Coach not Crutch -- Main Effect -- Main (#201140)

Author(s)
This pre-registration is currently anonymous to enable blind peer-review.
It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
This experiment investigates the effects of AI tools on learning.

All participants will learn about effective writing. They then will either skip the pratice phase (no practice condition) practice w/o AI, or practice w/ AI. Finally, all participants will all be tested on their writing skill without the use of any AI tools.

We hypothesize that the participants practicing with AI will significantly outperform those who were assigned to no practice and to practice without AI.

3) Describe the key dependent variable(s) specifying how they will be measured.
Our primary outcome will be the writing effectiveness of the test email, as measured by the unweighted average of 5 ratings from GPT-4o of 5 writing principles (Rogers & Lasky-Fink, 2023): less is more, make reading easy, design for easy navigation, use enough formatting but no more, and make responding easy.

4) How many and which conditions will participants be assigned to?
In this between-subjects experiment, participants will be randomly assigned to one of three conditions:
No practice: participants will skip the practice phase and go from the learning phase into the test immediately.
Practice w/o AI: participants will practice editing a cover letter email on their own
Practice w/ AI: participants will practice editing a cover letter email with an AI tool that provides an answer they can copy, paste, and submit

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will fit an ANCOVA model predicting writing effectiveness from condition, controlling for pretest score and baseline characteristics (age, gender, race/ethnicity, primary language, education level, motivation to improve writing skills, self-rated writing skill, experience with AI writings assistants, and baseline writing effectiveness). We will run two planned comparisons of the marginal means: (a) participants in the AI-practice condition are expected to outperform the practice w/o AI condition, and (b)participants in the AI-practice condition are expected to outperform the no practice condition. Because we expect that practicing with AI will improve writing effectiveness, we will use one-sided p-values.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will winsorize all outcome values at 3SD above or below the mean.
Time and keystrokes will be square-root and log-transformed, respectively.

We will exclude participants who complete the survey in less than 30 seconds.
We will exclude participants who have already taken part in other studies in this series.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect 750 participants per condition (3 x 750 = 2,250 participants).

We will start with a technical pilot with 300 participants. If there are no technical issues, these data will be included in our main analytic sample.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will fit an ANCOVA model predicting writing effectiveness from condition, controlling for pretest score and baseline characteristics (age, gender, race/ethnicity, primary language, education level, motivation to improve writing skills, self-rated writing skill, experience with AI writing assistants, and baseline writing effectiveness).

We will fit three ANCOVA models predicting subjective effort, time, and keystrokes, respectively–controlling for pretest score and baseline characteristics.

Likewise, we will fit ANCOVA models to compare changes in writing self-efficacy, perceived learning. In these models, baseline measures of outcomes will be included along with pretest score and baseline characteristics.

We will use logistic regression to predict whether participants chose to see optional feedback for their test from condition, controlling for pretest score and baseline characteristics.

We will calculate a learning rate variable for each condition by dividing the difference between test and pretest by each unit effort (subjective effort, time, and keystrokes). We will then use similar ANCOVA models to the ones described above to compare the effect of condition on learning rate, controlling for pretest scores and baseline characteristics.

As a robustness check, we will re-run the analyses in #5 above excluding participants who report having used outside resources while completing the survey (i.e., search engines, Grammarly, generative AI tools, consulting with other people, or any other self-disclosed use).

We are collecting participants formatted writing and interactions with the AI editor through an iframe, that saves the data to a Mongo Database.

Coach not Crutch -- Main Effect -- Pairwise Comparisons (#205316)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Emails written by participants in the practice with AI condition will be preferred over emails written in the practice without AI and no practice conditions.

3) Describe the key dependent variable(s) specifying how they will be measured.

The main dependent variable is the win-rate of each email. The win rate is defined as the number of times the target email was preferred over the comparison email divided by the total number of comparisons involving that email. Thus, it will be a number between 0 and 1, representing the probability of superiority.

4) How many and which conditions will participants be assigned to?

Raters will not be assigned to condition, but the emails they are rating come from a previous experiment in which participants were assigned to practice with AI, practice without AI, or to a no-practice control condition.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use beta regression to predict the win-rate based on the condition the email writer was originally assigned to.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will not exclude any data.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Raters will rate all emails. We will collect enough data so that each email is compared against two other emails from different conditions.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Raters will be offered a performance-based bonus if their ratings align with the majority consensus across raters, to incentivize careful and thoughtful responses.

Coach not Crutch -- Main Effect -- Pairwise Comparisons (#205316)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Emails written by participants in the practice with AI condition will be preferred over emails written in the practice without AI and no practice conditions.

3) Describe the key dependent variable(s) specifying how they will be measured.

The main dependent variable is the win-rate of each email. The win rate is defined as the number of times the target email was preferred over the comparison email divided by the total number of comparisons involving that email. Thus, it will be a number between 0 and 1, representing the probability of superiority.

4) How many and which conditions will participants be assigned to?

Raters will not be assigned to condition, but the emails they are rating come from a previous experiment in which participants were assigned to practice with AI, practice without AI, or to a no-practice control condition.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use beta regression to predict the win-rate based on the condition the email writer was originally assigned to.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will not exclude any data.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Raters will rate all emails. We will collect enough data so that each email is compared against two other emails from different conditions.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Raters will be offered a performance-based bonus if their ratings align with the majority consensus across raters, to incentivize careful and thoughtful responses.

Coach not Crutch -- Forecasts (#191800)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Participants will predict the results of an experiment with three conditions: practice alone, use AI, and no practice. The outcome is the amount of learning in each condition, operationalized as the results in a test with no AI tools.

We hypothesize that people will predict that participants using AI will learn less than those practicing alone.

3) Describe the key dependent variable(s) specifying how they will be measured.

The rank in which the forecaster predicted the AI condition and the practice alone conditions would have.

An binary variable indicating whether the forecaster ranked the AI condition below the practice alone condition

4) How many and which conditions will participants be assigned to?

No conditions

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct a proportion test to see whether the proportion of people who rank "practice alone" above "use AI" is greater than those that do so in the opposite order.

Additionally, as a robustness check, we will conduct a Wilcoxon signed rank test to compare the ranks in which people ranked the "practice alone" and "use AI" conditions.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

No data will be excluded from analysis

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will collect data from 150 participants.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Nothing else to pre-register.

Coach not Crutch -- Human Benchmarks -- Main (#239157)

Author(s)   Pre-registered on: 2025/07/20 - 11:54 PM (PT)
This pre-registration is currently anonymous to enable blind peer-review.
It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
This experiment investigates the effects of AI tools on learning.
In Phase 1, all participants will learn about effective writing. They will then practice writing a cover letter on their own for as long as they like. In Phase 2, participants will be randomized to practice with AI, view online examples, or receive feedback from a professional editor. Finally, all participants will be tested on their writing skills without the use of any AI tools.
We hypothesize that participants who practice with the AI tool will outperform those who view online examples and will not be inferior than those who receive feedback from professional editors.

3) Describe the key dependent variable(s) specifying how they will be measured.
Our primary outcome will be the writing effectiveness of the test email, as measured by the unweighted average of 5 ratings from GPT-4o of 5 writing principles (Rogers & Lasky-Fink, 2023): less is more, make reading easy, design for easy navigation, use enough formatting but no more, and make responding easy.

4) How many and which conditions will participants be assigned to?
In this between-subjects experiment, participants will be randomly assigned to one of three conditions:
Practice w/ AI: participants will practice editing a cover letter email with an AI tool that provides an answer they can copy, paste, and submit
Online examples: Participants will have access to an embedded search platform that replicates the first five pages of a Google Search for "cover letter examples."
Expert feedback: participants will see how a professional editor rewrote the letter they submitted in Phase 1.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will fit an ANCOVA model predicting writing effectiveness from condition, controlling for pretest score and baseline characteristics (age, gender, race/ethnicity, primary language, education level, motivation to improve writing skills, self-rated writing skill, experience with AI writings assistants, and baseline writing effectiveness). We will run two planned comparisons of the marginal means: (a) participants in the AI-practice condition are expected to outperform the online examples condition, and (b) participants in the AI-practice condition are expected to not be inferior to the expert feedback condition. Because we expect that practicing with AI will improve writing effectiveness, we will use one-sided p-values.
We will claim non-inferiority if we can rule out that expert feedback outperforms AI practice by more than d = 0.15. Specifically, non-inferiority will be concluded if the one-sided 95 % CI for **(Expert – AI)** is entirely below 0.15 SD (i.e., the upper bound of the difference is lower than 0.15).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will winsorize all outcome values at 3SD above or below the mean.
Time and keystrokes will be square-root and log-transformed, respectively.
We will exclude participants who complete the survey in less than 30 seconds.
We will exclude participants who have already taken part in other studies in this series.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit 1,000 participants per condition (3 × 1,000 = 3,000 participants).
The final sample size may be lower due to attrition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will fit three ANCOVA models predicting subjective effort, time, and keystrokes, respectively–controlling for pretest score and baseline characteristics. Likewise, we will fit ANCOVA models to compare changes in writing self-efficacy, perceived learning. In these models, baseline measures of outcomes will be included along with pretest score and baseline characteristics.
We will use logistic regression to predict whether participants chose to see optional feedback for their test from condition, controlling for pretest score and baseline characteristics.
As a robustness check, we will re-run the analyses in #5 above, excluding participants who report having used outside resources while completing the survey (i.e., search engines, Grammarly, generative AI tools, consulting with other people, or any other self-disclosed use).

Coach not Crutch -- Human Benchmarks -- Pairwise Comparisons (#243115)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

Pre-registered on: 2025/08/18 - 09:05 AM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Emails written by participants in the practice with AI condition will be preferred over emails written in the see Google examples condition.

Emails written by participants in the feedback from human editors will not be preferred over emails written in the practice with AI condition.

3) Describe the key dependent variable(s) specifying how they will be measured.

The main dependent variable is the win-rate of each email. The win rate is defined as the number of times the target email was preferred over the comparison email divided by the total number of comparisons involving that email. Thus, it will be a number between 0 and 1, representing the probability of superiority.

4) How many and which conditions will participants be assigned to?

Raters will not be assigned to a condition, but the emails they are rating come from a previous experiment in which participants were assigned to practice with AI, see feedback from professional human editors, or Google online examples of cover letters.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use beta regression to predict the win-rate based on the condition the email writer was originally assigned to.

To assess non-inferiority, we will use a margin of 3 p.p., meaning that the one-sided 95% confidence interval of the difference in win rates will lie completely to the left of that value

We will claim non-inferiority if we can rule out that expert feedback outperforms AI practice by more than 3 percentage points. Specifically, non-inferiority will be concluded if the one-sided 95 % CI for **(Expert – AI)** is entirely below 3 p.p. (i.e., the upper bound of the difference is lower than 3p.p.).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will not exclude any data.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Raters will rate all emails. We will collect enough data so that each email is compared against three other emails from different conditions.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Raters will be offered a performance-based bonus if their ratings align with the majority consensus across raters, to incentivize careful and thoughtful responses.

Coach not Crutch -- Human Benchmarks -- Counteractual Pairwise Comparisons (#255792)

Author(s)
This pre-registration is currently anonymous to enable blind peer-review.
It has one author.

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Emails edited by AI will not be preferred at lower rates than emails edited by professional human editors

3) Describe the key dependent variable(s) specifying how they will be measured.
1. Choice: a new sample of participants will see pairs of letters side by side, and express preference.
2. AI ratings: each text will be rated by GPT4o on 5 principles of effective writing.

4) How many and which conditions will participants be assigned to?
There are no conditions

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
For each stimulus pair, we will compute the proportion of judgments favoring the AI version (AI win rate). We will then perform a one-sided one-sample t-test on these item-level win rates to test the non-inferiority hypothesis that the mean AI win rate is at least 0.5 minus a non-inferiority margin $m = .05$ (test against .45). Non-inferiority will be concluded if the lower bound of the one-sided 95% CI for the mean AI win rate is strictly greater than 0.45.

I will analyze AI ratings with a paired t-test.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
No data will be excluded.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Each email pair will be rated three times.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Left/right presentation is randomized.

The data already exists, so this preregistration is for a new analysis on data from an existing preregistered experiment.

Coach not Crutch -- Willingness to Pay (#236487)

Author(s)
This pre-registration is currently anonymous to enable blind peer-review.
It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Do participants indicate a higher willingness to pay (WTP) for writing feedback from an AI tool or an experienced human editor?
We hypothesize that participants will report a higher willingness to pay for feedback from a human editor than from an AI tool.

3) Describe the key dependent variable(s) specifying how they will be measured.
Willingness to Pay (WTP), measured on a 0–100 cents slider for each feedback type (AI and Human), with randomized presentation order.

4) How many and which conditions will participants be assigned to?
This is a within-subjects design. Each participant will report how much of their bonus they would be willing to forgo for:
- Feedback from an AI tool
- Feedback from an Experienced Human Editor

They will only make hypothetical payment decisions.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will conduct a paired-sample t-test comparing WTP for AI vs. Human Editor and calculate Cohen's d as the effect size.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude participants who complete the study in 15 seconds.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect N = 150 participants. This provides >80% power to detect a small-to-medium within-subjects effect size (Cohen's d ≈ 0.30) at α = 0.05.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Nothing else to pre-register

Coach not Crutch -- Mechanism -- Main (#197704)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

This experiment investigates the effects of AI tools on learning.

All participants will learn about effective writing. They then will either practice w/o AI, practice w/ AI, or just see an AI-generated example without having the ability to edit it. Finally, all participants will all be tested on their writing skill without the use of any AI tools.

1) We hypothesize that the participants practicing with AI will significantly outperform those practicing without AI.
2) We hypothesize that participants seeing an AI-generated example will significantly outperform those practicing without AI.

One reason practicing with AI might improve writing skill relative to practicing alone is that learners see an excellent example of effective writing. Comparing practice w/ AI to just seeing an AI-generated example allows us to ask how much of the effect of AI practice is driven by exposure to an example.

3) Describe the key dependent variable(s) specifying how they will be measured.

Our primary outcome will be the writing effectiveness of the test email, as measured by the unweighted average of 5 ratings from GPT-4o of 5 writing principles (Rogers & Lasky-Fink, 2023): less is more, make reading easy, design for easy navigation, use enough formatting but no more, and make responding easy.

4) How many and which conditions will participants be assigned to?

In this between-subjects experiment, participants will be randomly assigned to one of three conditions:

Practice w/o AI: participants will practice editing a cover letter email on their own

Practice w/ AI: participants will practice editing a cover letter email with an AI tool that provides an answer they can copy, paste, and submit

AI-generated-example: participants will not get the opportunity to practice; they will only see an AI generated example but won't be given the opportunity to edit it and will not submit a practice email

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will fit an ANCOVA model predicting writing effectiveness, controlling for pretest score and baseline characteristics (age, gender, race/ethnicity, primary language, education level, motivation to improve writing skills, self-rated writing skill, experience with AI writings assistants, and baseline writing effectiveness). We will run two planned comparisons of the marginal means: (a) participants in the AI-practice condition are expected to outperform the practice w/o AI condition, and (b) participants in the AI-generated-example condition are expected to outperform the practice w/o AI condition. Because we expect that practicing with AI will improve writing effectiveness, we will use one-sided p-values.

To estimate how much of the effect of AI practice is driven by exposure to an example, we will compare the estimated marginal means of writing effectiveness for the practice w/ AI and the AI-example conditions, respectively.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will winsorize all outcome values at 3SD above or below the mean.

Time and keystrokes will be square-root and log-transformed, respectively.

We will exclude participants who complete the survey in less than 30 seconds.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will collect 750 participants per condition (3 x 750 = 2,250 participants).

We will start with a technical pilot with 100 participants. If there are no technical issues, these data will be included in our main analytic sample.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will fit three ANCOVA models predicting subjective effort, time, and keystrokes, respectively–controlling for pretest score and baseline characteristics.

Likewise, we will fit ANCOVA models to compare changes in writing self-efficacy, perceived learning. In these models, baseline measures of outcomes will be included along with pretest score and baseline characteristics.

We will use logistic regression to predict whether participants chose to see optional feedback for their test from condition, controlling for pretest score and baseline characteristics.

We will calculate a learning rate variable for each condition by dividing the difference between test and pretest by each unit effort (subjective effort, time, and keystrokes). We will then use similar ANCOVA models to the ones described above to compare the effect of condition on learning rate, controlling for pretest scores and baseline characteristics.

As a robustness check, we will re-run the analyses in #5 above excluding participants who report having used outside resources while completing the survey (i.e., search engines, Grammarly, generative AI tools, consulting with other people, or any other self-disclosed use).

Coach not Crutch -- Mechanism -- Followup (#199451)

Author(s)

This pre-registration is currently anonymous to enable blind peer-review.

It has one author.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

This experiment investigates the effects of AI tools on learning.

We sample participants who completed a study on effective writing. In it, all participants learned about effective writing, were assigned to practice w/o AI, practice w/ AI, or just see an AI-generated example without having the ability to edit it, and then completed a test without the use of any AI tools. This pre-registration concerns a one-day follow-up analysis.

1) We hypothesize that the participants practicing with AI will significantly outperform those practicing without AI on a one-day follow-up.
2) We hypothesize that participants seeing an AI-generated example will significantly outperform those practicing without AI on a one-day follow-up.

One reason practicing with AI might improve writing skill relative to practicing alone is that learners see an excellent example of effective writing. Comparing practice w/ AI to just seeing an AI-generated example allows us to ask how much of the effect of AI practice is driven by exposure to an example.

3) Describe the key dependent variable(s) specifying how they will be measured.

Our primary outcome will be the writing effectiveness of the follow-up email, as measured by the unweighted average of 5 ratings from GPT-4o of 5 writing principles (Rogers & Lasky-Fink, 2023): less is more, make reading easy, design for easy navigation, use enough formatting but no more, and make responding easy.

4) How many and which conditions will participants be assigned to?

In this between-subjects experiment, participants will be randomly assigned to one of three conditions:

Practice w/o AI: participants will practice editing a cover letter email on their own

Practice w/ AI: participants will practice editing a cover letter email with an AI tool that provides an answer they can copy, paste, and submit

AI-generated-example: participants will not get the opportunity to practice; they will only see an AI generated example but won't be given the opportunity to edit it and will not submit a practice email.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will fit an ANCOVA model predicting writing effectiveness, controlling for pretest score and baseline characteristics (age, gender, race/ethnicity, primary language, education level, motivation to improve writing skills, self-rated writing skill, experience with AI writings assistants, and baseline writing effectiveness). We will run two planned comparisons of the marginal means: (a) participants in the AI-practice condition are expected to outperform the practice w/o AI condition, and (b) participants in the AI-generated-example condition are expected to outperform the practice w/o AI condition. Because we expect that practicing with AI will improve writing effectiveness, we will use one-sided p-values.

To estimate how much of the effect of AI practice is driven by exposure to an example, we will compare the estimated marginal means of writing effectiveness for the practice w/ AI and the AI-example conditions, respectively.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will winsorize all outcome values at 3SD above or below the mean.

Time and keystrokes will be square-root and log-transformed, respectively.

We will exclude participants who complete the survey in less than 30 seconds.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will invite 300 people who completed the original study.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will fit three ANCOVA models predicting subjective effort, time, and keystrokes, respectively–controlling for pretest score and baseline characteristics.

Coach not Crutch -- Mechanism -- Pairwise Comparisons (#205315)

Author(s)
This pre-registration is currently anonymous to enable blind peer-review.
It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Emails written by participants in the practice with AI condition and See AI Example condition will be preferred over emails written in the practice without AI condition.

3) Describe the key dependent variable(s) specifying how they will be measured.
The main dependent variable is the win-rate of each email. The win rate is defined as the number of times the target email was preferred over the comparison email divided by the total number of comparisons involving that email. Thus, it will be a number between 0 and 1, representing the probability of superiority.

4) How many and which conditions will participants be assigned to?
Raters will not be assigned to condition, but the emails they are rating come from a previous experiment in which participants were assigned to see an AI example, practice with AI, or practice without AI.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will use beta regression to predict the win-rate based on the condition the email writer was originally assigned to.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will not exclude any data.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Raters will rate all emails. We will collect enough data so that each email is compared against two other emails from different conditions.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Raters will be offered a performance-based bonus if their ratings align with the majority consensus across raters, to incentivize careful and thoughtful responses.