Scholarship of Teaching and Learning in Psychology

Improving Students' Understanding Through Metacognition About Instructor Feedback: A Causal Modeling Approach

Sydney Y. Wood and Victoria L. Cross

Online First Publication, September 9, 2024. https://dx.doi.org/10.1037/stl0000412

AMERICAN PSYCHOLOGICAL ASSOCIATION

Scholarship of Teaching and Learning in Psychology

ISSN: 2332-2101

https://doi.org/10.1037/stl0000412

Improving Students' Understanding Through Metacognition About Instructor Feedback: A Causal Modeling Approach

Sydney Y. Wood and Victoria L. Cross

Department of Psychology, University of California, Davis

Previous research has documented learning benefits from metacognitive engagement and suggested many instructional practices to prompt metacognition in students. For example, a scaffolded curriculum that includes effective feedback may enable students to construct durable and nuanced knowledge. However, as ethical and logistical constraints prevent randomly assigning students to conditions or offering treatments in isolation from other confounding variables, true experiments are challenging to conduct in the scholarship of teaching and learning, and identifying causal relationships between teaching strategies and metacognition, or metacognition and learning, remains elusive. In this study, using an interrupted time series design and a series of structural equation models, we demonstrate the potential of causal modeling to analyze pedagogical data and evaluate a novel metacognitive intervention. A total of 445 undergraduate students in a psychological science research methods course at a large public university in the United States of America in 2023 were given the option to submit exam corrections based on feedback, formatted as an assignment designed to prompt metacognitive engagement. Exam performance over the academic term was compared between those who completed the intervention and those who opted out. Exam performance for those students who opted in was improved beyond what was predicted by their baseline trajectory. Through a series of eight nested structural equation models, this study presents evidence that the intervention was causally related to performance.

Keywords: metacognition, feedback, structural equation modeling, higher education, scholarship of teaching and learning

Supplemental materials: https://doi.org/10.1037/stl0000412.supp

Metacognition occurs when an individual uses external information or previous knowledge to evaluate and potentially change their own cognition. Engaging in metacognition allows a learner to understand what they know, identify what they do not know, and generate potential ways to bridge this gap. Metacognition is central to teaching and learning as some amount of metacognition is necessary for any amount of learning. Instructor-provided feedback is one source of external information that might prompt effective metacognition. However, identifying causal links between instructor feedback and student metacognition is challenging in an educational setting.

Research in educational settings has the benefit of authenticity but the limitation of an inability to control extraneous variables and identify causal relationships. Our ethical and instructional practices restrict our ability to run true experiments. Students are free to select which course or section they enroll in, so we cannot draw clear comparisons between sections. Treating students differently to manipulate an independent variable can lead to inequity in learning outcomes (Bunnell et al., 2022), potentially violating our ethical responsibility as instructors. Furthermore, ensuring effective manipulation in a classroom setting can be difficult as students engage in research procedures and course activities at varying levels of sincerity

Sydney Y. Wood https://orcid.org/0000-0002-5635-046X

Portions of these findings were presented at the 2023 Western Psychological Association Conference and at the 2023 Society for Teaching Psychology Annual Conference on Teaching.

Correspondence concerning this article should be addressed to Sydney Y. Wood, Department of Psychology, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States. Email: sywood@ucdavis.edu

and effort (Bernstein, 2018). Without being able to conduct true experiments and isolate causal relationships, we risk identifying what appear to be promising practices but are just spuriously related to learning outcomes (see Pogrow, 2017). This phenomenon can be seen where similar practices produce conflicting results, defying attempts to be concisely summarized in meta-analyses (Freeman et al., 2014; Kizilcec et al., 2020; Shi et al., 2020). If we abandon classroom studies and bring participants into a laboratory setting, we gain the ability to create equivalent groups through random assignment and the ability to manipulate variables, but we then struggle to find authentic ways to measure learning, especially longitudinally (Bernstein, 2018).

Scholarship in higher education often relies on quasi-experimental methods in authentic contexts (Bunnell et al., 2022; Divan et al., 2017; Felten, 2013). By implementing promising practices in a classroom setting, not only does the research maintain authenticity, but it also includes ongoing opportunities for data collection through assessments, coursework, and behavioral outcomes. Collecting data at multiple time points both before and after introducing an intervention also affords the opportunity to apply causal modeling techniques, such as structural equation modeling, to provide a richer analysis of quasi-experimental data when random assignment is not possible.

In the present study, we apply causal modeling to quasi-experimental data to evaluate a potential causal relationship between instructor feedback and student learning. In an *Introduction to Research Methods in Psychology* course, students were challenged to demonstrate their knowledge at multiple time points. Specifically, on five exams, students identified relevant information and wrote nuanced critiques of research summaries. As motivation for our study, we had anecdotally observed many of our students repeatedly making the same mistakes on these written exam prompts, despite the similarity of the contexts and the availability of feedback on previous exams. To encourage students to review and use their exam feedback effectively, we created an intervention designed to prompt metacognition through reviewing and responding to research-informed feedback.

The Role of Feedback and Metacognition in a Scaffolded Curriculum

Metacognition is the process by which an individual subjectively attends to their own knowledge, cognition, and learning (Flavell, 1979). Feedback provides additional information regarding performance on a task and is one mechanism to prompt metacognitive thought. Through metacognition, students not only learn the material but also learn about their thinking, memory, and knowledge. This self-knowledge and the ability to engage in metacognition becomes durable enough to be applied to other learning contexts in the classroom and beyond (Wiggins, 2012).

The Role of Metacognition in Learning

Metacognition can range from automatic to strategic. In lower-order learning, metacognition is often automatic, leaving the learner unaware that it is happening (Veenman et al., 2006). In contrast, higher-order learning is often characterized by strategic or conscious metacognition. Strategic metacognition is defined as engaging in some combination of eight pillars of cognition about cognition: academic knowledge of cognition, operational knowledge of cognition, self-monitoring, self-regulation, adaptation, recognition, discrimination, and mnemosyne (Drigas & Mitsea, 2020). When engaging in strategic metacognition, an individual may experience any combination of these pillars. However, the more pillars an individual engages in during the learning process, the more advanced the metacognition process is and, potentially, the more effective the learning.

The first three pillars are the underlying components of metacognitive awareness, a very well-researched domain within the scholarship of teaching and learning (SoTL; Cao & Nietfeld, 2007; Çini et al., 2023; Lee et al., 2010; Miller & Geraci, 2011). Metacognitive awareness is typically very developed among students studying psychology at a competitive university (Woods & Cross, 2020). These students have experienced much academic success (e.g., entrance to a competitive university) and are aware of the work they must complete to achieve their academic goals. However, research has shown that metacognitive awareness does not correlate strongly with self-regulatory study skills (e.g., pillars: self-regulation) or adaptation of new knowledge (e.g., pillars: adaptation). For example, Cao and Nietfeld (2007) documented that while students have an excellent academic understanding of metacognition, metacognitive awareness does not automatically translate into self-regulatory and

adaptive behaviors or other metacognitive skills. Our intervention aims to explicitly connect self-monitoring and self-regulatory behaviors, while promoting higher engagement with instructor feedback and course content through adaptation.

Developing Metacognition Through Feedback

We anecdotally observed many students repeatedly making similar mistakes as they learned to discuss and critique research. First, we ensured that we were providing effective feedback. In teaching, an instructor's primary goal for providing feedback is to provide the learner with additional information regarding their performance on a task. Previous research has documented variability in feedback quality and effectiveness, and identified guidelines to maximize efficacy (see Dawson et al., 2021; Hattie & Timperley, 2007; Ryan et al., 2024). To be effective in higher education, feedback must be able to be internalized, actively processed, and applied in multiple contexts (see Carless, 2022; Ryan et al., 2021; Wiggins, 2012). Students are most likely to engage with feedback that is goal-directed, explains how the response differs from a correct response, encourages students by acknowledging when students do well, refrains from using value-based language, and provides actionable advice (Dawson et al., 2021; Ryan et al., 2024; Wiggins, 2012). To reduce our concern that the quality of the feedback was the reason students made repeated mistakes, we created standardized exam rubrics using Wiggins's (2012) seven guidelines for giving good feedback.

Second, we promoted student engagement with and understanding of our feedback. Providing feedback would not produce metacognition and meaningful learning if the students are unable or unwilling to engage with it (Haddara & Rahnev, 2022). Evidence suggests that learners' understanding of their own agency in learning is essential to developing feedback literacy and their awareness of their learning process (Callender et al., 2016; Carless, 2022; Carless & Boud, 2018; Haddara & Rahnev, 2022; Sato & Loewen, 2018; Winstone et al., 2017, 2019; Yan & Carless, 2022). Therefore, we developed our intervention to promote students' feedback literacy by leveraging scaffolding and metacognition.

Scaffolding provides an opportunity for students to develop feedback literacy by modeling the metacognitive process (Ambrose et al., 2010; Finn & Metcalfe, 2010; Lovett et al., 2023; Tanner, 2012). Novice learners may struggle to extrapolate common factors and require multiple feedback iterations to abstractly consider the underlying principles (Ambrose et al., 2010; Bailey et al., 2017; Miller & Geraci, 2011; Sato & Loewen, 2018). Using a scaffolded curriculum, an instructor may start by providing assignment-specific or task-level feedback and gradually connect repeated feedback explicitly to provide process-level information (Ryan et al., 2021; Tempelaar, 2020). Through process-level information, students begin to independently recognize patterns and gain a deeper understanding of the underlying concepts (Nicol & McCallum, 2022; Ryan et al., 2024). They can then apply this new knowledge in similar but novel contexts (Corral & Carpenter, 2023). Once students begin making such connections across contexts, instructors can prompt metacognition through guided problem solving on complex tasks that explicitly incorporate different components of previous content. We created this scaffolding by designing this course with frequent exams that repeatedly challenged students to apply knowledge to novel research examples.

Past Interventions Promoting Metacognitive Engagement With Feedback

With mixed success, many practices have been used to encourage metacognitive engagement with instructor feedback (see Cohn & Stewart, 2016; Lee et al., 2010; Molin et al., 2020; Naujoks et al., 2022; Nicol & McCallum, 2022; Saenz et al., 2019). Despite the abundance of proposed strategies, the field lacks a comprehensive understanding of which approaches effectively foster sustained metacognitive engagement and feedback literacy among students. The intervention documented in this study combines two of the most promising practices: reflective writing and exam wrappers. The intervention comprises reflective writing elements adapted from Daniel et al. (2015) and Cohn and Stewart (2016), presented as an exam wrapper that allows students to correct exam answers by incorporating feedback.

Reflective writing is a commonly used way to engage students' metacognitive awareness. Research has documented that having students write reflections on their learning is associated with improved metacognition, enhanced self-

regulation, deeper understanding, and increased feedback engagement (Carless & Boud, 2018; Cohn & Stewart, 2016; Daniel et al., 2015; Ion et al., 2019; Sachar, 2020; Winstone et al., 2017). This strategy works best when the writing is low-stakes and includes specific questions to prompt self-reflection (Cohn & Stewart, 2016; Colthorpe et al., 2018). Often, reflective writing is paired with an assignment where feedback had previously been given to students. This practice encourages students to think deeply about their learning experiences, understand their thought processes, and connect their actions and outcomes (Carless & Boud, 2018; Cohn & Stewart, 2016; Daniel et al., 2015). Additionally, through reflective writing, students can set specific, actionable goals for their learning and improvement while documenting their steps toward attaining those goals. This forward-looking aspect helps students take control of their learning and set a direction for future courses.

Exam wrappers are a metacognitive tool to help students engage in self-assessment and reflective thinking about their performance on exams or other question-based assessments (Ambrose et al., 2010; Lovett, 2013; Naujoks et al., 2022; Owen, 2019; Rowell et al., 2023; Soicher & Gurung, 2017). In their most general form, exam wrappers are additional questions about the experience of preparing for and taking an exam that can prompt students to reflect on their performance, identify areas for improvement, and develop strategies for enhancing their learning and study techniques (Edlund, 2020; Owen, 2019; Rowell et al., 2023; Soicher & Gurung, 2017). By engaging in this reflective process, students become more aware of their strengths and weaknesses and can make informed decisions about improving their future performance (Basey et al., 2014; Lee et al., 2010). However, the efficacy of exam wrappers in strengthening students' exam performance has been inconsistent (LaCaille et al., 2019; Rowell et al., 2023; Soicher & Gurung, 2017). Our study presents a novel evaluation of an exam wrapper intervention that combines metacognitive prompting with active engagement with instructor feedback through an exam wrapper assignment.

The Present Study

In the present study, we created an intervention that instructed students to engage in reflective writing in combination with an exam wrapper that prompted them to make corrections based on instructor feedback. The intervention first promoted metacognitive awareness through recorded videos explaining how to engage in metacognition and the importance of metacognition in learning, then required students to engage in adaptation by correcting their written exam responses using feedback. Finally, it instructed students to engage in self-regulation and self-monitoring through writing a cover letter explaining their exam corrections process and the change in their understanding of the course topic overall. While this intervention is adapted from previously evaluated interventions, the strategic implementation and opportunity for causal modeling techniques provide new evidence for evaluating its effect.

We used a quasi-experimental interrupted time series with a comparison group design to assess whether engaging with the metacognition intervention improved the performance in subsequent exams. We were unable to require or prohibit students from participating in the intervention, so we could not randomly assign students to treatment and control conditions. However, we designed the study to have three preintervention and two postintervention exams. These repeated measures allow for an accurate assessment of the performance trajectories of both groups before and after the intervention. As participants self-selected to treatment groups, we fully expect underlying differences between the opt-in and opt-out groups (i.e., motivation, personality, past performance, time availability, etc.). However, if the growth trajectories for each group are equivalent before the intervention, we can be reasonably confident that any differences in the trajectories after the intervention suggest that the intervention impacted those who opted in. To estimate this potential impact, we used latent growth curve modeling, a form of structural equation modeling that can provide support for causal relationships when random assignment is not possible.

To support a causal interpretation of the metacognition intervention improving exam performance, our data must first pass various assumption checks in addition to identifying the hypothesized differences in performance. We wrote our hypotheses with these assumptions in mind to assess how well our study fulfills Mill's criteria for causation in the absence of random

assignment (Pearl, 2009; St.Clair et al., 2016). The overarching hypothesis is that even when controlling for demonstrated learning before the intervention, students who engage in metacognition through exam corrections and cover letter writing will demonstrate a better understanding of research validity, compared to those who did not do exam corrections (see preregistration at https://osf.io/a6gbn). This hypothesis can be broken down into four separate hypotheses. If the intervention had a causal impact on those who opted in, we will find the following:

> *Hypothesis 1 (nonlinear trends):* Due to the varying difficulty of each exam, performance trajectories will not be linear.
>
> *Hypothesis 2 (parallel trends assumption/ equivalent groups):* Before the intervention, students who opt-in will have the same expected performance trajectories as those who do not opt-in to the treatment.
>
> *Hypothesis 3 (covariance and temporal precedence):* After the intervention, students who opted into the treatment will show greater improvement in exam performance than students who opted out.
>
> *Hypothesis 4 (eliminating alternative explanations):* Inherent differences between groups will not be sufficient to explain the change in performance trajectories between groups after the intervention.

Method

Participants

Participants were 370 undergraduate students enrolled in an in-person research methods course offered at a large public university in the United States of America in 2023. In total, 445 students were enrolled in this class. Seventy students did not consent to have their data included in this study. Five additional students were excluded from the analyses as they missed more than one exam and did not complete the course.

Demographics

Participants self-reported their age, gender, parental education, ethnicity, familiarity with academic English, financial support, and admittance type. Ages ranged from 18 to 42 ($\bar{X} = 19.43$, $SD = 1.75$), with only one participant declining to state their age. Overall, 72.82% identified as women, 20.32% as men, and 4.22% as nonbinary or genderqueer. Many students (33.51%) indicated that they were among the first generation of their family to attend university (i.e., neither of their parents had earned a college degree), and more than one quarter (27.70%) identified themselves as belonging to a historically underrepresented ethnic minority (URM) group. Familiarity with academic English was estimated through responses to a 7-point scale ranging from 0 (*I am not confident reading in English*) to 6 (*I am confident reading very complex information in English*, e.g., college textbooks or published papers). The median response was the highest level of confidence in reading ability. Most students reported receiving significant financial support from their family of origin (e.g., tuition and housing) on a 7-point scale ranging from 1 (*I receive no financial support from my parents*) to 7 (*fully supported*; e.g., all necessities are provided or paid for). Most students started their university education at this institution (90.24%), with only 7.39% having transferred from a community college. There were no significant differences between the opt-in and opt-out groups regarding these student demographics (all $\chi^2 < 2.37$, all $p > .05$; see Table 1).

Course Context

The learning objectives for this required *Introduction to Research Methods in Psychology* course focus on scientific vocabulary and critical thinking about validity in research. The course spanned a 10-week term. The content was delivered through 3 hr of in-person lectures each week, readings from a required textbook, and homework assignments. Students enrolled in one of the two lecture sections, and all had access to the same course site on the learning management system. The sections had the same enrollment cap, were offered in the same room, and were scheduled consecutively (9–10:20 a.m. and 10:30–11:50 a.m.) on Tuesdays and Thursdays. A single instructor designed the course, managed the course learning management system site, wrote the assignments and exams, delivered all the lectures for both sections, held office hours each week, and worked closely with the seven

Table 1
Descriptive Statistics of Student Demographics

Variable	*n*	$\bar{X}$	*SD*	*Mdn*	MAD	Minimum	Maximum	Difference by intervention	
								U	*p*
Age	370	19.43	1.75	19	1.48	18	42	17,595	.62
Financial support	370	5.57	7.9	6	1.48	1	7	16,629	.93
English proficiency	370	5.19	0.98	6	0	0	6	16,147	.34

Variable	*n*	%	Difference by intervention	
			χ^2	*p*
URM	105	27.70%	2.23	.14
First generation	127	33.51%	0.41	.52
Transfer	28	7.39%	0.12	.73
International	54	14.25%	2.37	.12
Gender (level)			0.43	.81
Man	77	20.32%		
Woman	276	72.82%		
Gender-queer	16	4.22%		

Note. Due to significant skew in distributions, Mann–Whitney *U*-tests of median differences were used to determine any difference in the decision to opt-in to the intervention by continuous demographic variables. MAD = median absolute difference from the median; URM = underrepresented ethnic minority.

graduate student-teaching assistants (TAs). The TAs graded the exams, and each held office hours each week.

The curriculum included five exams, each with multiple-choice and short-answer questions. The multiple-choice questions assessed students' understanding of the vocabulary and concepts presented in the current unit. The written short-answer questions assessed students' ability to apply their understanding to novel research studies by prompting them to identify and evaluate the research design of a provided example. The multiple-choice questions were not cumulative and were not repeated measures; they solely tested understanding of the most recently presented course content. Therefore, this longitudinal research design preregistered the analysis of only the responses to the written portion of each exam, as these repeatedly measured the same ability.

The students were unaware of the research example they would be asked to evaluate before they began each exam. However, as many of the prompts (i.e., "Please identify and write in your own words the operational definition of the predictor variable") were repeated verbatim on each exam, the students were aware of the scope of the prompts. The prompts and rubrics were provided with practice examples well in advance of each exam (see Supplemental Materials for a sample practice exam). The students were repeatedly shown that many of the prompts were identical from one exam to another (though they would require different answers depending on the details from the research example) and explicitly told that the feedback on one exam would apply to the subsequent exams. The full text of the exam, the exam responses, and the graduate student TAs' feedback were returned to students before the next exam using Gradescope, an online grading tool. Gradescope was configured to display to each student which rubric elements they had satisfied as well as those they had not satisfied. Therefore, all the rubric elements were visible to all students.

The design of these rubrics and feedback were evidence-informed. In an attempt to reduce the negative emotions and avoidance documented by other researchers (e.g., Carless & Boud, 2018; Eskreis-Winkler & Fishbach, 2022), the rubrics primarily used positive grading with points awarded for correct elements and a few deductions for errors. To increase the likelihood that the feedback given to students was adequate, the rubrics were created using Wiggins's (2012) seven guidelines for providing good feedback. The feedback was *goal-referenced* as it was

explicitly connected to the learning objectives for that unit. Each student had *transparent* access to the entire rubric and scope of possible feedback, alongside the *tangible* feedback specific to their performance. Whenever possible, feedback was *actionable* by including instructions on producing a correct response to the prompt or providing examples of necessary details. Gradescope is a *user-friendly* grading technology that can be accessed on any internet viewing device (e.g., smartphone, tablet, or computer) and displays the feedback with the relevant prompt. We attempted to use *accessible* language and provided additional comments to elaborate on standardized rubric items. Feedback was *timely,* as it was returned at least a week before the next exam. Feedback was given on an *ongoing* basis, as the exams were scheduled frequently, and the feedback from one exam was directly relevant to portions of the subsequent exams. Graduate student TAs held calibration meetings for each exam to ensure that rubrics were *consistently* applied.

Materials

Assessments

Each of the five in-person exams presented a novel one-page summary of recent research followed by prompts requiring students to describe and evaluate that research. Novel examples were used on each exam to maximize the possibility of measuring actual comprehension of the material. Students could not simply regurgitate memorized responses; they must generate relevant answers during the exam. For Exam 1, students identified and explained some important information from the summary (i.e., variable labels, operational definitions, hypotheses, and results). Each subsequent exam included at least one identically worded prompt (i.e., paraphrasing the operational definition) from all previous exams and prompts relevant to new material (e.g., evaluate construct or internal validity). In this way, the prompts were repeated verbatim across exams, but the answers changed as they were specific to the research being described. The research summaries used on each exam were similar in complexity. All had only one predictor variable and one outcome variable. The variables were either categorical (with only two levels) or continuous, and were all clearly described as either measured or manipulated (see Supplemental Material for the repeated exam prompts).

Intervention

The intervention was a single extra credit assignment after Exam 3, inviting students to learn about metacognition, submit exam corrections, and demonstrate metacognition. Learning about metacognition was accomplished by providing links to two videos describing the process of metacognition and its importance in learning. The first video gave an overview of the process and benefits of metacognition with actionable advice on how students can practice metacognition (Peterson's Test Prep, 2020). The second video outlined research in metacognition and growth mindsets and promoted metacognitive skills by providing actionable advice and exercises for higher education learning (Orpurt, 2018). Completing the exam correction required identifying an incorrect response from either Exam 2 or Exam 3, copying the specific feedback given, and correcting the response using that feedback and other course resources. Demonstrating metacognition required writing a cover letter explaining the original misunderstanding alongside the metacognitive processes employed to reach the new understanding. Allowing corrections from two exams ensured that all students could opt in to this intervention. All students had taken at least one of these exams, and no student received perfect scores on both exams; therefore, all students had corrections they could have made.

Measures

Academic Performance

Scores on the five exams were downloaded from the course gradebook at the end of the term. Graduate student TAs graded the short-answer written responses using clear and strict rubrics. All rubric items were standardized for repeated prompts. The TAs met to determine the relevant prompt-specific details and calibrate their grading for each exam. Students were able to request regrades for any exam prompts. All regrade requests were resolved before the end of the term.

Demographic Survey

Students were rewarded for completing a brief demographic survey by earning 0.5% extra credit points toward their final grade. The beginning of the survey included a consent form. Students

were informed that the survey responses and performance on other assignments and assessments would be used for pedagogical research. Students could decline to be part of the study but still gain the extra credit and complete the intervention. All students who declined consent or did not take the survey were excluded from the analysis.

As listed in the preregistration (see preregistration: https://osf.io/a6gbn), data on more variables were collected but not included in this analysis.

Procedure

This data was gathered during a single academic term. There were no significant disruptions to the curriculum. Students completed Exams 1, 2, and 3 before the intervention. After the grades and feedback were returned for Exam 3, the optional metacognition intervention was offered as an opportunity to earn up to 0.5% extra credit toward their final grade. Students were given 5 days to complete the intervention. TAs who were not involved in the research graded the submissions and gave feedback on the exam corrections and cover letters. The grades and feedback were returned before the students sat for Exam 4. Exam 5 was taken at the end of the term and without additional metacognitive prompting.

Results

Of the students who consented to the use of their data, 180 opted out of the intervention and 190 opted in to the intervention. All hypotheses were evaluated with a series of nested structural equation models. We preregistered our significance level for these analyses to be $\alpha = .01$ to combat the inflated Type I error due to repeated testing. A power analysis showed we have over .95 power to detect an effect size of .1 at the .01 α level.

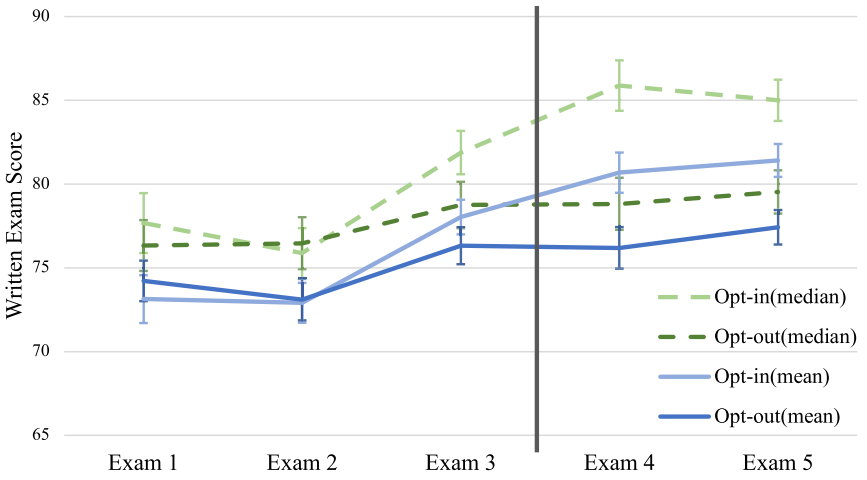
Descriptives: Exam Performance

The 370 students each had an opportunity to take five exams for a total of 1,850 possible exams. Overall, 29 exams were missed (1.57%). Three to eight students missed each exam. Students who missed more than one exam were excluded from the study, as they likely needed to retake the course. Therefore, these 29 missed exams are from 29 unique students.

Measures of central tendency and variability for exam performance over time by group are graphed in Figure 1. The median performance was consistently higher than the mean performance; as expected in assessment data, exam performance has a significant negative skew for

Figure 1

Mean and Median Performance by Exam and Group



Note. The error bars represent standard error of the mean and standard error of the median. The vertical gray line indicates when the intervention occurred. See the online article for the color version of this figure.

both groups over all five exams (Shapiro–Wilks Normality test all $p < .01$, see Supplemental Material for code and results). According to Levene's Test of Equality of Variance, the groups did not significantly differ in variances at each exam ($p > .05$). Regardless of group, students showed very little change in performance between Exam 1 and Exam 2—all four means or medians fall within the same margins of error. Visual inspection revealed that performance on Exam 3 shows a larger separation between those who later opted in and those who opted out, particularly when considering the median performance. On Exams 4 and 5, after the intervention, there are pronounced differences between the groups; there is no overlap in standard error bars for either the means or the medians. The medians, which are the more robust measure of central tendency for skewed data, overall showed a larger difference between group performances.

Latent Growth Curve Models

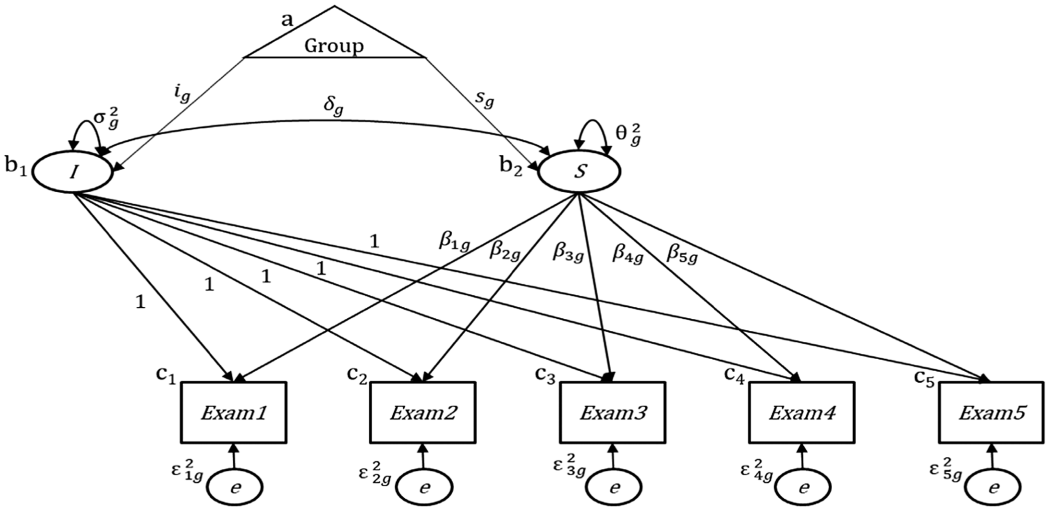
Structural equation modeling is a very versatile and powerful modeling technique. It uses multiple continuous measures (*items*) and categorical predictors to estimate a theoretical (*latent*) construct of interest (Bollen & Curran, 2006). In addition, structural equation modeling allows for flexibility in the data, especially regarding skewed and missing data (McArdle et al., 2009). We estimated our models using full information maximum likelihood, which allows the models to account for missing exam observations at the individual level (Enders & Bandalos, 2001). Of the 29 missed, nine exams were missed by the opt-in group (0.95%), and 20 were missed by the opt-out group (2.22%). This shows a marginally significant difference in willingness to miss an exam, $\chi^2(N = 370) = 5.20$, $p = .02$, prompting some concern about missing not-at-random data. However, the amount of missing data (1.57%) is negligible in the context of our sample size and estimated power. These rates of missing data are within the parameters accounted for by full information maximum likelihood. The structural equation modeling used here,latent growth curve and latent basis growth curve models, when specified based on preregistered theoretical expectations, allow researchers to make causal inferences using changes in shared variances (covariance) across time (temporal precedence) and compare the fit of the hypothesized relationships to the null, as well as alternative explanations of the data (elimination of alternative explanations; Mill, 1882; Pearl, 2009).

We conducted eight nested latent growth curve models and a comparison null model (see Figure 2 for model specification and parameters of interest). Three of these models are linear (see Table 2), and five are nonlinear (see Table 3). These models increase in complexity following theoretical justification. As each model is fit to the data, a chi-square test of differences evaluates the change in model fit statistics ($\Delta\chi^2$) and indicates if the more complex model provides a significantly better fit than the previous (simpler) model. If the fit is significantly improved, we reject the simpler model and are justified in holding more complex assumptions about the relationship between the variables. A significant improvement from one model to the next does not indicate that we have reached the best model. We continue through the nested models, systematically introducing more complex explanations based on theoretical predictions, until we reach a model that is not an improvement over the previous model (see Appendix for the model estimates for each of the following models).

Linear Models

The null model assumed no growth over time and no difference between groups. That is, it estimated each student's score on Exams 2 through 5 to be the same as his, her, or their score on Exam 1. Model 1 estimated a main effect of time: whether the findings could be explained by linear growth over time, with performance increasing at the same rate across all exams and with no difference between the groups (see Figure 3A for Model 1 estimated trajectory). Model 1 fit significantly better than the null model, $\Delta\chi^2(3) = 102.6$, $p < .001$. Model 2 estimated significantly different slopes and intercepts (i.e., mean structure) between students who opted in to the intervention and those who opted out; this is a linear regression with a categorical predictor, $\Delta\chi^2(2) = 73.28$, $p < .001$. Model 3 tested whether the groups had differing interdependence between their starting performance and growth trajectories. Model 3 was not significantly better than Model 2, $\Delta\chi^2(1) = .59$, $p > .05$. If we assume that the growth in exam performance is linear, there were no significant differences in the covariance between the mean

Figure 2
Latent Growth Curve Path Diagram With Parameter Labels



Note. Components of latent growth curve: (a) Categorical predictor variable-represents the impact of intervention. (b) Latent variables- I [b_1] represents the *intercept* or expected performance on Exam 1. The *paths* (arrows) from the *intercept* to each observation is always 1. S [b_2] represents the *slope* or change in performance over time. The paths from the *slope* to each exam observation represent each exam's effect in student performance. (c) Items—represent the raw observations for each individual. Each exam is an item in the model. Parameters: Each parameter can be *freed* to be estimated for each group (g). For models to converge, a minimum number of parameters must be *fixed* to specific values, usually 1 or 0. When parameters are *constrained* to be equal across groups, the subscript g is dropped and only one value is estimated for the entire sample. i_g—the estimated intercept represents the expected starting performance. σ^2_g—the estimated variance in the intercept represents the variability of scores on Exam 1. s_g—the estimated slope represents the expected rate of change in performance over time. θ^2_g—the estimated variance in the slope represents the variability in each individual's rate of change over time. δ^2_g—the estimated covariance between intercept and slope represents the codependence of growth on initial performance. β_{ig}—the estimated loading for each observation represents the effect of the exam on performance (after accounting for expected rate of change). ε^2_{ig} —the estimated error variances represent individual differences for each observation.

structures for either group. Therefore, Model 2 was the best-fitting linear model (see Figure 3B).

Nonlinear Models

A latent basis model estimates the loadings from the latent variable performance to each exam score. The loadings are multiplied by the slope and added to the intercept to give the expected score for each exam. Loadings are equivalent to acceleration; they estimate the speed at which the slope changes over time. Computationally, to estimate learning trajectories, two loadings must be fixed (set by the researcher) to specific values. Consistent with theory, we fixed the loading for Exam 1 to be 0 and the loading for Exam 3 to be 1. These loadings indicate no growth from the intercept at Exam 1 and the expected maximum growth achieved before the opportunity to engage in the intervention. By estimating the loadings for Exams 2, 4, and 5, we can compare the shape of estimated trajectories before and after the intervention (see supplemental analyses for the results of additional assumption tests).

To test the simplest of the latent basis assumptions, Model 4 introduced nonlinear loadings across time while returning to the assumption that there were no group differences. Model 4 fit better than the simplest linear model, Model 1; $\Delta\chi^2(3) = 12.08, p < .01$. This estimated that growth increased faster after Exam 3 than before Exam 3 and supported Hypothesis 1 that performance should not be assumed to have a linear trajectory (see Figure 3C).

Having established that growth was nonlinear, Model 5 introduced the possibility that a difference occurred between the groups *after* the intervention. Model 5 fit significantly better than Model 4, $\Delta\chi^2(2) = 12.16, p < .005$. This suggested that there are true differences between students who opted in

Table 2
Linear Models

Model	Growth over time	Group difference	Compared to model	χ^2 test of difference	p	Conclusion
0. Null	None: Flat slope	None				Rejected
1. Linear growth with no group differences	Linear: Same rate across all exams	None	Null	$\Delta\chi^2(3) = 102.6$	$p < .001$	Better fit than null model
2. Linear growth with intercept and slope (mean structure) differences by group	**Linear: Same rate across all exams**	**Groups may have different intercepts and slopes (not related to intervention)**	1	$\mathbf{\Delta\chi^2(2) = 73.28}$	$\boldsymbol{p < .001}$	**Better fit than Models 1 and 3** **Best linear model**
3. Linear growth with mean structure and covariance differences by group	Linear: Same rate across all exams	Groups may have differing interdependence between their starting performance and growth trajectories	2	$\Delta\chi^2(1) = 0.59$	$p > .05$	Rejected

Note. The bolded font indicates the best fitting model.

to the intervention and those who did not on both Exam 4 and Exam 5 (see Figure 3D). However, as this model only considered group differences after the intervention, we tested Model 6 to determine whether the assumption of parallel trends holds and determined whether there were significant differences in performance trajectory between the groups *before* the intervention. Critically for Hypothesis 2, we found that the parallel trends assumption held: Model 6 did not fit significantly better than Model 5, $\Delta\chi^2(1) = 0.1$, $p > .1$. This provided causal evidence of the impact of the intervention on exam performance of students who opted in to the assignment. However, Model 5 still assumes that the only difference between groups is the intervention, so we must test whether there are any mean structure differences between groups that are separate from the effect of the intervention.

Model 7 estimated slopes and intercepts for both groups under the parallel trends assumption. Model 7 fit significantly better than Model 5, $\Delta\chi^2(2) = 16.3$, $p < .005$, indicating inherent differences between groups not explained by the intervention alone. As the inherent differences between groups may invalidate the causal interpretation, we tested the alternative explanation that the intervention has no effect and that any differences in performance are due to inherent differences between the groups (mean structure differences). Model 8 constrains the nonlinear loadings to be the same across groups but estimates the mean structure to be different. Therefore, Model 8 is a less complex model than Model 7.

As predicted, we found that Model 7 fit significantly better than Model 8, $\Delta\chi^2(2) = 10.31$, $p < .01$ (see Figure 3F). Therefore, Model 7 explained significantly more variance, suggesting that the intervention positively impacted the expected exam performance ($E[Y]$) of students who opted in; see Table 4 for raw model coefficients[1] and Figure 3E for graphed trajectories. According to Model 7, both groups performed equivalently on Exam 1 ($E[Y]_{\text{opt-in,exam1}} = 72.80$, $SE = 1.42$ and $E[Y]_{\text{opt-out,exam1}} = 74.00$, $SE = 1.10$) and Exam 2 ($E[Y]_{\text{opt-in,exam2}} = 73.93$, $SE = 1.64$ and $E[Y]_{\text{opt-out,exam2}} = 74.03$, $SE = 1.32$). On Exam 3, the opt-in condition ($E[Y]_{\text{opt-in,exam3}} = 78.31$, $SE = 1.42$) performed slightly better than the

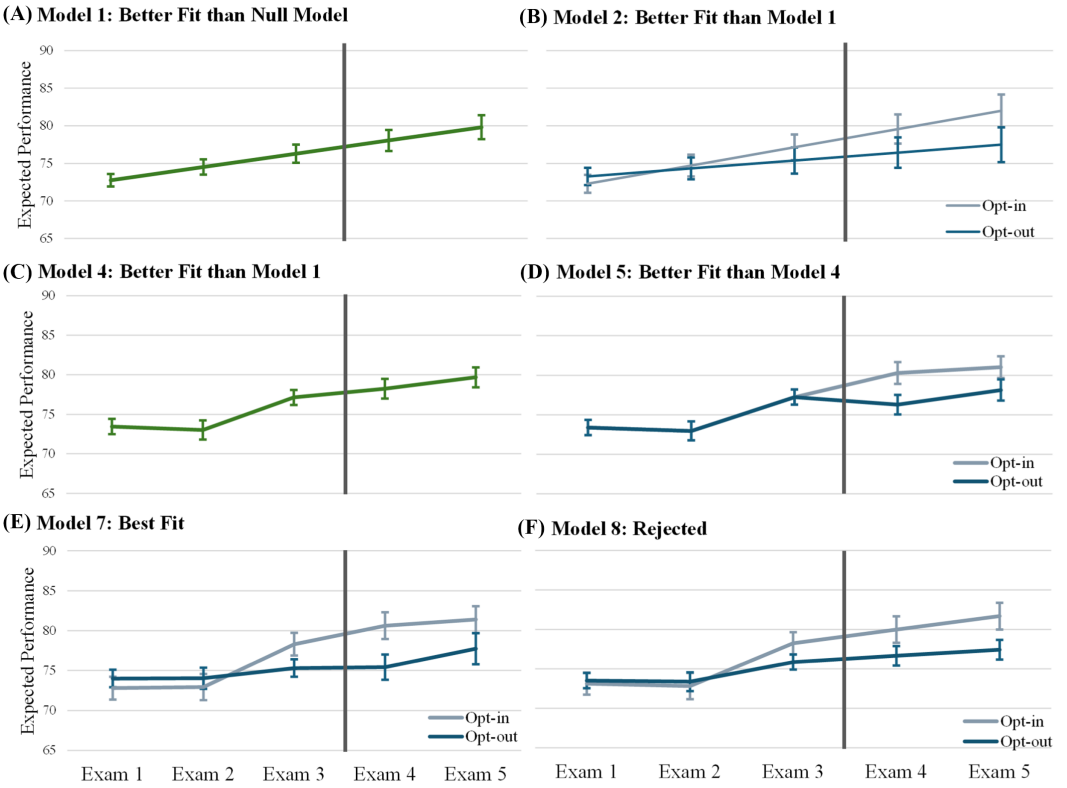
[1] As structural equation models are computationally equivalent to regressions, model estimates are the equivalent of effect size measures. They represent the expected change in scores for different conditions.

Table 3
Non-Linear Models

Model	Growth over time	Group difference	Compared to model	χ^2 test of difference	p	Conclusion
0. Null	None: Flat slope	None				Rejected
4. Latent basis (non-linear) model no group differences	Non-linear (e.g., different exams had different difficulty)	None	1	$\Delta\chi^2(3) = 12.08$	$p < .01$	Better fit than Model 1
5. Latent basis with group differences in post-intervention loadings (hypothesized model)	Non-linear	Differences in the trajectories occurred only *after* the intervention	4	$\Delta\chi^2(2) = 12.16$	$p < .005$	Better fit than Model 4
6. Latent basis with group differences in both pre- and post-intervention loadings (testing parallel trends)	Non-linear	Differences in trajectories could occur *before* and/or *after* the intervention	5	$\Delta\chi^2(1) = 0.1$	$p > .1$	Better fit than Model 5
7. Latent basis with group differences in post-intervention loadings and mean structure	**Non-linear**	**Inherent differences in slope and intercept between groups and an effect of the intervention**	**5** **8**	**$\Delta\chi^2(2) = 16.3$** **$\Delta\chi^2(2) = 10.31$**	**$p < .005$** **$p < .01$**	**Better fit than Models 5 and 8** **Best model**
8. Latent basis with group differences in mean structure only (no effect of the intervention)	Non-linear	Intervention has no effect. Inherent group differences explain the difference in trajectories				Rejected

Note. The italics are merely to emphasize the differences between models and the bold indicates the best fitting of the nested models.

Figure 3
Visualizations of Nested Model Estimates



Note. The error bars are the standard errors of the estimated intercept plus the standard error of the estimated loadings for each exam. The vertical gray lines indicate when the intervention occurred. See the online article for the color version of this figure.

opt-out condition ($E[Y]_{\text{opt-out,exam3}} = 75.31$, $SE = 1.10$). However, after the intervention, the opt-in condition ($E[Y]_{\text{opt-in,exam4}} = 80.62$, $SE = 1.68$) performed over five percentage points better on Exam 4 than the opt-out condition ($E[Y]_{\text{opt-out,exam4}} = 75.43$, $SE = 1.59$). On Exam 5, both groups performed better than on Exam 4, and the opt-in condition still performed marginally better ($E[Y]_{\text{opt-in,exam5}} = 81.4$, $SE = 1.68$) than the opt-out condition ($E[Y]_{\text{opt-out,exam4}} = 77.73$, $SE = 1.96$).

Discussion

Through longitudinal design and causal modeling, we feel justified in reporting that the reflective writing and exam wrapper intervention improved exam performance for the students who opted in. Students who engaged in the metacognitive intervention scored an estimated five percentage points higher on an exam given shortly after the intervention than those who opted out. Despite only a single intervention, those who opted in continued to score marginally higher on a subsequent exam more than 2 weeks later. Our nested models suggest that performance differences cannot solely be explained by underlying differences between groups. Overall, we found statistical support for all four of our hypotheses.

Specifically, comparing Models 0, 1, and 4 showed the nonlinearity of expected performance trajectories (Hypothesis 1), and the comparison of Models 5 and 6 provided evidence that students, regardless of eventual opt-in decisions, had similar growth trajectories before the intervention (Hypothesis 2). Models 5 and 7 also showed significant differences in performance trajectory between the two groups after the intervention (Hypothesis 3). Furthermore, the analysis revealed that the intervention improved performance for

Table 4
Model 7 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	0.02	0.22	†
b_3	1.00	0.00	NA
b_{4_0}	1.09	0.49	*
b_{5_0}	2.85	0.86	**
b_{4_1}	1.42	0.26	***
b_{5_1}	1.56	0.26	***
$i_{\_0}$	74.00	1.10	***
$s_{\_0}$	1.31	0.69	†
$i_{\_1}$	72.80	1.42	***
$s_{\_1}$	5.51	1.23	***
i_{var}	167.41	20.70	***
s_{var}	10.55	6.53	†
$i{\sim\sim}s$	−21.59	10.28	*
e_1	154.92	15.04	***
e_2	106.12	11.69	***
e_3	92.95	10.86	***
e_4	129.44	22.33	***
e_5	54.26	10.50	***

Note. $df = 23$; $\chi^2 = 25.58$; TLI = 0.997; CFI = 0.966; RMSEA = 0.025; AIC = 14485.78; BIC = 14552.31. All estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. ** $p < .01$. *** $p < .001$.

individuals who opted in to the assignment as Model 7 estimated the impact of the intervention (Hypothesis 3) and fit better than Model 8, which assumed no effect of the intervention (Hypothesis 4). Therefore, the analysis of these eight nested models provides evidence that the intervention had a causal impact on exam performance for those who took part in the intervention. However, due to underlying differences between students who opted in to the intervention and those who did not, we cannot assume that the intervention would have the same, or any, impact on students who opted out, should they do the assignment.

We aimed to promote student engagement with feedback at three levels. To ensure that students were looking at their feedback, the assignment required students to submit both their original exam answer and the feedback provided by the grader. To engage adaptation and ensure that students used the feedback to update their responses, the intervention asked students to provide a corrected response to the original prompt(s). Finally, to explicitly connect self-monitoring to self-regulation, the intervention required students to write a cover letter explaining their initial misunderstanding alongside the metacognitive processes employed to reach their new understanding (see Supplemental Materials). Additionally, the assignment explicitly promoted metacognitive awareness by including links to videos describing the process of metacognition and its importance in learning. However, we have no information about the engagement of each student with these videos, as they were hosted through YouTube.

The novelty of our findings stems from the intervention's combination of multiple practices to prompt strategic metacognition, each with inconsistent results in the literature (e.g., exam wrappers, reflective writing, and engagement with corrective feedback). Strategic metacognition requires learners' engagement with multiple pillars of metacognition (Drigas & Mitsea, 2020). Most prior metacognition interventions emphasize metacognitive awareness comprising the first three pillars (i.e., academic knowledge of cognition, operational knowledge of cognition, and self-monitoring), for example, studies that use exam wrappers to prompt evidence-based study strategies for multiple-choice exams or quizzes (Edlund, 2020; LaCaille et al., 2019; Rowell et al., 2023; Soicher & Gurung, 2017). While some of these studies emphasize metacognitive reflection (i.e., metacognitive awareness), none explicitly includes a reflective writing assignment or requires students to incorporate instructor feedback (i.e., self-regulation and adaptation; Edlund, 2020; Owen, 2019; Soicher & Gurung, 2017). Of the studies that do employ reflective writing cover letters to promote incorporating written feedback, many are limited by empirical designs and statistical analyses that cannot establish baseline trajectories before learners engage in reflective writing (Daniel et al., 2015; Duijnhouwer et al., 2012). In addition, many only provide descriptive analyses of small-scale implementations(Granville & Dison, 2009; Z. V. Zhang & Hyland, 2023). Our intervention attempts to engage metacognition at multiple levels by asking students to watch videos about metacognition (i.e., pillars: academic knowledge of cognition and operational knowledge of cognition); reflect on their past performance (i.e., pillar: self-monitoring); identify where their rationale went wrong and where they can improve (i.e., pillar: self-regulation); and correct their past answers based on instructor feedback (i.e., pillar: adaptation). We use reflective

writing to explicitly engage these behaviors in an observable and accountable way. We believe it is the combination of these practices that contributes to the positive impact observed in our study.

Limitations

In this quasi-experimental design, we cannot rule out other possible alternative explanations. However, as we compared students enrolled in the same course during the same academic term rather than attempting to compare between terms or between courses, systematic differences in experiences were minimized. Additionally, the data were gathered during a typical term with students enrolled in an in-person course at a primarily residential campus. There were no major disruptions to the academic calendar or delivery of the course. All these design choices help limit the possible alternative explanations to individual differences. We documented inherent differences between the two groups beyond the impact of the intervention (e.g., marginal difference in willingness to miss an exam). However, our model showed that these differences were insufficient to fully explain the increase in performance after the intervention.

While we found evidence that the intervention impacted performance trajectories for those who participated, we cannot claim that the intervention would similarly impact the students who did not do the assignment, or those who did not consent to have their data included. It remains a possibility that the students who opted out of the intervention or the study would not have benefited in the same way as those who opted in. However, without causal modeling, we could have only reported whether students who opted in were likely to do better on later exams compared to students who opted out, as we would have no evidence of whether performance differences could be tied to the intervention.

Though our quantitative analysis presents evidence that the assignment benefited the students who opted in, it does not explain the mechanism that led to that benefit. This assignment was designed to leverage students' grade-based motivation to not only attend to exam feedback, but also to engage actively with their past mistakes through metacognition. However, it is possible that the assignment itself did not improve students' ability to engage in, and learn from, feedback. Perhaps the assignment simply served to remind students to review their past performance and feedback, and it was merely the act of looking at feedback that led to higher exam performance. Therefore, our future research pipeline includes various evaluations to clarify the mechanism of improvement.

Future Directions

In outlining future directions for understanding the impact of feedback and metacognition on learning, several avenues emerge for exploration. Primarily, future studies could investigate the likelihood of metacognition being the mechanism that benefited learning by measuring and assessing the relationship between demonstrated metacognition and changes in performance. Furthermore, if metacognition continues to be identified as a causal mechanism, research could investigate the effects of each pillar of metacognition. Additionally, future research might attempt to disentangle the potential impact of metacognitive engagement with feedback from feedback-viewing behavior or placebo effects. Future investigations could systematically isolate the impact of these mechanisms by comparing the procedure used in this study to interventions that target either metacognition or feedback-viewing and to a control assignment. To ensure that the impacts of the intervention are not dependent on a specific course or instructor, future research could replicate this design in other universities and other course contexts.

Applicability to the Teaching of Psychology

SoTL research aims to advance the quality and efficacy of teaching. Testing teaching innovations in the classroom is inherently noisy and inconsistently allows for true experiments. Applying structural equation modeling to SoTL research allows for relatively strong causal conclusions to be drawn, even in the absence of random assignment to the levels of the treatment. We have made all methods, materials, and code used in this study available and easily adaptable in the hopes that they can provide infrastructure for other psychology educators and researchers to implement and evaluate this or other repeated measure interventions in their classrooms.

The Metacognitive Intervention

This intervention can easily be adapted for use in many course contexts. As the intervention

is designed to promote student engagement with detailed feedback, it requires an assessment where students demonstrate higher-order thinking, such as creative or evaluative writing. However, nonwritten creative assignments, such as multimedia demonstrations or the development of a product, would also be suitable. Another requirement for successful implementation is that students must have the opportunity to incorporate the feedback in a new submission, for example, through cumulative written exams or a rough draft and final draft submission model (Daniel et al., 2015).

Research Design

Classrooms are ideal for gathering longitudinal data such as those used in this quasi-experimental interrupted time series with a comparison group design. We have access to our research participants (students) for weeks or months and often already include multiple observations (summative and formative assessments) across that time. Including multiple observations allows us to not only practice research-informed course design, but also to gather multiple observations before and after the implementation of any intervention. Using an interrupted time series design, we can establish a baseline trajectory for each student and observe changes in continued performance. While including a nonrandomized comparison group cannot ensure equivalent groups, as discussed previously, true experiments in a classroom setting (and the causal evidence that they would provide) are rare due to various practical and ethical considerations (Bunnell et al., 2022; Regan et al., 2012; Stodder, 1998). In addition, due to the unpredictability of classroom settings and campus interruptions, confounds may occur even with random assignment to conditions. However, by combining the longitudinal design with the nonrandom comparison group, causal models can be used to estimate potential confounds that can explain changes in the trajectory and provide evidence for a potential causal link, as demonstrated in this study.

Causal Modeling

As a powerful tool, longitudinal structural equation modeling allows for more nuanced interpretations of measured variables and can provide causal evidence. This analysis method is not exclusive to interrupted time series data; it can be adapted and applied to other experimental and quasi-experimental designs, such as random assignment of different courses or sections. Due to a lack of external control, classroom data often suffer from unexpected or unidentifiable deviations from empirical design. Using causal modeling techniques can increase confidence in causal claims and provide clear theory-driven tests of specific hypotheses under such uncertainties. However, due to limits of degrees of freedom, the number of observations collected per individual limits the number of parameters that can be estimated. This computational complexity also requires higher statistical power to ensure model convergence, meaning relatively large sample sizes are required to estimate these models (Wolf et al., 2013). Though classroom research provides the benefit of a longitudinal relationship between the researcher and participants that affords multiple observations per individual, researchers with small classes or classes that do not have explicit scaffolding may encounter difficulty in applying longitudinal models. Pooling data from multiple sections, terms, and institutions may help with small sample sizes; such practices require models to specify additional predictors to estimate group-level invariance (see McArdle et al., 2009; McCormick et al., 2023; L. Zhang et al., 2014, for more information on applying structural equation models). In addition, theorized models should be preregistered and theoretically justified to allow for valid causal interpretations. To limit misspecification and misleading interpretations, researchers must also systematically test model assumptions by fitting nested models and comparing fit. Despite these barriers or potential difficulties, the increased insight and confidence provided through structural equation modeling make them worth the investment. Additionally, the investment is, to some extent, a one-time cost, as future research can leverage the analysis code and design principles. Once the design is made and the code is debugged, these models can be run to evaluate any future intervention with a comparable research design and data structure.

Conclusions

In conclusion, our comprehensive analysis employing latent growth curve modeling yielded support for the effectiveness of the metacognition

intervention. The systematic testing of nested structural equation models provided a detailed understanding of the intervention's impact on student exam performance, both shortly after the intervention and, to a lesser extent, after a 2-week delay. By identifying the best-fitting model, we confirmed a nonlinear growth trajectory and demonstrated that inherent differences between groups were insufficient to explain the improvement in the ability to identify and critique research methods for those students who completed the intervention. This study demonstrates that quasi-experimental design paired with longitudinal causal modeling allows for nuanced evaluations of interventions, even in classroom conditions where researchers are unable to exert control.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., Norman, M. K., & Mayer, R. E. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.

Bailey, E. G., Jensen, J., Nelson, J., Wiberg, H. K., & Bell, J. D. (2017). Weekly formative exams and creative grading enhance student learning in an introductory biology course. *CBE—Life Sciences Education*, *16*(1), Article ar2. https://doi.org/10.1187/cbe.16-02-0104

Basey, J. M., Maines, A., & Francis, C. (2014). Time efficiency, written feedback, and student achievement in inquiry-oriented biology labs. *International Journal for the Scholarship of Teaching and Learning*, *8*(2), Article 15. https://doi.org/10.20429/ijsotl.2014.080215

Bernstein, J. L. (2018). Unifying SoTL methodology: Internal and external validity. *Teaching and Learning Inquiry*, *6*(2), 115–126. https://doi.org/10.20343/teachlearninqu.6.2.9

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Wiley-Interscience.

Bunnell, S. L., Felten, P., & Matthews, K. E. (2022). Toward trust in SoTL: The role of relational ethics. In L. M. Fedoruk (Ed.), *Ethics and the scholarship of teaching and learning* (pp. 129–146). Springer. https://doi.org/10.1007/978-3-031-11810-4_9

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning*, *11*(2), 215–235. https://doi.org/10.1007/s11409-015-9142-6

Cao, L., & Nietfeld, J. L. (2007). College students' metacognitive awareness of difficulties in learning the class content does not automatically lead to adjustment of study strategies. *Australian Journal of Educational & Developmental Psychology*, *7*, 31–46.

Carless, D. (2022). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, *23*(2), 143–153. https://doi.org/10.1177/1469787420945845

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Çini, A., Järvelä, S., Dindar, M., & Malmberg, J. (2023). How multiple levels of metacognitive awareness operate in collaborative problem solving. *Metacognition and Learning*, *18*(3), 891–922. https://doi.org/10.1007/s11409-023-09358-7

Cohn, J., & Stewart, M. (2016). *Promoting metacognitive thought through response to low-stakes reflective writing*. *Journal of Response to Writing*, *2*(1), 58–74.

Colthorpe, K., Sharifirad, T., Ainscough, L., Anderson, S., & Zimbardi, K. (2018). Prompting undergraduate students' metacognition of learning: Implementing 'meta-learning' assessment tasks in the biomedical sciences. *Assessment & Evaluation in Higher Education*, *43*(2), 272–285. https://doi.org/10.1080/02602938.2017.1334872

Corral, D., & Carpenter, S. K. (2023). Long-term hypercorrection, return errors, and the transfer of learning in the classroom. *Journal of Applied Research in Memory and Cognition*, *12*(2), 208–229. https://doi.org/10.1037/mac0000048

Daniel, F., Gaze, C. M., & Braasch, J. L. G. (2015). Writing cover letters that address instructor feedback improves final papers in a research methods course. *Teaching of Psychology*, *42*(1), 64–68. https://doi.org/10.1177/0098628314562680

Dawson, P., Carless, D., & Lee, P. P. W. (2021). Authentic feedback: Supporting learners to engage in disciplinary feedback practices. *Assessment & Evaluation in Higher Education*, *46*(2), 286–296. https://doi.org/10.1080/02602938.2020.1769022

Divan, A., Ludwig, L., Matthews, K., Motley, P., & Tomljenovic-Berube, A. (2017). Research approaches in scholarship of teaching and learning publications: A systematic literature review. *Teaching and Learning Inquiry*, *5*(2), 16–29. https://doi.org/10.20343/teachlearninqu.5.2.3

Drigas, A., & Mitsea, E. (2020). The 8 pillars of metacognition. *International Journal of Emerging Technologies in Learning*, *15*(21), 162–178. https://doi.org/10.3991/ijet.v15i21.14907

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance.

Learning and Instruction, *22*(3), 171–184. https://doi.org/10.1016/j.learninstruc.2011.10.003

Edlund, J. E. (2020). Exam wrappers in psychology. *Teaching of Psychology*, *47*(2), 156–161. https://doi.org/10.1177/0098628320901385

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Eskreis-Winkler, L., & Fishbach, A. (2022). You think failure is hard? So is learning from it. *Perspectives on Psychological Science*, *17*(6), 1511–1524. https://doi.org/10.1177/17456916211059817

Felten, P. (2013). Principles of good practice in SoTL. *Learning Inquiry*, *1*(1), 121–125. https://doi.org/10.2979/teachlearninqu.1.1.121

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition*, *38*(7), 951–961. https://doi.org/10.3758/MC.38.7.951

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Granville, S., & Dison, L. (2009). Making connections through reflection: Writing and feedback in an academic literacy programme. *Southern African Linguistics and Applied Language Studies*, *27*(1), 53–63. https://doi.org/10.2989/SALALS.2009.27.1.5.753

Haddara, N., & Rahnev, D. (2022). The impact of feedback on perceptual decision-making and metacognition: Reduction in bias but no change in sensitivity. *Psychological Science*, *33*(2), 259–275. https://doi.org/10.1177/09567976211032887

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, *44*(1), 124–138. https://doi.org/10.1080/02602938.2018.1484881

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

LaCaille, R. A., LaCaille, L. J., & Maslowski, A. K. (2019). Metacognition, course performance, and perceived competence for learning: An examination of quiz and exam wrappers. *Scholarship of Teaching and Learning in Psychology*, *5*(3), 209–222. https://doi.org/10.1037/stl0000114

Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2010). Improving self-regulation, learning strategy use, and achievement with metacognitive feedback. *Educational Technology Research and Development*, *58*(6), 629–648. https://doi.org/10.1007/s11423-010-9153-6

Lovett, M. C. (2013). Make exams worth more than the grade: Using exam wrappers to promote metacognition. In N. Silver, M. Kaplan, D. LaVaque-Manty, & D. Meizlish (Eds.), *Using reflection and metacognition to improve student learning* (pp. 18–52). Routledge.

Lovett, M. C., Bridges, M. W., DiPietro, M., Ambrose, S. A., & Norman, M. K. (2023). *How learning works: Eight research-based principles for smart teaching* (2nd ed.). Jossey-Bass.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*(2), 126–149. https://doi.org/10.1037/a0015857

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023). The Hitchhiker's guide to longitudinal models: A primer on model selection for repeated-measures methods. *Developmental Cognitive Neuroscience*, *63*, Article 101281. https://doi.org/10.1016/j.dcn.2023.101281

Mill, J. S. (1882). *A system of logic, ratiocinative and inductive* (8th ed.). Harper & Brothers Publisher, Franklin Square.

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, *6*(3), 303–314. https://doi.org/10.1007/s11409-011-9083-7

Molin, F., Haelermans, C., Cabus, S., & Groot, W. (2020). The effect of feedback on metacognition—A randomized experiment using polling technology. *Computers & Education*, *152*, Article 103885. https://doi.org/10.1016/j.compedu.2020.103885

Naujoks, N., Harder, B., & Händel, M. (2022). Testing pays off twice: Potentials of practice tests and feedback regarding exam performance and judgment accuracy. *Metacognition and Learning*, *17*(2), 479–498. https://doi.org/10.1007/s11409-022-09295-x

Nicol, D., & McCallum, S. (2022). Making internal feedback explicit: Exploiting the multiple comparisons

that occur during peer review. *Assessment & Evaluation in Higher Education*, *47*(3), 424–443. https://doi.org/10.1080/02602938.2021.1924620

Orpurt, S. (2018, August 1). *Metacognition skills learning to learn* [Video]. YouTube. https://www.youtube.com/watch?v=CGouE8EQbO0

Owen, L. R. (2019). The exam autopsy: An integrated post-exam assessment model. *International Journal for the Scholarship of Teaching and Learning*, *13*(1), Article 4. https://doi.org/10.20429/ijsotl.2019.130104

Pearl, J. (2009). *Causality*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Peterson's Test Prep (Director). (2020, April 28). *Metacognition: The skill that promotes advanced learning* [Video]. https://www.youtube.com/watch?v=elZFL4FLVLE

Pogrow, S. (2017). The failure of the U.S. education research establishment to identify effective practices: Beware effective practices policies. *Education Policy Analysis Archives*, *25*, Article 5. https://doi.org/10.14507/epaa.25.2517

Regan, J.-A., Baldwin, M. A., & Peters, L. (2012). Ethical issues in pedagogic research. *Journal of Pedagogic Development*, *2*(3), 44–54. https://hdl.handle.net/10547/336064

Rowell, S. F., Cohen-Shikora, E. R., Walck-Shannon, E. M., Mazur, J., & Frey, R. F. (2023). Randomized study strategy intervention in a large introductory psychology course. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. https://doi.org/10.1037/stl0000360

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021). Designing learner-centred text-based feedback: A rapid review and qualitative synthesis. *Assessment & Evaluation in Higher Education*, *46*(6), 894–912. https://doi.org/10.1080/02602938.2020.1828819

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2024). Feedback in higher education: Aligning academic intent and student sensemaking. *Teaching in Higher Education*, *29*(4), 860–875. https://doi.org/10.1080/13562517.2022.2029394

Sachar, C. O. (2020). Revising with metacognition to promote writing achievement: A case study. *Journal of Scholarship of Teaching and Learning*, *20*(3), 49–63. https://doi.org/10.14434/josotl.v20i3.28675

Saenz, G. D., Geraci, L., & Tirso, R. (2019). Improving metacognition: A comparison of interventions. *Applied Cognitive Psychology*, *33*(5), 918–929. https://doi.org/10.1002/acp.3556

Sato, M., & Loewen, S. (2018). Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets: Metacognitive instruction and corrective feedback. *Language Learning*, *68*(2), 507–545. https://doi.org/10.1111/lang.12283

Shi, Y., Yang, H., MacLeod, J., Zhang, J., & Yang, H. H. (2020). College students' cognitive learning outcomes in technology-enabled active learning environments: A meta-analysis of the empirical literature. *Journal of Educational Computing Research*, *58*(4), 791–817. https://doi.org/10.1177/0735633119881477

Soicher, R. N., & Gurung, R. A. R. (2017). Do exam wrappers increase metacognition and performance? A single course intervention. *Psychology Learning & Teaching*, *16*(1), 64–73. https://doi.org/10.1177/1475725716661872

St.Clair, T., Hallberg, K., & Cook, T. D. (2016). The validity and precision of the comparative interrupted time-series design: Three within-study comparisons. *Journal of Educational and Behavioral Statistics*, *41*(3), 269–299. https://doi.org/10.3102/1076998616636854

Stodder, J. (1998). Experimental moralities: Ethics in classroom experiments. *The Journal of Economic Education*, *29*(2), 127–138. https://doi.org/10.1080/00220489809597946

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113–120. https://doi.org/10.1187/cbe.12-03-0033

Tempelaar, D. (2020). Supporting the less-adaptive student: The role of learning analytics, formative assessment and blended learning. *Assessment & Evaluation in Higher Education*, *45*(4), 579–593. https://doi.org/10.1080/02602938.2019.1677855

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Wiggins, G. (2012). Seven keys to effective feedback—Educational leadership. *Educational Leadership*, *70*(1), 10–16.

Winstone, N. E., Mathlin, G., & Nash, R. A. (2019). Building feedback literacy: Students' perceptions of the Developing Engagement With Feedback Toolkit. *Frontiers in Education*, *4*, Article 39. https://doi.org/10.3389/feduc.2019.00039

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Woods, V., & Cross, V. (2020, October). *Critical thinking, reflective practice, and metacognition: A SOTL approach* [Video]. Society for Teaching

of Psychology. https://www.youtube.com/watch?v=gMBG2cKocf8

Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, *47*(7), 1116–1128. https://doi.org/10.1080/02602938.2021.2001431

Zhang, L., Goh, C. C. M., & Kunnan, A. J. (2014). Analysis of test takers' metacognitive and cognitive strategy use and EFL Reading Test performance: A multi-sample SEM approach. *Language Assessment Quarterly*, *11*(1), 76–102. https://doi.org/10.1080/15434303.2013.853770

Zhang, Z. V., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, *58*, Article 100784. https://doi.org/10.1016/j.asw.2023.100784

(*Appendix follows*)

Appendix

Table A1
Null Model Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0	0	NA
b_2	0	0	NA
b_3	0	0	NA
b_4	0	0	NA
b_5	0	0	NA
i	76.88	0.67	***
s	0	0	NA
i_{var}	129.59	12.15	***
s_{var}	0	0	NA
$i{\sim\sim}s$	0	0	NA
e_1	188.47	18.87	***
e_2	144.04	15.09	***
e_3	88.79	11.26	***
e_4	134.15	21.97	***
e_5	87	9.41	***

Note. $df = 33.00$; $\chi^2 = 168.53$; TLI = 0.89; CFI = 0.82; RMSEA = 0.15; AIC = 14608.74; BIC = 14636.14. All estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
*** $p < .001$.

Table A2
Model 1 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	1.00	0.00	NA
b_3	2.00	0.00	NA
b_4	3.00	0.00	NA
b_5	4.00	0.00	NA
i	72.76	0.83	***
s	1.76	0.19	***
i_{var}	174.93	22.21	***
s_{var}	3.19	1.48	*
$i{\sim\sim}s$	−12.74	4.64	**
e_1	151.30	17.09	***
e_2	117.71	12.35	***
e_3	91.85	10.97	***
e_4	132.04	22.28	***
e_5	69.96	12.21	***

Note. $df = 30.00$; $\chi^2 = 65.95$; TLI = 0.967; CFI = 0.951; RMSEA = 0.08; AIC = 14512.15; BIC = 14551.28. All estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
* $p < .05$. ** $p < .01$. *** $p < .001$.

(*Appendix continues*)

Table A3
Model 2 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	1.00	0.00	NA
b_3	2.00	0.00	NA
b_4	3.00	0.00	NA
b_5	4.00	0.00	NA
$i_{\_0}$	73.28	1.15	***
$s_{\_0}$	1.05	0.29	***
$i_{\_1}$	72.29	1.19	***
$s_{\_1}$	2.42	0.25	***
i_{var}	173.87	22.04	***
s_{var}	2.50	1.52	†
$i{\sim\sim}s$	−12.01	4.64	*
e_1	152.02	16.88	***
e_2	117.59	12.31	***
e_3	91.58	10.95	***
e_4	131.08	22.55	***
e_5	68.45	12.47	***

Note. $df = 28$; $\chi^2 = 48.351$; TLI = 0.98; CFI = 0.972; RMSEA = 0.063; AIC = 14498.55; BIC = 14545.51. Subscript 0 is opt-out of the intervention, subscript 1 is opt-in to the intervention. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. *** $p < .001$.

Table A4
Model 3 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	1.00	0.00	NA
b_3	2.00	0.00	NA
b_4	3.00	0.00	NA
b_5	4.00	0.00	NA
$i_{\_0}$	73.28	1.15	***
$s_{\_0}$	1.05	0.29	***
$i_{\_1}$	72.29	1.19	***
$s_{\_1}$	2.42	0.25	***
i_{var}	175.06	22.04	***
s_{var}	2.45	1.53	†
$i{\sim\sim}s_{\_0}$	−10.89	4.83	*
$i{\sim\sim}s_{\_1}$	−13.21	4.98	**
e_1	151.37	16.87	***
e_2	117.74	12.31	***
e_3	91.49	10.96	***
e_4	131.22	22.62	***
e_5	68.80	12.51	***

Note. $df = 27$; $\chi^2 = 47.76$; TLI = 0.979; CFI = 0.972; RMSEA = 0.064; AIC = 14499.96; BIC = 14550.84. Subscript 0 is opt-out of the intervention, subscript 1 is opt-in to the intervention. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. ** $p < .01$. *** $p < .001$.

(*Appendix continues*)

Table A5
Model 4 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	−0.12	0.27	†
b_3	1.00	0.00	NA
b_4	1.30	0.29	†
b_5	1.69	0.31	*
i	73.46	0.97	***
s	3.68	0.92	***
i_{var}	163.98	20.36	***
s_{var}	10.89	6.52	†
$i{\sim\sim}s$	−21.98	10.39	*
e_1	161.43	15.59	***
e_2	103.41	12.70	***
e_3	91.42	11.09	***
e_4	132.40	22.78	***
e_5	70.72	12.49	***

Note. $df = 27$; $\chi^2 = 53.865$; TLI = 0.973; CFI = 0.963; RMSEA = 0.073; AIC = 14506.07; BIC = 14556.94. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. *** $p < .001$.

Table A6
Model 5 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	−0.11	0.25	†
b_3	1.00	0.00	NA
b_{4_0}	0.75	0.28	**
b_{5_0}	1.23	0.38	**
b_{4_1}	1.79	0.41	***
b_{5_1}	1.98	0.41	***
i	73.38	0.96	***
s	3.86	0.90	***
i_{var}	156.32	21.62	***
s_{var}	4.00	8.52	†
$i{\sim\sim}s$	−15.80	11.93	†
e_1	164.04	16.16	***
e_2	111.87	14.78	***
e_3	91.72	11.23	***
e_4	130.20	21.66	***
e_5	76.66	11.94	***

Note. $df = 25$; $\chi^2 = 41.71$; TLI = 0.982; CFI = 0.977; RMSEA = 0.06; AIC = 14497.91; BIC = 14556.61. Subscript 0 is opt-out of the intervention, subscript 1 is opt-in to the intervention. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. ** $p < .01$. *** $p < .001$.

(*Appendix continues*)

Table A7
Model 6 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_{2_0}	−0.06	0.32	†
b_{2_1}	−0.16	0.32	†
b_3	1.00	1.00	NA
b_{4_0}	0.76	0.29	*
b_{5_0}	1.24	0.36	**
b_{4_1}	1.78	0.41	***
b_{5_1}	1.98	0.41	***
i	73.38	0.97	***
s	3.85	0.90	***
i_{var}	155.97	21.56	***
s_{var}	3.83	8.31	†
$i{\sim\sim}s$	−15.51	11.87	†
e_1	164.39	16.22	***
e_2	112.00	14.70	***
e_3	91.59	11.22	***
e_4	130.24	21.69	***
e_5	76.73	11.79	***

Note. $df = 24$; $\chi^2 = 41.61$; TLI = 0.98; CFI = 0.976; RMSEA = 0.063; AIC = 14499.81; BIC = 14562.43. Subscript 0 is opt-out of the intervention, subscript 1 is opt-in to the intervention. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table A8
Model 8 Parameter Estimates

Parameter	Estimate	*SE*	*p*
b_1	0.00	0.00	NA
b_2	−0.06	0.24	†
b_3	1.00	0.00	NA
b_4	1.34	0.27	***
b_5	1.66	0.28	***
$i_{\_0}$	73.52	1.11	***
$s_{\_0}$	2.35	0.73	**
$i_{\_1}$	73.14	1.42	***
$s_{\_1}$	5.13	1.22	***
i_{var}	164.55	20.57	***
s_{var}	8.93	6.43	†
$i{\sim\sim}s$	−21.93	10.27	*
e_1	159.60	15.27	***
e_2	105.977	12.47	***
e_3	91.33	10.95	***
e_4	131.21	23.07	***
e_5	72.07	11.85	***

Note. $df = 25$; $\chi^2 = 35.89$; TLI = 0.988; CFI = 0.985; RMSEA = 0.049; AIC = 14492.09; BIC = 14550.79. Subscript 0 is opt-out of the intervention, subscript 1 is opt-in to the intervention. Estimates with an *SE* of 0 are fixed (not estimated). *SE* = standard error; *df* = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; NA = not applicable.
† $p > .05$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Received December 12, 2023
Revision received April 17, 2024
Accepted April 18, 2024 ■