

A Large-Scale Observational Study on Obtaining Lightweight, Randomized Weekly Student Feedback

Associations with End-of-Term Course Evaluations

Yunsung Kim
yunsung@stanford.edu
Stanford University
California, USA

Candace Thille
cthille@stanford.edu
Stanford University
California, USA

Hansol Lee
hansol@stanford.edu
Stanford University
California, USA

Chris Piech
cpiech@cs.stanford.edu
Stanford University
California, USA

Abstract

Conventional methods of obtaining student feedback on course experiences face a fundamental tradeoff between feedback frequency and quality: as feedback requests become more frequent, participation often declines, and responses become less thoughtful over time. To obtain both timely and thoughtful feedback from students, Kim and Piech [27] recently proposed a simple, lightweight course feedback mechanism: surveying each student a small number of times per term during randomly selected weeks. Termed *High-Resolution Course Feedback (HRCF)*, this method has been shown to elicit feedback that instructors find helpful without placing excessive survey burden on students.

An important question, however, remains unanswered: is the use of this simple method associated with measurable improvements in *students' actual course experiences*? We study HRCF use from Fall 2021 through Fall 2025 across 103 course offerings, representing 24,216 student enrollments in 42 unique computer science courses at an R1 institution. Using regression analyses of average student ratings for four end-of-term course evaluation items, we find that first-time use of HRCF is not associated with a measurable difference in average student ratings. However, among small- and medium-enrollment (<250 students) course offerings, continued HRCF use is associated with 0.045- to 0.048-point higher average ratings per additional term of use for learning-related items. We observe no statistically significant associations for large-enrollment (250 or more students) course offerings, nor for items measuring instructional quality and course organization. Together, these findings suggest that sustained HRCF use may support improvements in students' learning experiences, but that further design enhancements may be needed to produce measurable improvements in instruction quality and course organization.

CCS Concepts

• **Applied computing** → **Computer-assisted instruction.**



This work is licensed under a Creative Commons Attribution 4.0 International License. *L@S '26, Seoul, Republic of Korea*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2293-6/2026/06
<https://doi.org/10.1145/3774398.3811622>

Keywords

Course Feedback; Course Evaluation; Instructional Improvement; Weekly Course Feedback; Student Learning Experience; Student Course Experience; Student Evaluations of Teaching; SETs; Student Ratings; End-of-Term Course Evaluations

ACM Reference Format:

Yunsung Kim, Hansol Lee, Candace Thille, and Chris Piech. 2026. A Large-Scale Observational Study on Obtaining Lightweight, Randomized Weekly Student Feedback: Associations with End-of-Term Course Evaluations. In *Proceedings of the Thirteenth ACM Conference on Learning @ Scale (L@S '26)*, June 29–July 03, 2026, Seoul, Republic of Korea. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3774398.3811622>

1 Introduction

The journey of becoming an excellent teacher is a continuous process of reflection and improvement. One of the most widely used inputs for instructional improvement is students' own accounts of their course experiences and their evaluations of an instructor's teaching, which many instructors use to reflect on their teaching and identify target areas for improvement [21, 25, 32, 38]. Reflecting this role of student evaluations and feedback, many universities in North America routinely collect student evaluations of teaching to monitor instructional quality and to support instructional improvement [9, 33].

Student evaluations are most commonly obtained at the end of the academic term. This method, while useful for assessing the quality of instruction, is not well-suited to providing *specific and timely* feedback that supports instructional improvement *during* the course, nor does it benefit the learning experience of the students providing it [53]. To fill in this gap, many instructors additionally collect midterm student evaluations halfway through the course [22, 40] or opt to request feedback more frequently on a weekly basis [53]. With increased frequency of feedback, a fundamental tension arises between timeliness and quality. Frequent feedback requests enable a rapid feedback loop but also create excess burden on students such that providing feedback “soon becomes routine and too onerous a task” [53], leading to decreased participation and less thoughtful responses over time. This phenomenon is known as “survey fatigue” [24, 43], and it compromises the validity of the received feedback and results in a systematic nonresponse bias [3].

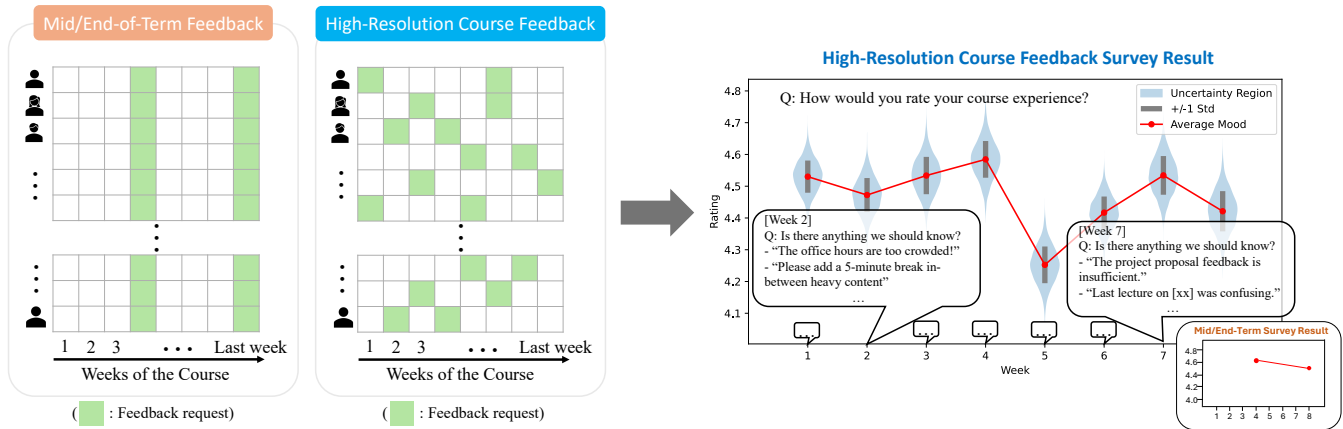


Figure 1: Overview of High-Resolution Course Feedback (HRCF) [27]. HRCF obtains high-quality and frequent student feedback by randomizing when each student receives feedback, unlike conventional feedback mechanisms that request feedback from all students at a few specific points in the course.

A simple, lightweight approach to resolving the tradeoff between timely feedback and survey fatigue has recently been proposed. “**High-Resolution Course Feedback**” (HRCF), first proposed at Learning@Scale 2023 [27], requests feedback from a small *random* subset of students each week, so that all students are surveyed the same, fixed number of times per term (typically twice) but on random weeks (Figure 1). By randomly distributing feedback requests across time, HRCF aims to obtain specific, time-sensitive feedback at a weekly cadence without increasing the burden on students relative to conventional feedback mechanisms that survey everyone at specific points in the course. Prior work [27] found that HRCF generates timely and constructive comments that instructors found useful in understanding the weekly sentiment of students and making adjustments to their instruction.

However, an important aspect of this approach still remains unexplored: whether HRCF use is associated with measurable improvements in *students’ actual course experiences*. Whether the effects of the instructional changes made by the instructors meaningfully propagate to students’ perceived learning experiences is an important indicator for the instructional value of HRCF and its broader adoption in courses. We conduct a large-scale observational study of HRCF use across four years from Fall 2021 to Fall 2025. During this period a total of 42 computer science courses at an R1 institution used HRCF for a total of 103 course offerings, with 24,216 total enrollments and 15,965 unique students. Using student course evaluation data for the same courses for all offerings between Fall 2018 and Fall 2025, we measured the association between HRCF use and average student ratings for 4 evaluation items (Table 5) regarding (1) the perceived amount of learning goals achieved (GOAL), (2) the perceived learning gains (LEARN), (3) the overall instructional quality (QUAL), and (4) the organizational quality of the course (ORG).

We use regression analyses to estimate the associations between HRCF use and average student evaluation ratings for each of the four course evaluation items. Specifically, we address the following research questions:

RQ1: Is the use of HRCF associated with students’ reported course experiences?

RQ2: Is continued use of HRCF across course offerings associated with students’ reported course experiences, and does this marginal association vary by enrollment size?

We find that the initial adoption of HRCF is not associated with a measurable change in average student ratings. However, among small/medium-enrollment (< 250) course offerings from courses that continue using HRCF, each additional term of HRCF use is associated with average rating increases of 0.045 (GOAL) and 0.048 (LEARN) points for learning-related items. We observe no significant associations for large-enrollment course offerings (≥ 250) or for the QUAL and ORG items for any enrollment size. Taken together, these results provide evidence that repeated HRCF use does have a positive association with improved student evaluations of their learning experience, but that HRCF needs to be enhanced to support broader measurable improvements in instruction quality and course organization.

2 Related Work

Educational institutions evaluate teaching effectiveness using a wide range of approaches including “peer visits, external and internal reviews of teaching portfolios, and qualitative feedback from students,” but these methods are often more demanding and time-consuming than student evaluations of teaching [44]. Consequently, student evaluations of teaching (SETs) have become the most common scalable instrument for monitoring perceived teaching and course quality in universities [44, 53]. Prior research emphasizes that SETs are used by various educational stakeholders for different purposes [32]: as input to administrative judgments such as tenure [41]; as a channel for students to share past course experiences and obtain information for making enrollment decisions [31–33]; and as signals for instructors and course designers to diagnose issues, iteratively improve courses, and make instructional improvements [19, 21, 32, 38]. High-Resolution Course Feedback is an example of the last “formative” use of student feedback.

Decades of prior research on student evaluations and feedback on teaching has explored what these instruments are capable of measuring and how they should be interpreted. Researchers have engaged in active debates about their validity as measures of teaching effectiveness [2, 28, 38, 50], their relationship to student learning and academic outcomes [14, 38], and the extent to which these scores may reflect factors and biases unrelated to teaching effectiveness, such as course structure and size, workload, instructional methods, grading leniency, instructor gender, and student demographic and background characteristics [5, 8, 10, 14, 20, 23, 34, 45, 46]. Overall, studies conclude that student evaluations can provide informative signals of instructional quality and actionable directions for improvement [12, 33, 38, 42] and capture diverse aspects of effective instruction [50]. Consistent with this, instructors often report that they find student evaluations to be a useful input for refining their teaching practice [4, 47].

Yet, the quality and utility of student evaluations and feedback depend crucially on the students' willingness to provide careful and authentic feedback. Prior work suggests that this motivation is closely tied to whether students believe that their evaluation and feedback will be valued by their instructors [6, 30, 50, 51]. For instance, [53] reports that students were more positive towards providing evaluations "if the lecturer discussed them and/or made changes to their future lectures." Similarly, [11] argues that students are motivated to provide evaluations when they see their actions result in a positive outcome. In line with this, [12] reports that students consider the most desirable outcome of student evaluations of teaching to be "improvement of teaching," with changes to course content and format as a secondary outcome. On the other hand, students often lack the confidence that their feedback will be taken seriously by the instructors, which can lead to apathy and less careful feedback [49].

In this light, the nature of end-of-term evaluations is largely summative and retrospective, and studies have noted their limitations in benefiting the learning experience of students who are providing the evaluations [39, 53]. Midterm feedback [1, 14, 26, 40] and weekly feedback [53] have been studied as alternatives to end-of-term evaluations for formative purposes. Studies provide evidence that formative feedback can help make improvements to diverse aspects of instruction, such as teaching practice [12] as well as course content and structure [18, 48]. The use of midterm feedback has also empirically been associated with higher end-of-term evaluation ratings [14, 40], improved learning gains and enhanced affective outcomes [40], and greater student satisfaction with the feedback process [1, 52]. Beyond these measurable outcomes, [17] demonstrates that feedback obtained during class can facilitate mutual communication between students and instructors around instructional decision making.

To maximize the benefits of formative feedback on improving teaching quality and student learning experiences, prior work emphasizes the value of eliciting feedback that is both timely and specific [15, 30, 35]. Weekly feedback was studied as an alternative that can provide continuous formative feedback to the instructors throughout the term [53]. When feedback requests were made weekly, however, [53] reports that students' willingness to complete the feedback survey varied and noticeably declined over the academic term. When interviewed, students also "almost unanimously"

did not want to fill out surveys every week noting that students would "get fed up quickly," and several suggested collecting feedback twice a term as a more motivating cadence. High-Resolution Course Feedback [27] was designed to resolve this tradeoff between feedback timeliness and survey fatigue through randomized sampling.

3 High-Resolution Course Feedback

This section provides a brief overview of the High-Resolution Course Feedback mechanism (HRCF) [27] and preliminary findings about its utility. HRCF is a formative course feedback mechanism designed to provide course instructors with specific and timely feedback and opportunities to adjust their instruction throughout the course. Instead of collecting feedback from all students at a few fixed points in the course (e.g., midterm and end of term), HRCF makes a fixed number of feedback requests to every student but in randomly selected weeks of the course. As a result, instructors receive feedback from a random subset of students each week. In the baseline setup, each student is invited to respond in $S = 2$ randomly chosen weeks, spaced at least two weeks apart.

Operationally, HRCF follows a 3-phase weekly cycle. On Monday of each week, the system emails feedback invitations to the subset of students scheduled for that week with a link to an anonymous survey page. The survey includes the four default questions from Table 1, along with optional custom questions of the week set by the instructors. The four default questions are designed to (1) gauge weekly course sentiment, (2) surface encouragement and appreciation for the teaching team, (3) prompt student reflection on their learning process, and (4) support open, ongoing communication between students and instructors. On Thursday, reminder emails are sent to students who have not completed the survey. The week's feedback results are collected and sent to the instructors on Sunday in a weekly digest email. The digest email reports (i) the weekly response rate, (ii) compiled feedback responses with visual summaries for rating items, and (iii) a temporal graph of "class mood" ratings calculated as the average ratings of the course experience from the default question (on the scale of "Excellent," "Good," "Average," "Below Average," and "Poor").

Prior work has explored the utility of HRCF from the perspective of course instructors. [27] found that HRCF elicits feedback that is both timely and instructionally useful. In their analysis of weekly HRCF responses from a large, representative course, they found that roughly 25% to 50% of comments surfaced ongoing, course-level issues that could be addressed during the term. Instructors reported that HRCF was useful for identifying student pain points and making concrete instructional adjustments to lecture pacing and structure, course content, office-hour logistics, and assignments. They also reported that weekly course experience ratings ("class mood") were useful for probing the dynamics of course sentiment throughout the course.

4 Study Method

4.1 HRCF Usage Data

Our analysis is based on observations of HRCF use primarily within computer science courses at an R1 institution operating on a quarter system. These observations span a total of 13 academic terms from

Table 1: Default High-Resolution Course Feedback Questions [27]

Question Text	Intended Outcome	Response type
How would you rate your course experience this week?	Overall weekly sentiment	5-point Likert
What did you like about the course so far?	Positive feedback and encouragement	Free text
Is anything from class still confusing to you?	Self-reflections on the learning process	Free text
Is there anything the teaching team should know?	Fostering open communication	Free text

Table 2: HRCF and End-of-Term Evaluation Dataset Detail

Metric	Value
Course Metrics	
Number of Unique Courses	42
Total Number of Course Offerings	310
Course Offerings with HRCF	103
Course Offerings without HRCF	207
HRCF Survey Metrics	
Total Class Size	24,216
Number of Unique Students	15,965
Total Feedback Requests	56,314
Total Responses Received	22,431
Overall Response Rate	39.8%
End-of-Term Evaluation Metrics	
Average Response Rate	68.3%
Average Ratings for Each Question	
Learning goals (GOAL)	3.98
How much learned (LEARN)	4.21
Overall quality (QUAL)	4.23
How organized (ORG)	4.14

Fall 2021 through Fall 2025. Among the courses that used HRCF, we filtered out course offerings with enrollment size smaller than 20, which roughly corresponds to fewer than 4 feedback requests per week on average. This resulted in 103 course offerings across 42 courses that used HRCF, with a total enrollment of 24,216 and 15,965 unique students. The overall HRCF response rate was 39.8%. Table 2 includes a summary of the HRCF usage data. Among the 42 courses, 21 courses used HRCF in more than one offering. The distribution of course sizes and number of prior HRCF uses per course offering can also be found in Tables 3 and 4.

4.2 End-of-Term Student Evaluations

For every course that used HRCF at least once, we collected end-of-term student evaluation reports for all course offerings from Fall 2018 through Fall 2025. Evaluation reports between Fall 2018 and Fall 2021 were included to serve as within-course control observations prior to HRCF adoption. This resulted in the inclusion of student evaluation reports for 207 non-HRCF course offerings in addition to the 103 HRCF course offerings. The end-of-term evaluation reports contained student responses to eight questions listed in Table 5. We used the first four items—“Learning goals”

Table 3: Distribution of Course Offerings by Enrollment

Enrollment	HRCF	Non-HRCF	Total
20–49	4	12	16
50–99	11	35	46
100–199	30	71	101
200–299	29	51	80
300–399	16	15	31
400–499	9	16	25
500+	4	7	11
Total	103	207	310

Table 4: Distribution of HRCF Course Offerings by Prior HRCF Use and Enrollment Size (< 250 vs. ≥ 250)

Prior HRCF Use	< 250	≥ 250	Total
0 (First Time Use)	35	7	42
1	10	11	21
2	9	6	15
3	4	7	11
4	0	5	5
5	1	2	3
6	0	3	3
7	0	1	1
8	0	1	1
9	0	1	1
Total	59	44	103

(GOAL), “How much learned” (LEARN), “Overall quality of instruction” (QUAL), and “How organized” (ORG)—as our target measures. Ratings used a 5-point Likert scale (1–5). Mean item ratings ranged from 3.98 to 4.23 (Table 2). Of the remaining items, the three questions on hours spent per week (HOURS) and percent of attendance in-person/online (INPERSON/ONLINE) were used as additional control variables for each course offering to be included in our analysis.

4.3 Analysis Model

We use an ordinary least squares (OLS) model to analyze the association between HRCF usage and each of the four target student evaluation ratings. Recall that the purpose of our study is to analyze whether student evaluations are associated with the use of HRCF (RQ1) and whether any association varies with repeated HRCF use and with enrollment size (RQ2). We operationalize these

Table 5: End-of-Term Course Evaluation Questions

Label	Question Text	Response type
GOAL	How well did you achieve the learning goals of this course?	5-point Likert
LEARN	How much did you learn from this course?	5-point Likert
QUAL	Overall, how would you describe the quality of the instruction in this course?	5-point Likert
ORG	How organized was the course?	5-point Likert
HOURS	How many hours per week on average did you spend on this course (including class meetings)?	Numeric
INPERSON	About what percent of the class meetings (including discussions) did you attend in person?	Percentage
ONLINE	About what percent of the class meetings did you attend online?	Percentage
	What would you like to say about this course to a student who is considering taking it in the future?	Free text

constructs with the following explanatory variables: (1) indicator variable $U_{ct} \in \{0, 1\}$ denoting whether course c used HRCF during its offering in term t , and (2) variable $K_{ct} \in \{0, 1, 2, \dots\}$ denoting the number of prior HRCF uses in the course. We allow the HRCF-use association to change with repeated use and to differ for large-enrollment course offerings.

Our complete regression model is the following:

$$\begin{aligned} \text{AverageRating}_{ct} = & \beta_0 + \delta_c + \delta_y \\ & + \beta_1 U_{ct} + \beta_2 (U_{ct} \cdot K_{ct}) + \beta_3 (U_{ct} \cdot K_{ct} \cdot \mathbf{1}\{\text{Enrollment}_{ct} \geq 250\}) \\ & + \beta_4 (\text{Enrollment})_{ct} + \beta_5 (\text{Avg Hours Per Week})_{ct} \\ & + \beta_6 (\text{Avg \% of Class Meetings Attended In Person})_{ct} \\ & + \beta_7 (\text{Avg \% of Class Meetings Attended Online})_{ct}. \end{aligned}$$

In this model, β_1 captures the association between first-time HRCF use (i.e., when prior use $K_{ct} = 0$) and average ratings. We allow the usage effect to vary with repeated use by including the interaction term $\beta_2 (U_{ct} \cdot K_{ct})$. Here, β_2 represents how the average ratings change per additional prior use. To understand how the association differs in large-enrollment settings, we further include $\beta_3 (U_{ct} \cdot K_{ct} \cdot \mathbf{1}\{\text{Enrollment}_{ct} \geq 250\})$. Coefficient β_3 measures the additional change in the repeat-use association for large-enrollment course offerings (≥ 250) relative to smaller course offerings (< 250). The enrollment threshold of 250 was chosen to be close to the median enrollment size (225) across HRCF course offerings while maintaining a balanced distribution of repeat-use observations across enrollment bins.¹

We control for course-specific (δ_c) and academic year-specific (δ_y) fixed effects. Even within the same course, several factors specific to each offering are known to affect student ratings, including enrollment size, subject area, method of instruction, workload, and grades [5, 8, 10, 34, 45]. To the greatest extent possible, we control for these variables using data available to us, which include enrollment size, average hours per week spent (HOURS item in Table 5), and percent of class meetings attended in person and online (INPERSON and ONLINE items in Table 5).

5 Results and Analysis

Table 6 presents the full regression results. We summarize the main technical observations below and provide a more in-depth analysis of these observations in the subsequent sections:

¹Using the median as the threshold produces slightly stronger results. We report the more conservative results with 250 as the threshold.

First-time HRCF use does not show a statistically significant association with student course evaluations. Across all four course evaluation items, no detectable change in average student ratings was observed for first-time HRCF use.

For LEARN and GOAL items, each additional term of prior HRCF use has a statistically significant positive association for small/medium-enrollment course offerings. In contrast to first-time use, repeat HRCF use is positively associated with student evaluations in small/medium-enrollment (< 250) course offerings with at least one prior HRCF use for “Learning goals” ($\beta = 0.044^*$) and “How much learned” ($\beta = 0.048^*$) items. On the other hand, for large-enrollment courses, the repeat-use associations are not statistically significant.

Repeat HRCF use does not show a statistically significant association for QUAL and ORG items regardless of enrollment size. Across these items, repeat HRCF use showed no statistically significant association with average ratings regardless of enrollment size condition.

5.1 Higher Ratings with Additional Prior HRCF Use in Small/Medium Course Offerings

Overall, the regression analysis suggests that adopting HRCF for the first time is not associated with measurable changes in students’ end-of-term evaluation ratings. In contrast, among small/medium-enrollment course offerings with prior HRCF use, each additional prior use is associated with meaningfully higher learning-related student ratings. The magnitudes of the estimated regression coefficients for “Learning goals” and “How much did you learn” items indicate that each additional prior HRCF use is roughly associated with a 0.045- to 0.048-point increase in average course evaluation ratings for small/medium-enrollment (< 250) course offerings from courses that continue to use HRCF. In practical terms, this means that, within the same course, an offering after two prior HRCF uses would be predicted to have GOAL/LEARN ratings roughly 0.1 points higher than the course’s first HRCF offering. Considering that average course ratings are already near or above 4 on a 5-point scale (see Table 2) and that HRCF surveys are low-intensity (two short surveys per student), these associations should be interpreted as meaningful improvements in student perceptions of their learning.

The positive repeat-use association for GOAL/LEARN items in small/medium-enrollment course offerings fits prior observations

Table 6: OLS estimates of associations between HRCF use and course evaluation ratings. Standard errors in parentheses. All models include year and course fixed effects. * $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.**

	(1) Learning goals (GOAL)	(2) How much learned (LEARN)	(3) Overall quality (QUAL)	(4) How organized (ORG)
HRCF Effects				
HRCF (first use)	−0.038 (0.029)	−0.038 (0.027)	−0.043 (0.042)	−0.024 (0.036)
HRCF Repeat (Enrollment < 250)	0.044* (0.021)	0.048* (0.020)	0.046 (0.031)	0.042 (0.027)
HRCF Repeat (Enrollment ≥ 250)	0.002 (0.011)	0.008 (0.010)	0.003 (0.016)	0.002 (0.014)
Control Variables				
Enrollment (per 100)	0.034* (0.017)	0.015 (0.016)	0.085*** (0.025)	0.058** (0.022)
Hours per week	−0.036*** (0.007)	−0.003 (0.007)	−0.037*** (0.011)	−0.029** (0.009)
Attend in person (%)	0.007*** (0.002)	0.006*** (0.002)	0.012*** (0.002)	0.010*** (0.002)
Attend online (%)	0.006*** (0.002)	0.005*** (0.001)	0.011*** (0.002)	0.010*** (0.002)
<i>N</i>	310	310	310	310
R^2	0.6183	0.7293	0.6908	0.7045
Adj. R^2	0.5338	0.6694	0.6224	0.6391

that HRCF provides useful formative feedback for making instructional improvements that address difficulties in student learning. The HRCF question “Is anything from class still confusing to you?” is specifically targeted at providing instructors with an understanding of which course content students need the most support with. In addition to this question, instructors reported using this feedback for making instructional changes that specifically address student pain points in learning [27], such as adding short recaps and clarifications in the beginning of lectures, providing more worked examples during lecture, allocating more time to address commonly confusing course content, providing additional learning materials or practice problems, and restructuring office hours. A finer-grained knowledge of content-specific student difficulties along with instructor-reported adjustments made to improve learning experiences offers a plausible mechanism for the observed association.

5.2 Learning Curve and First-Time Use

The null associations of initial adoption, when taken together with the positive repeat-use association, also align with long-running findings within the student evaluations literature that improvements in teaching require deliberate sense-making and structured follow-through in addition to merely collecting feedback [30, 42]. It is a well-known concept in feedback research more generally that

individuals vary in the capacities, dispositions, and prior experiences that shape how well they can interpret and use feedback in learning [7]. The same insight applies to instructors using feedback to learn about their instructional effectiveness and make improvements, and not all instructors have the capacity to process student feedback and translate it into effective instructional adjustments [13]. This could explain why first-time uses showed no statistically significant association, as instructors could have been facing a learning curve where they were initially unfamiliar with how to respond to student feedback.

This is also particularly relevant in the context of our current study design. Our study is an observational study that leverages naturally occurring variations in HRCF adoption and repeated use across academic terms to measure its association with student evaluations. Because the choice to adopt HRCF and continue using it is at the discretion of the instructor, it is plausible that courses that used HRCF more than once may partially reflect contexts in which instructors were more proficient in processing and interpreting student feedback to make effective instructional adjustments. This could further explain the positive associations with student ratings for small/medium-enrollment course offerings with prior HRCF use. These results suggest that structured support for helping instructors make meaningful interpretations of the feedback they receive and guide them in deriving effective instructional improvement decisions could help improve student learning experiences.

5.3 Constraints of Acting on Feedback at Scale

On the other hand, our analysis finds neither first-time nor repeated HRCF use to be associated with any student evaluation item in large-enrollment course offerings. This result is consistent with prior work on the practical constraints of large-enrollment instruction. Large courses face structural challenges of scale that small to medium-enrollment courses do not. These structural challenges include reduced opportunities for student–instructor interaction, limited capacity for individualized student support, greater classroom management and coordination demands, constraints on course structure and delivery, difficulties providing timely and high-quality assessment and feedback at scale, and greater heterogeneity in student preparedness [16, 36, 37]. These factors not only negatively affect student-reported learning and evaluations of teaching quality in large-enrollment courses [36], but also make instructional adaptation more difficult [37]. This rigidity is further compounded by the additional staff time and resources required to design, implement, and coordinate such changes [37]. Even when the instructors do make instructional improvements, the constraints of scale may attenuate how much change can propagate into noticeable differences for individual learners. Further research should investigate how the method and practice of receiving and delivering lightweight, weekly feedback could be improved with a focus on making tangible enhancements to the everyday learning experiences of individual learners.

5.4 Limited Associations with Global Composite Metrics of Course Quality

Unlike with LEARN and GOAL items, neither first-time nor repeat use had statistically significant associations with the ratings of the overall quality of instruction (QUAL) and organizational quality (ORG) regardless of enrollment size. These results suggest that the effect of the instructional adjustments that HRCF supports may be localized to students' perceptions of learning as opposed to broader perceptions of quality of instruction and course organization. Prior work on student evaluations of teaching emphasizes that composite evaluations such as evaluations of instruction and course quality often reflect various course-related factors other than learning and instruction alone [50], such as course design, assessment and grading practices, fairness of grading, grading leniency, and workload dynamics [9, 45]. Many of these structural factors, such as course design and schedule as well as workload, are typically determined in advance and are less amenable to change compared to instructional practices. Consistent with this interpretation, a recent study [29] finds that features of course design are stronger predictors of students' perceived course quality than instructor-related factors, suggesting that composite metrics like QUAL and ORG may be less responsive to incremental instructional changes during a term that the current design of High-Resolution Course Feedback seems to mainly provide. How High-Resolution Course Feedback can be improved to support meaningful enhancements in broader, structural elements of a course remains an interesting direction for future work.

6 Limitations and Future Work

Our findings should be interpreted in light of several limitations of the data and study design. First, our outcomes are based on end-of-term student evaluation ratings. While student evaluations can be meaningful measures of downstream impact of instructional improvements on student learning experiences, this outcome measure misses many other aspects of instructional improvement such as objective learning outcomes or alignment with known effective instructional practices. Other meaningful measures of HRCF's effectiveness include improvements in students' actual learning gains, evaluations from teaching specialists, peer evaluations, and indicators of student engagement, motivation, and confidence. Future work should consider a more comprehensive evaluation of the quality of instructional improvements enabled by HRCF and measure their impact on a diverse set of metrics.

Another limitation of our study is generalizability across domains. Our study relies on data collected from a single discipline (computer science) at a single R1 institution. While analyzing courses within a single discipline allows for more consistent comparisons across course offerings, it also narrows the contexts in which our findings can generalize to. Other disciplines with different pedagogical practices (e.g., studio- or lab-based instruction, writing-intensive instruction), different assessment structures, or different expectations around student-instructor interactions may result in different contexts and constraints of acting on HRCF feedback. Likewise, different institutional policies, teaching support resources, enrollment size distributions, and student cultures can shape both the types of feedback that students provide and the instructors' capacity to process and respond to feedback. More work is needed to collect evidence of the effects of HRCF use to provide a more generalizable understanding of its effectiveness.

Our results suggest that instructors first adopting HRCF may be facing a learning curve, and instructors who are proficient at processing and responding to feedback may benefit more from the tool compared to others who do not. Moreover, the associations observed for large courses may be limited by issues of scale. Future work should explore targeted enhancements to HRCF that support instructors' sense-making of the received feedback and make an impact in large courses. Concretely, HRCF may incorporate automated clustering and prioritization of recurring issues or use AI to generate helpful action plans that would support instructional improvements based on the received feedback. Additionally, since students' motivation to provide thoughtful feedback depends on their belief that their feedback will be valued, HRCF could be enhanced to support an interface where instructors respond directly to students' feedback. Implementing and evaluating these improvements via field experiments would be an important area of future work.

7 Conclusion

We studied the associations between using High-Resolution Course Feedback, a simple and lightweight feedback mechanism for obtaining timely and actionable feedback from students, and students' actual learning experiences as measured by course evaluation ratings. High-Resolution Course Feedback obtains weekly feedback from students without creating survey fatigue by requesting feedback

from a random sample of students each week, with every student getting the same, small number of feedback requests throughout the term. Through a large-scale observational study of 42 unique computer science courses and 103 course offerings that used HRCF over a 4-year span in an R1 institution, we analyzed the association between HRCF use and student evaluation ratings of their course experience, and whether this association varied by repeated use and enrollment size. We find that, while first-time HRCF use is not associated with a measurable change in student evaluation ratings, each additional use of HRCF has a statistically significant association with a 0.045- to 0.048-point increase in average ratings for learning-related evaluation items for course offerings with small/medium-enrollment (<250). No statistically significant associations were observed for large-enrollment (≥ 250) course offerings, and for evaluations of overall instructional quality and organizational quality. Overall, the results suggest that HRCF is a useful mechanism associated with improvements in students' perceived learning, but that it needs to be improved to have a broader impact on students' overall course experience. Future work should explore effective methods for supporting instructors' sense-making of the received weekly feedback and consider enhancements of HRCF that would result in measurable improvements in overall quality of instruction and course organization.

References

- [1] Robert D Abbott, Donald H Wulff, Jody D Nyquist, Vickie A Ropp, and Carla W Hess. 1990. Satisfaction with processes of collecting student opinions about instruction: The student perspective. *Journal of Educational Psychology* 82, 2 (1990), 201.
- [2] Philip C Abrami, Sylvia d'Apollonia, and Peter A Cohen. 1990. Validity of student ratings of instruction: What we know and what we do not. *Journal of educational psychology* 82, 2 (1990), 219.
- [3] Meredith JD Adams and Paul D Umbach. 2012. Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education* 53, 5 (2012), 576–591.
- [4] Linet Arthur. 2009. From performativity to professionalism: Lecturers' responses to student feedback. *Teaching in Higher Education* 14, 4 (2009), 441–454.
- [5] Larry A Braskamp and John C Ory. 1994. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. Jossey-Bass Higher and Adult Education Series. ERIC.
- [6] Michael J Brown. 2008. Student perceptions of teaching evaluations. *Journal of Instructional Psychology* 35, 2 (2008), 177–182.
- [7] David Carless and David Boud. 2018. The development of student feedback literacy: Enabling uptake of feedback. *Assessment & evaluation in higher education* 43, 8 (2018), 1315–1325.
- [8] John A Centra. 1993. *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. The Jossey-Bass Higher and Adult Education Series. ERIC.
- [9] John A Centra. 2003. Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in higher education* 44, 5 (2003), 495–518.
- [10] John A Centra and F Reid Creech. 1976. The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness. *Project report* 76, 1 (1976).
- [11] Yining Chen and Leon B Hoshower. 1998. Assessing student motivation to participate in teaching evaluations: An application of expectancy theory. *Issues in Accounting Education* 13, 3 (1998), 531.
- [12] Yining Chen and Leon B Hoshower. 2003. Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & evaluation in higher education* 28, 1 (2003), 71–88.
- [13] Peter A Cohen. 1980. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in higher education* 13, 4 (1980), 321–341.
- [14] Peter A Cohen. 1981. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of educational Research* 51, 3 (1981), 281–309.
- [15] K Patricia Cross and Thomas A Angelo. 1988. *Classroom Assessment Techniques. A Handbook for Faculty*. (1988).
- [16] Joe Cuseo. 2007. The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development* 21, 1 (2007), 5–21.
- [17] Miriam Rosalyn Diamond. 2004. The usefulness of structured mid-term feedback as a catalyst for change in higher education classes. *Active Learning in Higher Education* 5, 3 (2004), 217–231.
- [18] Laura A Driscoll and William L Goodwin. 1979. The effects of varying information about use and disposition of results on university students' evaluations of faculty and courses. *American Educational Research Journal* 16, 1 (1979), 25–37.
- [19] Howard Ebmeier. 2003. How supervision influences teacher efficacy and commitment: An investigation of a path model. *Journal of Curriculum and supervision* 18, 2 (2003), 110–141.
- [20] Kenneth A Feldman. 1992. College students' views of male and female college teachers: Part I—Evidence from the social laboratory and experiments. *Research in Higher Education* 33 (1992), 317–375.
- [21] Jonas Flodén. 2017. The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education* 42, 7 (2017), 1054–1068.
- [22] Pamela Gravestock and Emily Gregor-Greenleaf. 2008. *Student course evaluations: Research, models and trends*. Higher Education Quality Council of Ontario Toronto.
- [23] Anthony G Greenwald and Gerald M Gillmore. 1997. Grading leniency is a removable contaminant of student ratings. *American psychologist* 52, 11 (1997), 1209.
- [24] Robert M Groves, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey methodology*. John Wiley & Sons.
- [25] Alastair Irons and Sam Elkington. 2021. *Enhancing learning through formative assessment and feedback*. Routledge.
- [26] Carolin S Keutzer. 1993. Midterm evaluation of teaching provides helpful feedback to instructors. *Teaching of psychology* 20, 4 (1993), 238–240.
- [27] Yunsung Kim and Chris Piech. 2023. High-resolution course feedback: Timely feedback mechanism for instructors. In *Proceedings of the Tenth ACM Conference on Learning@ Scale*. 81–91.
- [28] James A Kulik. 2001. Student ratings: Validity, utility, and controversy. *New directions for institutional research* 2001, 109 (2001), 9–25.
- [29] Henrik Levinsson, August Nilsson, Katarina Mårtensson, and Stefan D Persson. 2024. Course design as a stronger predictor of student evaluation of quality and student engagement than teacher ratings. *Higher Education* (2024), 1–17.
- [30] Karon G Lewis. 2001. Using midsemester student feedback and responding to it. *New Directions for Teaching and Learning* 2001, 87 (2001), 33–44.
- [31] Herbert W Marsh. 2007. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. *The scholarship of teaching and learning in higher education: An evidence-based perspective* (2007), 319–383.
- [32] Herbert W Marsh and Lawrence Roche. 1993. The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American educational research journal* 30, 1 (1993), 217–251.
- [33] Herbert W Marsh and Lawrence A Roche. 1997. Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American psychologist* 52, 11 (1997), 1187.
- [34] Kathleen E McKone. 1999. Analysis of student feedback improves instructor effectiveness. *Journal of Management Education* 23, 4 (1999), 396–415.
- [35] Milbrey Wallin McLaughlin and R Scott Pfeifer. 1988. *Teacher Evaluation: Improvement, Accountability, and Effective Learning*. Teachers College Press.
- [36] James Monks and Robert M Schmidt. 2011. The Impact of Class Size on Outcomes in Higher Education. *The BE Journal of Economic Analysis & Policy* 11, 1 (2011).
- [37] Catherine Mulryan-Kyne. 2010. Teaching large classes at college and university level: Challenges and opportunities. *Teaching in higher education* 15, 2 (2010), 175–185.
- [38] Harry G Murray. 1997. Does evaluation of teaching lead to improvement of teaching? *The International Journal for Academic Development* 2, 1 (1997), 8–23.
- [39] Kasturi Narasimhan. 2001. Improving the climate of teaching sessions: the use of evaluations by students and instructors. *Quality in Higher Education* 7, 3 (2001), 179–190.
- [40] JU Overall and Herbert W Marsh. 1979. Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of educational psychology* 71, 6 (1979), 856.
- [41] Angela R Penny. 2003. Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in higher education* 8, 3 (2003), 399–411.
- [42] Angela R Penny and Robert Coe. 2004. Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of educational research* 74, 2 (2004), 215–253.
- [43] Stephen R Porter, Michael E Whitcomb, and William H Weitzer. 2004. Multiple surveys of students and survey fatigue. *New directions for institutional research* 2004, 121 (2004), 63–73.
- [44] William J Read, Dasaratha V Rama, and K Raghunandan. 2001. The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business* 76, 4 (2001), 189–192.

- [45] Richard Remedios and David A Lieberman. 2008. I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal* 34, 1 (2008), 91–115.
- [46] John TE Richardson. 2005. Instruments for obtaining student feedback: A review of the literature. *Assessment & evaluation in higher education* 30, 4 (2005), 387–415.
- [47] Liora Pedhazur Schmelkin, Karin J Spencer, and Estelle S Gellman. 1997. Faculty perspectives on course and teacher evaluations. *Research in Higher Education* 38 (1997), 575–592.
- [48] Ronald D Simpson. 1995. Uses and misuses of student evaluations of teaching effectiveness. *Innovative Higher Education* 20, 1 (1995), 3–5.
- [49] Karin J Spencer and Liora Pedhazur Schmelkin. 2002. Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education* 27, 5 (2002), 397–409.
- [50] Pieter Spooren, Bert Brockx, and Dimitri Mortelmans. 2013. On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research* 83, 4 (2013), 598–642.
- [51] Marilla D Svinicki. 2001. Encouraging your students to give feedback. *New Directions for Teaching and Learning* 2001, 87 (2001), 17–24.
- [52] Ann Veeck, Kelley O'Reilly, Amy MacMillan, and Hongyan Yu. 2016. The use of collaborative midterm student evaluations to provide actionable results. *Journal of Marketing Education* 38, 3 (2016), 157–169.
- [53] Maxwell K Winchester and Tiffany M Winchester. 2012. If you build it will they come?; Exploring the student perspective of weekly student evaluations of teaching. *Assessment & evaluation in higher education* 37, 6 (2012), 671–682.