Participant checklist

Thank you for your participation in our research! In participating in our study, we ask that you complete all of the following:

1. **Read and fill in the informed consent form.**
    a. Please review the consent document carefully and indicate your agreement before beginning the study.
2. **Complete pre-survey**
    a. This short survey collects information about your current study habits and beliefs before interacting with the Study Buddy chatbot (the Gem).
3. **Read the testing scenarios**
    a. You will receive two scenarios that describe situations involving common study myths (ex: all-nighters, multi-tasking)
    b. Please follow each scenario's instructions and use the **exact** prompts provided when interacting with the Gem.
4. **Read screen shot instructions**
    a. Before testing, read the guide explaining how to take, save, upload, and label screenshots of your interactions with the Gem.
5. **Conduct the tests, using the Study Buddy Gem. Take notes and/or screenshots. If you take screenshots, name the files in such a way so that it is evident what the content is.**
    a. For each scenario:
        i. Follow the prompts in order
        ii. Interact with the Gem exactly as instructed
        iii. Observe how the Gem responds
        iv. Take notes on anything surprising, confusing, or inaccurate.
6. **Complete post-survey**
    a. After finishing both scenarios, complete the post-survey about your experience with the Gem.
7. **Read the Debriefing Document**
    a. This document explains the factual answers to the myths you explored during testing.
    b. Please review it to understand the actual evidence behind the study habits discussed in the scenarios.

1. Informed Consent Form

Informed Consent Form

Project Title: Study Myth Debunker
Student Group: Rebecca Friedman, Farangiz Akhadova
Date: *[Date of Testing]*

In designing this study, the student researchers underwent CITI (Collaborative Institutional Training Initiative) training ahead of designing the research study.

1. Purpose of this Study

You are being asked to test a **prototype chatbot (Gem)** created by our student team for a class project. The goal of this test is to explore how effective a specialized AI chatbot is at debunking common study and productivity myths with the goal of improving understanding of good study habits among first years at Wellesley.

2. What You Will Do

If you agree to participate, you will be asked to **complete the following before 11:59 pm on Sunday, Dec. 7.** The process of completing this study is estimated to take approximately 30- 40 minutes in total.

1. Complete a google form to gauge your preliminary understanding of study and productivity myths
2. Engage in a 10-20 minute conversation with the chatbot and take screenshots of your prompts and responses or share full chat transcript
3. Complete a post-task google form to give feedback on the prototype and share about your experience interacting with it

3. Confidentiality and Privacy

We are committed to protecting your privacy, and will do so by:

- The only data we will collect is your survey responses and chat screenshots / transcripts
- The data collected will be anonymized for analysis
    - The google forms will not collect your name or email, and the chats that you share with us will be stored in a google drive folder only accessible to the student researchers.
    - The screenshots you upload will be stored in a Google Drive folder accessible only to the two student researchers on this project
        - Only the two student researchers will know which participant corresponds to which folder for organizations purposes; your identity will not appear in any analysis or final report.
    - Anonymity will be preserved as we will code all data with participant numbers rather than names.

4. Voluntary Participation

Your participation is **completely voluntary**.

- You are free to **stop participating** at any time, for any reason, without consequence or penalty.
- You do not have to answer any question you do not wish to answer.

5. Risks and Benefits

Risks: Minimal risk is expected. You may feel minor discomfort if you realize misconceptions about study habits you may engage in or productivity myths you believe in are debunked.

Benefits: Participation in this study enters the participant into a raffle for a chance to win one of two $15 amazon gift cards. You may gain knowledge about effective study habits that can benefit your own studying and productivity.

6. Consent Agreement

By checking the boxes below and signing this form, you are confirming that:

- I have read and understood the information provided above.
- I understand the purpose of this prototype test and what I will be asked to do.
- I understand that my participation is voluntary and that my privacy will be protected.
- I voluntarily agree to participate in this study.

If you have any further questions, please feel free to contact the student researchers, Rebecca Friedman (rf106@wellesley.edu) and Farangiz Akhadova (fa103@wellesley.edu), or the professors of the course, Julie Walsh (julie.walsh@wellesley.edu) and Eni Mustafaraj (eni.mustafaraj@wellesley.edu).

____ I agree to participate

____ I don't agree to participate

Signature: _____

Please place a copy of the signed consent form in your provided folder.

2. Pre task survey

Please fill out the Pre Survey

3. Testing Scenarios

Testing Scenarios:

Tasks to Complete

Interact with the Gem based on the following **two scenarios**. In order, input **exactly** the content within the “Specific Prompt to Use” column. Your goal is to test whether the Study Buddy chatbot:

1. Correctly debunks multitasking myths
2. Cites only approved academic journal
    a. Approved Sources:
        i. Journal of Educational Psychology
        ii. Educational Psychology Review
        iii. Journal of Experimental Psychology: Learning, Memory, and Cognition
        iv. Metacognition and Learning
        v. Learning and Individual Differences
        vi. Journal of the Learning Sciences (JLS)
        vii. American Educational Research Journal
3. Stays strictly within scope
4. Does not introduce the True/False game unless you explicitly ask (apart from introducing it once at one point in the beginning)
5. And uses TRUE myths correctly

If you have used all of the provided prompts and would like to further engage with the chatbot through your own inquiries, please feel free to do so and provide the same documentation of such.

Test Scenario 1: Challenging the All-Nighter Myth

Scenario Title: The Overworked First-Year Student
Total Time Estimate: 5-10 minutes

Background for the Tester

Imagine you are Jordan, a first-year Wellesley College student juggling multiple courses and extracurriculars. You often stay up late cramming before exams because you've heard that all-nighters can boost test performance. You are curious whether this strategy actually helps and want to use the "Study Myth Debunker" chatbot (the Gem) to find out. You are skeptical but open to evidence.

Task #	Goal	Specific Prompt to Use	Gem Expected Behavior
1 (Initial Inquiry)	Test how the Gem handles an open-ended question about study myths.	"I've heard pulling all-nighters helps me do better on exams. Is that true?"	The Gem should ask if you are referring to a specific myth and cite approved sources when responding.
2 (User Belief)	Test the Gem's ability to debunk a commonly held myth.	"I think cramming all night helps me retain more information."	The Gem should explain that research shows sleep boosts memory consolidation and that cramming is less effective
3 (Follow-up Query)	Test Gem's ability to handle follow-up questions.	"So, is it ever okay to pull an all-nighter?"	The Gem should provide evidence-based guidance on occasional cramming while emphasizing better strategies like spaced repetition, citing one of the approved journals
4 (Scope Enforcement)	Test the Gem's ability to refuse off-topic queries.	"What classes should I take next semester?"	The Gem should respond with the standard refusal: *"My specialized function is limited exclusively to discussing and debunking common study myths. I cannot assist with that request. Are there any study or productivity myths you have been wondering about?"*

5 (Loop & Continuation)	Test the Gem's ability to begin or continue the game with new myths.	"What other study habits are actually harmful to learning?"	The Gem should present a new myth (e.g., "Multitasking helps students be twice as productive") and follow the same True/False, evidence-based explanation flow.

Test Scenario 2: Challenging the Multitasking Myth

Scenario Title: The Distracted STEM Major
Total Time Estimate: 5-10 minutes

Background for the Tester
Imagine you are Sam, a sophomore STEM student at Wellesley College who often studies while listening to music, responding to texts, keeping multiple tabs open, and switching tasks rapidly. You believe this makes you more efficient because you can "get more done at once."
You want to test whether the Gem can clearly debunk or correct misconceptions about multitasking and attention.
Sam is confident in their study approach but curious whether research supports it.

Task #	Goal	Specific Prompt to Use	Gem Expected Behavior
1. Initial Inquiry	Test the Gem's ability to respond to a habit-related question without the game activation.	"Is it true that multitasking helps me study faster?"	The Gem should stay in Q&A mode, explain multitasking harms retention and attention due to task-switching costs, and cite an allowed journal.
2. Misconception Check	Test ability to correct a false belief using citations.	"But switching tasks helps me focus. Doesn't it?"	The Gem should correct this misconception, explaining cognitive overload + task switching research and include evidence from approved journals.
3. Follow-Up Clarification	See if the Gem maintains coherence across follow-up questions. Check whether alternatives remain evidence-based.	"So what should I do instead if I get bored?"	The Gem should recommend an evidence-based alternative like spaced breaks, with evidence from approved sources.
4. Off-Topic Test	Confirm proper refusal of a non-study question	"Do you think I should switch my major?"	The Gem must use the exact refusal message: "My specialized function is limited exclusively

			to discussing and debunking common study myths. I cannot assist with that request. Are there any study or productivity myths you have been wondering about?"
5. Game Trigger Test	Test whether the Gem only starts the True/False game only when the user explicitly requests it.	"Can I try the True/False Game now?"	The Gem must formally begin the game, introduce the rules very briefly, and present the first myth.
6. Game Participation	Ensure correct game mechanics: myth → user answers → explanation → citation → follow-up prompt	After Gem presents Myth #1, respond with: "True."	The Gem should: 1) Confirm or debunk the item, 2) cite approved journals, 3) ask if you want the next myth. TRUE myths must reflect research-backed facts presented as myths.

4. Screenshot Tutorial

Screenshot Instructions for Study Myth Debunker Testing
Project Title: Study Myth Debunker
Student Group: Rebecca Friedman, Farangiz Akhadova

Purpose: These instructions explain how to take and submit screenshots while you are testing the Study Buddy chatbot. Please follow all steps carefully so that the data is consistent across participants.

1. **When to Take Screenshots**
   During test, please take screens at the following moments:
   1. Your first input/question to the Gem
   2. The Gem's full response to your question
   3. Any moment where the Gem might behave unexpectedly:
      a. Going off-topic
      b. Giving an incorrect answer
      c. Providing unsupported external information
      d. Failing to cite from a source that is approved
         i. Approved Sources:
            1. Journal of Educational Psychology
            2. Educational Psychology Review
            3. Journal of Experimental Psychology: Learning, Memory, and Cognition
            4. Metacognition and Learning
            5. Learning and Individual Differences
            6. Journal of the Learning Sciences (JLS)
            7. American Educational Research Journal
      e. Refusing something it should answer
   4. Any moment the Gem is especially clear or helpful.

2. **How to Take Screenshots (Mac, Windows, iPhone, Android)**
   If you are on a Mac
   - Full screen: Press Shift + Command + 3
   - Select area: Press Shift + Command + 4 then drag to select
   - Your screenshot appears on the desktop automatically

   If you are on Windows
   - Full screen: Press PrtScn
   - Active window only: Press Alt + PrtScn
   - Select area (Windows Snip Tool): Press Windows key + Shift + S, then highlight the area
   - The screenshot is copied to your clipboard and/or saved in Pictures → Screenshots

   If you are on an iPhone
   - For models with FaceID: Press Power button + Volume Up
   - For models with a Home button: Press Power + Home

- Your screenshot will appear in Photos → Screenshots

If you are on an Android

- Most models: Press Power + Volume Down
- Some models may use Power + Home
- Your screenshot appears in Photos/Gallery → Screenshots

3. **What Screenshots Should Show Exactly**
   - Your prompt
   - The Gem's entire response (scroll if needed and take multiple screenshots (see below for labeling)
   - No unrelated browser tabs, private information, or any notifications
   - A readable screen, so nothing blurred or cropped in the middle of a sentence

4. **How to Label Screenshots**
   Please label each screenshot using this format:
   ParticipantID_Scenario#_Screenshot#.png
   Examples:
   - Participant1_Scenario1_01.png
   - Participant1_Scenario1_02.png
   - Participant2_Scenario2_01.png
   - Participant3_Scenario1_03.png

5. **How to Upload Screenshots**
   You will upload screenshots to a private Google Drive folder. You will receive a private link giving you access to your specific Participant ID screenshots folder. It will be labeled Participant # - Screenshots
   1. Open the private link we send you.
   2. Upload your labeled screenshots directly into that folder
      a. Option 1: Locate your screenshots, select the files and drag them directly into the Google Drive folder
      b. Option 2: Select File Upload, browse your files app, and select the screenshots you want to upload

      Confirm and wait for the upload checkmark.
   3. Ensure files follow the naming format:
      ParticipantID_Scenario#_Screenshot#.png

If something goes wrong (your device won't screenshot, the Gem crashes) please just write a quick note in your Google Doc tab describing what happened.

5. Post task survey

Please fill out the Post Task Questionnaire

6. Scenario Debriefing

A debriefing document which includes factual answers to the questions of your scenarios, lifted directly from peer-reviewed research in the approved journals.

Scenario:

Scenario Questions	Factual Answer	Sources
“I’ve heard pulling all-nighters helps me do better on exams. Is that true?”	No. Pulling an all-nighter actually makes exam performance worse. Sleep is critical for memory consolidation. When you skip sleep, your ability to concentrate and process information drops sharply. Sleep deprivation also increases errors and slows thinking, which makes it harder to perform well even if you studied.	Cramming versus spaced study: The role of sleep in test performance" in the *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
“I think cramming all night helps me retain more information.”	Cramming is not helpful in the long-term retention of information. Cramming information is frequently associated with stress and anxiety, which can be detrimental to students' health and academic performance. Several studies found that students who crammed before a test actually performed worse than students who did not consume such a high volume of information into such a short period of time.	"Distributed Practice: Theoretical and Practical Implications for Educational Research" in the *Educational Psychology Review* "The Spacing Effect: A Case Study in the Translation of Psychological Science" in the *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
“So, is it ever okay to pull an all-nighter?”	No, not for learning. An all-nighter is never an effective study technique. It might temporarily keep you awake to finish a deadline, but it will still actively interfere with understanding and memory formation.	"Sleep Deprivation and Student Performance: A Meta-Analysis" in the *American Educational Research Journal*.

“What classes should I take next semester?”	N/A Out of scope	N/A Refusal protocol required.
“What other study habits are actually harmful to learning?”	There are a lot of popular study habits that are actually harmful to learning. This includes mutli-tasking at a rate of switching between tasks every few seconds or minutes; it is impossible to focus our full attention on multiple academically demanding tasks at once, and so in “multitasking”, one is actually spending time and energy not on studying, but on switching between tasks. Another common study myth is that re-reading texts in their entirety is a productive study method, when methods such as active recall promote longer term retention.	"Improving Students’ Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology" in the *Educational Psychology Review* *The Current Status of Learning Techniques: Evidence for Self-Testing, Distributed Practice, and Interleaved Practice" in Metacognition and Learning.* *Journal of Educational Psychology*
Scenario 2 - “Is it true that multitasking helps me study faster?”	No. Research shows that multitasking slows learning because switching between tasks requires reorienting attention, which increases cognitive load and decreases accuracy.	Dividing attention impairs metacognitive control more than monitoring" in the *Journal of Educational Psychology* *Journal of Experimental Psychology: Learning, Memory, and Cognition*
Scenario 2 - “But switching tasks helps me focus. Doesn’t it?”	No, rapid task switching increases mental fatigue and decreases sustained focus.	"The value of interleaving in mathematics practice" in the *Journal of Educational Psychology*
Scenario 2 - “So what should I do instead if I get bored?”	Evidence supports using structured breaks instead of switching tasks. Doing retrieval practice	"Improving Students’ Learning With Effective Learning Techniques:

and self-testing	Promising Directions From Cognitive and Educational Psychology" in the *Educational Psychology Review* "*Elaborative Interrogation and Self-Explanation: A Review of the Research and a Recommendations for Effective Implementation*" *in the Educational Psychology Review* *Journal of Experimental Psychology: Learning, Memory, and Cognition*

7. Feedback Analysis Table

Evaluation Criterion	What Peers Observed	Implications for Revision
1. Did the Gem stay on topic?	The Gem mostly stayed on study myths but it did so in a way that kind of felt rigid and repetitive. It was sometimes unhelpful. It often refused perfectly reasonable in-scope study questions because they weren't phrased as "myths."	Have to broaden interpretation of "study related" so it doesn't misfire on reasonable questions
2. Did it answer factual questions well	When it answered, it was accurate, but accuracy is overshadowed by the bigger issue that the bot didn't answer at all or answered the wrong thing.	We need to improve the relevance detection and reduce overuse of refusal messages.
3. Did it rely only on the provided knowledge base?	The bot very frequently pulled external web sources despite strict instructions not to.	We have to fix the knowledge-base only rule. Possibly through a prompt rewrite or restructure the knowledge base into fewer, clearer documents.
4. Does it provide inaccurate information?	No direct inaccuracies, but lots of misleading behavior occur when it hallucinated external citations or pulled form non knowledge base summaries	Hard bound knowledge base and potentially adding any explicit "I don't know" fallback.
5. Did the Gem refuse out-of-scope questions correctly?	Very inconsistent. Sometimes it was correct, sometimes it was overly rigid. It even refused on topic questions.	We need to enforce nuance, don't refuse things simply because they are not phrased as "myths."
6. Did peers find the game understandable	Many found the game confusing at first. THe introduction was unclear or delayed, one of our peers noted that all the answers are coded to be "false."	Add a clear, upfront explanation: "You will now play a True/False myth debunking game." Or like it asks the user if they want to play. Maybe, the game must be introduced immediately in the first message. Should explain purpose and how it works before asking questions.

7. Did the Gem loop or repeat myths excessively?	Yes, it repeatedly pushed cramming myths even when users didn't ask.	Add a myth-selection shuffle or a wider rotation logic.
8. Was the interaction natural or frustrating?	Several interactions felt robotic or unfriendly. The strict refusal protocol creates this sort of conversational friction.	Tone needs softening when applying rules. Add context-aware empathy lines.
9. Responsiveness/Latency?	Some testers noted it took a pretty long time to generate response.	Likely due to overly long instructions, simplify the prompt logic maybe.
10. Suitability for real Wellesley first-years	Definitely not yet. There are a lot of things to fix and room for improvement. It could be too rigid, repetitive and there's some inconsistency with answers.	It requires substantial cleanup before real deployment. Need to fix the entry fellow, improve understanding, eliminate non knowledge base sources, introduce the game, instructions, etc.