107303349

[19] Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). USA: Free Press.

[20] Oreg, S. (2003). Resistance to change: Developing an individual differences measure. *Journal of Applied Psychology*, *88*(4), 680–693. https://doi.org/10.1037/0021-9010.88.4.680

[21] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

[22] Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research: Vol. 29* (pp. 279–326). Springer.

[23] Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*, 153–166. https://doi.org/10.1023/A:1008168421283

[24] Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*(3), 281–309. https://doi.org/10.3102/00346543051003281

[25] Davies, M., Hirschberg, J., Lye, J., Johnston, C., & Mcdonald, I. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers*, *46*(1), 18–38. https://doi.org/10.1111/j.1467-8454.2007.00303.x

[26] Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395–430. https://doi.org/10.3102/00346543059004395

[27] Schiekirka, S., & Raupach, T. (2015). A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. *BMC Medical Education*, *15*, 30. https://doi.org/10.1186/s12909-015-0311-8

[28] 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

[29] MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303. https://doi.org/10.1007/s10755-014-9313-4

[30] Richmond, A. S., Berglund, M. B., Epelbaum, V. B., & Klein, E. M. (2015). a + (b_1) Professor–student rapport + (b_2) humor + (b_3) student engagement = ($\hat{Y}$) student ratings of instructors. *Teaching of Psychology*, *42*(2), 119–125. https://doi.org/10.1177/0098628315569924

[31] Young, S., & Duncan, H. E. (2014). Online and face-to-face teaching: How do student ratings differ? *MERLOT Journal of Online Learning and Teaching*, *10*(1), 70–79.

[32] Hessler, M., Pöpping, D. M., Hollstein, H., Ohlenburg, H., Arnemann, P. H., Massoth, C., . . . , & Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, *52*(10), 1064–1072. https://doi.org/10.1111/medu.13627

[33] Benton, S. L., Duchon, D., & Pallett, W. H. (2013). Validity of student self-reported ratings of learning. *Assessment & Evaluation in Higher Education*, *38*(4), 377–388. https://doi.org/10.1080/02602938.2011.636799

[34] Kyriakides, L., Creemers, B. P. M., Panayiotou, A., Vanlaar, G., Pfeifer, M., Cankar, G., & McMahon, L. (2014). Using student ratings to measure quality of teaching in six European countries. *European Journal of Teacher Education*, *37*(2), 125–143. https://doi.org/10.1080/02619768.2014.882311

[35] Boysen, G. A. (2015). Uses and misuses of student evaluations of teaching: The interpretation of differences in teaching evaluation means irrespective of statistical information. *Teaching of Psychology*, *42*(2), 109–118. https://doi.org/10.1177/0098628315569922

[36] Hornstein, H. A., & Law, H. F. E. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, *4*(1), 1304016. https://doi.org/10.1080/2331186X.2017.1304016

[37] Vaillancourt, T. (2013). Students aggress against professors in reaction to receiving poor grades: An effect moderated by student narcissism and self-esteem. *Aggressive Behavior*, *39*(1), 71–84. https://doi.org/10.1002/ab.21450

[38] Burke, L. A., Karl, K., Peluchette, J., & Evans, W. R. (2014). Student incivility: A domain review. *Journal of Management Education*, *38*(2), 160–191. https://doi.org/10.1177/1052562913488112

[39] Davidovitch, N., & Eckhaus, E. (2019). Teaching students to think – Faculty recommendations for teaching evaluations employing automated content analysis. *International Journal of Higher Education*, *8*(3), 83–93.

[40] Fisher, W., Orr, J., Page, J., Pelizzon, A., & Walsh, H. (2020). Student evaluations: Pedagogical tools or weapons of choice? *Legal Education Review*, *30*(1), 1–28.

[41] Rodriguez, J. (2019). The weaponization of student evaluations of teaching: Bullying and the undermining of academic freedom. *AAUP Journal of Academic Freedom*, *10*, 1–16.

[42] Feldmann, L. J. (2001). Classroom civility is another of our instructor responsibilities. *College Teaching*, *49*(4), 137–140. https://doi.org/10.1080/87567555.2001.10844595

[43] Balasubramanian, A. (2003). *From the student's view: Student course evaluations*. Retrieved from: https://cft.vanderbilt.edu/articles-and-essays/the-teaching-forum/from-the-students-view-student-course-evaluations/

[44] May, A., & Tenzek, K. E. (2018). Bullying in the academy: Understanding the student bully and the targeted 'stupid, fat, mother fucker' professor. *Teaching in Higher Education*, *23*(3), 275–290. https://doi.org/10.1080/13562517.2017.1379482

[45] Nordstrom, C. R., Bartels, L. K., & Bucy, J. (2009). Predicting and curbing classroom incivility in higher education. *College Student Journal*, *43*(1), 74.

[46] Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. USA: SAGE Publications.

[47] Patton, M. Q. (1990). *Qualitative evaluation and research methods*. USA: SAGE Publications.

[48] Hao, Y. (2016). Exploring undergraduates' perspectives and flipped learning readiness in their flipped classrooms. *Computers in Human Behavior*, *59*, 82–92. https://doi.org/10.1016/j.chb.2016.01.032

[49] Wilson, K. (2023). What does it mean to do teaching? A qualitative study of resistance to flipped learning in a higher education context. *Teaching in Higher Education*, *28*(3), 473–486. https://doi.org/10.1080/13562517.2020.1822312

[50] Burke, A. S., & Fedorek, B. (2017). Does "flipping" promote engagement?: A comparison of a traditional, online, and flipped class. *Active Learning in Higher Education*, *18*(1), 11–24. https://doi.org/10.1177/1469787417693487

[51] Roehling, P. V., Root Luna, L. M., Richie, F. J., & Shaughnessy, J. J. (2017). The benefits, drawbacks, and challenges of using the flipped classroom in an introduction to psychology course. *Teaching of Psychology*, *44*(3), 183–192. https://doi.org/10.1177/0098628317711282

[52] Boud, D. (2001). Making the move to peer learning. In D. Boud, R. Cohen & J. Sampson (Eds.), *Peer learning in higher education: Learning from and with each other* (pp. 1–20). Routledge.

[53] Grebennikov, L., & Shah, M. (2013). Monitoring trends in student satisfaction. *Tertiary Education and Management*, *19*(4), 301–322. https://doi.org/10.1080/13583883.2013.804114

[54] Loughry, M. L., & Tosi, H. L. (2008). Performance implications of peer monitoring. *Organization Science*, *19*(6), 876–890. https://doi.org/10.1287/orsc.1080.0356

[55] Trevino, L. K., & Victor, B. (1992). Peer reporting of unethical behavior: A social context perspective. *Academy of Management Journal*, *35*(1), 38–64.

[56] Etzioni, A. (1968). Social control: Organizational aspects. In D. L. Sills & R. K. Merton (Eds.), *International encyclopedia of social sciences, Vol. 14* (pp. 369–402). Macmillan.

[57] Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, *9*(2), 169–191. https://doi.org/10.5465/amle.9.2.zqr169

[58] Beatty, J. E. (2004). Grades as money and the role of the market metaphor in management education. *Academy of Management Learning & Education*, *3*(2), 187–196. https://doi.org/10.5465/amle.2004.13500516

[59] Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276–294. https://doi.org/10.1080/01973533.2020.1756817

[60] Pérez, O. E. G., & Trevino, J. P. (2019). Learning-oriented assessment in action: Impact on students of physics for engineering. *International Journal on Interactive Design and Manufacturing*, *13*(4), 1485–1501. https://doi.org/10.1007/s12008-019-00606-2

[61] Shibukawa, S., & Taguchi, M. (2019). Exploring the difficulty on students' preparation and the effective instruction in the flipped classroom: A case study in a physiology class. *Journal of Computing in Higher Education*, *31*(2), 311–339. https://doi.org/10.1007/s12528-019-09220-3

[62] McNatt, D. B. (2010). Negative reputation and biased student evaluations of teaching: Longitudinal results from a naturally occurring experiment. *Academy of Management Learning & Education*, *9*(2), 225–242. https://doi.org/10.5465/amle.9.2.zqr225

[63] Karanicolas, S., & Snelling, C. (2010). Making the transition: Achieving content connectivity and student engagement through flexible learning tools. In *Proceedings of the Distance Education Association of New Zealand (DEANZ) Conference*, 1–13.

[64] Borch, I., Sandvoll, R., & Risør, T. (2022). Student course evaluation documents: Constituting evaluation practice. *Assessment & Evaluation in Higher Education*, *47*(2), 169–182. https://doi.org/10.1080/02602938.2021.1899130

How to Cite: Fisher, R., Perera, C., & Laferriere, R. (2024). Unintended Consequences: When Innovation in Pedagogy Impacts Student Evaluations. *International Journal of Changes in Education*, *1*(4), 169–176. https://doi.org/10.47852/bonviewIJCE42023045