rppl

Annenberg Institute
BROWN UNIVERSITY

Does Providing Educators Choice over Feedback Improve the Effectiveness of Self-directed Professional Learning?

May 2026

Heather C. Hill
Harvard Graduate School of Education

Dora Demszky
Stanford Graduate School of Education

Eric S. Taylor
Harvard Graduate School of Education

Ashlee Kupor
Stanford University

Deepak Varuvel Dennison
Cornell University

Chris Piech
Stanford University

Introduction

Scholars and practitioners often recommend that educators have choice over their professional learning (PL) experiences – that is, that educators should shape when and how to engage in these experiences, and provide input about the focus of these experiences. Advantages of allowing educators this kind of choice include the potential for better alignment with their needs, enhanced engagement, and increased follow-through in terms of changes to their practice (Kennedy, 2016a). Weaving choice into learning opportunities has been endorsed by many leading scholars of professional learning (Carter Andrews & Richmond, 2019; Zeichner, 2019), and reflects a view of educators as knowledgeable professionals who can best identify their own learning needs.

However, granting educators more agency over the content and format of their PL comes at some cost. Differentiating learning pathways requires maintaining more PL offerings, and updating them periodically in response to teachers' changing needs and interests. Some also worry about teachers making inefficient choices, for instance by enrolling in professional learning opportunities that reflect interests or even strengths rather than areas of need in their practice.

Remarkably, despite the widespread belief that choice can enhance educators' engagement with and learning from PL, there exists little empirical evidence on the topic. What evidence does exist consists largely of illustrative cases or teacher self-reports (Brodie, 2021; Martin, Kragler, Quatroche & Bauserman, 2019; Philpott & Oates, 2017) rather than carefully controlled comparative or randomized studies. To address this gap, we conducted an experiment within an online course, giving instructors choice over the type of feedback they received about recorded lessons to determine how that choice impacts their engagement with the feedback, changes in practice, and student outcomes. Specifically, our study seeks to answer the following research questions:

(1) Does choice over feedback impact instructors' engagement with the feedback, perception of the feedback or their teaching practice?

(2) Does choice over feedback for instructors impact their students' outcomes?

(3) How do treatment effects vary by instructor demographics and whether the instructor engaged self-directed professional learning beyond automated feedback (training modules, teaching simulations)?

Methods

We conducted the study as part of Code in Place (CiP), a free, online, six-week-long introductory programming course created in 2020 by Stanford faculty members and delivered to an estimated 40,000 total students worldwide. While CiP material is primarily delivered asynchronously, volunteer instructors lead synchronous one-hour discussion sections each week to groups of about 10 learners.

Sections take place on Zoom and are recorded and automatically transcribed by Zoom's built-in automated speech recognition service. The scale at which CiP operates, the ability to vary key professional learning features like agency, and the automatic generation of outcome measures based on Zoom recordings make this an excellent example of a researcher-tech firm partnership (more information on these partnerships here).

Prior to the start of the course, CiP instructors are provided opportunities for learning, including training modules and teaching simulations (delivered via ChatGPT), emphasizing instruction that elicits students' ideas and then synthesizes and builds upon them. To this end, all instructors also receive feedback on their sections. Feedback is based on one of five "talk moves" that facilitate students' active participation per the Accountable Talk Framework (Michaels et al, 2008). The talk moves were divided into three types: *getting ideas on the table* (eliciting and revoicing student ideas), *building on student ideas* (adding on and probing ideas), and *orienting students to one another* (connecting student ideas). This feedback is generated automatically based on the Zoom transcripts by Large Language Models trained on classroom talk; this feedback also included experimental evidence about student curiosity. For each week's talk move, instructors received a) a summary statistic describing its frequency; b) a comparison of its frequency in the current week to prior weeks; c) examples of the move from the lesson's transcript; and d) a link to a training module on the specific talk move (see Figure 1). All instructors received emailed automated feedback the weekend after each class. This private feedback was intended to be non-judgmental, concise, actionable, and to encourage self-reflection.

Half of the 583 participating instructors[1] – a randomly assigned treatment group – got to choose the type of feedback they received. Such instructors could select the type and pattern of feedback they wanted across each pair of weeks (weeks 1-2, weeks 3-4, and weeks 5-6). The other half – the control group, also randomly assigned – did not get to choose their feedback type. Importantly, we matched the actual feedback received by the treatment and control group. For example, 36% of the treatment group chose the following pattern: two weeks on Getting ideas on the table, two weeks on Building on Ideas, and two weeks of feedback on student curiosity. Thus, 36% of the control group was assigned to that same pattern. This ensured that the only difference between the control and treatment group, in expectation, was whether the instructor had the option to choose their feedback pattern. Moreover, to reinforce the effect of the treatment, treatment group instructors were reminded that they had chosen their feedback in the email link to the feedback. About 80% of the treatment group instructors made a choice of feedback topic. The 20% of treatment group instructors who did not choose feedback were assigned feedback with the same weighted random assignment method as the control group, and remained part of the treatment group.

[1] Our participant sample were mostly young technology professionals with limited teaching experience. 66% of instructors were male, 32% were female, 1% were nonbinary and 1% preferred not to disclose their gender. Instructors' average age was 30, though they ranged from 18 to 75 years old. 48% were based in the United States, with the remaining 52% located in 70 unique countries.

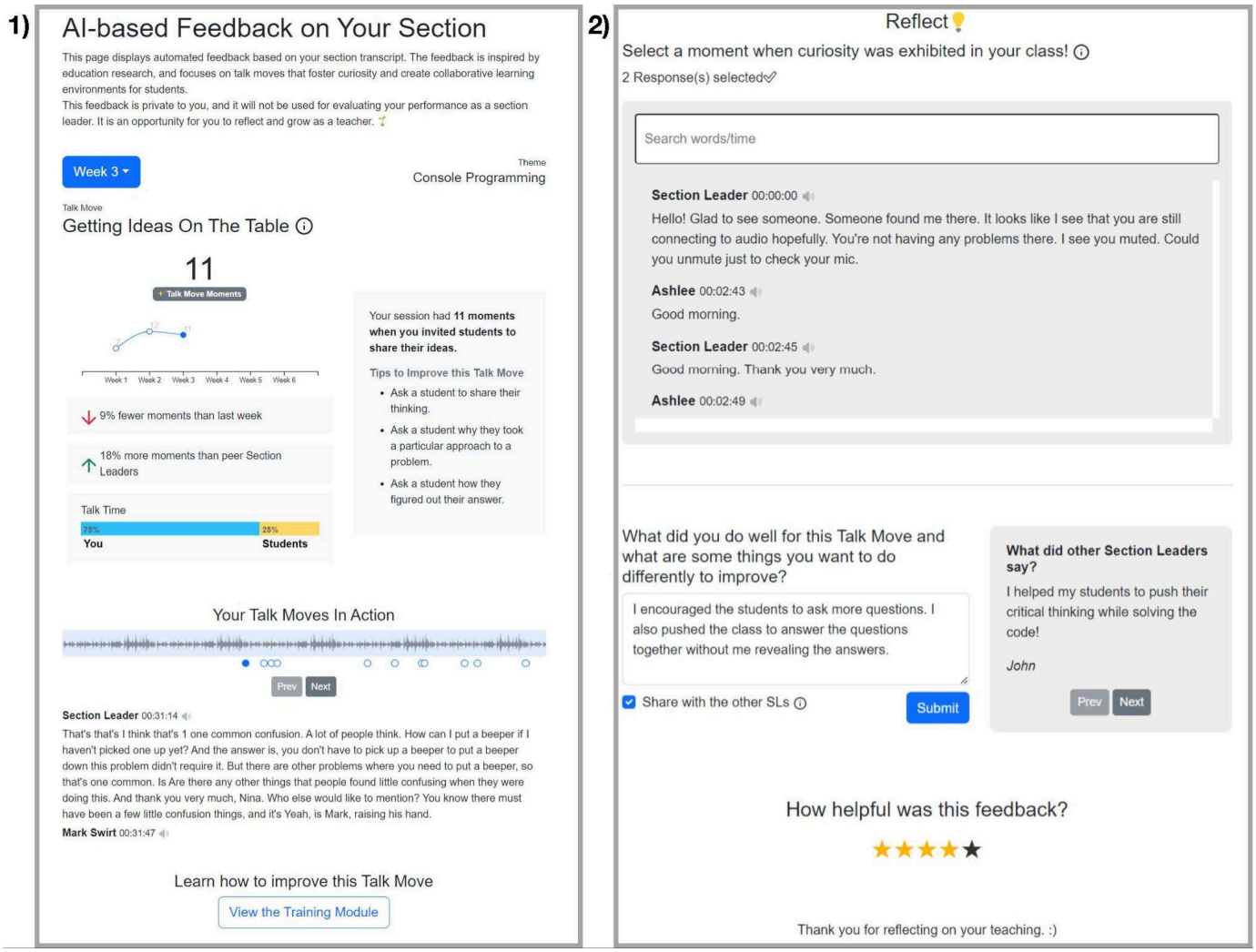

Figure 1: Screenshot of the feedback page, with a focus on the *Getting ideas on the table* talk move.

We measured four key types of outcomes to evaluate the impact of giving instructors choice over their feedback. First, we measured instructors' engagement with feedback by tracking whether they viewed the feedback as well as the total number of seconds they spent doing so. Second, we measured instructors' perceptions of the feedback, including their likelihood of recommending the system (calculated as a net promoter score) and perception of the feedback's utility. Third, we measured whether the feedback changed the frequency of instructors' talk moves. Because instructors started with varied teaching backgrounds and abilities, we controlled for the number of talk moves in the first session to observe relative changes in talk moves after receiving feedback in weeks two through six. Fourth, we measured student outcomes[2] by tracking the number of sessions attended and the number

[2] Our analytic sample also includes 8,254 students. Students were more balanced in terms of gender than instructors, with 52% of them identifying as female, 45% as male, 1% as non-binary and 2% as other/prefer not to say. Their average age was 31 years old. 28% of students were located in the United States, with the remaining 72% located across 145 unique countries. Our analytic sample also included 8,254 students who were taught by these instructors. Students were more balanced in terms of gender than instructors, with

of assignments completed. We chose these measures to capture both immediate instructor responses to having choice and downstream effects on teaching practice and student outcomes.

Findings

Research Question One – Did choice over feedback impact instructors' engagement with the feedback, perception of the feedback or their teaching practice?

We find that the treatment did not, on average, significantly impact instructors' engagement with the feedback, perception of the feedback, or their teaching practice. As shown in Table 1, across all outcome measures, the effects were statistically insignificant and relatively small in magnitude.

However, examining patterns over time reveals interesting trends. As shown in Figure 2, treatment group instructors tended to use more talk moves than control group instructors, particularly in the first three weeks of the course. This pattern was most pronounced for the "orienting students to one another" move, where treatment instructors maintained consistently higher rates throughout the course. While these differences did not reach statistical significance, they suggest that having choice over feedback may have had a positive impact on some aspects of instructional practice.

	Engagement		Perception		Practice	
	(1) Ever Viewed	(2) Seconds Spent	(3) NPS	(4) Overall Perception	(5) Wk 1 Talk Move Rate (std)	(6) Wk 2+ Talk Move Rate (std)
Treatment	−0.019 (0.027)	19.713 (69.283)	0.018 (0.393)	0.084 (0.126)	0.084 (0.071)	−0.001 (0.047)
Control Mean	0.876	483.979	6.03	3.505	−0.014	−0.014
R2	0.065	0.088	0.131	0.114	0.029	0.032
Observations	567	567	193	193	1611	7686

Table 1: Standard errors are in parentheses. These models estimate the effect of choice over the feedback (treatment) on instructor outcomes. Models 1–5 are at the instructor level, and Model 6 is at the instructor-week level. All models include instructor covariates (age, is female, is returning, in United States) and student covariates averaged within section (mean age, proportion female, proportion in United States). Models 5 and 6 additionally include dummy fixed effects for whether the given talk move was eliciting, building, or orienting. Model 6 also includes controls for baseline discourse features rates for eliciting, building, and orienting, and dummy indicator variables for the week of the session, rating between 2 and 6.

52% identifying as female, 45% as male, 1% as nonbinary, and 2% as other or "prefer not to say." Students were on average 31 years old, and they were located in 145 unique countries, with about 28% in the United States.

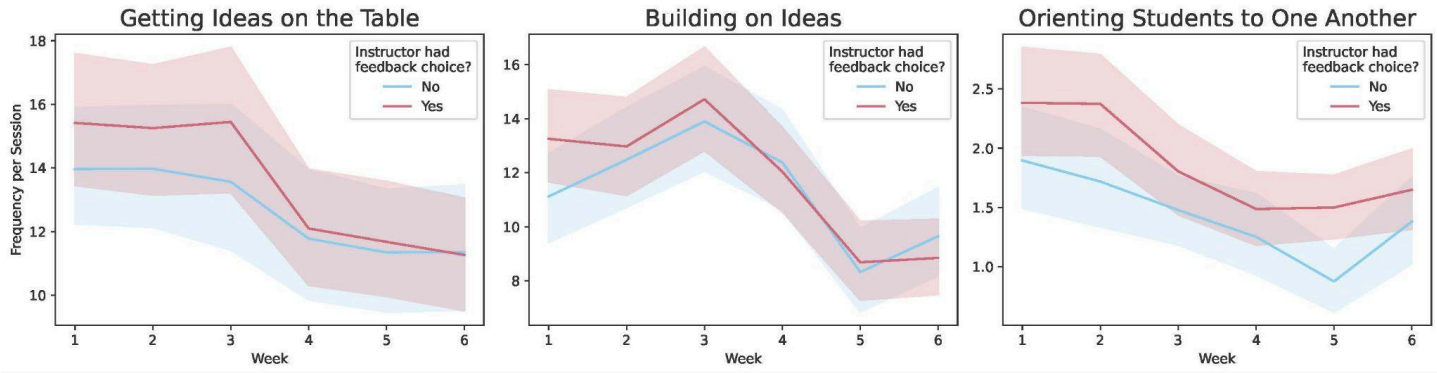

Figure 2: Talk move rates over time split by condition. Shaded bands represent 95% confidence intervals.

Research Question Two – Does choice over feedback for instructors impact their students' outcomes?

Next, we examined whether the treatment impacted student outcomes, including attendance and assignment completions. For assignment completions, we found no significant treatment effects — the coefficients were consistently positive but small for each assignment. Unexpectedly, however, we find a small but statistically significant positive effect on student attendance. As shown in Table 2, students whose instructors had choice over feedback attended on average 0.112 more sections compared to the control group. This represents a meaningful increase of about 3.1% over the control mean of 3.575 sections.

Given that students were unaware of the intervention, this effect may have been mediated through changes in instructor behavior. While we did not detect significant changes in our measured instructional practices, the treatment may have influenced other aspects of instruction that impacted student engagement.

	(1) Num. Sessions Attended	(2) Num. Assignments Completed
Treatment	0.112* (0.048)	0.050 (0.052)
Control Mean	3.575	3.434
R2	0.042	0.013
Observations	8254	8254

Table 2: Standard errors are in parentheses. These models estimate the effect of choice over the feedback (treatment) on student outcomes, at the student level. Both models include instructor covariates (age, is female, is returning, in United States) and student covariates (age, is female, in United States). Standard errors are clustered at the instructor level. * $p < 0.05$

Research Question Three – How do treatment effects vary by instructor demographics and whether the instructor engaged self-directed professional learning beyond automated feedback?

Finally, we turn to examining heterogeneous treatment effects. We did not find any notable variation in treatment effects based on instructor demographics. However, we found that giving instructors choice over feedback was most beneficial for instructors who also engaged with other self-directed professional learning resources (training modules and GPTeach).

As shown in Table 3, instructors who both received the treatment and completed training modules showed the strongest positive outcomes. These instructors had significantly higher baseline talk move rates in Week 1 (+0.198 SD, $p<0.05$), maintained higher talk move rates in subsequent weeks (+0.135 SD, $p<0.10$) and had significantly better student outcomes, with their students attending more sections (+0.260 sections, $p<0.01$) and completing more assignments (+0.176 assignments, $p<0.05$). Notably, these effects were greater than those observed for instructors who only completed modules (without choice) or only received choice (without completing modules). We observed similar patterns for instructors who used GPTeach to practice instruction.

These analyses suggest that providing instructors with choice over feedback was most effective when instructors were also intrinsically motivated to engage in other forms of professional learning. The combination of choice and engagement with supporting resources appears to have created a conducive environment for instructional improvement, leading to measurable benefits for students.

	Engagement		Perception		Practice		Students	
	(1) Ever Viewed	(2) Seconds Spent	(3) NPS	(4) Mean of Items	(5) Wk 1 Talk Move Rate	(6) Wk 2+ Talk Move Rate	(7) Num. Sessions Attended	(8) Num. Assn. Complete
(a) Treatment=0#Compl. Module=1	0.078* (0.037)	390.388** (94.403)	1.322* (0.602)	0.401* (0.191)	0.158 (0.105)	0.125+ (0.069)	0.113 (0.070)	0.058 (0.076)
(b) Treatment=1#Compl. Module=0	−0.007 (0.040)	143.350* (66.800)	0.740 (0.733)	0.261 (0.224)	0.114 (0.092)	0.018 (0.060)	0.081 (0.064)	−0.005 (0.071)
(c) Treatment=1#Compl. Module=1	0.041 (0.037)	240.703** (90.675)	0.796 (0.593)	0.347+ (0.193)	0.198* (0.099)	0.135+ (0.070)	0.260** (0.069)	0.176* (0.076)
(c) − (a)	−0.037	−149.685	−0.526	−0.054	0.040	0.010	0.147*	0.118
(c) − (b)	0.048	97.353	0.056	0.086	0.084	0.117	0.179**	0.181*
Control Mean	0.827	273.056	5.410	3.311	−0.070	−0.070	3.468	3.378
R2	0.074	0.113	0.155	0.138	0.033	0.026	0.043	0.013
Observations	567	567	193	193	1611	7992	8254	8254

Table 3: Treatment effects based on whether instructors completed any of the four optional training modules (moderator). Standard errors are in parentheses. Estimates are compared to the control group where the moderator variable is 0. The model specifications are the same as in Tables 1 and 2, with an additional interaction term as noted. Above we report the estimated total effect for each of the three groups relative to the control group, which did not complete any modules. The difference between point estimates was calculated with a robust Wald test. *p<0.05, **p<0.01, ***p<0.001

Limitations

Several limitations should be considered when interpreting these results. First, our study was conducted in the unique context of Code in Place; the effects of choice might differ in traditional education settings with trained educators and institutional incentives to participate in professional learning (Nezhad and Stolz, 2024). Whether these results generalize to K-12 settings is unknown. Second, while our automated measures of talk moves were validated, they are imperfect (Kupor, Morgan & Demszky, 2023) and may not capture all relevant aspects of instructional quality. The positive effects we observed on student outcomes despite limited changes in measured talk moves suggest there may be unmeasured changes in instruction that our current tools do not detect. Third, our intervention was relatively minimal, offering a one-time choice over feedback. It would be valuable to investigate the impact of providing instructors with more extensive opportunities to customize feedback. Finally, the

relatively short duration of our study (six weeks) limits our ability to observe longer-term effects of choice; future work should examine the impact of choice over extended periods.

Practical implications

The lack of a widespread benefit of choice to instructors raises practical questions about the optimal integration of educator choice into professional learning programs. Choice may be most efficient in contexts where participants are genuinely motivated and have the capacity to self-direct (O'Brien & Reale, 2021). On one hand, in elective professional learning—where instructors already have a clear interest in professional growth—increased autonomy can enhance motivation and ownership and potentially result in downstream impacts for students. For instance, by giving instructors choice over feedback, programs can better align professional learning content with immediate needs, which can improve both buy-in and teaching practice (Molla & Nolan, 2020; O'Brien & Reale, 2021). On the other hand, in mandated or large-scale K–12 professional learning contexts, educators are often time-constrained or less intrinsically motivated to experiment with new practices. As a result, providing extensive choice might not pay off in terms of engagement or outcomes (Lynch et al., 2019), especially when it creates logistical complexity for administrators.