Supplemental Material for

The Paradox of Explaining: When Feeling Unknowledgeable Prevents Learners from Engaging in Effective Learning Strategies

Exclusions and Attrition

Exclusions were largely in accordance with our pre-registration. All deviations from pre-registration are noted below. Pilot 1a was not pre-registered.

Pilot 1a

Session 1: Four-hundred participants successfully completed Session 1 and were eligible to participate in Session 2. Additional participants were excluded: 115 did not spend sufficient time reading the study text (<15 seconds for the first five sections and <10 seconds for the final, shorter section), 36 failed one or more attention check, and one was not interested in being contacted for Session 2. Attrition within Session 1 (i.e., participants starting but not finishing the study session) differed between conditions, $b = 1.85$, $z = 6.71$, $p < .001$, being higher in the WR condition ($N = 94$) than the control condition ($N = 18$), which led us to implement a short writing task as a screening task in future studies. An additional 13 participants quit the study before being assigned to condition.

Session 2: Three-hundred and sixty-seven participants ($N_{Control} = 204$; $N_{Written\ Reflection} = 163$) successfully completed Session 2 and were included in the analyses. Four additional participants were excluded because they reported looking up at least one of the answers to the quiz questions. Attrition among participants who started Session 2 was very low and similar between conditions ($N_{Written\ Reflection} = 2$, $N_{\text{Control}} = 1$, $N_{Unknown} = 4$ participants opened the survey but did not enter their information and therefore cannot be linked to their original condition). Attrition from Session 1 to 2 (i.e., participants completing Session 1 but not participating in Session 2) did not significantly differ between conditions ($N_{Written\ Reflection} = 17$, $N_{\text{Control}} = 16$), $b = .28$, $z = .78$, $p = .43$.

Pilot 1b

Session 1: Five-hundred and seventeen participants completed Session 1 and were eligible to participate in Session 2. Additional participants were excluded based on pre-registered criteria: 135 did not spend sufficient time reading the screening text (<25 seconds, pre-registered) or watching the videos (video duration minus 10 seconds, pre-registered), and 38 failed one or more attention checks (failed a video check, did not report subjective knowledge, failed to correctly answer a question about the study topic, failed to follow instructions to attention check). The following participants were excluded based on non-pre-registered criteria: 19 reported not being native English speakers, 2 participated from outside the U.S. or used a VPN despite the study specifying participants must be U.S.-based, and 3 were not interested in being contacted for Session 2. Attrition after assignment to condition was low overall but higher in the WR condition ($N = 10$) than the control condition ($N = 2$), $b = 1.60$, $z = 2.06$, $p = .04$. See additional information below on how we dealt with this differential attrition.

Session 2: Four-hundred and fifty ($N_{Control} = 220$; $N_{Written\ Reflection} = 230$) successfully completed Session 2 and were included in the analyses. Additional participants were excluded based on pre-registered criteria: 20 participants reported looking up at least one of the answers to the quiz questions and two failed an attention check. The following participants were excluded based on non-pre-registered criteria: five participants participated despite reporting not wishing to be recontacted after session 1 (this number is higher than the number cited for Session 1 because some of these participants were excluded for multiple reasons, e.g., did not as to be recontacted and were also not U.S.-based).

Note that the pre-registration for Pilot 1b is an updated version that we submitted after conducting Session 1 but before conducting Session 2. Our original plan was to not run Session 2 if there was differential attrition across conditions in completing Session 1. Specifically, we were

concerned that participants who understood the material less well would drop out of the study in the WR condition when asked to explain, but not the control condition. Our worry was that this would interfere with random assignment, selecting those with superior knowledge into the WR condition. However, even with differential attrition, we decided to run Session 2 because the actual number of participants who dropped out after assignment to condition was low (N = 12 out of 529 eligible; 10 in the WR condition and 2 in the control condition). We found no significant differences in reported knowledge for those who dropped out compared to those who did not, $b = .14$, $t(527) = .40$, $p = .69$. Given this and given the small number of participants who dropped out, we do not believe that losing eight more people in the WR condition compromises our conclusions.

As an additional check, we pre-registered re-conducting our analyses for quiz scores after dropping eight people from the control condition who reported the lowest subjective knowledge. Even after dropping these eight, people in the WR condition scored marginally significantly higher than those in the control condition, $b = .55$, $t(440) = 1.83$, $p = .067$. We also pre-registered checking whether the number of participants who completed Session 1 but not Session 2 was different between conditions. We found it was, ($N_{Written\ Reflection} = 13$, $N_{Control} = 28$), $b = .83$, $z = 2.39$, $p = .02$. However, these numbers were small relative to the number of participants who did complete both sessions. Finally, we also pre-registered checking whether participants who did versus did not complete Session 2 differed significantly in their subjective knowledge from Session 1, and found that they did not (participants who did Session 1 only: $M = 4.93$, $SD = 1.36$; participants who did both sessions: $M = 4.66$, $SD = 1.18$), $b = .30$, $t(518) = 1.57$, $p = .12$.

Study 1

The following participants were excluded based on pre-registered criteria: 7 participants failed an attention check and/or English proficiency check. The following participants were excluded based on non-pre-registered criteria: 15 participants did not complete one of the key measures (did not provide a ranking for the three activities). In addition, we mistakenly pre-registered excluding participants whose IP check placed them outside of the United States, when we intended to include participants from both the United States and the United Kingdom. Indeed, our screening criteria on Prolific allowed participants from both the US and the UK. Therefore, participants from the UK were not excluded.

Study 2

Session 1: Eight hundred and thirty-four participants successfully completed Session 1. Additional participants were excluded based on pre-registered criteria: 115 participants failed an attention check and/or English proficiency check. The following participants were excluded based on non-pre-registered criteria: 86 failed one or more checks for attention, bots, or English proficiency, 28 failed the video check, 16 reported not being native English speakers, 10 indicated they did not wish to complete Session 2, 9 participated from outside the U.S. despite the study specifying U.S.-based participants, and 3 were duplicates. Additional participants were blocked from participating because we could not verify their location ($N = 34$) or use of VPN ($N = 31$) or because the study quota had been met ($N = 52$). Finally, we pre-registered excluding participants who used a mobile device, but this information was not successfully recorded, and we pre-registered using an attention check in which participants are asked to ignore a scale and enter text into a textbook, but we did not end up including this check in the survey. Attrition within Session 1 did not differ between conditions ($N_{Written\ Reflection} = 8$, $N_{Control} = 7$, $N_{\text{Pre-assignment}} = 171$), $b = .09$, $z = .18$, $p = .85$.

Session 2: Seven-hundred and nineteen participants completed Session 2 and were included in the analyses. Additional participants were excluded based on pre-registered criteria: 25 people reported looking up at least one of the answers to the quiz questions, and one person failed an attention check. Attrition within Session 2 did not differ between conditions ($N_{Written\ Reflection}$ = 15, $N_{Control}$ = 18), b = .22, z = .64, p = .52. Attrition from Session 1 to 2 did not significantly differ between conditions ($N_{Written\ Reflection}$ = 54, $N_{Control}$ = 49), b = .07, z = .32, p = .75.

Study 3

The following participants were excluded based on pre-registered criteria: 84 participants failed an attention check and/or English proficiency check, and 5 participated from outside the U.S. or used a VPN. We pre-registered excluding participants who used a mobile device, but this information was not successfully recorded. The following participants were excluded based on non-pre-registered criteria: 4 reported not being native English speakers, 4 indicated they did not wish to complete a second session (this session did not actually take place). Additional participants were blocked from participating because their location could not be verified or because they failed the video check. Attrition after condition assignment was low and did not significantly differ between the low subjective knowledge condition (N = 8) and the high subjective knowledge condition (N = 2), b = 1.33, z = 1.67, p = .10.

Study 4

The following participants were excluded based on pre-registered criteria: 17 participants failed the video and audio check, 11 self-reported an insufficient level of English proficiency, 6 failed an attention check and/or English proficiency check, and 3 participated from outside the U.S. or U.K. We pre-registered excluding participants who used a mobile device, but this information was not reliably recorded. Attrition after condition assignment was low and did not significantly differ based on condition between the pleasant (N = 5) and the effective (N = 5) condition, b = .04, z = .06, p = .94., nor between the control condition (N = 2) and the pleasant, b = .91, z = 1.08, p = .28, or the effective, b = .95, z = 1.13, p = .26, conditions.

Study 5

The following participants were excluded based on pre-registered criteria: 12 students failed to complete the full study. The following participants were excluded based on non-pre-registered criteria: two did not sign the consent form, and one completed the wrong condition.

Measures and Materials

Explaining Instructions from Pilots a-b

These are the instructions participants saw when being asked to explain what they had learned. The instructions in the following studies were either identical or similar with minor changes in phrasing.

> Your goal now is to explain how GPS works to someone else. Please take a moment to think back to what you've learned in preparation for explaining it to someone else.
> You should try to explain the information assuming that the other person will get tested on what they learn. Therefore, your goal is to explain it so the other person understands it **really well**. Try to explain it the best way you can so the other person understands it as well as possible.
> Make sure that you include as much information as you can. At the same time, try making your explanation clear and understandable.

On the next page, you will be asked to write your explanation to someone else. You will have 5 minutes to do so. Once the time is up, you will automatically advance to the next page.

Subjective Knowledge Across Studies

We used similar but not identical measures of subjective knowledge across studies. In Pilots 1a and 1b, we asked: **"**How knowledgeable do you feel about GPS?" "If you had to rate your feeling of knowledge about GPS, how would you rate it?" "In general, how knowledgeable do you think you are about GPS?" "How comfortable would you feel explaining how GPS works to other people?" "Imagine you had to take an important test on GPS and you wanted to do well. How ready for that test would you feel right now?".

In Studies 1-4, the questions were the same as in the Pilots except with the following additional question: "Imagine you had to take an important test on GPS and you wanted to do well. How much more preparation do you think you would need in order to do well?" (reversed scored).

In Studies 5 and 1S, we asked: "How well do you feel you understand glycolysis?" "How knowledgeable do you feel about glycolysis?" "How comfortable would you feel explaining how glycolysis works to other people?" "Imagine you had to take an important quiz on glycolysis and you wanted to do well. How ready for that test would you feel right now?" "If you had to rate your feeling of knowledge about glycolysis, how would you rate it?"

In Study 2S, the questions were identical to those in Studies 5 and 1S, except they asked about the topic "How Edwin Hubble discovered that our Universe is expanding."

Manipulation Text in Study 4

Text in the Effectiveness Condition.

Explaining helps you understand things better!

Writing an explanation about material you just learned helps you understand the material better. Although people are sometimes worried that explaining what they learned won't improve their understanding of the material, we actually find that writing an explanation increases people's score on an upcoming test.

In several studies, we asked people on Mechanical Turk to listen to information about how GPS works, just like we asked you to do. We then asked half of the people to write an explanation about what they learned. The other half of people wrote about something else. A week later, everyone took a quiz about GPS. We found that people who wrote an explanation scored higher on the quiz – just writing an explanation helped them to answer more questions correctly!

Even people who felt like they didn't understand the material about GPS very well scored better on the test after writing an explanation. The process of writing an explanation helped people understand the material better, regardless of whether they knew a lot or a little.

Thus, even if writing an explanation about what you learned isn't very fun, the process of writing an explanation still turns out to help you understand and remember the material better.

Text in the Pleasantness Condition.

Explaining is fun!

Writing an explanation about material you just learned is an enjoyable experience. Although people are sometimes worried that explaining what they learned will be unpleasant, people actually find the experience quite pleasant. They report that writing an explanation of what they just learned is fun and exciting.

Even people who feel like they don't understand the topic very well find explaining to be fun. It can be fun to discover what you already know, but also exciting to discover what you don't quite understand yet. In other words, people find it fun and exciting to figure out what they already know and what they can still learn more about.

Thus, even if explaining doesn't help you understand the topic better, the process of writing an explanation still turns out to be enjoyable and exciting.

Supplemental Results and Figures For Main Studies

Pilots

Participants' Confidence and Performance Estimate.

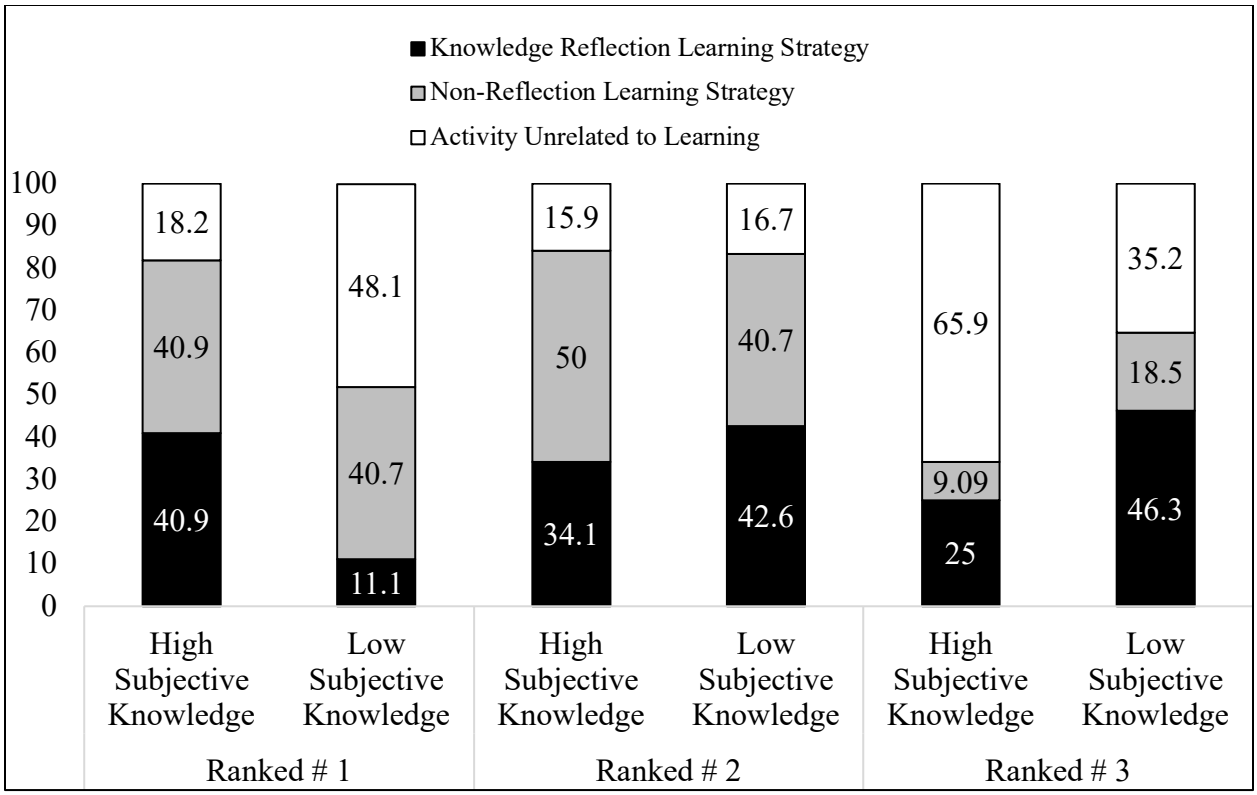
After answering each quiz question, participants reported their confidence in their answer. After finishing the quiz, they estimated the total number of quiz questions they had gotten right Participants who engaged in written reflection were marginally significantly and non-significantly more confident in their answers (M_{Pilot1} = 4.61, SD = 1.01; M_{Pilot2} = 4.90, SD = 1.07) than participants who did not engage in written reflection (M_{Pilot1} = 4.41, SD = 1.00; M_{Pilot2} = 4.79, SD = 1.03) in Pilot 1a, b = .20, t(367) = 1.90, p = .06, and Pilot 1b, b = .11, t(466) = 1.19, p = .23, respectively. A similar pattern emerged for participants' estimates of their total score on the quiz, with participants in the written reflection condition estimating their scores significantly and non-significantly higher (M_{Pilot1} = 9.76, SD = 4.00; M_{Pilot2} = 11.00, SD = 4.35) than participants who did not engage in written reflection (M_{Pilot1} = 8.82, SD = 3.71; M_{Pilot2} = 10.50, SD = 4.42) in Pilot 1a, b = .94, t(367) = 2.33, p = .02, and Pilot 1b, b = .53, t(466) = 1.32, p = .19, respectively.

Participants underestimated their overall quiz performance (M_{Pilot1} = -0. 69 SD = 3.73; M_{Pilot2} = -1.64, SD = 3.95) in Pilot 1a, t(368) = -3.57, p < .001, and Pilot 1b, t(467) = -8.94, p < .001. There was no significant difference between conditions in this underestimation either in Pilot 1a, b = .28, t(367) =.71, p = .47, or Pilot 1b, b = -.10, t(466) = -.28, p = .78.

Study 1

Figure 1S

Percent Participants Ranking of Each Activity Type as a Function of Subjective Knowledge in Study 1



Note. The percentage of low and high subjective knowledge participants (1 SD below and above the mean, respectively) ranking each activity type as # 1 (two leftmost bars), # 2 (two middle bars) and # 3 (two rightmost bars). Knowledge reflection learning strategy was explaining; non-reflection learning strategy was re-watching one of the video sections and typing notes on it; activity unrelated to learning was writing about one's favorite season.

Study 5

In addition to obtaining students' quiz scores, we also obtained their grades on the final exam. There was only one question about glycolysis on the exam, and students' performance on it did not vary significantly by conditions (control vs. written reflection), $b = .11$, $z = .49$, $p = .68$, and neither did students' total score, $b = 1.87$, $t(459) = 1.65$, $p = .1$. However, adding the exam question to the quiz questions on glycolysis (adding one if it was answered correctly and zero if it were not) results in the same pattern that was observed for the quiz questions, i.e., students who were assigned to explain glycolysis scored higher on glycolysis questions than students who were assigned to review, $b = .17$, $t(452) = 2.30$, $p = .022$.

Supplemental Studies

Pilot 1S

In this pilot study, students reported the learning strategies they have used recently. They also rated a list of learning strategies based on how much knowledge reflection each one required.

Method

Participants. Participants were recruited through Prolific and were limited to U.S.-based students. Forty-seven students successfully completed the study and were included in the analyses (21 women, 25 men, 1 non-binary; mean age = 24.81, SD = 8.02; 33 White, 5 Black or

African American, 4 multiracial, 3 Asian, 1 American Indian or Alaska Native, and 1 did not report race). Additional participants were excluded: two reported not having studied in the past month, and one was not U.S.-based or was using a VPN.

Design and Procedure. Participants reported whether they were students and whether they had studied in the past month, for example for an exam or quiz. If they answered Yes to both, they listed all learning strategies they had used in the past month. After submitting their answer, they were shown a list of 15 learning strategies, 5 in each category: requiring no, some, and high levels of knowledge reflection (see list in Table S1). They self-coded their self-generated list, i.e., they selected which of these 15 learning strategies were on the list they had just made. They were then shown the same list again and selected any learning strategies they had used the previous month but had forgotten to include in their initial list. They then rated all 15 learning strategies based on how much each strategy would require them to reflect on their knowledge of the material, from 1 ("Not at all") to 6 ("To a large extent"). Finally, they answered an attention check question and reported their demographic information.

Results

The results are summarized in Table 1S. Participants' reflection ratings largely matched our a-priori assignment of learning strategies to low, some, and high reflection; "high reflection" strategies were seen as requiring more reflection than those in the "some reflection" category, $b = .65$, $t(9.87) = 2.99$, $p = .01$, and "no reflection" categories, $b = 1.28$, $t(9.87) = 8.80$, $p < .001$, and "some reflection" strategies as requiring more reflection than "no reflection" strategies, $b = .63$, $t(9.79) = 3.18$, $p = .01$.

Although participants self-coded which of the 15 strategies were on their list, we also asked two undergraduate research assistants to code the lists. They marked which of the fifteen strategies were included in each participants' list. The final coding was additive: we coded a particular strategy as matching the list if one or both research assistant marked it. These results are reported in the Table 1S.

Next, we examined participants' self-coding of their initial list. They reported using high reflection learning strategies less than strategies requiring some reflection, $b = -.50$, $z = -1.92$, $p = .05$, and significantly less than strategies requiring no reflection, $b = -2.04$, $z = -8.35$, $p < .001$. Participants used strategies requiring some reflection less than strategies requiring no reflection, $b = -1.49$, $z = -7.02$, $p < .001$. Looking at participants' final list, which also includes strategies they had initially forgotten, reveals the same pattern: participants reported using high reflection learning strategies less than strategies requiring some reflection, $b = -.57$, $z = -2.59$, $p = .009$, and less than strategies requiring no reflection, $b = -1.97$, $z = -8.66$, $p < .001$. Participants used strategies requiring some reflection less than strategies requiring no reflection, $b = -1.42$, $z = -6.76$, $p < .001$.

Table 1S

Strategy Use and Average Reflection Ratings of Each Strategy (Pilot 1S)

	Use of Strategy (RA Coding)	Use of Strategy (Self-Coding)	Use of strategy (Self-Coding + Forgotten)	Reflection Rating
	% Participants	% Participants	% Participants	*M* (*SD*)
High Reflection				
1. Explained the material to yourself out loud or in writing from memory	0.02	0.17	0.36	4.00 (1.04)
2. Explained the material to someone else from memory	0.00	0.19	0.34	4.23 (1.11)
3. Answered essay practice questions without consulting notes or other materials	0.00	0.17	0.3	3.96 (1.16)
4. Generated new examples using the material	0.02	0.06	0.15	3.45 (1.12)
5. Came up with questions that may be on the exam and tried to answer them without using notes	0.00	0.13	0.23	4.13 1.01)
Average		0.14	0.28	3.95 (.83)
Some Reflection				
6. Rehearsed the material by speaking out loud or writing it down from memory	0.36	0.21	0.4	3.98 (1.07)
7. Answered practice questions with the help of notes or other materials	0.09	0.43	0.64	3.19 (1.10)
8. Asked someone else (instructor, classmates) about material that you did not understand	0.00	0.15	0.43	2.77 (1.34)
9. Tried to tie the material to your own life	0.04	0.11	0.26	3.40 (1.33)
10. Prepared a visual representation of the material	0.06	0.15	0.17	3.15 (1.16)
Average		0.21	0.38	3.30 (.64)
No Reflection				
11. Re-read your notes	0.72	0.7	0.77	2.77 (1.48)
12. Re-read the textbook or other class materials (e.g., class slides)	0.62	0.66	0.81	2.47 (1.4)
13. Highlighted key information in your notes	0.04	0.28	0.53	2.66 (1.39)
14. Highlighted key information in textbook, class slides, etc.	0.00	0.32	0.51	2.51 (1.46)
15. Took notes on the material (based on textbook, class slides, etc.)	0.72	0.70	0.83	2.94 (1.31)
Average		0.53	0.69	2.67 (1.23)

Pilot 2S

In this pilot study, participants imagined learning about GPS. They saw a list of activities and rated each one based on effort, knowledge reflection, and effectiveness. Our goal was to identify three learning strategies that required little reflection and three activities that were unrelated to studying that matched in effort the three high reflection learning strategies we had already selected.

Method

Participants. U.S.-based participants were recruited through Prolific. Thirty-nine participants successfully completed the study and were included in the analyses (21 women, 17 men, 1 non-binary; mean age = 41.08, SD = 15.49; 11 current students and 28 non-students; 35 White and 3 Asian). An additional six participants were excluded for failing an attention check.

Design and Procedure. Participants imagined learning about how GPS works using video clips in the same way that participants learned in Pilot 1b and Studies 1-4. They were then presented with twenty activities in random order. These were a random subset of forty activities, which comprised three high reflection learning strategies (the ones for which we were searching for a match), thirty-two common learning strategies that required little or no reflection, and five activities unrelated to studying. The full list is in Table 2s. Participants imagined that they would spend ten minutes engaging in an activity on the list, and rated how much effort each activity would require, from 1 ("Very little effort") to 5 ("A lot of effort"). They then saw the same list again and rated how much each activity would require them to reflect on their knowledge of GPS, from 1 ("No reflection on my knowledge") to 5 ("A lot of reflection on my knowledge"). Finally, they saw the list a third time and rated how effective each activity would be for preparing them for a test about GPS, from 1 ("Not at all effective") to 5 ("Very effective"). Finally, they answered an attention check question and reported their demographic information.

Results

The results are summarized in Table 2S. We selected three learning strategies that, based on two-sided paired t-tests, require little reflection but were matched in effort to our three high reflection learning strategies. Specifically, "Re-watch one of the video sections and memorize it" (little reflection) required less reflection than "Write an explanation of how GPS works as if you were explaining it to someone else" (high reflection), $t(8) = -4.53$, $p = .002$, but required similar effort, $t(8) = -.77$, $p = .46$. "Re-watch one of the video sections, taking handwritten notes as you watch" (little reflection) required less reflection than "Write what you understood from the videos about how GPS works." (high reflection), $t(7) = -6.15$, $p < .001$, but required similar effort, $t(7) = -1.53$, $p = .17$. "Re-watch a video section and create a visual representation of the information on a blank piece of paper" (little reflection) required less reflection than "Answer an essay question about how GPS works" (high reflection), $t(9) = -4.02$, $p = .003$, but required similar effort, $t(9) = -1.24$, $p = .24$.

Similarly, we selected three activities that did not involve learning and therefore did not require knowledge reflection, but required similar effort as the high reflection learning strategies. "Write an essay about a TV show you recently watched" (non-studying) required less reflection than "Write an explanation of how GPS works as if you were explaining it to someone else" (high reflection), $t(9) = -4.09$, $p = .003$, but required similar effort, $t(9) = -.18$, $p = .86$. "Watch a video about mental health, taking notes while you watch it" (non-studying) required less reflection than "Write what you understood from the videos about how GPS works" (high reflection), $t(9) = -9.86$, $p < .001$, but required similar effort, $t(9) = -.45$, $p = .66$. "Write an essay

about your favorite season" (non-studying) required less reflection than "Answer an essay question about how GPS works" (high reflection), $t(7) = -3.05$, $p = .01$, but required similar effort, $t(7) = -2.05$, $p = .08$.

Finally, we note that written reflection ("Write an explanation of how GPS works as if you were explaining it to someone else.") was rated as the single most effective activity for test preparation on the list, M = 4.42 on a 1-5 scale.

Table 2S

Average Effort, Reflection, and Effectiveness Ratings of Each Activity (Pilot 2S)

	Effort	Reflection	Effectiveness
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
High Reflection			
1. * Write an explanation of how GPS works as if you were explaining it to someone else.	3.68 (1.06)	4.53 (0.77)	4.42 (0.69)
2. * Write what you understood from the videos about how GPS works.	2.89 (1.20)	4.47 (0.70)	4.11 (0.99)
3. * Answer an essay question about how GPS works.	3.63 (1.07)	4.37 (0.83)	3.84 (0.90)
Little to No Reflection			
4. * Re-watch one of the video sections and memorize it.	3.74 (1.10)	3.21 (1.32)	2.95 (0.91)
5. * Re-watch one of the video sections, taking handwritten notes as you watch.	2.84 (0.96)	2.37 (1.07)	3.37 (1.07)
6. * Re-watch a video section and create a visual representation of the information on a blank piece of paper.	3.56 (0.92)	3.33 (1.46)	3.50 (0.99)
7. Read an explanation from a textbook about how GPS works, highlight at least 12 sentences, and then transcribe those 12 sentences exactly.	3.36 (1.29)	2.68 (1.09)	3.00 (1.20)
8. Read a summary of the videos prepared by another participant, memorize the summary, and then write the summary from memory.	4.42 (0.84)	3.47 (1.02)	3.68 (0.82)
9. Read a transcript of the 7 videos sections, choose one section, and then record yourself narrating that portion of the video.	3.55 (1.19)	2.75 (1.37)	2.90 (1.37)
10. Re-watch one of the video sections and prepare flash cards for it (write a concept on one side and its meaning on the other), and then practice them.	3.57 (0.98)	3.52 (1.03)	3.95 (1.10)
11. Re-watch one of the video sections, taking notes in your own words.	2.68 (1.00)	2.58 (1.12)	3.21 (1.03)
12. Read a transcript of the 7 video sections and condense it by summarizing.	3.33 (1.14)	2.78 (0.94)	3.72 (0.96)
13. Read notes that another participant prepared on the videos and condense them by summarizing.	3.21 (0.79)	2.89 (1.15)	3.11 (1.05)
14. Re-watch one of video sections and come up with mnemonic devices and acronyms to help remember the information.	4.05 (0.89)	3.40 (1.23)	3.70 (1.08)
15. Read notes that another participant prepared on the videos and rewrite them in your own words.	3.21 (1.36)	2.74 (1.15)	3.11 (1.23)
16. Make a mind map to visualize what you learned about GPS (on a piece of blank paper, draw circles around the main topics and then circles for concepts	3.33 (1.02)	4.38 (1.07)	4.10 (1.14)

or subtopics, using lines to show how different topics and concepts relate to each other).			
17. Re-watch one of the video sections, then read a textbook section related to it, and then create an outline of all the concepts.	3.85 (1.14)	3.45 (1.05)	3.85 (0.88)
18. Watch a new video lecture that explains the same content and transcribe it exactly.	4.00 (1.05)	2.38 (1.53)	3.29 (1.23)
19. Read a summary of the videos prepared by another participant and generate questions to ask based on it.	3.26 (0.86)	2.96 (1.26)	3.22 (1.09)
20. Re-watch one of the video sections, take notes as you watch, and then record yourself reading your notes aloud.	3.31 (1.08)	2.69 (1.08)	3.00 (1.03)
21. Re-watch one of the video sections and come up with five questions for a quiz on it.	3.18 (1.18)	3.27 (1.16)	3.27 (0.88)
22. Re-watch one of the video sections and transcribe it exactly.	3.95 (1.10)	2.35 (1.31)	2.80 (1.28)
23. Re-watch two of the video sections about GPS, taking notes as you watch.	2.50 (0.83)	2.15 (0.93)	3.35 (0.93)
24. Read an explanation from a textbook about how GPS works, taking notes as you read.	2.95 (1.24)	2.62 (1.02)	3.86 (0.73)
25. Read a transcript of the 7 videos sections, taking notes as you read.	3.35 (1.00)	2.71 (1.21)	3.41 (0.87)
26. Read another participant's summary of the GPS videos, taking notes as you read.	2.57 (0.98)	2.62 (1.07)	3.10 (1.18)
27. Read the transcript of all 7 videos section.	3.00 (1.20)	2.09 (1.02)	3.00 (1.27)
28. Re-watch two of the video sections.	1.53 (0.80)	1.59 (0.94)	2.82 (0.95)
29. Read an explanation from a textbook about how GPS works.	2.59 (1.28)	2.47 (1.23)	3.65 (1.11)
30. Re-watch one of the video sections.	1.22 (0.55)	1.83 (1.04)	2.61 (1.14)
31. Read the transcript of all 7 videos section, highlighting important information as you read.	2.90 (0.97)	2.40 (0.99)	3.20 (1.11)
32. Read a summary that another participant wrote after they watched the GPS videos, and proofread it for typos and grammatical errors.	2.55 (1.19)	2.25 (1.33)	2.25 (1.16)
33. Create a study guide for the information you learned.	3.89 (0.76)	4.78 (0.55)	4.39 (0.70)
34. Re-watch one of the video sections, writing down key terms and concepts as you watch.	2.25 (1.02)	2.90 (1.12)	3.65 (1.09)
35. Watch a new video lecture that explains the same content.	2.16 (1.01)	1.95 (0.97)	3.53 (1.02)
Non-Studying			
36. * Write an essay about a TV show you recently watched.	3.85 (1.23)	2.55 (1.67)	1.70 (1.34)

37. * Watch a video about mental health, taking notes while you watch it.	2.59 (1.00)	1.47 (0.62)	1.47 (0.94)
38. * Write an essay about your favorite season.	3.82 (1.18)	2.45 (1.71)	1.64 (1.26)
39. Watch a video about mental health and transcribe it exactly.	4.28 (0.89)	1.83 (1.20)	1.39 (0.98)
40. Read a text about mental health and count all the vowels in it.	4 .00(1.17)	1.50 (1.15)	1.20 (0.70)

Note: Activities with an asterisk were selected to be used in Study 1.

Study 1S

In this study, we tested whether we could increase students' beliefs that written reflection is pleasant and effective by having them read a brief text to that effect (pleasant condition) vs. not (control condition), and whether increasing their beliefs would, in turn, lead them to choose to engage in written reflection more.

Method

The study was pre-registered at https://aspredicted.org/YGZ_Y4S.

Participants. One-hundred and ninety-nine students successfully completed the study (138 women, 60 men, 1 non-binary; 152 White, 28 Asian, 2 Black, 9 multiracial, 1 unknown, 1 chose "Other", 6 did not report race; modal year in college: 1st Year), of whom 192 took the quiz.

Design and Procedure. The study took place in the Spring semester of 2022 with students enrolled in the class Introductory Biology. Because the classes in Study 5 and this study were equivalent, there was no overlap in students. Like Study 5, the study was conducted during the class covering glycolysis. After explaining glycolysis, the instructor paused lecture for 15 minutes for the study, after which lecture resumed. As in Study 5, study packets were interleaved with respect to condition to ensure random assignment.

Students rated their subjective knowledge of glycolysis in the same way as in Study 5 (Cronbach's $\alpha = .95$) and their motivation to do well on the quizzes in the class. If they were assigned to the pleasant condition, they read the following text about written reflection being a fun activity:

Participants read the following text entitled "Explaining is fun!"

> Writing an explanation about material you just learned is a fun experience. Although people are sometimes worried that explaining what they learned will be unpleasant, people actually find the experience quite pleasant. They report that writing an explanation of what they just learned is fun and exciting.
>
> Even people who feel like they don't understand the topic very well find explaining to be more fun than they thought. It can be nice to discover what you already know, but also exciting to discover what you don't quite understand yet. In other words, people find it fun and exciting to figure out what they already know and what they can still learn more about.
>
> Thus, in addition to helping you understand the material better, the process of writing an explanation turns out to be enjoyable and exciting.

If they were assigned to the control condition, they did not read this. Students then reported their beliefs that engaging in written reflection would be effective and pleasant in the same way as in Study 5. They rated their interest in engaging in written reflection and in the control activity (reviewing notes) as in Study 5, but with the understanding that the activity they rated higher would be the one they engaged in for the next 5-7 minutes. If they rated both the same, they had to select the one they preferred on the following page. They then engaged in their selected activity. Finally, they reported their demographic information.

The instructors made the quiz available to students seven days after the study session through the class website. The quiz included three questions related to glycolysis and eight

questions unrelated to glycolysis, with each question being worth 2 points. The quiz made up about 0.5% of students' final grade in the class.

Results

As in previous studies, subjective knowledge predicted interest in engaging in written reflection, $b = .38$, $t(195) = 4.10$, $p < .001$, even controlling for motivation to do well on the quiz, $b = .38$, $t(194) = 4.03$, $p < .001$ (again, motivation did not predict interest in written reflection $b = .16$ $t(195) = .84$, $p = .40$). Once again, subjective knowledge did not significantly predict interest in reviewing one's notes on the material, $b = -.07$, $t(197) = -.88$, $p = .38$. Interest in written reflection was also predicted by people's beliefs that it would be fun, $b = .44$, $t(197) = 5.01$, $p < .001$, and marginally by their beliefs that it would be helpful, $b = .15$, $t(197) = 1.93$, $p = .054$.

Most students selected to review their notes (73%) rather than engage in written reflection (27%). Students who felt more knowledgeable about glycolysis were more likely to choose to engage in written reflection, $b = 29$, $z = 2.11$, $p = .03$. Students were also more likely to select written reflection the more they thought written reflection would be fun, $b = 0.46$, $z = 4.11$, $p < .001$, and helpful, $b = 0.45$, $z = 3.13$, $p = .002$. When both are in the model, only beliefs about fun remain significant predictors of choice, $b = 0.37$, $z = 2.98$, $p = .003$, whereas beliefs that WR is helpful are not, $b = 0.24$, $z = 1.47$, $p = .14$.

The manipulation check indicated that the manipulation was *un*successful; students who read that written reflection was fun did not rate written reflection as significantly more pleasant ($M = 3.60$, $SD = 1.28$), or more helpful ($M = 5.51$, $SD = 1.20$), than students in the control condition ($M_{pleasant} = 3.53$, $SD = 1.80$; $M_{helpful} = 5.51$, $SD = 1.48$), $b_{pleasant} = .07$, $t(197) = .32$, $p = .75$, $b_{helpful} = -.05$, $t(197) = -.25$, $p = .80$. Not surprisingly, then, students who read that written reflection was fun did not express any more interest in engaging in written reflection ($M = 3.76$, $SD = 1.57$) than students in the control condition ($M = 3.76$, $SD = 1.72$), $b = 0.00$, $t(195) = .03$, $p = .98$, and were no more likely to choose written reflection over reviewing their notes (29%) than students in the control condition (25%), $b = .21$, $z = .65$, $p = .51$. There were no significant interactions between condition and subjective knowledge to predict any of these dependent variables, $|t|$s $< .66$, ps $> .5$.

Students' scores on the glycolysis quiz questions did not differ significantly by condition (pleasant condition: $M = 5.90$, $SD = .44$; control condition: $M = 5.83$, $SD = .64$), $b = .07$, $t(190) = .94$, $p = .35$, nor did their scores on the non-glycolysis questions, (pleasant condition: $M = 13.00$, $SD = 2.52$; control condition: $M = 13.40$, $SD = 2.58$), $b = -.43$, $t(190) = -1.17$, $p = .24$. The null difference on quiz score is not surprising, given that the manipulation failed to increase their choice to engage in written reflection. Note that we pre-registered using the final exam scores because we planned to run this study during the same semester as study 5, but we ended up running it the following semester, so it made more sense to use the quiz scores to be consistent with study 5.

Study 2S

In this study, we tested whether we could increase students' subjective knowledge about the material by having them answer a few very easy questions about it, and whether increasing their subjective knowledge would, in turn, lead them to choose to engage in written reflection more.

Method

This study was pre-registered at https://aspredicted.org/G8V_NB2

Participants. Two-hundred and ninety-six students successfully completed the study (150 women, 138 men, 4 non-binary, 2 selected "other" and 2 did not wish to report; 214 White, 65 Asian, 11 Black, 3 American Indian or Alaska Native, 2 unknown, 2 chose "Other", 7 did not report race (students could select multiple races); modal year in college: 1st Year). Two participants were excluded from all analyses: one who did not rate their interest in WR, making it unclear which task they selected, and one who did not sign the consent form.

Design and Procedure. The study took place in the Fall semester of 2022 with students in an introductory astronomy class titled The Evolving Universe: Stars, Galaxies, and Cosmology. The study was conducted in both sections of the course during the class session covering "how Edwin Hubble discovered that our universe is expanding". After the instructor finished explaining this topic, students were given the opportunity to participate in the study for extra credit in the course. Study packets were interleaved with respect to condition to ensure random assignment.

Students in the subjective knowledge condition answered four multiple choice questions about the topic, whereas students in the control topic did not. These questions were constructed with the help of the two instructors, with the aim that students would be able to answer them very easily after listening to the lecture. All students then rated their subjective knowledge of the topic in the same way as in Studies 1S (Cronbach's $\alpha = .95$) and their motivation to perform well in the class.

Then students reported their beliefs that engaging in written reflection would be fun and pleasant in the same way as in Study 1S. They rated their interest in engaging in written reflection and in a control activity – reviewing notes – as in Study 1S. They then engaged in the activity they rated higher for the next 5-7 minutes. Finally, they reported their demographic information.

Results

As in previous studies, subjective knowledge predicted interest in engaging in written reflection, $b = .52$, $t(292) = 8.66$, $p < .001$, even controlling for motivation to do well on the quiz, $b = .47$, $t(291) = 7.39$, $p < .001$. Also consistent with the previous studies, subjective knowledge but did not predict interest in reviewing one's notes, $b = -.01$, $t(294) = -.19$, $p = .85$. Unlike in previous studies, motivation on its own also predicted interest in written reflection $b = .35$ $t(292) = 4.92$, $p < .001$.

However, our manipulation was unsuccessful. Students' performance on the four "easy" questions suggest that the questions were far more difficult than anticipated: only 56.7% answered all four questions correctly. Indeed, the manipulation check indicated that answering the "easy" questions was not successful in increasing students' subjective knowledge; students in the subjective knowledge condition did not feel significantly more knowledgeable about it ($M = 4.33$, $SD = 1.21$) than students in the control condition ($M = 4.12$, $SD = 1.27$), $b = .21$, $t(294) = 1.47$, $p = .14$.

In our pre-registration we predicted that should the manipulation not affect subjective knowledge, it would also not affect choice to engage in written reflection, and this was indeed the case. Most students selected to review their notes (80%) rather than engage in written reflection (20%), and this did not vary by condition (19.3% in the subjective knowledge condition vs. 20.5% in the control condition), $b = -.08$, $z = -.26$, $p = .79$. Within the subjective knowledge condition, there was a significant effect of performance on the "easy" questions on choice, $b = 1.06$, $z = 2.26$, $p = .02$, such that those who did not answer all four questions correctly were less likely to select written reflection (10.8%) than those who answered all questions

correctly (25.9%). Students who answered all questions correctly also had higher subjective knowledge (M = 4.66, SD = 1.16) than those who did not (M = 3.89, SD = 1.15), b = .77, $t(148) = 4.03$, $p < .001$, and those in the control condition (M = 4.12, SD = 1.27), b = .54, $t(229) = 3.24$, $p = .001$. This provides suggestive evidence that if subjective knowledge is successfully increased, it could prompt more openness to written reflection.

As in previous studies, students' beliefs that written reflection would be fun, b = .57, $t(292) = 13.00$, $p < .001$, and helpful, b = .25, $t(292) = 4.05$, $p < .001$, predicted their interest in it. These did not vary by condition, $b_{fun} = -.02$, $t(294) = -.09$, $p = .93$, $b_{helpful} = .04$, $t(294) = .30$, $p = .77$.