



## Supplementary Materials for

### **Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence**

Shakked Noy and Whitney Zhang

Corresponding author: [snoy@mit.edu](mailto:snoy@mit.edu)

#### **The PDF file includes:**

Materials and Methods  
Supplementary Text  
Figs. S1 to S21  
Tables S1 to S3  
References

# Contents

<b>A</b>	<b>Methods</b>	<b>2</b>
A.1	Experimental Methodology . . . . .	2
A.2	Analysis Details . . . . .	7
A.3	Deviations from Preregistration and Mistakes . . . . .	8
<b>B</b>	<b>Materials</b>	<b>10</b>
B.1	Task Screener Survey . . . . .	10
B.2	Task Survey . . . . .	12
B.3	Grading Screener . . . . .	39
B.4	Grading Survey . . . . .	45
B.5	14-Day Followup Survey . . . . .	53
B.6	Tasks . . . . .	55
<b>C</b>	<b>Supplementary Text</b>	<b>82</b>
C.1	Treatment-on-Treated Effects . . . . .	82
C.2	Task Structure and Skill Demand . . . . .	82
C.3	Job Satisfaction and Self-Efficacy . . . . .	83
C.4	Self-Reported and Objective Measure of ChatGPT Usage and Post-Usage Editing . . . . .	83
C.5	Comparison to Pure ChatGPT Output . . . . .	87
C.6	Other Descriptive Figures . . . . .	87
C.7	Robustness Checks . . . . .	88
<b>D</b>	<b>Supplementary Figures</b>	<b>89</b>
D.1	Methods: Grader Characteristics . . . . .	89
D.2	Additional Results . . . . .	90
D.3	Robustness Checks . . . . .	101
<b>E</b>	<b>Supplementary Tables</b>	<b>107</b>
	<b>Bibliography</b>	<b>109</b>

# A Methods

## A.1 Experimental Methodology

Our experimental design was preregistered at the AEA RCT Registry (AEARCTR-0010882); see Section A.3 for a list of our (predominantly very minor) deviations from the preregistration protocol.

**Participants** We recruit participants on Prolific, an online survey platform widely used in social science research (21). We screen participants on their self-reported occupation and preferentially recruit those with college degrees and in fulltime employment.

**Recruitment** Screened-in participants are invited to complete a study involving two 20-30 minute writing tasks designed to resemble tasks performed by people in their occupation, as well as an online signup task. They are informed that the survey will take about an hour and they will be paid \$10 for completing it, plus up to \$14 in bonus payments, with the average respondent earning \$8 in bonus payments. The expected earnings of \$18 for an hour of work substantially exceeds the Prolific standard of \$12/hour. Participants are told the survey is part of a study conducted by MIT researchers investigating the determinants of productivity on writing tasks.

Main data collection ran from the 27th of January to the 21st of February. The survey was mainly active between 5pm and 12am EST, when ChatGPT signups were almost always up, ensuring a 92% successful signup rate. About 8% of treated participants nevertheless experienced idiosyncratic technical difficulties that prevented them from signing up.

**Survey Structure** The full questionnaires of all surveys are available in Section B, and Figure S.1 visualizes the survey flow. Survey respondents are asked for their consent, then given a comprehensive set of instructions for the writing tasks. They are required to correctly answer two comprehension questions about the instructions and respondents who fail the comprehension checks twice (in accordance with Prolific’s requirements) are screened out.

Respondents answer a short list of questions about their demographic characteristics, skills, and awareness and usage of different technologies (including ChatGPT). They then complete the first task and are asked followup questions about how much time they spent, their job satisfaction and self-efficacy, and the realism of the task.

Subsequently, they complete the signup task (signing up for ChatGPT or Overleaf, depending on which group they are in, and submitting screenshots as proof). Respondents in the treatment group are asked to enter several sample prompts into ChatGPT and submit screenshots of the output. The sample prompts are designed to communicate ChatGPT’s range of capabilities: one involves pasting in a prompt for a professional writing task, another involves reprompting ChatGPT to expand on or rephrase its output, a third involves varying the wording of a prompt, and a fourth involves pasting in several paragraphs and asking ChatGPT to summarize them.

Respondents then complete the second task, answering more followup questions, including three questions about their beliefs about automation. Afterwards, all respondents (both treatment and control) are asked whether they used ChatGPT on the second task; if they did, they are asked a battery of followup questions about how they used it. The survey then ends.

**The Tasks** Each occupation has two associated tasks, and respondents see the two tasks in a random order. We designed the tasks to imitate real tasks in these occupations, based on conversations with

people in the occupations and pilot surveys on Prolific. The full list and text of the tasks is available in Section B.6. Briefly, the marketers in our experiment write a press release about a hypothetical product; grantwriters write a cover letter for a grant application; managers and HR professionals write a long company-wide email on a sensitive topic; data analysts write an analysis plan in code-notebook format; and consultants write a short report based on three provided sources. Each task is designed to take 20-30 minutes and involve writing about 400 words.

After respondents complete the first task, we ask them if this task realistically imitates tasks that people in their occupation do, on a scale from “very unrealistic” to “very realistic.” The results are plotted in Figure S.6: 85% of respondents report the tasks are “realistic” or “very realistic,” 10% say “neutral”, and 5% report “unrealistic” or “very unrealistic.” We also ask whether they have completed a similar task before in their job; 68% say yes.

**Incentive Arms** At the beginning of the survey, respondents are randomized into three incentive arms. All respondents know each of their essays will be graded on a 1-7 point scale by an experienced professional in their occupation. In the “linear incentives” group (40% of respondents), participants receive a bonus \$1 for each point they receive on each essay (up to \$14 total). In the “convex” group (40% of respondents), they receive the linear payment plus a \$3 payment for each time they score a 6 or 7. In the “exact time” group (20% of respondents), respondents receive the linear payment and are forced to spend exactly 15 minutes on the task: respondents cannot voluntarily click onto the next page: instead, the page auto-advances after 15 minutes.<sup>1</sup> We emphasize to these respondents that we will be monitoring their essay text box throughout the 15 minutes and will not approve bonus payments to respondents we judge to be noncompliant (for example, because they are idle for several minutes at a time). Respondents are consequently highly compliant with the 15-minutes instructions according to our objective measures of time active on the task, as shown in Figure S.8.

We cross-randomize incentive schemes for a few reasons. First, any single incentive scheme we chose would have been somewhat arbitrary and non-organic, so we wanted to show robustness across multiple different incentive schemes. Second, each arm serves a specific purpose. The baseline linear incentive scheme is easy to understand and induces high effort (the 7-point grading scale was chosen so that we could pay \$1 per point, taking advantage of left-digit bias while remaining within our budget). The convex arm lets us see what happens to ChatGPT usage when there is a special emphasis on producing very high-quality output. Finally, the exact-time arm lets us hold effort fixed between the treatment and control groups, allowing us to interpret any difference in grades as an effect of ChatGPT on productive capacity. We chose 15 minutes so that respondents were somewhat constrained and because we expected compliance with the time requirement to decline substantially with longer time periods (e.g., 25 minutes).

Effort turns out to be not much higher in the convex incentives group than the linear group, suggesting that the linear incentive scheme already elicits maximum feasible effort.

**The Retry Arm** 30% of treatment group respondents go through one final step before the end of the survey. They are re-shown the first task prompt and the essay they wrote, and told that they have the option to edit or replace their output, and may use ChatGPT to do so if they wish. If they do, they will be paid based on the grade received by their edited/replaced output instead of their original essay. Respondents are forced to spend at least 2 minutes on this page (as a nudge away from clicking through impatiently to the end of the survey), but are free to do whatever they want. This arm was

---

<sup>1</sup>In this group the payments are linear for each point received *above 1*, which lets us subtract a base \$2 from the payment since this group spends much less time on the survey on average.

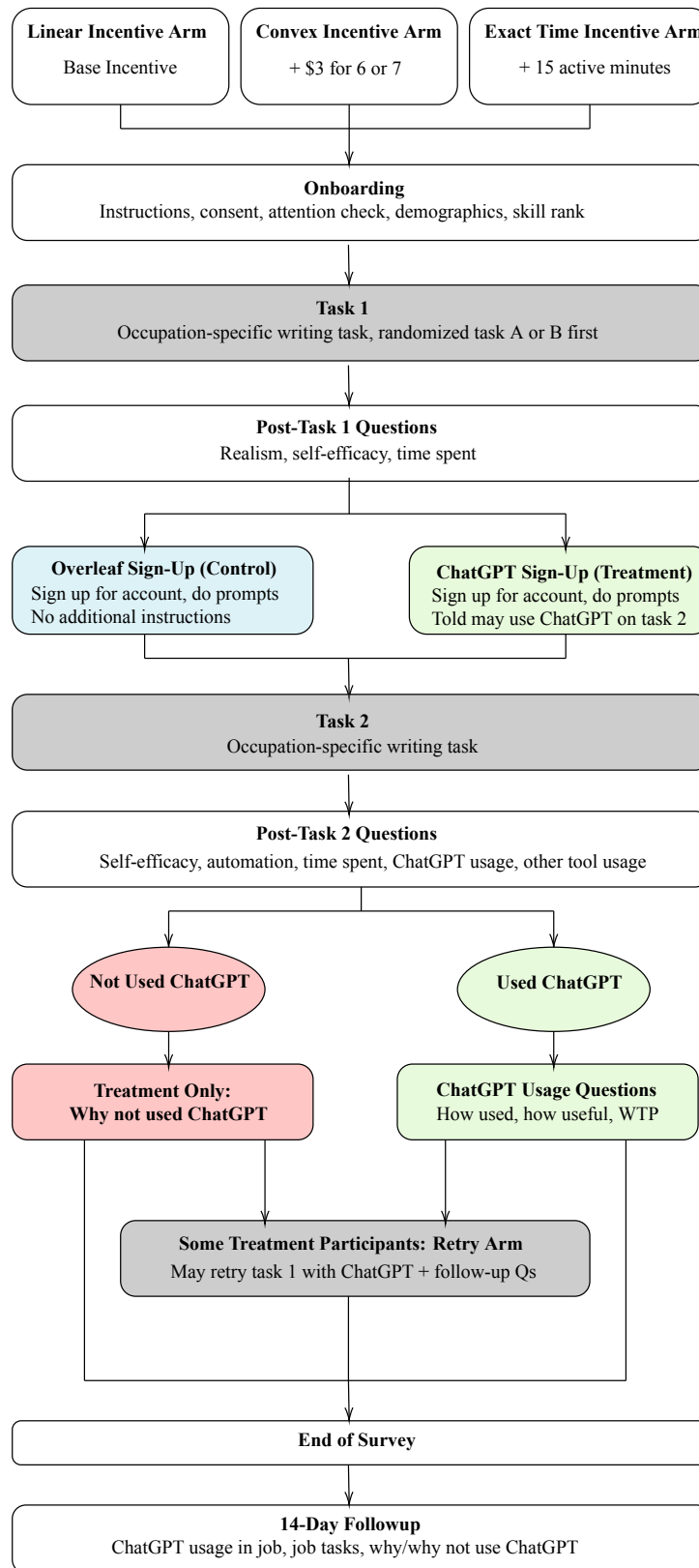


added about halfway through data collection, based on a suggestion we received, and all subsequent treatment group respondents in the linear and convex incentive arms are in this arm.<sup>2</sup>

---

<sup>2</sup>Thanks to Lucas Barros for suggesting this arm.

Figure S.1: Experimental Design



**Grading** We recruit graders on Prolific as well, imposing stricter screening criteria. We require graders to have done a task similar to the experimental tasks in their job in the past year. We also prioritize graders who have greater occupational tenure and higher wages. Graders have an average occupational tenure of 14.5 years (IQR [5, 21]). Figure S.2 shows the distribution of grader salaries. We exclude participants in our main survey. Graders complete a grading survey lasting around 1.5 hours; they see both the first prompt and second prompt corresponding to their occupation-specific task, and grade both the first task and second task for up to 7 respondents.<sup>3</sup> To aid in grading, they are provided with two examples per prompt, one example of a high grade and one of a low grade. Graders are paid a base rate of \$16 and may receive bonuses of up to \$8. They are told that their bonuses will depend on how correlated their grades are with those of other graders seeing the same essays, as well as two randomly selected essays that we may double-check.

At the grader by participant observation level, the grader characteristics are similar in the treatment and control groups. The distribution of salaries is plotted in Figure S.3. In the control group, the mean years of tenure is 14.5, SD 11.28. In the treatment group, the mean years of tenure is 14.93, SD 11.53.

**Bonus Payments: Planned and Actual** We instructed evaluators to do their best to give the median essay a score of 4 out of 7, and told respondents that they should expect to receive \$8 of bonus payments on average (4 per essay; respondents in the convex group knew they could get an extra \$3 per essay on top of this). For each participant, we randomly picked one of the participant’s evaluators and used that evaluator’s grades for the participant’s bonus payments. This resulted in an average actual bonus payment of \$8.7 including convex bonuses, and an average total payment of \$18.3 (recall that respondents in the linear and convex conditions receive a \$10 base payment and respondents in the exact-time condition receive an \$8 base payment). With an average survey completion time of 65 minutes, this comes out to a roughly \$17/hour average rate, substantially above the Prolific standard of \$12/hour.

**Followup Surveys** We launch the two-week and two-month followup surveys on Prolific and use custom allowlists so that only employed participants in the original survey can see the followup surveys. The connection between the followup surveys and original survey is obfuscated as much as possible: participants just see the survey as one of the many surveys they are eligible to participate in on their standard Prolific homescreen, they do not know there is a custom allowlist governing who can see the survey, and we launch the two-week followup survey under a different account name than the main survey.

In the case of the two-week survey, we add respondents to the allowlist on a rolling basis as soon as two weeks have passed since they completed the main survey. (Since this was done manually, sometimes the delay is a bit longer than two weeks). The two-week survey is open for about a month after the main survey finishes collecting responses, but the vast majority of responses to the two-week survey come in soon after respondents become eligible to complete it. In the case of the two-month survey, we invite everyone who completed the main survey, starting on the 19th of April and closing the survey on the 29th of April.

As one sidenote, note that the obfuscation of the two-week followup survey lets us verify that respondents were accurately reporting their occupation in the original screeners. This is an obfuscated followup weeks later that asks about what kind of writing tasks you do in “your job,” and the answers match up perfectly with the occupation we record from our screening surveys weeks earlier.

---

<sup>3</sup>Some graders see 7 human responses. Some graders see 6 responses, plus a “raw” ChatGPT output to each prompt. A small subset of graders see fewer than 6 responses. The pay does not differ by number of responses

## A.2 Analysis Details

**Sample Restrictions** If a participant somehow completes the main survey multiple times, we keep only their first response. If a participant completes one of the grading surveys after doing the task, we drop all grades they gave. Similarly, if a participant does the task after grading, we drop them from the task sample. (We did our best to manage the invite lists on the Prolific surveys so that task participants did not see the grading surveys and vice versa, but since this required periodic manual updates of allowlists, some people slipped through the cracks).

A few of our evaluators clearly do not pay attention in the grading survey. In an attempt to objectively screen out such evaluators, we drop every evaluator who has a correlation of  $<0.1$  between their grades and the grades of other evaluators who see the same essays as them. This drops 23 of our 148 evaluators. We view this as a reasonable main specification. However, this turns out not to matter at all for our results—the results look quantitatively very similar if we don’t do this—as we show in the Robustness Checks.

Our main figures in the paper use only the linear and convex incentive schemes. Separate figures in these Supplementary Materials report results for the exact-time group. The job satisfaction, self-efficacy, and beliefs about automation results use all participants.

**Regression Specifications** We have two main regression specifications, one for when the outcome is evaluator grades (since there we have multiple grade observations per task participant), and one for when the outcome is anything else.

First, consider regressions where the outcome variable is a non-grade outcome. Here we use a participant-level dataset; let  $i$  denote participants. Let  $\Delta y_i$  denote the difference between the post-treatment outcome and the pre-treatment outcome for participant  $i$ . Then the regression is

$$\Delta y_i = \beta_0 + \beta_1 D_i + \theta_{\text{occ}(i)} \times \alpha_{\text{first}(i)} + \pi_{\text{inc}(i)} + \varepsilon_i \quad (\text{S.1})$$

where  $D_i$  is a dummy for the tasker being treated,  $\theta_{\text{occ}(i)}$  are occupation fixed effects,  $\alpha_{\text{first}(i)}$  is a dummy for which of the two occupation-specific tasks the respondent saw first,  $\pi_{\text{inc}(i)}$  are incentive scheme fixed effects, and  $\varepsilon_i$  is the error term. We use robust standard errors.

Second, consider regressions where the outcome is evaluator grades. Let  $i$  denote a participant completing the task and let  $j$  denote an evaluator. Note that when an evaluator sees a participant’s pre-treatment task, the evaluator also sees that participant’s post-treatment task later in the same grading survey (though is not told that both essays were produced by the same person). An observation in this dataset is hence a participant-evaluator pair. Define  $\Delta y_{ij}$  as the difference between the grade that evaluator  $j$  gave to participant  $i$ ’s post-treatment task and the grade that evaluator  $j$  gave to participant  $i$ ’s pre-treatment task. The regression equation is

$$\Delta y_{ij} = \beta_0 + \beta_1 D_i + \gamma_j + \theta_{\text{occ}(i)} \times \alpha_{\text{first}(i)} + \pi_{\text{inc}(i)} + \eta_{ij} \quad (\text{S.2})$$

where in addition to the variables above,  $\gamma_j$  are evaluator fixed effects and  $\eta_{ij}$  is the error term. Here we cluster standard errors at the participant level.

**Standardization of Regression Coefficients** Our main figures report regression coefficients standardized to be expressed in standard deviations of the outcome variable. To do this, we take the regression coefficient and standard error from the regression equations described above, and divide both by the standard deviation of the outcome variable  $y_i$  in the pre-treatment period.

**Inequality Results** As explained in the figure note, these figures are created from a dataset at the participant-evaluator level, and the dots and lines plot raw slopes. The printed coefficients, meanwhile, come from a regression of the form

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 D_i + \beta_3 (D_i \times x_{ij}) + \gamma_j + \nu_{ij} \quad (\text{S.3})$$

where  $y$  is the y-axis variable and  $x$  is the x-axis variable from the plots. We cluster at the participant level. As usual,  $D_i$  is a treatment dummy and  $\gamma_j$  are grader fixed effects. The “control slope” printed in the plot is  $\beta_1$  and the “difference in slopes” is  $\beta_3$ .

## A.3 Deviations from Preregistration and Mistakes

### A.3.1 Deviations from Preregistration

Here we describe aspects of data collection that were amended from the preregistered protocols or were not described in those protocols.

**Sample Size** We preregistered a sample size of 600 based on rough budgeting projections, but ended up spending substantially more than expected on screening and grading, meaning we adjusted down and ultimately ran out of money at 450 responses.

**Randomization Procedure** For our first 189 responses, participants were randomized into incentive-arm groups using a Qualtrics function that generates a random integer between 1 and 10. We assigned respondents with an integer 1-4 to the linear group, 5-8 to convex, and 9-10 to exact-time. However, we noticed this was not exactly achieving the desired 40/40/20 proportions even as our sample size grew—at 189 responses, we had 90 linear, 69 convex, and 30 exact responses. It turns out the Qualtrics random integer generator does not seem to sample uniformly (see e.g. [here](#) or [here](#)) despite Qualtrics claiming it is a Mersenne Twister which should sample uniformly. Following this, we changed the procedure to use Qualtrics’s “randomly present element” feature and changed allocations to 36% linear, 42% convex, 22% exact to achieve the ultimately desired 40/40/20 split. Adding fixed effects for before/after this change does not quantitatively affect our results at all.

**Data Analyst Instructions and Screening** After collecting our first 20 data analyst responses, we noticed that there was a lot of variance in the structure of responses, and a couple of respondents wrote in open-text feedback that they don’t analyze data in their job and mainly clean and build datasets. Following this, we tightened our data analyst screening procedure to screen out analysts who did not tick “writing code to analyze data” as one of the main things they do in their job, and added some extra instructions to the task to clarify what the output format should look like. In the Robustness Checks section we report results excluding data analysts.

**Dropping Bad Faith Responses** We manually drop responses where the respondent clearly did not exert effort on the first task (e.g., left it blank, only wrote 1-2 sentences). We did not explicitly preregister this. All our manual dropping decisions can be inspected in our uploaded data and code.

**Employment Status** After collecting 204 responses, we realized that Prolific’s employment-status metadata on respondents is not very reliable (it is missing for many respondents), so we added a question about employment status to the main survey and launched a retroactive survey asking for employment status, inviting the respondents who had already completed our survey. The response rate to this retroactive survey is 96% so we are only missing employment data for a few responses.

**Editing Essays to Remove Identifying Information** We manually edited some essay responses to remove potentially real names respondents used in signing off the hypothetical emails.

**Added Question** For respondents in the treatment group who didn’t use ChatGPT, we added a question asking why, midway through response collection.

**Added Retry Arm** We added the retry arm (where, at the very end of the survey, treated respondents are given the opportunity to replace or edit their first-task essay using ChatGPT) about halfway through data collection; a colleague suggested this arm to us after an internal presentation of initial survey results.

**Added Extra Signup Prompt** Midway through data collection, we added a fourth prompt to the ChatGPT signup page, instructing participants that rewording prompts to ChatGPT may affect its output. This was done to address concerns that individuals were simply pasting our prompt because they did not know that changing the wording of the prompt would affect its output. We added a fourth prompt to the Overleaf signup page as well.

**Dropping Uncorrelated Graders** Our grading scheme was a rough draft in the preregistration; we mentioned checking how correlated graders are with each other, and as one assessment method suggested dropping graders uncorrelated with everyone else. We ended up doing this in our main analyses, but it doesn’t make a difference to our results anyways as is shown in the Robustness Checks.

### A.3.2 Mistakes

**Copywriters to Grant Writing Task** We realized after data collection that we had been assigning people who self-designated their occupation as “copywriters” to the grant writing task, despite intending to assign them to the marketing task (which would be more suitable). This affects only 2% of our respondents, dropping them from the dataset does not affect our results, and 50% of them have nevertheless completed tasks similar to the grant writing task in their job before.

**Mistakes in Working Paper Version** The working paper versions of this paper, the first of which was released on March 2 2023, had two mistakes that were caught pre-publication.

- The “time taken” histogram in the March 2 version of the paper was miscoded (we were assigning responses to histogram bins using weak inequalities on both ends of the bins, so responses on the boundaries were coded as belonging to two bins, and the columns hence clearly added up to more than 100%). This was caught by a reader of the paper and fixed between the March 2 and March 10 versions of the working paper.

- The “how did the respondent use ChatGPT” variable was miscoded in both the March 2 and March 10 versions of the working paper; we messed up the mapping from Qualtrics’s numerical answer codes to the answers respondents saw. (Because, when you add new options to a multi-choice question in Qualtrics and then change the order of the answers, the answers keep their original numeric codes despite being in a different order now.) This mistake led to us overstating the percentage of respondents who said they submitted ChatGPT’s output without editing, and understating the percentage of respondents who said they submitted after editing. We caught this mistake while preparing the replication package before publication.

Between the working paper and published version we also added some data and tweaked the analysis in various small ways.

## **B Materials**

### **B.1 Task Screener Survey**

#### **B.1.1 Welcome**

Welcome to this short 1-minute screening survey. We will ask you a few questions about yourself. Depending on your responses, we may ask you if you want to be invited to a 1-hour followup survey within the next week.

First, what is your Prolific ID?

## B.1.2 Occupation

Which of the following, if any, **best describes your occupation?** (Please tick "Other" if none of them describe your occupation).

Content Creator or Blogger	Administrator
Policy Analyst	Grant/Proposal Writer
Manager (non-office job)	Technical Writer
Medical Worker	Manager (office job)
Freelance Writer	HR Professional
Lawyer	Engineer/Software Developer
Public Relations Professional	Marketer
Business Strategist	Teacher
Data Analyst or Data Scientist	Management Consultant
Copywriter	Other (please specify)
	<input type="text"/>

## B.1.3 Data Analyst Follow-Up (Data analysts only)

Which of the following tasks do you mainly do in your job? Please tick all that apply.

Analyzing data using a spreadsheet interface
Writing code that analyzes data
Pulling/tabulating data



## B.1.4 Consent (Screened-in occupations only)

### **Congratulations! You are eligible for participation in the followup survey.**

Below, we **describe the followup survey** and then ask if you would be interested in participating.

#### **Followup Survey Details**

The followup survey will take **about 1 hour** and has a few parts. We will ask you to complete **two 20-30 minute writing tasks** designed to mimic tasks that people in your occupation do in their jobs. These tasks will be graded. We will also ask you to **sign up for an online account** (at a well-known, reputable website, as you will be able to verify). Finally, we will ask a few other questions throughout.

We are looking for respondents who are willing to **take the writing tasks seriously and do their best to produce good output**. These tasks **will be graded**, and we will **pay substantial bonuses for high grades**, as we explain below. We are also looking for people who are **willing to sign up for an online account**. We cannot tell you beforehand what website the account will be for, but it will be a well-known website and you will be able to verify its safety.

#### **Payment for Followup Survey**

We will pay a base of \$10 for completion of the survey. In addition, each of your two writing tasks will be graded on a scale of 1 to 7 by another professional in your occupation. You will receive \$1 in bonus payments for each point you receive, up to \$14 of bonus payments (\$24 total) across both tasks. The average grade, among respondents who put in a committed effort, will be about 4.

We therefore expect the **average respondent** who puts in a **committed effort on the writing tasks** to earn about **\$18 in total**, for around 1 hour of work.

#### **Additional Information**

The followup survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks.

Would you like us to invite you to participate in the followup survey, sometime over the next week?

Yes

No

## B.1.5 Invite (Screened-in occupations only)

Thanks for your interest!

**The followup survey will be active on Prolific at various times over the next week**, usually in the afternoons and evenings. It will pop up in your Prolific feed, as we will custom-invite everyone who consents in this screening survey. The title of the study will be something like "Complete two writing tasks and a signup task (big bonus payments)."

Please click onto the next page to be redirected back to Prolific with your completion code. You will be paid \$0.20 for your participation in this screening survey.

## B.2 Task Survey

The pages shown below are for managers. For other occupations, all instances of "manager" are replaced by that occupation, and the tasks are replaced by occupation-specific tasks. For the full list of

tasks, see section B.6.

## B.2.1 Consent (Exact time arm)

### Welcome!

This survey has a few parts. We will ask you to complete **two 15-minute writing tasks** designed to mimic tasks that people in your occupation do in their jobs. **We will require you to spend exactly 15 minutes on each task, and will track whether you are on-task.** We will also ask you to **sign up for an online account** (at a well-known, reputable website, as you will be able to verify). Finally, we will ask a few other questions throughout.

We are looking for respondents who are willing to **take the writing tasks seriously and do their best to produce good output.** To this end, we will give **substantial bonus payments for good performance on the writing tasks.** Overall, we expect the **average respondent who puts in a committed effort** on the writing tasks to earn about **\$6 in bonus payments,** on top of the \$10 base pay for this study (\$16 overall). You can earn up to \$12 of bonus payments (\$22 overall) for getting perfect grades on the tasks. We expect the whole survey to take about 45 minutes on average.

### Additional Information

This survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks. Your participation in this study is completely voluntary and you can choose to withdraw at any time without any penalty or consequence. If you volunteer to participate, we will assign you questions and writing tasks as described above, and ask you to sign up for an online account. We do not anticipate any risks or discomforts in the survey. The research may involve risks that are currently unforeseeable. We anticipate the study will provide benefits to society by enabling a better understanding of the determinants of productivity.

If you have any concerns or comments about this study, you can contact the researchers at [snoy@mit.edu](mailto:snoy@mit.edu) or [zhangww@mit.edu](mailto:zhangww@mit.edu). You can contact the MIT Committee on the Use of Humans as Experimental Subjects at [couhes@mit.edu](mailto:couhes@mit.edu).

Data from this survey may be made public. We will remove Prolific IDs and all identifying information before posting the data, to maintain confidentiality.

Are you willing to participate in this survey? Please click "Yes" only if you are willing to **take the writing tasks seriously** and are **willing to sign up for an online account.**

Yes

No

## B.2.2 Consent (Linear arm)

### Welcome!

This survey has a few parts. We will ask you to complete **two 20-30 minute writing tasks** designed to mimic tasks that people in your occupation do in their jobs. We will also ask you to **sign up for an online account** (at a well-known, reputable website, as you will be able to verify). Finally, we will ask a few other questions throughout.

We are looking for respondents who are willing to **take the writing tasks seriously and do their best to produce good output**. To this end, we will give **substantial bonus payments for good performance on the writing tasks**. Overall, we expect the **average respondent who puts in a committed effort** on the writing tasks to earn about **\$8 in bonus payments**, on top of the \$10 base pay for this study (\$18 overall). You can earn up to \$14 of bonus payments (\$24 overall) for getting perfect grades on the tasks. We expect the whole survey to take about 60 minutes on average.

You may also be randomly offered additional bonus payments for other aspects of task performance.

#### Additional Information

This survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks. Your participation in this study is completely voluntary and you can choose to withdraw at any time without any penalty or consequence. If you volunteer to participate, we will assign you questions and writing tasks as described above, and ask you to sign up for an online account. We do not anticipate any risks or discomforts in the survey. The research may involve risks that are currently unforeseeable. We anticipate the study will provide benefits to society by enabling a better understanding of the determinants of productivity.

If you have any concerns or comments about this study, you can contact the researchers at [snoy@mit.edu](mailto:snoy@mit.edu) or [zhangww@mit.edu](mailto:zhangww@mit.edu). You can contact the MIT Committee on the Use of Humans as Experimental Subjects at [couhes@mit.edu](mailto:couhes@mit.edu).

Data from this survey may be made public. We will remove Prolific IDs and all identifying information before posting the data, to maintain confidentiality.

Are you willing to participate in this survey? Please click "Yes" only if you are willing to **take the writing tasks seriously** and are **willing to sign up for an online account**.

Yes

No

## B.2.3 Instructions (Exact time arm)

### Instructions

Thanks for agreeing to participate! This page contains more detailed instructions about the content of this survey. **Please read this information carefully; at the bottom, there are three comprehension questions that you must answer correctly to proceed into the study.**

### The Writing Tasks

This survey contains two writing tasks that are designed to resemble real tasks that people in your occupation perform. We would like you to **take them seriously** and **treat them as if they are part of your real job.**

We will pay substantial bonus payments for your performance on these tasks. Specifically, both of your writing pieces will be **graded by another professional in your occupation.** This professional will assign on grades on a scale of 1 to 7. This means you can receive between 2 and 14 points across both essays. For **each point** you earn above 1, **we will pay you \$1.** So, for example, if both of your essays receive a score of 2/7, you will receive a \$2 bonus payment. If both essays receive a score of 6/7, you will receive a \$10 bonus.

You may also be randomly offered **additional bonus payments** for performing a task in a certain way.

### Grading

Your essays will be graded by another professional in your occupation. The grader will **imagine seeing your piece of writing in a real work setting**, and **evaluate it on that basis.** The grader will pay attention to the quality of the content, the quality of your writing, and your originality/creativity.

### Timing

We will require you to spend **exactly 15 minutes** on each task. There will be a timer on each task page; **you will not be allowed to advance before 15 minutes, and the page will advance automatically at 15 minutes.**

To **track whether you are on task**, we will **take a snapshot of your essay box each minute** and **look at how your essay evolves.** If we detect that you are off-task for several minutes at a time, or **pretending to be on-task** by randomly editing a couple of words every now and then, **you will not receive your bonus payment for that essay.**

## Comprehension Questions

Please answer the following comprehension questions. They should be easy if you have read the instructions above. If you fail these questions twice, you will not be able to participate in the study.

Which of the following is true?

These tasks are generic writing tasks originally designed for high schoolers.

The tasks are designed to mimic real tasks for people in my occupation. I should treat them as if they were part of my work.

The tasks are hypothetical new assessments that are being tested on people to see how they perform.

Which of the following correctly describes the bonus payments system for this study?

I will be paid \$1 for each point I receive on each of my writing tasks, up to \$12 total.

If I complete this survey, I will be entered into a lottery for a \$50 bonus payment.

I will receive a \$10 payment if I am in the top 20% of performers on the writing tasks.

The correct answers are “These tasks are designed to mimic real tasks for people in my occupation, and I should treat them as if they were part of my work” and “I will be paid \$1 for each point I receive on each of my writing tasks, up to \$12 total.”

## B.2.4 Instructions (Linear arm)

### Instructions

Thanks for agreeing to participate! This page contains more detailed instructions about the content of this survey. **Please read this information carefully; at the bottom, there are three comprehension questions that you must answer correctly to proceed into the study.**

### The Writing Tasks

This survey contains two writing tasks that are designed to resemble real tasks that people in your occupation perform. We would like you to **take them seriously** and **treat them as if they are part of your real job**.

We will pay substantial bonus payments for your performance on these tasks. Specifically, both of your writing pieces will be **graded by another professional in your occupation**. This professional will assign grades on a scale of 1 to 7. This means you can receive between 2 and 14 points across both essays. For **each point** you earn, **we will pay you \$1**. So, for example, if both of your essays receive a score of 2/7, you will receive a \$4 bonus payment. If both essays receive a score of 6/7, you will receive a \$12 bonus.

You may also be randomly offered **additional bonus payments** for performing a task in a certain way.

### Grading

Your essays will be graded by another professional in your occupation. The grader will **imagine seeing your piece of writing in a real work setting**, and **evaluate it on that basis**. The grader will pay attention to the quality of the content, the quality of your writing, and your originality/creativity. Be sure to meet the word count.

## Comprehension Questions

Please answer the following comprehension questions. They should be easy if you have read the instructions above. If you fail these questions twice, you will not be able to participate in the study.

Which of the following is true?

These tasks are generic writing tasks originally designed for high schoolers.

The tasks are designed to mimic real tasks for people in my occupation. I should treat them as if they were part of my work.

The tasks are hypothetical new assessments that are being tested on people to see how they perform.

Which of the following correctly describes the bonus payments system for this study?

I will be paid \$1 for each point I receive on each of my writing tasks, up to \$14 total.

If I complete this survey, I will be entered into a lottery for a \$50 bonus payment.

I will receive a \$10 payment if I am in the top 20% of performers on the writing tasks.

The correct answers are “These tasks are designed to mimic real tasks for people in my occupation, and I should treat them as if they were part of my work” and “I will be paid \$1 for each point I receive on each of my writing tasks, up to \$14 total.”

### B.2.5 Attention Warning (If wrong answer)

**At least one of your answers is wrong. Please try again.**

If at least one of your answers is wrong again, you won't be allowed to participate in this study.

Participants are then redirected back to the Instructions page and given a second chance to answer. If they answer incorrectly a second time, they are not allowed to continue the survey.

### B.2.6 Demographics

First, we'd like to ask you a few questions about yourself.

Please enter your Prolific ID.

What is your current employment status?

Employed fulltime

Employed part-time

Unemployed and looking for work

Not looking for work

What is your annual salary in your main job? (Or your last main job if you are unemployed. If you are paid hourly, please convert to a rough annual salary by multiplying your hourly rate by the number of hours you typically work).

dollars

How many hours a week do you typically work in your main job? (Or your last main job if you are unemployed).

hours

How many years of experience do you have in your occupation?

years

Which of the following software **are you aware of?** Select all that apply.

Google Drive
Tableau
ChatGPT
Overleaf
Jasper
Grammarly

Which of the following software **have you used?** Select all that apply.

Google Drive
Tableau
ChatGPT
Overleaf
Jasper
Grammarly

--

### B.2.7 Skill Rank

Think about the tasks you perform at work. Consider three different skills:

- Being a good communicator (writing well, speaking/presenting effectively).
- Being creative and original (coming up with lots of new ideas, even if they aren't all the best).
- Reliably coming up with good solutions to problems (even if they aren't very original/creative).

If you had to **rank these skills according to how good you are at them**, how would you rank them?

	1	2	3
Being a good communicator	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reliably coming up with good solutions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being creative and original	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## B.2.8 Task Instructions (Linear and convex arms)

### Task Instructions

On the next page, we will present your **first writing task**.

In each of the tasks, we would like you to keep track of roughly **how much time you spend on the task**. Specifically, we will ask you about:

- How much time you spent **brainstorming**.
- How much time you spent **writing a rough draft**.
- How much time you spend **editing/improving your draft**.

You may find it helpful to use a timer, or just keep a rough track mentally of how much time you are spending on each component.

Remember that your submission will be **graded by another professional in your occupation** on a scale of 1 to 7, and you will receive a **bonus \$1 for each point you receive**. The grader will also determine whether you made a good-faith attempt at completing this task. If the grader decides you did not, you will not receive any bonus payment. So please make an honest attempt.

## B.2.9 Task Instructions Continued (Exact time arm)

# Congratulations!

You have been randomly selected to receive an **additional bonus payment of \$3** if you receive a grade of 6 or 7 on your task.

This means you will earn only \$5 for a grade of 5, but \$9 or \$10 for a grade of 6 or 7, respectively.



## B.2.10 Task

### Task

Imagine you are a senior manager working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to “sit together” in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

-----

The exact time arm then sees the following:

Your response will be graded by a professional in your occupation on a scale of 1-7. The grader will **imagine seeing your piece of writing in their real work setting, and evaluate it on that basis.**

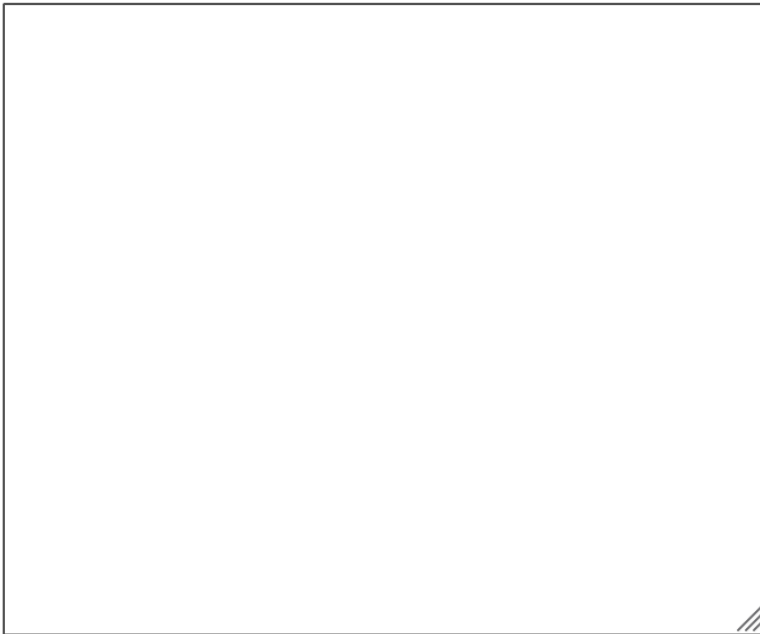
For every point you earn, you will receive a bonus payment of \$1.

**This page will auto-advance after 15 minutes** (see the timer below). You must be on-task for the full 15 minutes.

To **track whether you are on-task**, we will **take a snapshot of your essay box each minute** and look at how your essay evolves. If we detect that you are **off-task** for several minutes at a time, or **pretending to be on-task** (for example, by randomly editing a word or two occasionally), **you will not receive your bonus payment for this essay.**

Please write here. Separate your paragraphs with blank lines.

Current word count: 0



000102

The linear arm then sees the following:

-----  
Your response will be graded by a professional in your occupation on a scale of 1-7. The grader will **imagine seeing your piece of writing in their real work setting, and evaluate it on that basis.**

For every point you earn, you will receive a bonus payment of \$1.

Please write here. Separate your paragraphs with blank lines.

Current word count: 0

The convex arm sees the same as the linear arm, but with the added line “If you receive a score of 6 or 7, you will receive an additional bonus payment of \$3.” prior to “Please write here.”

### B.2.11 Time Spent on Task 1

How many minutes did you spend on the following components?

Your answers should sum up to the total amount of time you spent on the task.

brainstorming what to write

minutes

writing a rough draft

minutes

editing/improving your rough draft

minutes

other (please specify)

minutes

Total

### B.2.12 Efficacy Following Task 1

On a scale of 1 to 10, how much did you **enjoy doing** the previous task?

Not enjoyable at all: 1	2	3	4	5	6	7	8	9	Enjoyable: 10
-------------------------	---	---	---	---	---	---	---	---	---------------

On a scale of 1 to 10, how **skilled/effective did you feel** while doing the previous task?

Not skilled/effective at all: 1	2	3	4	5	6	7	8	9	Very skilled/effective: 10
---------------------------------	---	---	---	---	---	---	---	---	----------------------------

### B.2.13 Realism

On a scale of 1-5, **how realistically does this sample task imitate real tasks that managers do?**

Please give us your honest assessment. Please also ignore whether the specific situation is realistic: we are interested in whether the overall task of writing a mass email to employees to persuade on a sensitive topic is realistic, not whether the specific topic is realistic.

Very Unrealistic	Unrealistic	Neutral	Realistic	Very Realistic
------------------	-------------	---------	-----------	----------------

Have you ever done a task similar to this in your job?

Yes
No

In the past year, how many times have you done a task similar to this?

0
1-3
4-10
more than 10

## B.2.14 Sign-Up (Control)

We have now reached the **online signup** portion of this survey.

Please **sign up for an Overleaf account**. Overleaf is a scientific document editor, similar to Word or Google Docs. You can read about it on Wikipedia [here](#).

Navigate to <https://www.overleaf.com/register> and follow the instructions to create an account. If you already have an Overleaf account, you can proceed immediately to the second step.

Once you are logged into your Overleaf account, **click "New Project"**. Name it whatever you want. Then, for each of the following blocks of text, paste that block in between "`\section{Introduction}`" and "`\end{document}`", and click "Recompile" in the top left. **Take a screenshot of the output, and then attach the output below**. If you run into technical problems, just attach a screenshot of the error you're getting.

If you run into technical difficulties, take a screenshot, attach it, and let us know below.

**Text 1:** "The cow jumped over the moon."

**Text 2:** "She sells seashells by the seashore.."

**Text 3:** "How much wood would a woodchuck chuck if a woodchuck could chuck wood?"

**Text 4:** "Peter Piper picked a peck of pickled peppers"

The screenshots should look roughly like this:



Text 1 Screenshot

Drop files or click here to upload

Text 2 Screenshot

Drop files or click here to upload

Text 3 Screenshot

Drop files or click here to upload

Text 4 Screenshot

Drop files or click here to upload

Let us know of any technical difficulties you encountered:

## B.2.15 Sign-Up (Treatment)

We have now reached the **online signup** portion of this survey.

Please **sign up for a ChatGPT account**. ChatGPT is an AI chatbot produced by OpenAI; you can read about it on Wikipedia [here](#).

Navigate to <https://openai.com/blog/chatgpt/> and click "Try ChatGPT." Follow the instructions to create an account. If you already have a ChatGPT account, you can proceed immediately to the second step.

Second, **paste the following 3 prompts into ChatGPT, take a screenshot of the output, and then attach the output below**. This activity is meant to **show you the kinds of things that ChatGPT is capable of**, so please **pay attention to the prompts and outputs**.

**What if the ChatGPT website is down?** Sometimes, the ChatGPT signup website is overloaded and signing up for an account is not possible. **If this is the case**, please take a screenshot of the signup website and attach it for each of the three prompts. Also, let us know in the "technical difficulties" box at the bottom of this page. We will verify whether ChatGPT was in fact down when you submitted your response.

**Prompt 1:** *Paste the following into ChatGPT's prompt bar, then hit enter:*

### **Task**

You are working for a local community center in the year 2030. The center is submitting a grant application to the state government in an attempt to secure funding for a proposed virtual-reality community gym. The gym will consist of a bunch of exercise machines augmented with virtual-reality headsets (for example, a treadmill with a headset that allows the user to feel like they are running in any location they want—a beach in Hawaii, an Antarctic glacier, or the surface of Mars).

Your job is to write a 400-word cover letter for this grant application. Your letter should briefly summarize the case for funding this project and should convey the community center's enthusiasm about the project. Please note the following:

- The community center wants you to explain why a community-provided gym is necessary even though there are already private gyms in the area.
- The community center wants you to clearly communicate why a virtual-reality gym is preferable to a normal gym.

**Prompt 2:** *First, enter:*

"I have to write a paragraph discussing whether it is better to invest in the stock market or put money in a savings account. Please write this paragraph for me. Your paragraph should discuss the role of age, risk tolerance, and macroeconomic conditions in this decision."

*After ChatGPT has produced output for that prompt, tell it:* "Expand your discussion to 3 paragraphs, and add more detail."

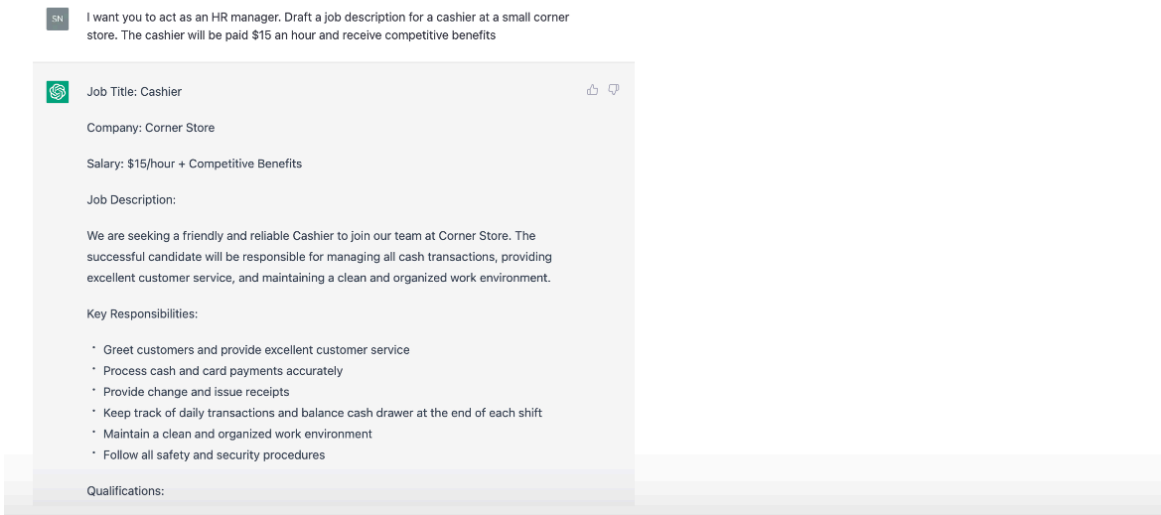
*After ChatGPT has produced output for that prompt, tell it:* "Rewrite the first paragraph you produced, in simpler language."

*(Take a single screenshot of the end of this conversation; it's okay if your screenshot doesn't cover it all).*

**Prompt 3:** *ChatGPT may deliver different responses depending on how the prompt is worded. Enter:* "Imagine that you are a financial advisor. Inform a client whether it is better for them to invest in the stock market or put money in a savings account. You should include information about age, risk tolerance, and macroeconomic conditions. Limit your response to 200 words."

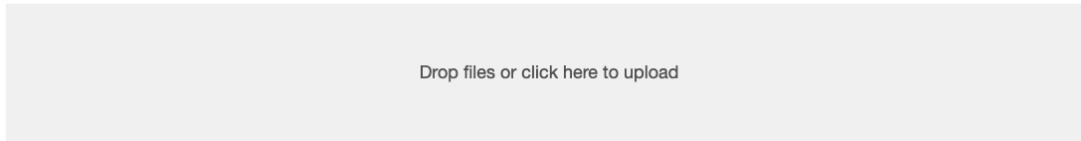
**Prompt 4:** *Enter:* "I want you to summarize the following [Paste in the first 3 paragraphs of the Wikipedia page for World War II]"

The screenshots should look roughly like this:

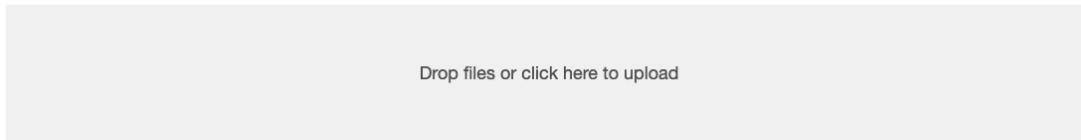




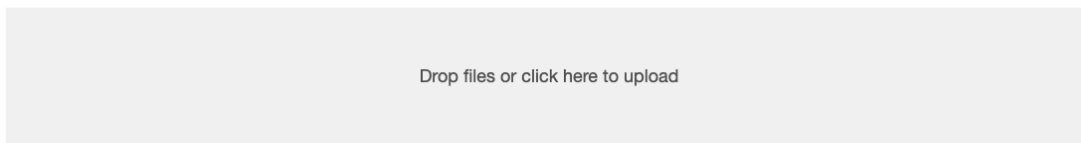
Prompt 1 Screenshot



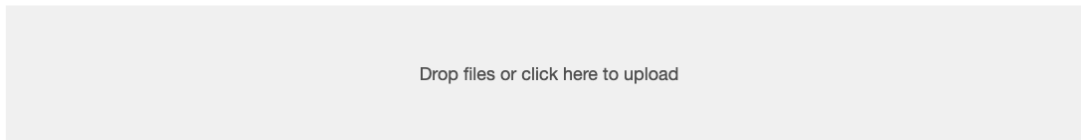
Prompt 2 Screenshot



Prompt 3 Screenshot



Prompt 4 Screenshot



Let us know of any technical difficulties you encountered:

## B.2.16 Further Instructions (Treatment Only)

ChatGPT can also find resources, summarize text, edit text, and do other writing-based tasks. However, factual responses that ChatGPT provides may not always be accurate.

**Feel free to take some time to play around with ChatGPT, as it may be helpful for the next question.** Note that you can ask ChatGPT to revise or expand on its output, or ask it to include specific things in its output.

## B.2.17 Task 2 Begins

Thank you. We will now proceed to the second task.

### B.2.18 Task 2 Additional Instructions (Treatment Only)

You may use ChatGPT to help you complete this task, if you find it useful. There is **no bonus or penalty for using ChatGPT**; you will be graded and paid purely on your output.

As you saw during signup, there are many ways to use ChatGPT. For example:

- You could paste in the full task prompt, then edit what ChatGPT has written.
- You could paste in the full task prompt, then ask ChatGPT to make various changes to what it has written.
- You could brainstorm by asking ChatGPT questions or asking it for ideas.
- You could ask ChatGPT to edit/simplify things you have written.

If you do end up using ChatGPT, please **keep the tab open** after finishing the task.

### B.2.19 Task 2

#### Task

Imagine you are a senior manager working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to "sit together" in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

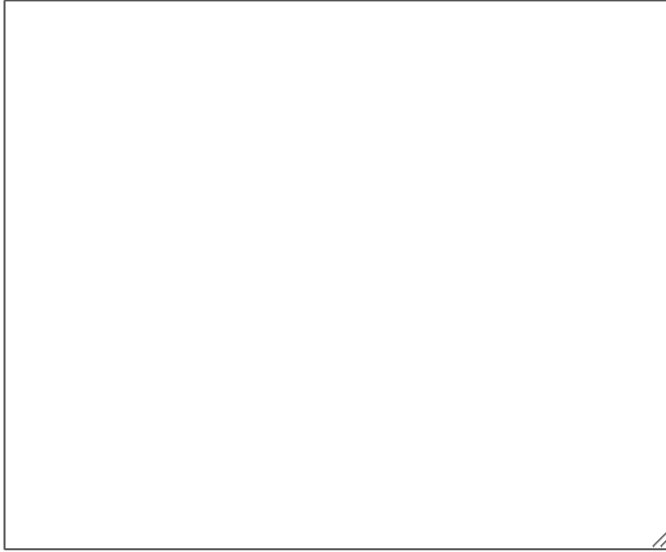
-----

Your response will be graded by a professional in your occupation on a scale of 1-7. The grader will **imagine seeing your piece of writing in their real work setting, and evaluate it on that basis.**

For every point you earn, you will receive a bonus payment of \$1.

Please write here. Separate your paragraphs with blank lines.

Current word count: 0



Identical to Task 1, there are minor deviations to this page dependent on the incentive arm.

### B.2.20 Time Spent on Task 2

How many minutes did you spend on the following components?

Your answers should sum up to the total amount of time you spent on the task.

brainstorming what to write

minutes

writing a rough draft

minutes

editing/improving your rough draft

minutes

other (please specify)

minutes

Total

### B.2.21 Efficacy on Task 2

On a scale of 1 to 10, how much did you **enjoy doing** the previous task?

Not enjoyable at all: 1	2	3	4	5	6	7	8	9	Enjoyable: 10
-------------------------	---	---	---	---	---	---	---	---	---------------

On a scale of 1 to 10, how **skilled/effective did you feel** while doing the previous task?

Not skilled/effective at all: 1	2	3	4	5	6	7	8	9	Very skilled/effective: 10
---------------------------------	---	---	---	---	---	---	---	---	----------------------------

### B.2.22 Automation

On a scale of 1 to 10, how worried are you about **workers in your occupation** being replaced by AI?

Not worried at all: 1	2	3	4	5	6	7	8	9	Very worried: 10
-----------------------	---	---	---	---	---	---	---	---	------------------

On a scale of 1 to 10, how **optimistic are you** that AI may make workers in your occupation more productive?

Not optimistic at all: 1	2	3	4	5	6	7	8	9	Very optimistic: 10
--------------------------	---	---	---	---	---	---	---	---	---------------------

On a scale of 1 to 10, **how do you feel** about the impacts of future advances in AI?

Very pessimistic: 1	2	3	4	5	6	7	8	9	Very optimistic: 10
---------------------	---	---	---	---	---	---	---	---	---------------------

### **B.2.23 ChatGPT Use**

Did you use ChatGPT on the second writing task? (Please answer truthfully: there is no bonus or penalty for having used it.)

Yes

No

Don't know what ChatGPT is

### **B.2.24 Why Not ChatGPT (If treatment and did not use ChatGPT)**

Why did you not use ChatGPT?

## B.2.25 Why ChatGPT (If used ChatGPT)

How useful was ChatGPT on the task?

Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful
-------------------	-----------------	-------------------	-------------	------------------

Please list all prompts you entered into ChatGPT. (We are curious about how people use it).

If you closed the ChatGPT tab, you can return to <https://chat.openai.com/>. Your chat will be saved.

How did you use ChatGPT? Please tick all that apply.

Used it to brainstorm ideas
Used it to summarize text
Asked it to write a draft, then directly submitted the draft
Asked it to write a draft, then edited the draft
Asked it to write a draft, then asked it to make changes
Used it to edit/simplify something I wrote
Other (please specify)
<input type="text"/>

What aspects of the task (if any) did you feel **ChatGPT was good at** doing?

What aspects of the task (if any) did you feel **ChatGPT was bad at** doing?

Did you use ChatGPT on the **first task** (before the signup portion)? (Again, please answer truthfully; there is no bonus or penalty).

Yes
No

### B.2.26 Willingness to Pay (If used ChatGPT)

Imagine that ChatGPT was not free, and you could **pay a monthly subscription fee** to be able to **use ChatGPT in your job**. What is the **highest monthly fee you would be willing to pay?**

(Note: we are not affiliated with the company that makes ChatGPT and this information will not be used in any pricing decisions. We are just curious how much you value it).

dollars

How often would you use ChatGPT in your job, if you had access to it?

Every day

A couple of times a week

A couple of times a month

Almost never

### B.2.27 Other Tool Use

Did you use any software other than ChatGPT on either of the writing tasks? (Please answer truthfully: there is no bonus or penalty for having used these.)

AI writing tool (ie. Jasper, Ryter, etc.)

Google Search

Grammarly

Overleaf

Other (please state)

### B.2.28 Retry Intro (If retry arm — treatment + linear/convex arms only)

Congratulations! You are eligible for one final opportunity before the survey ends.

On the next page, we will show you **the first task you completed**. You may either **directly copy-paste and resubmit** your original response. Or, **if you want, you may edit or replace your submission**, and you **may use ChatGPT to assist you if it is helpful**.

Remember that there are no bonuses or penalties for using ChatGPT; your essay will be graded by someone in your occupation who does not know how the essay was produced. So you should edit or replace your essay **only if you expect this will increase its grade**.

We will require you to spend 2 minutes on the next page, but you should feel free to not edit your submission at all or to take as long as you want.



**B.2.29 Retry Instructions (If retry arm, linear arm)**

Here is what you submitted for task 1:

---

---

You may directly **copy-paste and resubmit** your original response. Or, you may **edit or replace your submission**, including using ChatGPT, if you believe it will be helpful.

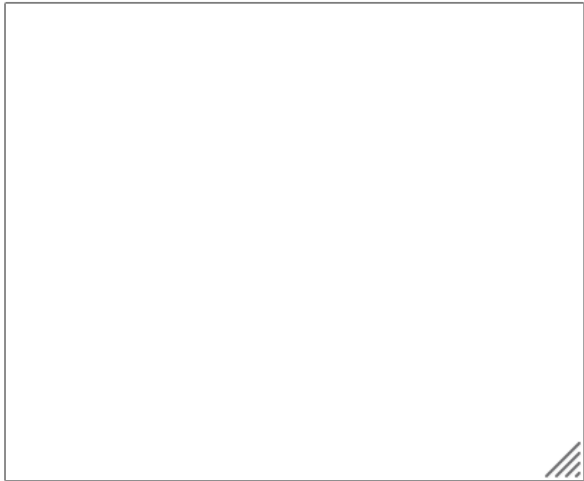
Your response will be graded on a scale of 1-7 by a professional in your occupation. Your grade will not in any way depend on your use of ChatGPT.

For every point you earn, you will receive a bonus payment of \$1.

Please write your final response here. Separate your paragraphs with blank lines.

**The "next page" button will appear after 2 minutes.**

Current word count: 0



The convex arm sees the same as the linear arm, but with the added line “If you receive a score of 6 or 7, you will receive an additional bonus payment of \$3.” prior to “Please write here.”

### B.2.30 Action (If retry arm)

Did you...

Copy-paste and submit your original response

Replace your original response completely

Edit your original response

### B.2.31 ChatGPT Replace (If replaced)

Did you **use ChatGPT** to replace your original response?

Yes, I took ChatGPT output and **did not** edit ChatGPT output

Yes, I took ChatGPT output and **made my own edits** to ChatGPT output

No

### B.2.32 ChatGPT Edit (If edited)

Did you **use ChatGPT** to edit your response? (Select all that apply)

Yes, I put my response into ChatGPT and asked ChatGPT to edit it

Yes, I received output from ChatGPT and combined it with my existing response

No

**B.2.33 Why used ChatGPT (If used ChatGPT to replace or edit)**

Why did you use ChatGPT to replace or edit your response?

**B.2.34 Why not use ChatGPT (If did not use ChatGPT to replace or edit)**

Why did you not use ChatGPT to replace or edit your response?

**B.2.35 Why not edit or replace (If did not replace or edit)**

Why did you not edit or replace your response?

**B.2.36 Completion**

Thank you for completing our survey! We will approve the base payment ASAP. It might take slightly longer for your essay to be graded but we promise to send that part of the bonus payment within the next week. **Is there any feedback you would like to give us about the survey and tasks?**

When you are finished, click to the next page to be redirected to Prolific with a completion code.

## B.3 Grading Screener

### B.3.1 Welcome

Welcome to this short 1-minute screening survey. We will ask you a few questions about yourself. Depending on your responses, we may ask you if you want to be invited to a 1.5-hour followup survey within the next week.

First, what is your Prolific ID?



### B.3.2 Occupation

Which of the following, if any, **best describes your occupation?** (Please tick "Other" if none of them describe your occupation).

Business Strategist	Lawyer
Grant/Proposal Writer	Journalist
Teacher	Copywriter
Manager	Data Analyst or Data Scientist
Management Consultant	HR Professional
Freelance Writer	Medical Worker
Policy Analyst	Public Relations Professional
Content Creator or Blogger	Administrator
Marketer	Engineer/Software Developer
Technical Writer	Other (please specify)

Only those who are in our occupations of interest proceed.

### B.3.3 Demographics

What is your annual salary in your main job? (If you are paid hourly, please convert to a rough annual salary by multiplying your hourly rate by the number of hours you typically work).

< \$20,000
\$20,000-39,999
\$40,000-59,999
\$60,000-79,999
\$80,000-99,999
\$100,000-119,999
\$120,000+

How many hours a week do you typically work in your main job?

How many years of experience do you have in your occupation?

What is your highest level of completed education?

High school
Some college
College (BA)
Post-graduate (MA, PhD)

The completed education question is only included for those who were not education pre-screened on Prolific.

### B.3.4 Task Experience

Imagine the following task: Write a company/department-wide email on a sensitive topic (such as adoption of a new system/technology or an organizational change) to employees.

Have you ever done a task similar to this in your job?

Yes

No

In the past year, how many times have you done a task similar to this?

0

1-3

4-10

more than 10

Participants are invited for certain if they work more than 20 hours a week, have an annual salary of at least \$60,000, and have done the task at least once in the past year.

Participants are may be invited if they are within the occupation but do not fulfill the criteria.

## B.3.5 Consent (Certain Invite)

### Congratulations! You are eligible for participation in the followup survey.

Below, we **describe the followup survey** and then ask if you would be interested in participating.

#### Followup Survey Details

The followup survey will take **about 1.5 hours** and has two parts. We will ask you to read **2 prompts for ~400 word writing tasks** designed to mimic tasks that people in your occupation do in their jobs. We will then ask you to **grade responses to these prompts** on a 1-7 scale on their overall quality, as well as on additional criteria.

We are looking for respondents who are willing to **take grading seriously and do their best to produce good output**. Your grades will be compared to those of experienced professionals in this occupation, and we will **pay bonuses for diligent grading**, as we will explain below.

#### Payment for Followup Survey

We will pay a base of \$16 for completion of the survey. In addition, we will compare some of your ratings to those of experienced professionals in this occupation. You will receive bonus payments depending on how close you are to the average experienced professional grade, **up to \$8 of bonus payments (\$24 total)**.

#### Additional Information

The followup survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks.

Would you like us to invite you to participate in the followup survey, sometime over the next week?

Yes

No



### B.3.6 Consent (Maybe Invite)

## You may be eligible for participation in the followup survey.

Below, we **describe the followup survey** and then ask if you would be interested in participating.

### Followup Survey Details

The followup survey will take **about 1.5 hours** and has two parts. We will ask you to read **2 prompts for ~400 word writing tasks** designed to mimic tasks that people in your occupation do in their jobs. We will then ask you to **grade responses to these prompts** on a 1-7 scale on their overall quality, as well as on additional criteria.

We are looking for respondents who are willing to **take grading seriously and do their best to produce good output**. Your grades will be compared to those of experienced professionals in this occupation, and we will **pay bonuses for diligent grading**, as we will explain below.

### Payment for Followup Survey

We will pay a base of \$16 for completion of the survey. In addition, we will compare some of your ratings to those of experienced professionals in this occupation. You will receive bonus payments depending on how close you are to the average experienced professional grade, **up to \$8 of bonus payments (\$24 total)**.

### Additional Information

The followup survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks.

Would you like us to invite you to participate in the followup survey, sometime over the next week?

Yes

No

### B.3.7 End (Certain Invite)

Thanks for your interest!

**The followup survey will be active on Prolific at various times over the next week.** It will pop up in your Prolific feed, as we will custom-invite everyone who consents in this screening survey. The title of the study will be something like "Grade responses to writing tasks (Big Bonus)."

Please click onto the next page to be redirected back to Prolific with your completion code. You will be paid \$0.20 for your participation in this screening survey.

### **B.3.8 End (Maybe Invite)**

Thanks for your interest!

If you are invited, the survey will pop up in your Prolific feed. The title of the study will be something like "Grade responses to writing tasks (Big Bonus)."

Please click onto the next page to be redirected back to Prolific with your completion code. You will be paid \$0.20 for your participation in this screening survey.

## **B.4 Grading Survey**

This survey is for managers. For other occupations, all instances of “manager” are changed to that occupation. There are other occupation-specific deviations that will be noted below.

## B.4.1 Consent

Thank you for helping us with our research study!

What is your Prolific ID?

### Instructions and Consent

**Please read these instructions carefully; we will ask comprehension questions afterwards.**

This survey involves **grading pieces of writing** written in response to a **task that we assigned**. The task is meant to **imitate real tasks that managers do in their jobs**.

We will first ask you a few short questions about yourself.

Next, we will show you the task instructions. **Please read the task instructions carefully**; we will ask you comprehension questions afterwards.

Lastly, we will show you 9 completed writing pieces. We will ask you to assign each writing piece an **overall grade between 1 and 7**. To decide on this grade, you should **imagine seeing this piece of writing in a real-life work setting**, and think about **how good it would be as a piece of work**.

You should feel free to use the full range, from 1 to 7. You should assign roughly as many grades below 4 as above 4. We will give you some examples of graded pieces, for guidance.

We will also ask you to assign 3 additional grades: for writing quality, content quality, and creativity/originality. We will explain these more on a later page.

**It is very important to us to get accurate grades for these tasks**, for our research study and because the grades you assign will be used to pay the people who completed the tasks. Please **read and think carefully about each piece of writing**. You should aim to spend **4 minutes per writing piece**.

To **reward you for grading carefully**, we are **offering bonus payments up to \$8**. Specifically, we will **compare your grades to the grades of other graders seeing the same tasks**, and may also **randomly choose 2 of your grades to double-check ourselves**. The closer your grades match the grades of other graders, as well as the more we agree with the grades we double-check, the higher your bonus payment will be.

Overall, you will be paid a base rate of \$16, plus bonus payments up to \$8 (**total \$24**). We expect the average respondent who makes a committed effort to read the tasks to spend an 1.5 hours and earn \$22.

#### Additional Information

This survey is part of a research study conducted by Shakked Noy and Whitney Zhang at the Massachusetts Institute of Technology (MIT) Economics Department. The research aims to understand the determinants of people's productivity on writing tasks. Your participation in this study is completely voluntary and you can choose to withdraw at any time without any penalty or consequence. If you volunteer to participate, we will assign you questions and writing tasks as described above, and ask you to sign up for an online account. We do not anticipate any risks or discomforts in the survey. The research may involve risks that are currently unforeseeable. We anticipate the study will provide benefits to society by enabling a better understanding of the determinants of productivity.

If you have any concerns or comments about this study, you can contact the researchers at [snoy@mit.edu](mailto:snoy@mit.edu) or [zhangww@mit.edu](mailto:zhangww@mit.edu). You can contact the MIT Committee on the Use of Humans as Experimental Subjects at [couhes@mit.edu](mailto:couhes@mit.edu).

Data from this survey may be made public. We will remove Prolific IDs and all identifying information before posting the data, to maintain confidentiality.

Are you willing to participate in the survey?

Yes

No

## B.4.2 Attention

Please read these comprehension questions carefully and answer them. If you fail twice to answer them both correctly, you will not be eligible to participate in the survey.

Which of the following is correct?

The grading scale is between 1 and 7. My grades should mostly be in the 4-7 range, with 1-3 reserved for unusually bad essays.

The grading scale is between 1 and 7. I should assign almost all essays to be in the 3-5 range.

The grading scale is between 1 and 7. I should feel free to use the full range and should assign about as many grades above 4 as below 4.

Which of the following is correct?

These tasks are meant to mimic real tasks done by managers. I should assign grades based on how I would feel seeing this piece of writing in a real work setting.

This is an educational experiment focused on writing quality. I should just pay attention to the grammar and vocabulary of the pieces of writing.



### The correct answers are

Please read these comprehension questions carefully and answer them. If you fail twice to answer them both correctly, you will not be eligible to participate in the survey.

Which of the following is correct?

The grading scale is between 1 and 7. My grades should mostly be in the 4-7 range, with 1-3 reserved for unusually bad essays.

The grading scale is between 1 and 7. I should assign almost all essays to be in the 3-5 range.

The grading scale is between 1 and 7. I should feel free to use the full range and should assign about as many grades above 4 as below 4.

Which of the following is correct?

These tasks are meant to mimic real tasks done by managers. I should assign grades based on how I would feel seeing this piece of writing in a real work setting.

This is an educational experiment focused on writing quality. I should just pay attention to the grammar and vocabulary of the pieces of writing.



## B.4.3 Retry (Incorrect Attention)

**At least one of your answers is wrong. Please try again.**

If at least one of your answers is wrong again, you won't be allowed to participate in this study.



## B.4.4 Task 1 Prompt

Congratulations on correctly answering the comprehension questions! We will now proceed to the grading. First, **read the prompt that the responses answer:**

### Task

Imagine you are a senior manager working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to "sit together" in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

Which of the following are reasons employees currently do not use the VR headsets? (Select all that apply)

Price of the headsets

Privacy preferences

Uncomfortable headsets

Interferes with doing work tasks

The email should do which of the following? (Select all that apply)

Encourage employees to begin using headsets

Explain why the company thinks it's important for employees to be using the virtual office space

Be clear about punishments for employees who do not apply with the virtual office policy

Address employee concerns



Task comprehension questions for other occupations are available in section B.6.

## B.4.5 Task 1 Experience

Have you ever done a task similar to the prompt in your job?

Yes

No

If yes, how many times have you done a task similar to the prompt in the past year?

0

1-3

4-10

more than 10



## B.4.6 Grading Instructions

For each response, you will first provide an overall grade based on **the quality of the response in a real-life work setting**.

Additionally, you will be asked to grade on the following criteria:

Writing quality

- Is the text well-structured and easy to follow?
- Are the paragraphs and sentences logically organized?
- Is the language clear and easy to understand?
- Are technical terms and jargon explained or avoided?
- Is the text free of errors in grammar, punctuation, and spelling?

Content quality

- Does the text effectively answer the prompt provided?
- Is the information accurate and comprehensive?
- Would the intended audience be receptive to the text?
- Is the text engaging?

Originality

- Are the ideas presented original and creative

On the next two pages, you will read two example graded responses.



## B.4.7 Task 1 Example 1

This is an example of a 2 response  
Hello Team,

Good afternoon. There's been some important changes in the past which have not picked up the way we expected. While we think we may know why that is, we still are working to have a concrete solution towards that problem. There are some things we expect from you as a team and that is what I would like to speak with you today about. While we value all your unique skills and abilities, we feel the environment has not been as cohesive of lately.

With the introduction of the virtual reality office we expected a boost in motivation and effort due to the ability to be able to work at home while maintaining that office feel. We see that did not turn out and have been trying to find a way to motivate people to try using the virtual workspace again. We understand the topic about being uncomfortable and we are looking into it immediately for similar models that may be more comfortable. However, we hope that is not the sole factor for the low attendance in the virtual workspace.

My goal is to obtain from each of you what you would like to see change in the workspace as well as what kept you from attending. We understand it's not attractive to work in an empty office space and that's clearly what we are trying to focus on changing. It's hard to incentivize without direct input from the team. We are hoping for some creative ideas that will help us be able to bring the teams a bit closer with communication and effort.

We know the changes are also difficult to adapt to which is why we want to do our best to take criticism to try and improve the current state of things. While we are going to try our best we are going to ask the same from you. Going forward this week we would like for you to make a top 5 list of things you appreciated the most and did not appreciate from the work space. This will allow us to evaluate the state in which the workspace is in regards to being organic and feasible. Please make sure you are honest and critical in your lists, simple aspects or factors will be overlooked, especially if they are too common and many people are selecting them.

Together with your efforts I believe the virtual workspace will be a fun and interesting place. Look forward to changes in the environment, the headset, and our policies. We're going to try to make the best of the situation and will not be able to do so without your help. I very much appreciate you all, the HR team is thrilled to work with each and every one of you and would very much like to see this project work out. Let's make this upcoming week a great one by changing the pace and focusing on our new top priority on our agenda. Thank you all for your time, take care.

Kind regards,

Notice how the writing has **poor flow**, is **often vague**, does **not address privacy concerns**, and **does not sufficiently explain why the company thinks it is important for employees to be using the virtual office space**.



Task examples for other occupations are available in section B.6.

## B.4.8 Task 1 Example 2

This is an example of a 6 response

Hello,

Last month you should have received a virtual reality headset from the company IT team. If you have not received your package, please reply to this email immediately so that I can track down what may have happened with your parcel.

If you have received it, I'd love some feedback as to how you're enjoying the unit and the virtual office space? It was designed to help our company stay connected during a time of remote-first work schedules.

Studies show that many people who work from home can feel isolated and disconnected from their peers and lacking in the engagement of company culture. To prevent that feeling of disconnectedness, the company invested in the virtual office space to preserve the teamwork and well-being of our employees.

The headsets should be used regularly to foster a sense of connection and for team building exercises. It's important to remain in regular communication with your coworkers and managers to stay abreast of the latest developments, company news, DE&I training and to feel a part of a bigger plan.

I've noticed a fairly limited use of the system to date and have been made aware that the device isn't very comfortable. I have reached out to the developer, and was pleased that there are adjustments that can be made to create a more optimal experience.

First, the band on the headset is fully adjustable to fit different headshapes and accommodate eyeglasses. Inside the packaging there is also optional foam padding can be a great help in making the headset feel more natural during use and alleviate any pressure points from wearing the unit or from the headband. If you need more information on setting up your unit, please refer to the video tutorial link below.

As you may know, each team is expected to meet in the virtual office space at least twice a week to collaborate and engage with one another. As a people-first organization, we feel it's important to keep up with the well-being of our employees and this is one way we can accomplish that goal. Staying up to date with the pulse of our crew is one way we strive to do better and keep putting our people first.

Feel free to collaborate with your team leader to pick dates and times for your virtual meeting. Get creative and have fun with the experience. Pick a lunch time and order DoorDash (on the company!) to enjoy a working lunch with your team.

If I can help in any way, please let me know. I'll touch base at the end of the month to see if the above was helpful or if you have any further concerns. My door is always open as well.

Notice how the response **explains the company's motivation for the virtual office space**, **comprehensively addresses employee concerns**, displays **empathy and concern for employees**, and has **good flow and word usage**. It also displays **thoughtfulness and originality** with the first sentence and by encouraging employees to enjoy a working lunch.





## B.4.9 Task 1 Grading Page

Please spend **4 minutes** (timer below) carefully reading and grading the response on this page. Pay attention to the content of the response and imagine seeing it in a real work setting.

If your grades are validated by other graders, you will **receive bonus payments of up to \$8** and **may be re-invited for more surveys**.

000005

Review Prompt

### Response

Dear colleagues,

I'm writing you today to talk to you about the future. Your future, the company, and your future at the company. The company has given you a piece of that future. Whether we like it or not virtual reality is the future and it's for the better. No longer do we have to commute to work, wear uncomfortable clothes or wake up early to rush to work all because of VR.

I know that the headset can be uncomfortable but I promise you'll get used to it. Thirty days is all you need to use it before it becomes a habit. I also know that most of us would rather work alone in the privacy of our homes but we're social creatures that need human interaction.

I don't know about all of you but I would rather put a headset on than getting up three hours early and make sure that I was presentable enough to come into the office just to stare at a computer screen for the whole day. Forgive but that's nonsense I would rather put on a headset than do that. Plus you are still at home which that during lunchbreak you can eat anything you want instead of having to pack a lunch or visit a store that you might not even like.

Now let's talk about the virtual office space. In the real world, you can't decorate your space with amazing things like stars for instance but in VR you can (as long as it's within company policy of course) there is so much you can do with your office space that you can't do in the real world.

I know that the future can be sometimes unknown but I

Did the respondent put in good faith effort?

Yes	No
-----	----

**Remember: If your grades are validated by other graders, you will receive bonus payments of up to \$8 and may be re-invited for more surveys.**

Please **imagine seeing this piece of writing in a real-life work setting** and give it an **overall grade**. Use the full range and assign roughly as many grades below 4 as above 4.

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Justify your grade above in 2 sentences.

Rate the response according to the additional criteria.

	1	2	3	4	5	6	7
Writing Quality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Content Quality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Originality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



The grading page is repeated for each prompt.  
All task 1 pages are repeated for task 2.

## B.4.10 Finish

Thank you! Your completion code is CG3D55ZJ

Do you have any feedback on this task (pay structure, instructions, format, etc.)?

## B.5 14-Day Followup Survey

### B.5.1 Welcome

# Welcome!

In this survey, we will ask you a few questions about your job. Your responses are fully anonymous; we may post the data from this survey publicly, but will remove Prolific IDs and all other personal information before doing so.

Since we may post the data publicly, in the text-box responses, please do not state any identifying or confidential information about yourself or your company (e.g., company name).

What is your Prolific ID?

### B.5.2 Satisfy

On a scale of 1 to 10, how satisfied are you with your job?

1: Extremely dissatisfied	2	3	4	5	6	7	8	9	10: Extremely satisfied
---------------------------------	---	---	---	---	---	---	---	---	-------------------------------

### B.5.3 Writing

Can you briefly describe the kinds of **writing tasks** you typically do in your job?

### B.5.4 ChatGPT

Have you used ChatGPT (or a similar chatbot) in your job in the last week?

Yes

No

### B.5.5 Why Not Use ChatGPT Followup (If not use)

What is the main reason you do not use ChatGPT in your job?

Not useful for my job

Workplace doesn't allow it

Didn't think of that/don't know ChatGPT/don't have an account

ChatGPT is usually down during the workday

Other (please specify)

### B.5.6 Why Not Use ChatGPT Followup (If not use)

Why is ChatGPT not useful for the writing tasks you do in your job?

### B.5.7 How Use ChatGPT (If use)

Can you describe, in a few sentences, what kinds of tasks you use ChatGPT (or a related chatbot) for and how you use it?

How useful is ChatGPT for your job?

Not at all useful

Slightly useful

Moderately useful

Very useful

Extremely useful

Imagine that ChatGPT was not free, and you could **pay a monthly subscription fee** to be able to **use ChatGPT in your job**. What is the **highest monthly fee you would be willing to pay (in dollars)**?

(Note: we are not affiliated with the company that makes ChatGPT and this information will not be used in any pricing decisions. We are just curious how much you value it).

## B.6 Tasks

The following are tasks, task comprehension questions, and task examples, sorted by occupation.

## B.6.1 Manager Task A

### Task

Imagine you are a senior manager working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to "sit together" in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

Which of the following are reasons employees currently do not use the VR headsets? (Select all that apply)

Price of the headsets

Privacy preferences

Uncomfortable headsets

Interferes with doing work tasks

The email should do which of the following? (Select all that apply)

Encourage employees to begin using headsets

Explain why the company thinks it's important for employees to be using the virtual office space

Be clear about punishments for employees who do not apply with the virtual office policy

Address employee concerns

Correct answers to the first question are "Privacy preferences" and "Uncomfortable headsets".

Correct answers to the second question are "Encourage employees to begin using headsets," "Explain why the company thinks it's important for employees to be using the virtual office space," and "Address employee concerns."

This is an example of a 2 response

Hello Team,

Good afternoon. There's been some important changes in the past which have not picked up the way we expected. While we think we may know why that is, we still are working to have a concrete solution towards that problem. There are some things we expect from you as a team and that is what I would like to speak with you today about. While we value all your unique skills and abilities, we feel the environment has not been as cohesive of lately.

With the introduction of the virtual reality office we expected a boost in motivation and effort due to the ability to be able to work at home while maintaining that office feel. We see that did not turn out and have been trying to find a way to motivate people to try using the virtual workspace again. We understand the topic about being uncomfortable and we are looking into it immediately for similar models that may be more comfortable. However, we hope that is not the sole factor for the low attendance in the virtual workspace.

My goal is to obtain from each of you what you would like to see change in the workspace as well as what kept you from attending. We understand it's not attractive to work in an empty office space and that's clearly what we are trying to focus on changing. It's hard to incentivize without direct input from the team. We are hoping for some creative ideas that will help us be able to bring the teams a bit closer with communication and effort.

We know the changes are also difficult to adapt to which is why we want to do our best to take criticism to try and improve the current state of things. While we are going to try our best we are going to ask the same from you. Going forward this week we would like for you to make a top 5 list of things you appreciated the most and did not appreciate from the work space. This will allow us to evaluate the state in which the workspace is in regards to being organic and feasible. Please make sure you are honest and critical in your lists, simple aspects or factors will be overlooked, especially if they are too common and many people are selecting them.

Together with your efforts I believe the virtual workspace will be a fun and interesting place. Look forward to changes in the environment, the headset, and our policies. We're going to try to make the best of the situation and will not be able to do so without your help. I very much appreciate you all, the HR team is thrilled to work with each and every one of you and would very much like to see this project work out. Let's make this upcoming week a great one by changing the pace and focusing on our new top priority on our agenda. Thank you all for your time, take care.

Kind regards,

Notice how the writing has **poor flow**, is **often vague**, does **not address privacy concerns**, and **does not sufficiently explain why the company thinks it is important for employees to be using the virtual office space**.



This is an example of a 6 response

Hello,

Last month you should have received a virtual reality headset from the company IT team. If you have not received your package, please reply to this email immediately so that I can track down what may have happened with your parcel.

If you have received it, I'd love some feedback as to how you're enjoying the unit and the virtual office space? It was designed to help our company stay connected during a time of remote-first work schedules.

Studies show that many people who work from home can feel isolated and disconnected from their peers and lacking in the engagement of company culture. To prevent that feeling of disconnectedness, the company invested in the virtual office space to preserve the teamwork and well-being of our employees.

The headsets should be used regularly to foster a sense of connection and for team building exercises. It's important to remain in regular communication with your coworkers and managers to stay abreast of the latest developments, company news, DE&I training and to feel a part of a bigger plan.

I've noticed a fairly limited use of the system to date and have been made aware that the device isn't very comfortable. I have reached out to the developer, and was pleased that there are adjustments that can be made to create a more optimal experience.

First, the band on the headset is fully adjustable to fit different headshapes and accommodate eyeglasses. Inside the packaging there is also optional foam padding can be a great help in making the headset feel more natural during use and alleviate any pressure points from wearing the unit or from the headband. If you need more information on setting up your unit, please refer to the video tutorial link below.

As you may know, each team is expected to meet in the virtual office space at least twice a week to collaborate and engage with one another. As a people-first organization, we feel it's important to keep up with the well-being of our employees and this is one way we can accomplish that goal. Staying up to date with the pulse of our crew is one way we strive to do better and keep putting our people first.

Feel free to collaborate with your team leader to pick dates and times for your virtual meeting. Get creative and have fun with the experience. Pick a lunch time and order DoorDash (on the company!) to enjoy a working lunch with your team.

If I can help in any way, please let me know. I'll touch base at the end of the month to see if the above was helpful or if you have any further concerns. My door is always open as well.

Notice how the response **explains the company's motivation for the virtual office space**, **comprehensively addresses employee concerns**, displays **empathy and concern for employees**, and has **good flow and word usage**. It also displays **thoughtfulness and originality** with the first sentence and by encouraging employees to enjoy a working lunch.



## B.6.2 Manager Task B

The second prompt is as follows:

### Task

Imagine that you are a senior manager working for a company, AccureCo, in 2025. A new CEO, Sarah Davidson, recently took over the company and is planning to reorganize its internal hierarchy.

Specifically, the CEO believes that the company is currently too vertically hierarchical: there are too many different tiers of authority (the company has 6 levels of managers). The CEO believes this harms company performance by slowing down decision-making (since decisions have to trickle through multiple levels of authority), and by creating jealousy and disharmony among the company's employees, since none of them feel equal to each other. The CEO wants to announce a switch to a flatter structure with only 3 tiers of managers. This will require effectively demoting some of the intermediate managers to lower positions with less authority.

The CEO has instructed you to draft a company-wide email that does the following:

- Briefly announces that the company will be shifting to a flatter organizational structure, without giving concrete details (say that the details are still being worked out).
- Explains the rationale for the change and the benefits this will bring to most employees.
- Tactfully addresses the concerns of employees who might now be facing demotion.

The email should be about 400 words long.

Why does the CEO want to reduce the number of tiers of authority? (Select all that apply)

Cuts company costs

Slows down decision-making

Prevents the company from having to layoff workers

Creates jealousy and disharmony among the company's employees

The email should do which of the following? (Select all that apply)

Provide concrete details about the flatter organizational structure

Explain the rationale for the change

Address the concerns of employees who might now be facing demotion

Announce that the company will be shifting to a flatter organizational structure

Correct answers to the first question are “Slows down decision-making” and “Creates jealousy and disharmony among the company’s employees”.

Correct answers to the second question are “Explain the rationale for the change,” “Address the concerns of employees who might now be facing demotion,” and “Announce that the company will be shifting to a flatter organizational structure.”



This is an example of a 2 response

Dear Company,

As you may know, within the last half-decade, our company underwent times of significant change. First, there was a pandemic that caused distress and uncertainty about our company's future. Next, there was a recession that proceeded the pandemic that we are still recovering from. As a result of this, our company needs to optimize the way we function as a whole.

With today's announcement, we will be able to carry out our strategy to provide even better services to our customers. Some of you may know that recently a new CEO has taken over the company. Under the new CEO's leadership, there will be many core changes to our company, one of them being an organizational change. The top management has concluded that our company is too vertically hierarchal, in other words, there are too many tiers of authority in the company. Currently, in our company, there are six levels of management. We believe that this creates many issues. Firstly, being too vertically hierarchal creates confusion and performance issues. In terms of performance issues, there are slowdowns in decision-making as each decision has to trickle down multiple levels of authority. Additionally, six levels of authority create jealousy and disharmony in our company's workforce.

In light of the new changes, I understand that many of you have grand concerns. I would like to begin by addressing demotions. By changing the organizational structure from six levels of authority to three, there are bound to be demotions. Rest assured that top management will conduct a fair and equitable assessment of each employee to ensure that you are placed where you are meant to be. To ensure that this process is completely equitable, employees that are dissatisfied with where they end up as a result of the restructuring process will be able to request a second assessment. Lastly, I want to ensure that there will be no drastic demotions in the company.

We felt that this was the best way to approach optimization. It was either laying off workers or restructuring the company. We decided to go with the latter part as we want to retain everyone. I want to reassure you that in the long run, this will leave a positive impact on our company creating more jobs and a more powerful workforce.

I am aware that today's reorganization news has been challenging. There is no way to sugarcoat it. However, the harsh reality is that either spending had to go down or revenue had to expand more quickly. Taking a glance over the balance sheets, the math just didn't add, and the status quo couldn't continue. Even when the truth is painful or hard to deal with, we must be honest with ourselves and each other if we are to succeed.

In my perspective, accountability is the first step toward winning. You to each other, me to you, and me to you. As a group, we either sink or swim. Restructuring now serves as a reminder that outcomes matter. Never before has it been more crucial for us to be realistic about our abilities and follow through on our commitments. It has to do with managing expectations and accountability.

Notice how the response displays the change in the company's organization structure as an alternative to a layoff and necessary for the company's finances, even though this was **not a reason listed in the prompt**. The response **does not display empathy with employees** and **does not tactfully address concerns**. There is **incorrect and awkward word usage** ("proceeded the pandemic", "me to you, and me to you").

This is an example of a 6 response

Greetings Team,

I hope you are doing well, I would like to present you with some good news that I think will put a smile on your face. While we value you and every member of the team tremendously, we know the stress behind the corporate hierarchy and the factors hindering decision making processes. Well that is going to change because we are working on a new system that will be more efficient and level the playing field. This will allow people to work without the pressures that currently exist.

While I do not have the specifics to present to you at this current moment we will have something sent out officially in the near future. The reason I wanted to present this information now is to keep you motivated and make sure you know that your input is valued. When we can make changes to improve everyone's daily workflow I think we can all agree that's a net positive.

If there are some questions or concerns you may have about what will happen, feel free to let me know. I will answer in as much detail as possible for the current moment. There will be some management changes during these processes and the shifts within the organization will be reflected upon immediately. All information and instructions will be provided directly to each and everyone of you after the official announcement is made.

Also I would like to note that the same performance reviews from the past will still apply. While some people had more difficulties than others performing, the low performers will now have a chance to show an improvement in these upcoming weeks. To be fair I think is to be neutral which is why there is still an appreciation for everyone regardless which end of the threshold they are for performance.

We will also look forward for input on the changes after they are announced. Everything that is being done is for the best interests of the employees and the organization. Either way I feel everyone will appreciate the coming changes and I do look forward to hearing about how we can further improve things moving forward.

Finally, again a great amount of genuine appreciation from the CEO to each and everyone working their jobs making sure that they fulfill their tasks. Without you there wouldn't be a business and you can't be thanked enough. There was a large amount of communication in the past about the struggles and frustrations some employees were having and the CEO took the best course of action that he saw fit after analyzing and strategizing with different departments. His actions will be a direct reflection of how much he appreciates employees in different positions and how everyone is part of the team which can not function without each other. Once again thank you for your time at the organization as well as your time reading this message about the future changes. I will follow up with you all soon. Take care.

Kind regards,

Notice how the response **positively emphasizes the benefits** this will bring to most employees and makes the **audience feel appreciated and included** by mentioning the CEO's appreciation and collaboration with different departments. The response also has a **personable tone** and **flows well**. There are some minor instances of awkward sentence structure, but these do not detract from the overall comprehensiveness and quality of the content.

### **B.6.3 HR Professional Task A**

#### **Task**

Imagine you are an HR professional working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to "sit together" in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

The example and comprehension questions are the same as for managers.

## **B.6.4 HR Professional Task B**

### **Task**

Imagine you are an HR professional working for a company, WorkCo, in 2025. Most of the company's employees work from home.

In an attempt to recreate the collaborative office environment while allowing employees to work from home, the company has recently paid for the creation of a virtual office space and sent each employee a virtual-reality headset that allows them to connect to the space. The headsets allow employees to "sit together" in a digital open-plan office, enabling (the company hopes) casual conversations, feelings of connection, and the development of a fun office culture. The company views these as essential to collaboration, innovation, and motivation.

However, since receiving the virtual-reality headsets, employees have not used them very much. Some investigation by the HR team suggests this is because employees find the headsets slightly uncomfortable and because they prefer the privacy of working alone at home. Also, because very few employees are currently using the headsets, the virtual office is mostly empty so it is not yet attractive to work there.

Write a draft of a company-wide email that gently but firmly encourages employees to begin using their headsets. Your email should be about 400 words long and should explain why the company thinks it's important for employees to be using the virtual office space, and should address employee concerns.

The example and comprehension questions are the same as for managers.

## B.6.5 Data Analyst Task A

### Task

Imagine that you are a data analyst working at a bank. The bank is launching a customer outreach campaign to reduce customer churn (when a customer stops using a company's products or services), and your team is tasked with determining which customers should be targeted.

Your manager has asked you to determine which customers should be targeted. The dataset contains the following information:

- Customer ID
- Surname
- Credit Score
- Country
- Gender
- Age
- Tenure
- Balance
- Number of Products
- Has Credit Card
- Whether they are an Active Member
- Estimated Salary
- Whether they have exited

**Write a “code notebook” as if you were going to analyze this dataset to determine which customers should be targeted.** You may write pseudocode but do not write any code. Be specific about what analysis you would do, including what figures or tables you would produce, and what modeling techniques you would use.

Write as many steps and outputs as you feel is necessary. The steps can be written in bullet points. **The output should be about 400 words.**

Congratulations on correctly answering the comprehension questions! We will now proceed to the grading. First, **read the prompt that the responses answer:**

The prompt says to do which of the following? (Select all that apply)

Write a "code notebook" (without code) to determine which customers should be targeted to market a new product

Write a memo on which customers should be targeted to reduce churn

Write a memo on which customers should be targeted to market a new product

Write a "code notebook" (without code) to determine which customers should be targeted to reduce churn

Which of the following are data variables? (Select all that apply)

Purchase Frequency

Estimated Salary

Marital Status

Balance

Correct answers to the first question are “Write a ‘code notebook’ (without code) to determine which customers should be targeted to reduce churn.”

Correct answers to the second question are “Balance” and “Estimated Salary.”

## This is an example of a 2 response

Reducing customer churn is a critical challenge for banks, as it can have a significant impact on the bottom line. By identifying which customers are most likely to leave and targeting them with retention efforts, banks can not only retain valuable customers but also improve their overall profitability. Data analysis is a powerful tool that can be used to tackle this problem.

The first step in using data analysis to reduce customer churn is to define the problem. In this case, the target variable is whether a customer has exited. Once the problem has been defined, the next step is to prepare the data. This involves cleaning and preprocessing the dataset, removing any missing or irrelevant data.

After the data has been prepared, it is important to explore it in order to gain insights into customer behavior. Visualization techniques, such as histograms and scatter plots, can be used to understand patterns in the data. For example, a histogram of the tenure variable can show the distribution of customers by how long they have been with the bank. A scatter plot of balance versus credit score can show whether there is a relationship between these variables and customer churn.

Once the data has been explored, it is time to move on to feature engineering. This involves creating new variables that may be useful in identifying customers at risk of churn. For example, the tenure-to-age ratio or the balance-to-credit score ratio may be useful in identifying which customers are most likely to leave.

After the features have been engineered, the next step is to build a model that can predict which customers are most likely to leave. Machine learning algorithms, such as decision trees, random forests, or logistic regression, can be used to build such a model. These algorithms can analyze the data and identify patterns that are associated with customer churn.

Once a model has been built, it is important to evaluate its performance. Metrics such as accuracy, precision, and recall can be used to evaluate the model's ability to correctly identify customers who are at risk of leaving.

After the model has been evaluated, it can be deployed to target customers who are at high risk of leaving. Retention strategies, such as personalized offers or loyalty programs, can be implemented to keep these customers from leaving.

It is important to continuously monitor and improve the model as new data becomes available. This can be done by re-evaluating the data and retraining the model as needed. Additionally, it may be helpful to segment the customers by their demographics such as country, age, gender, and credit score, and analyze the behavior of each segment separately. This can help identify specific segments that are more likely to leave, and target them more effectively.

In conclusion, data analysis can be a powerful tool for reducing customer churn at a bank. By identifying which customers are most likely to leave and targeting them with retention efforts, banks can not only retain valuable customers but also improve their overall profitability. It is important to continuously monitor and improve the model as new data becomes available to keep the customer retention strategy effective.

The text is **not specific enough** in saying **how they will use** the different figures/tables and modeling techniques. Overall, there are many **vague ideas** presented, rather than a specific plan for how to **analyze this specific customer dataset**.

## This is an example of a 5 response

The goal is to determine which customers are to be targeted to reduce customer churn, but at its core the goal is to identify the demographics that would be most receptive to retention strategies and then target people with those characteristics. Credit score, country, gender, age, tenure, products, etc. that are not randomly assigned like ID numbers may give insight into the best optimization strategies. I am cleaning and aggregating the data for a large data set before starting the first step of explanatory analysis.

I am comparing people, using Excel, who have and have not exited. I am creating profiles of people who have exited, and note if there are any characteristics with majorities or pluralities, for example, a similar length of tenure or range of balance. Then I am forming similar profiles for those who have not exited, and comparing and contrasting the two groups. I am extracting the differences and using said differences to identify qualitative and quantitative features.

Afterwards, I am using Tableau (some people use Power BI, but I prefer Tableau) to create a data visualization so I may extract further insights. I am beginning with worksheets with simple tables such as in Excel, varying the types of dimensions, and then using bar charts to compare subsections or those who have exited and those who have not. I am creating a series of smaller bar charts for each dimension, breaking out side by side of active vs. not, and exited vs. not, and see if any particular dimension or cluster has a glaring visual difference.

I will use various Python libraries to create a time series model and, since this is the banking industry, use indices such as the CPI and VIX, to see if churn is related to greater macroeconomic conditions. I will also do a regression analysis in R between the different variables and the acts of being active and of exiting.

All in all I will use Excel tables grouped by data set dimension, i.e. the list mentioned above excluding randomly assigned markers, a Tableau dashboard with multiple bar charts, keeping only the most relevant ones with glaring differences, a time series model to map churn over time as compared to greater economic issues, and an analysis in R to more or less confirm the statistical significance seen said earlier analyses. I ideally hope to find what distinguishes inactive people who have exited vs. active people vs. inactive people who have not exited, to try to engage this latter group.

Although this is not formatted in the style of a code notebook, the text is **clear** about what steps would be taken to analyze the data and **states specific visualizations** that would be produced. The text presents **originality and context-specific knowledge** in discussing data on macroeconomic conditions. The **writing could be improved** (issues with verb tense, word usage and complexity, organization).

## B.6.6 Data Analyst Task B

### Task

Imagine that you are a data analyst working at Instacart. Instacart is considering launching a push notification campaign to encourage shoppers to buy, and your team is tasked with determining to whom and when notifications should be sent.

The dataset contains the following information:

- Customer ID
- Gender
- Age bin (18-24, 24-30, 30-40, 40-50, 50-60, 60+)
- Date and time of order
- Total cost of items ordered
- UPC codes of items ordered
- Type of item ordered (household good, grocery, beauty, etc.)
- Store ordered from
- Whether the customer used the “re-order item” option in placing an order

Write a “code notebook” as if you were going to analyze this dataset to determine which customers should be targeted. You may write pseudocode but do not write any code. Be specific about what analysis you would do, including what figures or tables you would produce, and what modeling techniques you would use.

Write as many steps and outputs as you feel is necessary. The steps can be written in bullet points. **The output should be about 400 words.**

The prompt says to do which of the following? (Select all that apply)

- Write a “code notebook” (without code) to determine how to target advertising to gain new customers
- Write a memo on which customers should be targeted to market a new product
- Write a “code notebook” (without code) to determine which customers should be targeted in a push notification campaign
- Write a memo on how advertising should be targeted

Which of the following are data variables? (Select all that apply)

- UPC codes of items ordered
- Marital status
- Gender
- State

Correct answers to the first question are “Write a ‘code notebook’ (without code) to determine which customers should be targeted in a push notification campaign.”

Correct answers to the second question are “UPC codes of items ordered” and “Gender.”



This is an example of a 2 response

To analyze this data, we'll start by creating a spreadsheet of all of this data. We'll need multiple fields for things like the UPC codes of items ordered and type of item ordered, because there could be multiple entries on one order. When we have all this data formatted correctly, we should create a table that focuses on the total cost of items ordered versus gender and another table for cost versus the age of the customer. This will be a good starting point to determine if there are some easy segments to recognize that we should target.

Once we have that out of the way, we should next focus on creating a table that filters out the types of items ordered by age and gender and we can start identifying how we could target different groups of people based on the types of things that they purchase. It is also important to identify repeat customers, so we could use the re-order field option to identify whether target demographics like age or gender are more likely to re-order products. I do not think individual results, such as the customer ID, will be as helpful for us because I want to focus on larger demographic groups to target our efforts. Once we have these tables it is important to create a visual representation that is easy to view and analyze. For that purpose, I would take those tables and create bar graphs that could help show us the results easily and then we could compare the different segments to determine which segment is the best customers.

One other thing I want to look at is changes over time, so we could create a line graph to look at how total amounts changed over time using the date and time information from the initial spreadsheet. Going back to the bar graphs, I think that targeting by age and gender is the easiest way to go, and will provide us with some information about the most likely groups to target. I think using the UPC codes in a chart could help us identify the most popular products that we could focus our advertising on to get people into the store. An important point here is that we can do all this and keep our results very simple. In most cases, we're only relying on two or three pieces of information to help inform our conclusions, which will help us make clear decisions. I would use Excel to analyze the data and create the results that we can work from, probably eventually putting it in a PowerPoint presentation to show it clearly.

The text has long sentences and long paragraphs that make it **hard to read**. There are **not sufficient details on what modeling techniques they will use**.

This is an example of a 5 response

Import and clean the data, ensuring that it is in a format that can be easily analyzed.  
Create a histogram of the distribution of ages among customers to get a sense of the age demographics of our customer base.  
Create a bar chart of the distribution of gender among customers.  
Create a bar chart of the distribution of type of items ordered by customers.  
Create a bar chart of the distribution of store ordered by customers.  
Create a bar chart of the distribution of the use of "re-order item" option among customers.  
Create a scatter plot of total cost of items ordered against the time of order to identify any patterns in purchase behavior.  
Use the UPC codes of items ordered to classify each order as a "household good" purchase, "grocery" purchase, "beauty" purchase, etc.  
Create a pivot table to show the frequency of purchases by customer ID, type of purchase, and time of purchase.  
Use a clustering algorithm (such as K-means) to group customers based on their purchase behavior.  
Use this information to create segments of customers, for example, "frequent household good buyers" or "infrequent grocery buyers"  
Use this information to create a heatmap to identify the peak days and hours when customers are making purchases.  
Use this information to create a decision tree model to predict the likelihood of a customer making a purchase based on their demographics, purchase history and purchase behavior.  
Use the decision tree model to create a new customer segmentation, for example, "likely grocery shoppers", "likely beauty product shoppers", etc.  
Use these customer segments to target specific groups of customers with personalized push notifications.  
Create A/B testing groups to test the effectiveness of the push notifications on customer behavior and use this information to optimize the campaign.

One approach to determining which customers to target with push notifications could be to use a decision tree model. First, we would segment the customer data by demographic and purchase behavior. Next, we would use the "re-order item" option as the target variable and the other features such as gender, age bin, total cost of items ordered, UPC codes of items ordered, type of item ordered, store ordered from as predictors. We would then train the model on a subset of the data and use the remaining data for testing. After evaluating the model's performance, we would use it to predict which customers are likely to use the "re-order item" option. Customers who are predicted to be likely to use this option would be targeted with push notifications.

The response states **what visualizations and analysis steps** will be conducted. There are **context-specific details** about the types of purchase categories and customer segments. The last paragraph **clearly explains how** the decision tree model works. But, there could be **more variation in the types of visualizations** used.

## B.6.7 Marketer Task A

### Task

Write a 400-word press release for the following product: a self-driving electric bicycle called the SmartBike. Your goal is to maximize how attention-grabbing and attractive to potential consumers the press release is. The raters will grade your press release on this basis. Please keep in mind the following facts:

- The bicycle is being produced by a company, E-Bike Inc, that prides itself on the convenience and environmental friendliness of its products. Its customer base consists of mostly middle-class city-dwellers.
- E-Bike Inc is hoping to market the SmartBike to three populations: existing bicycle users, people who currently commute by walking, and people who currently commute by car. Your press release should discuss the advantages of the SmartBike relative to all three of these alternative commuting methods.
- In focus groups, many participants expressed concerns about the safety of a self-driving bicycle. Make sure to address these concerns in your press release by referencing the bicycle's spotless track record in testing and E-Bike Inc's care for the safety of its customers.
- The SmartBike has the following features:
  - Up to 60 hours of battery life.
  - Extremely advanced obstacle navigation, speed control, and braking systems.
  - An integrated light system to signal turns and lane changes to drivers and enhance visibility in the dark.
  - A lightweight and strong carbon-fiber frame. Integration with smartphones and watches.
  - A remote-control mode that allows users to, for example, summon the bike from where it is parked.

Image of the SmartBike:



Congratulations on correctly answering the comprehension questions! We will now proceed to the grading. First, **read the prompt that the responses answer:**

E-Bike Inc is hoping to market to which of the following populations? (Select all that apply)

Existing bicycle users

People who currently commute by walking

People who currently commute by car

Students who live near their college campus

The SmartBike has which of the following features? (Select all that apply)

Up to 60 hours of battery life

Ability to be summoned from where it is parked

Advanced obstacle navigation, speed control, and braking systems

Turn signals

Correct answers to the first question are “Existing bicycle users,” “People who currently commute by walking,” and “People who currently commute by car.”

Correct answers to the second question are “Up to 60 hours of battery life,” “Ability to be summoned from where it is parked,” “Advanced obstacle navigation, speed control, and braking systems,” and “Turn signals.”

This is an example of a 2 response

Good day ladies and gentlemen. Ebike Inc is excited and wishes to announce the release of SmartBike: a self-driving electric bicycle. SmartBike is for anyone and everyone but will come in very handy mostly for middle-city-dwellers. Let us tell you more about this amazing product.

SmartBike is convenient and environmentally friendly due to its carefully selected manufacturing products. of its product. Its primary users will consist of mostly middle-class city dwellers because it can be used as a fast and reliable means of commuting to work and avoiding and escaping traffic.

We hope to market the SmartBike to three populations: existing bicycle users, people who currently commute by walking, and people who currently commute by car.

Electric bikes have proven to be a convenient micro-mobility solution for the modern adult. As discussed above, you can travel farther and faster on an eBike than a conventional bike, allowing you to both cover more distance and slash commute times whether you're headed to work or your child's Little League game.

Similarly, SmartBike holds several advantages over traditional gas-powered cars. You can dodge traffic jams using bike routes. An eBike can go pretty much anywhere a conventional bike can. Plus, you'll never need to spend time circling the parking lot over and over for an open spot again. Trade in your car keys for a bike helmet and skip the crowded parking lots altogether.

The SmartBike has the following features:

Up to 60 hours of battery life.

Extremely advanced obstacle navigation, speed control, and braking systems.

An integrated light system to signal turns and lane changes to drivers and enhance visibility in the dark.

A lightweight and strong carbon-fiber frame. Integration with smartphones and watches.

A remote-control mode that allows users to, for example, summon the bike from where it is parked.

Riding SmartBike is a great way to seamlessly integrate cardio exercise into your daily routine, no matter what your fitness level is. Even in pedal-assist mode, you are still using human power to get moving!

You can also still use an electric bike like a regular bike in pedal-only mode with no electric power if you want to opt for a more traditional cycling workout. Plus, there is nothing like enjoying the fresh air and beautiful scenery wherever you ride while getting your daily physical activity in at the same time.

with all these in mind, we urge you to choose SmartBike. Thank you.

Notice that the response **mostly copies from the prompt** without much extra detail, explanation, or detail added. The writing is generally **bland** and **does not catch the reader's attention**.

This is an example of a 6 response

Did you know patents for electric powered bikes date back to the 1890's and the original bike back to 1819? Since then there have been numerous improvements made in design, efficiency, and manufacturing, but at its core the usage of a bike has remained virtually the same for nearly 200 years!

That changes today, with the introduction of E-Bike Inc's SmartBike! The first of its kind and the next great evolution of the bike as we know it! The SmartBike incorporates state of the art technology supporting fully autonomous navigation. SmartBike is the perfect transportation solution for anyone whether you currently bike, walk, or drive for your daily commute.

Existing bike owners will find a comfortably familiar form factor as the technology integrates seamlessly into a traditional lightweight and strong carbon-fiber bike frame. The user has full control over what functionality they want to use, when they want to use it. Fully autonomous, electric, and manual modes are supported depending on the riders preference. Built in integration with smartphones and watches is another added benefit, for those looking to get the most out of their commute.

SmartBike is also the perfect option for those that currently walk. Perhaps you prefer the safety walking provides over vehicle transportation. Safety was one of E-Bike Inc's top priorities for the SmartBike, which features a spotless track record after extensive real world testing. Extremely advanced obstacle navigation, speed control, and braking systems keep the rider safe at all times. An integrated light system to signal turns and lane changes to drivers and enhance visibility in the dark have also been included to ensure safety in low light conditions.

For those that drive, SmartBike is the perfect blend of environmentally friendly and convenient. A typical passenger vehicle emits about 4.6 metric tons of carbon dioxide per year, where SmartBike's highly efficient electric motor provides zero emissions. In addition it features a 60 hour battery life so you will never need to worry about going without power. SmartBike also eliminates the stress and frustration of trying to park in crowded metro areas. It even features a remote-control mode that allows users to summon the bike from where it is parked.

Safety, convenience, and the environment are all at the heart of E-Bike Inc's design philosophy and the SmartBike as a product. By introducing the most advanced bike to the market in nearly 200 years, E-Bike inc is looking to revolutionize the future of transportation.

Notice how the writing has an **exciting hook**. The text provides a **clear description** of the features of the bike. The text also addresses **how the bike is beneficial** to those that currently walk and those that currently drive. It **crisply provides relevant details**.

## B.6.8 Marketer Task B

### Task

Write a 400-word press release for the following product: augmented-reality glasses that allow a user to overlay digital images onto the real world. Your goal is to maximize how attention-grabbing and attractive to potential consumers the press release is. The raters will grade your press release on this basis.

Please keep in mind the following facts:

- The glasses are being produced by a tech startup called VAR Inc, whose first product was the GameSet, a very popular and successful virtual-reality headset for video gaming. The press release should include a callback to the GameSet, so that consumers associate the glasses with the highly popular GameSet.
- VAR Inc want to focus on two main uses of the glasses: enhancing productivity at work or in business settings, and fun recreational usage. The press release should include discussion and examples of both uses.
- Previous augmented-reality glasses have been commercially unsuccessful because users felt they were uncool or nerdy and were reluctant to wear them. The press release should try to make the glasses seem cool and trendy.

Image of the augmented reality glasses:



The second prompt is as follows:

VAR Inc wants to focus on what kinds of usage for the glasses? (Select all that apply)

<input type="checkbox"/> Recreational usage
<input type="checkbox"/> Monitoring during sports training
<input type="checkbox"/> Supporting safety in trades jobs
<input type="checkbox"/> Enhancing productivity at work

The press release should also do which of the following? (Select all that apply)

<input type="checkbox"/> Make the glasses seem cool and trendy
<input type="checkbox"/> Describe the materials used to make the glasses
<input type="checkbox"/> Call back to the GameSet
<input type="checkbox"/> Discuss which ages would be best able to utilize the glasses

Correct answers to the first question are “Recreational usage” and “Enhancing productivity at work.”

Correct answers to the second question are “Make the glasses seem cool and trendy,” and “Call back to the GameSet.”

### This is an example of a 2 response

The management and team of VAR Inc. wish to announce the release of our newest product: augmented-reality glasses that allow a user to overlay digital images onto the real world. We appreciate the acceptance of GameSet, a virtual-reality headset for video gaming which was highly rated and patronized. The much-anticipated release of the augmented-reality glasses is the product of dedication, excellence in innovation, and the implementation of customer demands and requests.

The augmented reality glasses come with two main uses: enhancing productivity at work or in business settings, and fun recreational usage.

The augmented reality glasses have made internal service easier and more convenient as New employees to your business can watch training videos and interact with them in a hands-free way. With these smart glasses, they can take notes and participate in exercises in real-time.

In the case of factory workers, smart glasses increase order-picking accuracy and volume. Additionally, they help improve warehouse safety by giving the option of hands-free control. This in turn means that warehouse staff no longer require an RF gun in hand. By this, picking will be a hands-free operation. Furthermore, the use of voice technology allows users to easily navigate AR and Vision Picking apps with voice commands.

Many efforts were put into the aesthetics of these glasses to match the current trends and tastes in eyewear. Previous augmented-reality glasses have been commercially unsuccessful because users felt they were uncool or nerdy and were reluctant to wear them. These glasses are far from nerdy and uncool as they fit into all and any occasion. The universal nature and dynamism are out of this world.

Also, in customer relations, customers are able to both customize and view products in the company's catalog. Augmented Reality enables design capabilities for customers. Actualizing a product at their place of business or in the colors they choose helps the decision process move along.

In hotel and tourism, augmented reality glasses augment reality to create a futuristic city that will walk attendees through an experience of their stay in the hotel and a glimpse of the various activities they can engage in.

For marketing and advertisement purposes, augmented reality is a handy tool for explaining and engaging the target audience of products. Businesses that have objectives related to either of these should consider an Augmented Reality solution. AR experiences are also likely to be highly shareable, so they are well-suited for spreading the word about events and entertainment.

In summary, VAR Inc can assure the general public of the authenticity of the newly released augmented reality glasses and encourages you all to continue supporting and patronizing our tech products as we have your requests, suggestions, and comments as a top priority. Thank you.

Notice how the text is **longwinded, clunky, and difficult to read**. There is **little structure** and it **does not engage the audience**. There is **no explanation of how the glasses are "far from nerdy and uncool"**. There are many industry uses, but **little mention of recreational uses**.

### This is an example of a 6 response

Introducing VAR Inc's latest innovation: the augmented-reality glasses. These revolutionary glasses allow users to overlay digital images onto the real world, enhancing productivity and recreational experiences like never before.

For the ones who are looking for a new level of immersion and interaction, VAR Inc is the right choice. We are the pioneers of virtual and augmented reality technology, and we are constantly pushing the boundaries of what is possible. With our latest product, the augmented-reality glasses, we are taking the concept of augmented reality to the next level.

Building on the success of our first product, the GameSet virtual-reality headset, we at VAR Inc are proud to bring you a new level of immersion and interaction. These glasses take the concept of augmented reality to the next level, allowing you to seamlessly blend the digital and physical worlds.

With the new augmented-reality glasses from VAR Inc, you can revolutionize the way you work and play. Whether you're in a business setting or enjoying some leisure time, these glasses will enhance your experiences in ways you never thought possible. Imagine being able to review and edit important work documents right in front of you, without having to constantly switch between different screens or devices. This will save you time, increase your productivity and allow you to be more efficient.

But the benefits of these glasses don't stop there. They also offer endless opportunities for fun and entertainment. You can overlay digital games and graphics onto your surroundings, bringing your favorite games to life in a whole new way. Imagine playing a game of virtual chess in the park, or exploring a virtual world as you walk through your neighborhood. The possibilities are endless, and these glasses will open up a whole new world of entertainment options.

Not only are they functional and versatile, but these glasses are also sleek and fashionable, so you can wear them with confidence and style. No more feeling self-conscious or awkward wearing bulky, nerdy-looking augmented-reality glasses. These glasses are designed to be cool and trendy, so you can wear them with pride.

So whether you're looking to enhance your productivity at work, or looking for new and exciting ways to play, the new augmented-reality glasses from VAR Inc are the perfect solution. With endless possibilities and a sleek, fashionable design, these glasses are sure to change the way you see the world. So, don't wait, get your hands on these glasses and explore the world in a whole new way!

Notice how this response is **creative, exciting** and has a **clear call to action**. The writing **flows well** and is **easy to digest**. It also gives **clear examples of use cases** and states that **the glasses are sleek and fashionable**.

## B.6.9 Consultant Task A

### Task

You are a management consultant working for a **US fast casual Tex-Mex restaurant chain that is deciding whether to expand its operations to China**. The chain is seeking advice about the potential benefits and risks of expanding operations to China.

Your colleague has compiled three sources to help answer this question:

- McDonalds 2020 Annual Report: Operations Section
  - Read [here](#). Scroll to the page of the report with heading “OPERATIONS” (page 31 or PDF page 36).
- Mordor Intelligence China Foodservice Market Report
  - Read [here](#)
- HBR on KFC in China
  - Read [here](#)

Please use the sources to write a 400 word summary of the potential benefits and risks of expanding operations to China.

Congratulations on correctly answering the comprehension questions! We will now proceed to the grading. First, **read the prompt that the responses answer:**

According to Mordor Intelligence, which of the following are top players in the China Foodservice Market? (Select all that apply)

Starbucks Corporation

Yum! Brands Inc.

PepsiCo

McDonald's

The prompt says to do which of the following? (Select all that apply)

Give the client a yes/no answer about whether to expand to China

Summarize McDonalds 2020 Annual Report

Advise on the potential benefits and risks of expanding operations to China

Write a summary of KFC's expansion strategy in China

Correct answers to the first question are “Starbucks Corporation,” “Yum! Brands Inc,” and “McDonald’s.”

## Correct answers to the second question are “Advise on the potential benefits and risks of expanding operations to China.”

This is an example of a 2 response

As we all know, international businesses contribute to Chinese economic development in two significant ways, making their products or services available to average Chinese person and employing Chinese workers, even if the company doesn't have a market presence in the country.

If your business is evaluating expanding into China in either of these ways, it's important to understand the pros and cons of doing business there. China is an attractive locale for global businesses large and small. I'll outline and explain a few of the most compelling benefits of doing business in China.

A. A growing middle class

Two major factors have influenced China's ever-growing middle class:

1. As more international companies hire in China as part of expansion, they must provide competitive wages to remain attractive to employees with multiple job options. Thus, Chinese workers' wages and benefits continue to increase, giving them more spending power.
2. Urban development has skyrocketed in China throughout the last decade. As domestic and global employment has increased, Chinese workers have flooded to cities and urban areas to access job opportunities, requiring significant investments in metropolises.

B. Increasing reforms improve the ease of doing business

These reforms include:

- Streamlined business registrations and licensing
- Simplified permitting processes for construction and infrastructure projects
- Greater utility access and ease of setting up electrical and plumbing connections
- Improvements to the corporate tax system that prioritize small business savings
- Increased ease of importing and exporting goods
- Enhanced ease of insolvency proceedings for companies that struggle financially

C. Cultural commitment to innovation

D. Significant current and predicted economic growth

Although business opportunities in China abound and the growing economy continues to attract international interest, there are drawbacks. Let's review some notable disadvantages of expanding to China or hiring within the Chinese labor force.

1. Intellectual property protection gaps

Businesses inventing and developing new products, systems and services typically seek protection for their valuable innovations. Intellectual property (IP) laws serve this purpose by legally prohibiting theft of proprietary ideas.

2. Prioritization of domestic businesses

After a long history of economic lethargy and a weak presence in the global market, China has become very protective of its current domestic prosperity

3. Market break-in difficulties

Success stories of foreign brands in China certainly exist. For instance, both Nike and Walmart have expanded very successfully in the country. But as China continues to grow and prioritize domestic development, it's increasingly difficult for international companies to overcome Chinese market entry hurdles.

**The formatting, grammar, and overall readability are poor.** The text presents some broad points, but **does not specifically address the firm in the prompt**, a Tex-Mex restaurant. Additionally, the text **does not reference the sources**.

This is an example of a 5 response

Expanding operations in China is a high-risk, high-reward opportunity.

From a growth perspective, China offers one of the world's largest and fastest growing markets. With a large and growing middle class, these Chinese market contains a significant number of potential customers for a product as ubiquitous as food. While China has remained under some of form of lockdown due to the Covid-19-related restrictions for the past several years, policy makers have recently been publicly relaxing those measures and there is reason to be optimistic that the economy may experience a broader reopening. This presents an even more accelerated growth opportunity in the near term as the restaurant industry was acutely impacted by the nature of the pandemic policy responses and as a result has a diminished level of competition. Successful expansion in China would likely result in a relatively profitable operation for the company with an enormous potential for rapid future growth.

There are, however, several significant risks to expansion in Chinese markets. While competition may be diminished compared to before the pandemic, the Chinese market now has a proliferation of "Western" style food options available and it is no longer has the perception of novelty that once served as a driver of demand. Successful brands have not simply taken a pre-packaged model that worked in the United States and transplanted it overseas, they have generally had to rethink, redesign, and reengineer elements of their marketing and operations to become more appealing to the Chinese consumer. The sum of these factors increases the costs of attempting to expand in Chinese markets since there is an incrementally higher burden for any type of differentiated branding to truly resonate with potential customers. While it is true that policy makers are signaling a relaxation to pandemic-related restrictions, this is not the first time such indications have been made and there is a nonzero possibility this happens to a lesser degree or more slowly than expected. Regulators in China have historically targeted individual industries with disruptive policy actions which presents a potential source of risk with little recourse.

While expansion in Chinese markets, if successful, would provide the company with a presence in a robust market with an incredible growth potential, it would also require a significant financial investment with a highly uncertain probability of success given the depth of competition already entrenched in the market and the different tastes of the Chinese consumer.

**The risks and benefits are clearly outlined.** The text is **well-written** and **thoughtfully organized**. However, there is **little evidence cited** by the author to further strengthen their claim, and they **do not specifically address the restaurant industry**.



## B.6.10 Consultant Task B

### Task

You are a management consultant working for a **US meat producer that is deciding whether to begin investing in R&D for cultured meat production**. The producer is seeking advice about potential benefits and risks of making this investment.

Your colleague has compiled three sources to help answer this question:

- Zacks Small-Cap Research on Stakeholder Foods: Risks Section
  - Read [here](#). Scroll to the page of the report with heading “RISKS” (page 7).
- Cultured Meat Market Report Summary
  - Read [here](#)
- Nestle’s Partnership with Future Meat Technologies
  - Read [here](#)

Please use the sources to write a 400 word summary of the potential benefits and risks of investing in R&D for cultured meat production.

According to Zacks Small-Cap, which of the following are core components of Steakholder Food's processes? (Select all that apply)

Differentiating them into fat and muscles cells

Growing cell cultures

Creating structured cell-based meat products

Collecting animal cells

The prompt says to do which of the following?

Give the client a yes/no answer about whether to invest in R&D for cultured meat production

Write a comparative analysis of the strategies of Steakholder foods and Future Meat Technologies

Advise on the potential benefits and risks of making an investment in R&D for cultured meat production

Write a summary of the state of the cultured meat market

Correct answers to the first question are “Differentiating them into fat and muscles cells,” “Growing cell cultures,” and “Creating structured cell-based meat products.”

Correct answers to the second question are “Advise on the potential benefits and risks of making an investment in R&D for cultured meat production.”

This is an example of a 2 response

Lab grown meat, also known as cultured meat as good as it may seem , has both advantages and disadvantages that investors need to be aware of

1. It is more sustainable

There is no doubt that lab grown meat is a more sustainable solution. From an environmental standpoint this means less water is used to produce meat, less methane gas is put into the atmosphere, and it's overall a much cleaner solution than factory farming. Some of the Advantages includes

2. The animals suffer less (or not at all)

It is true that you can create cultured meat from an animal that does not end up going to slaughter. However, the reality is that many of the animals that will be used for research and development purposes are probably animals that are on the slaughter lines anyways. With that being said, we can have a future where we are eating meat and no new animals will be killed for each bite we take.

3. It has less bacteria

I'll wrap this one into the there are no antibiotics discussion. Since cultured meat, or for this purpose we will call it clean meat, has is grown in a sterile environment, there is little bacteria that is produced when producing the meat. This means that the meat that we eat will be free of many potential diseases. A lot of the antibiotics and drugs that are used to treat sick livestock animals will no longer be necessary.

Some of the risk includes

1. We are many years from producing large quantities

According to Cultured Meat Market Report Summary, Many of the companies that are pioneering the field, such as Mosa Meats, claim that we are still quite a few years out from creating cultured meat. Some of the companies are mentioning that 2020 or 2021 will be the year that they start producing cultured meat for the public. However, this is a vague statement and even by 2021 we are not sure what types of quantities will be released. Surely not even a fraction of the amount of meat such as beef or poultry produced from current animal agriculture.

2. It is very expensive to produce

Right now, there are many companies in the research and development phases of creating cultured meat.

3. We don't know how to regulate it

Another issue that we are currently facing is that we are not sure how to regulate it. In the United States, there are several organizations that are fighting over what it should be called, who tests for safety, and which governing body can regulate it. This could push development even farther back, especially if the meat and dairy industry want to push against it or if the animal welfare organizations don't make a strong enough case for the entire nation.

The text summarizes the benefits and risks of the product but **does not address the question** of whether an investment should occur. The text is generally **disjointed, looks unprofessional, and has many grammatical errors.**

This is an example of a 5 response

Investing in R&D for cultured meat production would be an important step toward increasing presence in a rapidly growing market with numerous drivers for future demand. However, its success would not be guaranteed and represents an additional financial burden going into a period where it is not certain how much funding will be readily available from capital markets.

The cultured meat market is growing at double-digit annual rates and has several societal tailwinds behind it that suggest robust future growth. Consumers are increasingly adopting protein-rich diets and relying on omega-3 fatty acids to meet nutritional needs. Concerns are rising regarding food safety standard and enforcement in the traditional food sector. Consumers are increasingly health conscious and have an increasing preference for health sources of food. The experience of the pandemic further accelerated this trend. Rising demand to acquire essential nutritional ingredients in developing countries offer an additional source of potential demand and future growth opportunities. Attitudes that are increasingly intolerant of the traditional meat industry's impact on environmental sustainability and animal welfare represent another source of incremental demand for cultured meat as a substitution. Compared to the traditional meat industry, cultured meat manufacturing offers zero animal cruelty, less deforestation, a minimization of antibiotic usage and is generally less taxing on the environment. Cultured meat also minimizes the greenhouse gases released during manufacturing which offers an important source of potential demand as policy makers look to impose restrictions on markets over time.

While the growth opportunity is significant, competition exists and will only become more fierce as the high growth rates for this market attract major players to expand offerings into the space. Nestle, for example, has already begun R&D related to cultured meat and represents a substantial threat as their deep resources mean they can outspend the company on both marketing and research. This will limit the potential market share the company can earn while also increasing the expense of gaining each incremental customer. Further, the payoff to R&D is uncertain and the company is already in an unprofitable condition with significant debt on the balance sheet. Overall economic growth conditions are slowing and it is unclear whether capital markets will be available should the need for capital funding arise, which we know will happen in the relatively near future. This would be an expense with a hopeful future payoff but would represent an additional burden in the near term.

The response is **well written and directly addresses the investment question.** However, it could have **used more detail and direct quantifiable data from the sources** to strengthen and clarify reasoning.

## B.6.11 Grant Writer Task A

grantwriting\_task1

### Task

You are working for a local community center in the year 2030. The center is submitting a grant application to the state government in an attempt to secure funding for a proposed **virtual-reality community gym**. The gym will consist of a bunch of exercise machines augmented with virtual-reality headsets (for example, a treadmill with a headset that allows the user to feel like they are running in any location they want—a beach in Hawaii, an Antarctic glacier, or the surface of Mars). **Your job is to write a 500-word cover letter for this grant application.** You can view an example grant cover letter [here](#). Your letter should briefly summarize the case for funding this project and should convey the community center's enthusiasm about the project. Please note the following:

- The community center wants you to explain why a community-provided gym is necessary even though there are already private gyms in the area.
- The community center wants you to clearly communicate why a virtual-reality gym is preferable to a normal gym.

Congratulations on correctly answering the comprehension questions! We will now proceed to the grading. First, **read the prompt that the responses answer:**

Which of the following will be in the gym? (Select all that apply)

Hologram personal trainers

Exercise machines augmented with virtual-reality headsets

Red-light saunas

The community center wants you to do which of the following? (Select all that apply)

Communicate why a virtual-reality gym will ultimately be cheaper than a regular gym

Communicate why a virtual-reality gym is preferable to a normal gym

Explain why a community-provided gym is necessary even though there are already private gyms in the area

Explain the budget for the virtual-reality gym set up costs

Correct

answers to the first question are “Exercise machines augmented with virtual-reality headsets.”

Correct answers to the second question are “Communicate why a virtual-reality gym is preferable to a normal gym” and “Explain why a community-provided gym is necessary even though there are already private gyms in the area.”

This is an example of a 2 response

Dear [Grant Committee],

We at the [Community Center] are thrilled to submit our grant application for funding for our proposed virtual-reality community gym. Our goal is to provide a unique and innovative fitness experience for members of our community, and we believe that a virtual-reality gym is the perfect solution.

As you may know, there are already private gyms in the area, but we believe that a community-provided gym is necessary for a number of reasons. First, private gyms can be expensive, and not everyone in the community can afford to join. By providing a community gym, we can ensure that fitness opportunities are available to everyone, regardless of their financial situation. Additionally, a community gym can serve as a hub for community members to come together and connect with one another. It can be a place for people to make friends, share tips and advice, and create a sense of community.

A virtual-reality gym is preferable to a normal gym for a number of reasons. First and foremost, it provides a unique and immersive experience that traditional gyms simply cannot match. By using virtual-reality headsets, users can feel as though they are working out in any location they can imagine from a beach in Hawaii to the surface of Mars. This creates a sense of excitement and motivation that traditional gyms can lack. Additionally, virtual-reality gyms can provide a wider range of workout options, as users can choose to run on a virtual track, swim in a virtual ocean, or even climb a virtual mountain. This allows for a more varied and interesting workout experience.

Furthermore, virtual-reality technology can also provide users with a more personalized workout experience. By tracking users' progress and providing real-time feedback, virtual-reality gyms can help users achieve their fitness goals more efficiently and effectively. Additionally, virtual-reality technology can also be used to provide users with virtual personal trainers, who can guide them through their workout and provide encouragement and support.

We at the [Community Center] are confident that a virtual-reality community gym would be a valuable asset to our community. We believe that it would provide an exciting and unique fitness experience for members of our community, and we are eager to bring this innovative technology to our community center. We would be honored to receive funding for this project, and we thank you for your consideration.

Sincerely,  
[Your name]  
[Your title]  
[Community Center]

The text **mostly repeats the prompt** without adding much. There is **no community support expressed**, there are **no additional details**, and the **feasibility of the project is not addressed**.

This is an example of a 5 response

Re: Virtual-Reality Community Gym

To whom it may concern,

I am pleased to present this proposal for your review, on behalf of our community center's virtual-reality gym. We serve low-income members of our community and look forward to partnering with you to bring our patrons fitness services aimed at improving both their physical and mental health.

For many low-income individuals, traditional gym memberships remain out of reach due to the expensive monthly fees that they charge. Poverty has many ill effects on health, which become compounded by lack of regular exercise. Low-income individuals often must make choices between buying healthy food and paying utility bills or even rent. As a result, many of them, by necessity, eat cheap but unhealthy foods, including those high in fat, sugar, and carbs. This poor diet affects every aspect of their future health and can lead to serious, life-threatening conditions like diabetes, heart disease, and stroke. And the constant stress caused by choices they must make in poverty can take a dire toll on mental health, leading to increased risks for depression, anxiety, and suicide.

Our virtual-reality gym aims to address this two-fold problem by making high-quality, virtual fitness experiences available to low-income individuals. The mitigating effect that routine exercise has on chronic health conditions is already well-documented. It has an equally profound effect on depression, anxiety, and many other mental illnesses. Routine access to high-quality fitness equipment can have amazing health benefits for those living under the constant strain of poverty, but remains beyond their ability to pay for, even at regular gyms with monthly membership fees. Our community gym offers them access to exercise equipment without adding additional worries about money to their lives.

Research has shown that travel can also greatly improve mental health because it offers individuals novel and diverse experiences they cannot get by always remaining in the same place. Unfortunately, traveling is well beyond the means of most low-income individuals. We believe that varied, pleasurable experiences like travel should be accessible even to those living in poverty because of the positive effects it has on mental health. We cannot send each of our community members to Hawaii or an Antarctic glacier, but we can offer them the next best thing: virtual-reality travel experiences.

At our gym, members will be able to run on the beaches of Hawaii or even across the surface of Mars. Our virtual-reality gym will pair needed physical fitness with unique and thrilling virtual experiences, so that those living in poverty can reap the mental and physical benefits they need to improve their overall health, without having to worry about whether their gym membership will make them unable to pay their next electric bill. Our proposal requests \$30,000 to acquire exercise machines and equipment as well as virtual reality headsets and software and make this vision a reality.

We appreciate your interest in help our community members live healthier and happier lives, no matter their income level. Please feel free to contact us if you need any additional information or have any questions about our proposed virtual-reality community gym.

Thank you

The text provides **creative and innovative uses** for the virtual gym. The writing is **generally solid and convincing**. It provides economic, social, and health-related arguments. However, it **does not do enough to address why VR equipment, rather than normal gym equipment, is needed**.

## B.6.12 Grant Writer Task B

grantwriting\_task2

### Task

You are working for a local school in the year 2030. The school is submitting a grant application to a nonprofit that provides funding for educational technologies. The school is asking for funding for an AI tutoring service. The tutoring service will provide on-demand help for students in the form of a chatbot (for example, a student can provide a math problem and their attempt at a solution to the AI tutoring service, and the AI will provide guidance on how to solve the problem and fix the mistakes the student has made). Your job is to write a 400-word cover letter for this grant application. You can view an example grant cover letter [here](#). Your letter should briefly summarize the case for funding this project and should convey the school's enthusiasm about the project. Please note the following:

- The school wants you to explain why AI tutors are necessary even though each class of 30-40 students has a teacher.
- The school wants you to clearly communicate why hiring this AI tutoring service is preferable to hiring in-person tutors or teacher aides.

What are features of the AI tutoring service? (Select all that apply)

A monitoring service for parents

Assistance lesson planning for teachers

Guidance on how to solve problems

On-demand help for students

You should explain which of the following? (Select all that apply)

Why AI tutoring is preferable to hiring in-person tutors or teacher aides

A breakdown of the costs of AI tutors

Why AI tutors are necessary

The impacts of AI tutors on teachers' jobs

Correct answers to the first question are “Guidance on how to solve problems” and “On-demand help for students.”

Correct answers to the second question are “Why AI tutoring is preferable to hiring in-person tutors or teacher aides” and “Why AI tutors are necessary.”

This is an example of a 2 response

Good day,

We are thrilled to share this application with you to request a grant for an AI tutoring service. We believe that this service will be a groundbreaking advancement in education, and will be in every school in America by the year 2040. To have it in our school this early would be a crowning achievement for our staff, students, and teachers.

In today's modern society, educators find themselves overworked, overstressed, and unable to give each student the direction that they require. With this AI tutoring service, the difficult and often overwhelming task of giving students individual help is taken off their crowded work schedule. What's more - it's highly effective and average student test scores have increased by 5-10% in our example group. Feedback has shown that students who use the service have a better job retaining the new lessons that they learn via the AI's help. Furthermore, they begin to feel a bond with the AI they're working with, one that's less formal than with their teachers but still has a clear educator-learner balance.

Aside from the help it will give our staff, the competitive and reasonable pricing model reduces annual budgets significantly. With just this one grant, we will be able to make several expensive temporary faculty positions redundant and therefore clear up costs for future projects, including our new proposed theater facility. Our school spends a significant amount of money every year hiring tutors for students, and this will save all of that. What's more, this service will save families in our community resources, as private tutors are often hired to help students with homework difficulties.

We're happy to have an in-person meeting with your board to further discuss any and all questions you may have. While our proposal goes into detail regarding the effectiveness of the AI program, to see the service in-action is its own unique experience. We would love for you to visit our campus and see a presentation!

With respect,  
Local School

The text **does not sufficiently explain the benefits** to students. It **does not sufficiently address why the AI tutoring service is necessary even with a teacher or preferable to a teacher's aide.**

This is an example of a 6 response

Re: AI Tutoring Service

To whom it may concern,

I am pleased to present to you this proposal for our school's AI tutoring service. We are passionate about improving educational outcomes for all of our students and look forward to partnering with you.

Our AI tutoring service will offer students on-demand access to help with their homework and test preparation. Traditional tutoring services, while a valuable resource for students, operate around the fixed schedules of tutors and can be quite expensive. The expense alone puts tutoring services out of reach for many students, especially those who come from low-income families. Our AI tutors makes academic success attainable for all students, no matter their families financial situation.

By utilizing our AI tutors, students can get help with homework assignments as they're completing them. The AI offers feedback on students' answers and can even guide them through an mistakes they might have made and show them the proper way to reach correct answers. Such a service levels the academic playing field for students who could not otherwise access tutors and may not be able to receive guidance from their parents while completing assignments.

Students can access this AI tutoring service at any time of the day, so they no longer have to worry about fitting in to a regular tutors schedule. Many students work jobs, attend clubs, and play for sports teams. Their schedules are often irregular, which adds an additional roadblock to receiving regular tutoring services. But with our AI, students can attend clubs and events that make their college applications stronger while still receiving the academic guidance they need to succeed in school.

In-the-moment feedback also improves students academic performance during the school year. They're able to receive guidance on their mistakes and corrections to their problem-solving methods before turning homework in. Historically, many low-income students have performed poorly on homework and tests compared to their high-income counterparts because their families cannot afford the academic support they need outside of school. Our AI improves these students chances to perform well in school and eventually attend college, and a college degree can help lift them out of the cycle of poverty. Our proposal requests \$5,000 in funding to bring this AI to our students via a chatbot service.

We want to offer all of our students an equal chance to succeed in school and hope that you'll work alongside us to bring these students a fair chance at success in school and life. Please feel free to contact us if you need any additional information or have an questions about our AI tutoring service.

Thank you

The text **generally flows well** and has **good tone and word choice**. The **benefits for students are well-laid out** and it is clear about **why the service is preferable** to a regular tutor. The text **sounds enthusiastic and is engaging**.

## C Supplementary Text

In this section we report supplementary analysis extending our main results, as well as a battery of tests and alternative specifications that ensure our main results are robust.

### C.1 Treatment-on-Treated Effects

Our experiment does not feature “perfect compliance:” it is not the case that 0% of control participants and 100% of treated participants use ChatGPT on the second task. Instead, 10-20% of control participants and 80% of treated participants do. In the main paper, we report “intent-to-treat” effects that do not adjust for imperfect compliance, in part for simplicity and in part because we think of “treatment” as being given the *option* to use ChatGPT rather than being forced to use ChatGPT (so we want to factor imperfect compliance in the treatment group into our estimates). Here, we also report results that adjust for imperfect compliance: we estimate the effects of using ChatGPT on our four main outcomes (time, grades, job satisfaction, and self-efficacy), instrumenting for ChatGPT usage with treatment assignment. Formally, we estimate our normal regression equation

$$\Delta y_i = \beta_0 + \beta_1 D_i^{GPT} + \theta_{\text{occ}(i)} \times \alpha_{\text{first}(i)} + \pi_{\text{inc}(i)} + \varepsilon_i \quad (\text{S.4})$$

by 2SLS, where the dummy  $D_i^{GPT}$  denotes using ChatGPT and we instrument for it with the treatment dummy  $D_i$ , and analogously for the regression equation when grades are the outcome. We report the results in tables S.1 and S.2.

### C.2 Task Structure and Skill Demand

**Task Structure** ChatGPT substantially changes the structure of writing tasks. Figure S.4 Panel A shows that prior to the treatment, participants spend about 25% of their time brainstorming, 50% writing a rough draft, and 25% editing. Post-treatment, the share of time spent writing a rough draft falls by more than half and the share of time spent editing more than doubles.

**Skill Demand** If ChatGPT is especially helpful to those with weak writing and communication skills, it could have major labor market implications. This could expand the available occupational choices and possibly raise the earnings of individuals that have strong idea-generation skills but struggle to effectively communicate those ideas in writing. We perform several tests of this hypothesis. Two measures of a person’s relative writing skills are constructed. First, at the beginning of the experiment, we asked participants to rank (from 1 to 3) their skills at communication (writing and speaking), problem solving, and creativity. The first column of Figure S.4 Panel (b) divides people by whether they ranked their writing skill first, second, or third. Second, in addition to assigning overall grades, evaluators separately assessed each piece of output based on writing quality, content quality, and originality; the gap between a person’s first-task overall score and their writing score affords another measure. The second column of Figure S.4 Panel (b) divides people by whether their writing score exceeded, equalled, or fell under their overall score.

Similarly, we construct two measures of the individual-level benefits of ChatGPT. First, at the end of the experiment, we asked treatment-group participants how much they would be willing to pay on a monthly basis to access ChatGPT in their jobs. This is the y-axis variable in the first row of Figure S.4 Panel (b). Second, we measure how much each treatment participant’s grade increased from the first to the second task. This is the y-axis variable in the second row of Figure S.4 Panel (b).

We find no clear evidence for the aforementioned hypothesis. Figure S.4 Panel B shows that average willingness to pay for ChatGPT is flat across the terciles of both our measures of writing skill: respondents, regardless of their writing skills, are willing to pay about 0.5% of their monthly salary for a monthly subscription to ChatGPT. Grade gains from ChatGPT are also roughly flat across both measures of relative writing skills: people with comparatively poor writing skills do not experience unusually large grade gains.

However, this analysis has important limitations. Mainly, it is not clear how meaningful the self-reported skill rankings are, and the two different skill measures (self-reported writing skill and writing-overall grade gap) are not correlated with each other. The willingness-to-pay variable and change in grades are also quite noisy as outcome variables, so these analyses are not very well-powered.

### **C.3 Job Satisfaction and Self-Efficacy**

Access to ChatGPT could affect job satisfaction in our experiment. For example, it could make participants happier by automating tedious or annoying components of the task or allowing them to finish tasks more quickly. Alternatively, it could make the experience less enjoyable by quickly automating the most enjoyable parts of the task. It could similarly either boost self-efficacy by giving participants access to a complex and powerful tool that enhances their capabilities, or lower it by making participants feel superfluous. We measure job satisfaction with a question, after each task, about how much participants enjoyed the task, and self-efficacy with a question about how skilled/effective they felt while completing the task, both on 1-10 Likert scales.

Results on job satisfaction and self-efficacy are depicted in Figure 3 in the main text. ChatGPT substantially increases job satisfaction, by about 0.50 standard deviations ( $p < 0.001$ ). It mildly and imprecisely increases self-efficacy, by 0.20 standard deviations ( $p = 0.05$ ), despite the fact that participants are mostly using it to substitute for their own effort. However, the self-efficacy result is not robust (see Robustness Checks section). Moreover, the self-efficacy measure we use is pretty limited and much less rich than more complex measures popular in the self-efficacy literature.

Qualitative feedback from participants (in an open-text box at the end of the survey) suggested that many enjoyed discovering and working with this tool.

### **C.4 Self-Reported and Objective Measure of ChatGPT Usage and Post-Usage Editing**

**Big Pasting Events** We want an objective measure of whether people used ChatGPT in the control group or on the pre-treatment task, to supplement our self-reported measure. We define a “big pasting event” as an instance where at least 100 words are modified in the respondent’s text box in a single minute, within the first 15 minutes of the task. Our objective measure of ChatGPT usage is the occurrence of a big pasting event, and we define the time of ChatGPT usage as the first occurrence of a big pasting event in the first 15 minutes of activity. Several pieces of evidence corroborate our interpretation of “big pasting events” as indicative of ChatGPT usage.

- On the pre-treatment task, such an event occurs 0 times for 363 of the respondents, once for 83 respondents, twice for 6 respondents, and thrice for 1 respondent. It hence seems to be an event that happens rarely without explicit ChatGPT usage instructions and, when it happens, happens only once.



- Among respondents who self-reported using ChatGPT on the first task, 63% experience a big pasting event on the first task, compared to 16% of those who did not report this.

**ChatGPT Usage Pre-Treatment and in the Control Group** After the second task, we ask all respondents whether they used ChatGPT on the second task. Respondents who answer “yes” are asked whether they also used it on the first task. Respondents are informed there are absolutely no consequences for reporting ChatGPT usage on either task and encouraged to answer honestly.

In the control group, 9% of respondents self-report using ChatGPT on the second task. This may be an underestimate because control respondents may be afraid to report usage despite our assurances. A more accurate self-reported measure might be the percentage of treatment group respondents who admit to having used ChatGPT on the first task (as, now that they have been explicitly allowed to use ChatGPT, the stigma associated with ChatGPT usage should be gone). Here, the self-reported usage rate on the first task is 12%.

As for our objective measure, 20% of control group respondents experience a big pasting event on the first task, which we view as an approximate upper bound on ChatGPT usage in the group (since this measure is a noisy proxy for ChatGPT usage).

**Post-Pasting Activity** We can get a rough sense of how much treatment respondents edit ChatGPT’s output by looking at how long they are active after they paste in the ChatGPT output, and at how many words they modify. 54% of treated participants who report using ChatGPT had a unique “big-paste” event; we restrict analysis to these respondents.

One might worry that this restriction selects for respondents who submitted ChatGPT’s response without editing, hence understating the total amount of editing that goes on in the treatment group. But in fact, this restriction overwhelmingly selects respondents who say they *did* edit ChatGPT’s output before submitting it. Specifically: we can separate the treatment group who used ChatGPT into three groups: those who say they submitted ChatGPT’s output without editing (58 people), those who say they took ChatGPT’s output and edited it (92 people), and those who reported doing something else, such as brainstorming with ChatGPT (23 people). This restriction selects 40% of the first group, 67% of the second group, and 35% of the last group, so actually positively selects for people who say they edited.

Hence we can safely restrict to these respondents. These respondents were active (modified >3 words in a given minute) for only 2.6 minutes on average post-paste, compared to 16 minutes total. If we restrict to respondents who reported editing ChatGPT’s output, this number is 3.3 minutes.

We can also get a sense of “how much” they modified ChatGPT’s output by computing the edit distance between their ChatGPT-pasted output and their final submission. This distance equals the minimum number of words that need to be added, subtracted, or modified to get from the ChatGPT-pasted output to the final output.<sup>4</sup> Among respondents who reported editing ChatGPT’s output, on average the Levenshtein distance is 30% of their final word count.

**Examples of Changes to ChatGPT Output** Below are three examples of how participants edited ChatGPT output. The percentage shown is the Levenshtein distance as a percentage of the final word count. Example 2 is representative of the typical kinds of edits made.

---

<sup>4</sup>We ignore case changes, punctuation changes, or line changes. Changes in word spelling/typos are counted as word changes. We drop observations where individuals paste ChatGPT output twice, since we cannot tell if they are iterating. We also drop a handful of observations where we believe the participant pasted ChatGPT output again a second time, but whose second paste is under 100 words.

### Example 1 (9.5% Edit)

Subject: Let's Reconnect with Our Virtual Office Space

Dear WorkCo employees,

~~I hope this email finds you well.~~ As you know, WorkCo has recently invested in a virtual office space, which was created to allow us to connect and collaborate as a team while working from home. With the aim of fostering casual conversations, fun office culture, and a sense of connection, the company has provided each of you with a virtual-reality headset that allows you to access ~~the~~ ~~this~~ ~~virtual~~ ~~office~~ space.

However, we have noticed that not many employees are using their headsets to access the virtual office. After investigating, we discovered that some employees find the headset slightly uncomfortable ~~and while some prefer~~ the privacy of working alone. We understand your concerns, and we are here to address them.

We want to remind you of the importance of using the virtual office space. The virtual space is designed to recreate the collaborative office environment that we all miss, and we believe that this is crucial for fostering innovation and motivation. It also helps us to stay connected with each other and maintain a sense of unity as a team.

We understand that the headsets may not be the most comfortable, but please be assured that we are working on making necessary improvements to enhance your experience. Also, please keep in mind that the virtual office is a space where you can work together with your colleagues and have casual conversations, just as you would in a physical office. It is a space where we can collaborate and learn from each other, which will ultimately benefit the company as a whole.

We urge you to give the virtual office a try and use your headset to connect with your colleagues. If you encounter any technical difficulties, please reach out to the IT support team, who will be happy to help you.

In conclusion, we encourage you to embrace the virtual office space and use it as a tool to enhance your collaboration, productivity, and motivation. Your participation in the virtual office is important to make it a success, and we look forward to seeing you there soon.

Thank you for your attention.

Best regards,

[Name] HR Professional, WorkCo

### Example 2 (28%)

Dear AccureCo Employees,

I hope this email finds you well! I am writing today to share some exciting news about our company. As many of you are aware, Sarah Davidson has recently taken over as CEO and ~~she~~ has been working hard to chart a new direction for AccureCo. One of the key initiatives she is working on is to reorganize our internal hierarchy to make it more effective and efficient.

~~We believe that~~ The current structure has become too vertically hierarchical and is slowing down decision-making and creating disharmony among employees. ~~By~~ Streamlining our

management tiers, we hope to improve collaboration and speed up the decision-making process, ultimately leading to better results for the company and its employees.

I understand that this change may raise concerns for some of our employees, particularly those who may face demotion. Please know that this decision was not taken lightly and was made with the best interests of the company and its employees in mind. ~~Please r~~Rest assured that the details of the change are still being worked out and everyone affected will be given the necessary support and resources during this transition. ~~Lets make sure to support our team members and one change details have been formed, we will address any concerns you felt have been left open ended~~

~~We are confident t~~That this new structure will bring benefits to most employees, including increased opportunities for growth, enhanced collaboration, and a more inclusive and equal workplace. This is an exciting time for AccureCo and I can't wait to see the positive impact that these changes will bring. ~~We are confident in this change!~~

~~As always, T~~thank you for your dedication and hard work. If you have any questions or concerns, please do not hesitate to reach out to HR or your manager.

Best regards,

[Your Name]

Senior Manager, AccureCo

### Example 3 (82% edit)

Dear Colleagues,

We hope you are enjoying your week ~~Remote working has become normal for most of us, and it is something most, if not all, of us enjoy. In order to As we all continue to work from home, it is important that we~~ maintain work relationships and a sense of community ~~and collaboration within our company,~~ we would like to encourage you to use the virtual-reality headset we recently sent. This easily ~~—With this in mind, we recently provided each of you with a virtual-reality headset that~~ allows you to connect to our virtual office space.

While we understand ~~that the new technology may seem a bit uncomfortable at first and that you may prefer the privacy of working alone at home~~most of you enjoy privacy and being comfortable in your own home even while working, it is ~~erueial~~vital that we maintain these parts of community within our company to optimize our efficiency and overall work performance. ~~make the most of our virtual-office space.~~ This space will allow us to communicate, collaborate, and even unwind. While working remotely, you will find the coworker relationships and sense of a work community within this work space, which can be vital when keeping you focused on work tasks, seeking help from others, or simply just avoiding the 'alone' on an island work feeling~~provides us with an opportunity to feel connected, to have casual conversations, and to create a fun office culture – all of which are essential to collaboration, innovation, and motivation.~~

~~In order to make the virtual office more attractive and effective, we need your participation. By simply logging into the virtual office and using your headset for a few hours a day, you can make a significant impact on our overall work culture. Plus, we are always looking for ways to improve the technology, so if you have any feedback on the headset's comfort,~~

~~please let us know.~~ You can think of this space like a break room or these sessions like coffee breaks. Just by a few hours a day, you can significantly impact your overall work culture along with overall company culture and community. We will all benefit from collaboration, innovation, and motivation.

~~We understand that some of you may be hesitant to use the virtual office, and that is okay. However, we encourage you to give it a try and see how it can benefit you and your colleagues. We believe that by working together in this virtual space, we can continue to build a strong community and achieve great things as a team.~~ We ask that you give this a try for at least a week and let us know honest feedback. If you find you are having issues with the software, headset itself, or have other concerns to voice about this space, please let us know. We are eager to understand how you feel but also how we can best tackle work unity and community even while working remotely.

Thank you for ~~your taking the time, to read this email,~~ and we all ~~We~~ look forward to seeing you in the virtual office.

Best regards,

[Your Name] HR Professional, WorkCo.

## C.5 Comparison to Pure ChatGPT Output

We ask graders to grade pure ChatGPT output, as produced by us. We begin with a blank chat, directly paste in the full prompt, and take multiple versions of the output (either through the “Regenerate response” option or by re-pasting the prompt), to get some variance so that graders are not literally seeing identical essays. For the consulting task, we paste in the full text of each document as well. In most grading sessions, graders see two ChatGPT responses produced by us, of their total 14 essays. (We introduced this format partway through the study, so some grading sessions have no pure ChatGPT responses, and a few grading sessions have multiple pure ChatGPT responses.) Table S.3 shows summary statistics on pure ChatGPT grades.

On average, the pure ChatGPT responses have slightly higher grades than that of the humans using ChatGPT due to worse performance in the data analyst, grant writer, and marketer occupations. As shown in Figure S.5, this is driven by individuals who are using ChatGPT and modifying the output to be worse. Qualitatively, most edits are minor (see Examples of Changes to ChatGPT Output in C.4 Self-Reported and Objective Measures of ChatGPT Usage); some appear to introduce errors, such as typos or grammatical mistakes, and flow less well than the original output.

## C.6 Other Descriptive Figures

**Task Realism** Figure S.6 plots answers to a question we asked on the pre-treatment task about how realistically the task imitates real tasks performed by people in the respondent’s occupation, on a 1-5 point scale. It also reports the percentage of respondents who said they had completed a similar task before in their job.

**Task Compliance: Time Spent and Word Counts** Figures S.7-S.9 check that our participants are complying with task instructions and not, for example, spending a lot of time off-task. We separate respondents in the linear/convex incentive groups from respondents in the exact-time group, who are required to spend exactly 15 minutes on each task page.

We plot 3 measures of time spent on the tasks: self-reported time, Qualtrics’s record of how long the respondent spent on the task page, and the number of minutes “active,” where a respondent is defined to be active in a given minute if they modify at least 3 words in the text box. Minutes active can be less than total self-reported time because respondents are taking time to read the prompt, because they are spending time brainstorming (literally) outside of the box, or because they are overreporting their time taken and are in fact off-task for some of the time they’re on the page.

We also plot word counts on the tasks, both distributions of final wordcounts and the evolution of wordcounts in the treatment and control groups before and after treatment.

**Cross-Evaluator Reliability** Figure S.10 plots the average correlation between the grades of different evaluators seeing the same essay. See figure note for calculation details. This figure plots the results after we drop graders who are completely uncorrelated (correlation  $< 0.1$ ) with other graders seeing the same essays, which is what we do for our main results (see associated robustness check below). Figure S.17 plots the correlation *including* the bad graders.

**How People Use ChatGPT** Figure S.11 plots people’s responses to a question we asked, after the post-treatment task, of everyone (in both treatment and control groups) who reported using ChatGPT on the second task. We asked how they used ChatGPT and gave a list of options; note people can select multiple options, and this figure plots the percentage of users who ticked each option.

**Results for Exact-Time Group** The figures in the main text use the linear and convex incentive groups, which constitute 80% of respondents. The remaining 20% of respondents are required to spend exactly 15 minutes on each task, holding effort fixed. The treatment effects for their grades and wordcounts are presented in Figure S.12. Note that due to the small sample size in this arm, there is an imbalance in pre-treatment outcomes: average grades are different between treatment and control. As long as we assume that the treatment and control groups would have experienced identical trends in grades between Task 1 and Task 2 in the absence of treatment (i.e., the difference-in-differences identifying assumption), the estimates in this figure should still reflect the causal effects of the treatment.

**Results for Other Grade Categories** In addition to overall grades, our evaluators give each essay a separate grade for writing quality, content quality, and originality. Figure S.13 plots treatment effects for these separate grades.

**Alternative Versions of Inequality Figures** In the main text Figure 2, we report binned scatterplots of Task-2 grades on Task-1 grades, and Task-2 time on Task-1 time. For completeness, we report the analogous pairs in Figure S.14: Task-2 time on Task-1 time, and Task-2 grades on Task-1 time.

## C.7 Robustness Checks

**Selective Attrition** Treatment respondents are slightly less likely to complete the survey than control respondents, resulting in a small treatment-control imbalance in terms of employment status and occupational composition. Note firstly that given that we control for Task-1 outcomes, attrition would need to be selective on *trends* to distort our results, which seems implausible. Nevertheless, in Figure S.15 we directly test whether our results are robust to selective attrition, in two ways: we report our main results controlling for the imbalanced characteristics, and we report Lee bounds (24) on our main results.

**Grade Results, Not Dropping “Bad” Graders** In our main results, we drop evaluators whose grades are completely uncorrelated (correlation  $<0.1$ ) with the grades of other evaluators seeing the same essays, as a proxy for evaluators who aren’t paying attention or taking the task seriously. This turns out not to really matter for our results: we report our main grade results (treatment effect on grades, and grade inequality), as well as the inter-grader reliability, when we do *not* drop bad graders, in Figures S.16 and S.17.

**Main Results, Restricting to Experienced Taskers** 68% of our respondents said, after the pre-treatment task, that they had completed similar tasks in their real job before. In Figure S.18 we replicate our main results restricting to these respondents.

**Main Results, Excluding Managers** Our most common occupation was managers, who constituted around 40% of our sample. In Figure S.19 we replicate our main results excluding managers.

**Main Results, Excluding Data Analysts** We adjusted the data analyst task midway through data collection for reasons described in Section A.3 and some of the early data analyst task submissions are of dubious quality. In Figure S.20 we replicate our main results excluding data analysts.

**Objective Time Measure** In Figure S.21 we replicate our treatment effects on time, using our objective measure of “time active” (number of minutes in which the respondent added or modified at least 3 words) instead of the self-reported time measure used in our main analyses.

## D Supplementary Figures

### D.1 Methods: Grader Characteristics

Figure S.2: Grader salary distribution

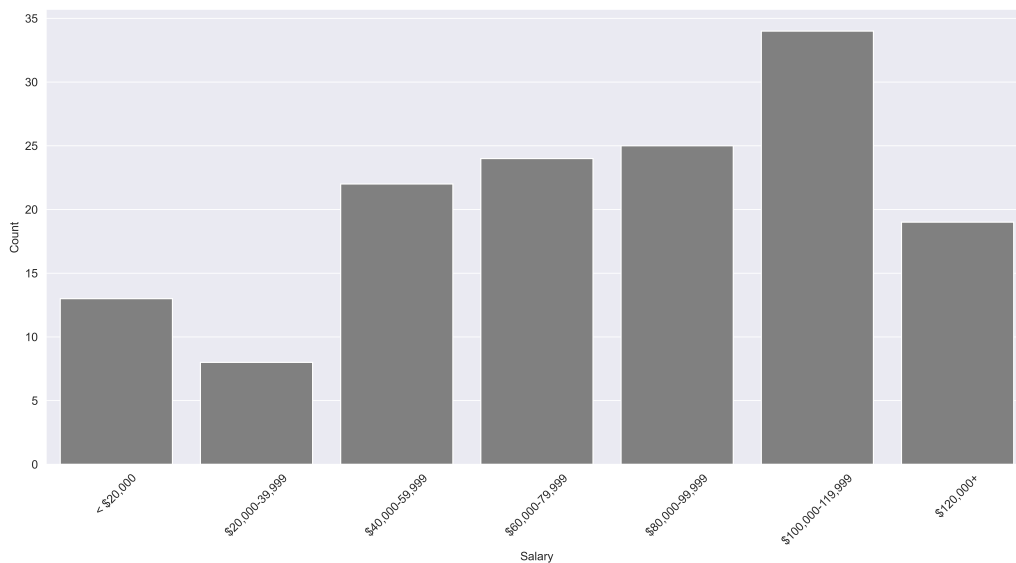
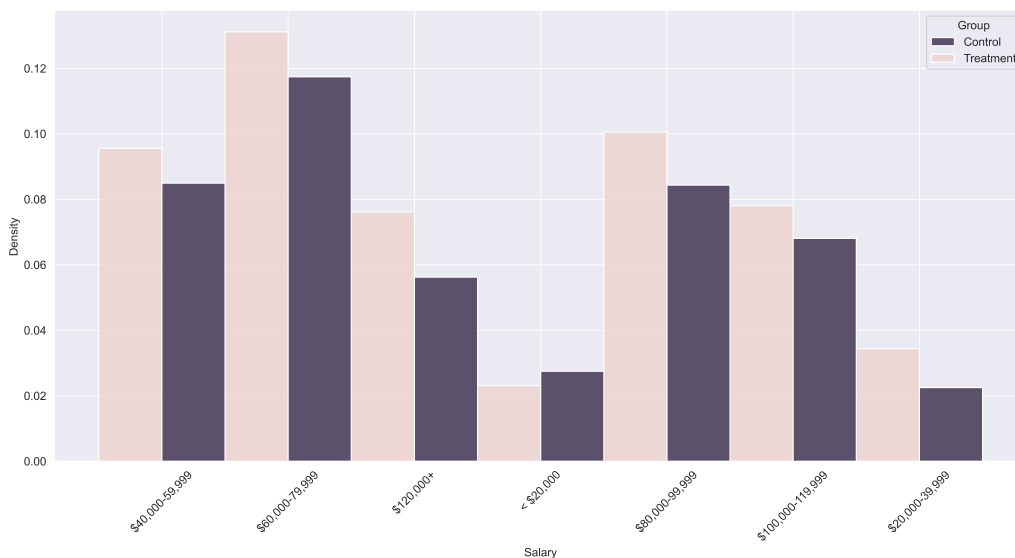
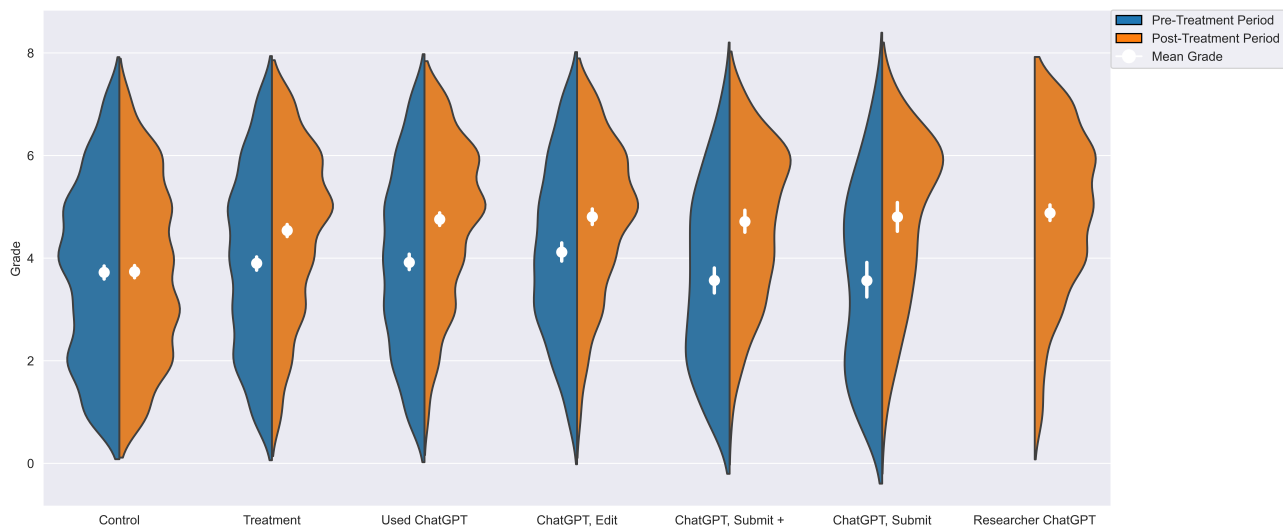


Figure S.3: Grader salary distribution by group



## D.2 Additional Results

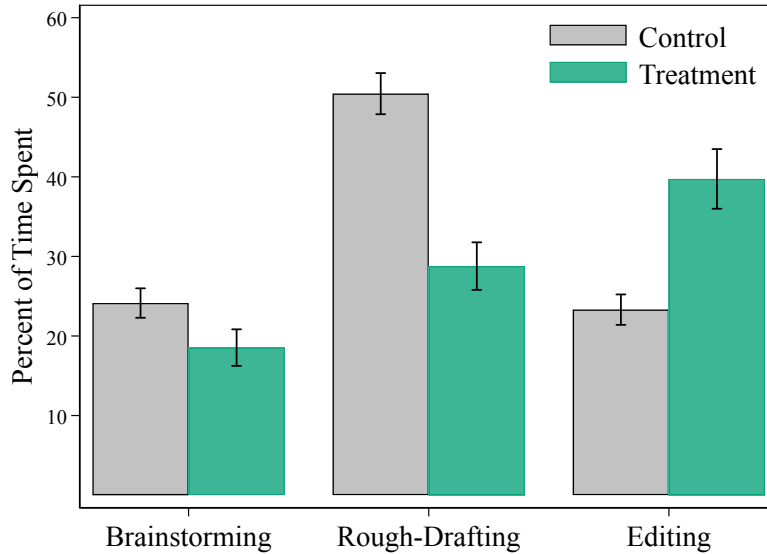
Figure S.5: Grade Distributions



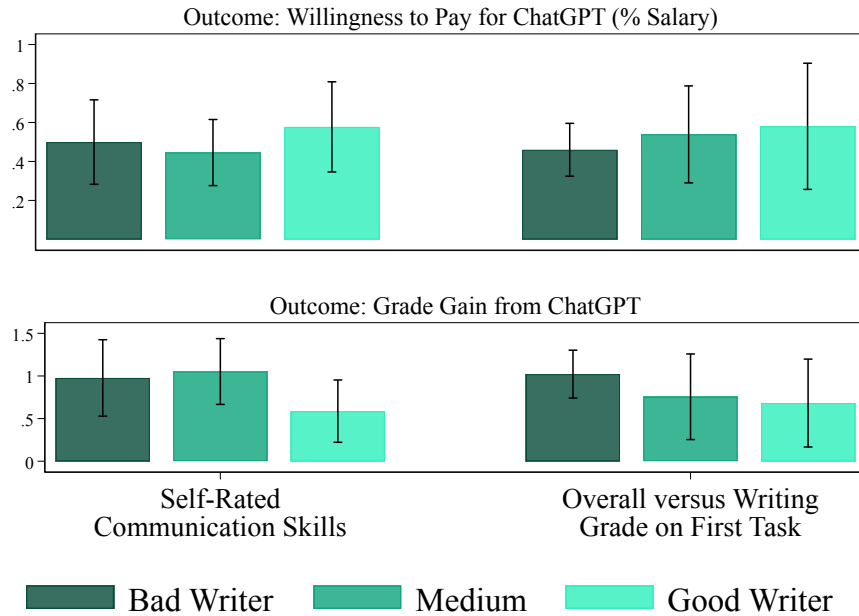
*Note:* This figure shows the distribution of grades across different groups. “Control” is all control group responses, “Treatment” is all treatment group responses, “Used ChatGPT” is all responses that used ChatGPT, “ChatGPT, Edit” is all responses that used ChatGPT and edited its output, “ChatGPT, Submit+” is all responses that used ChatGPT and stated that they directly submitted ChatGPT output (but may have selected other options), “ChatGPT, Submit” is all responses that only selected that they directly submitted ChatGPT output, and “Researcher ChatGPT” is the ChatGPT responses we produced. The blue distributions are the distribution of grades in the pre-treatment period, and the orange distributions are the distribution of grades in the post-treatment period. The mean grade in each distribution is shown with a white dot with 95% confidence interval error bars.

Figure S.4: Effects on Task Structure and Skill Demand

(a) Effects on Task Structure



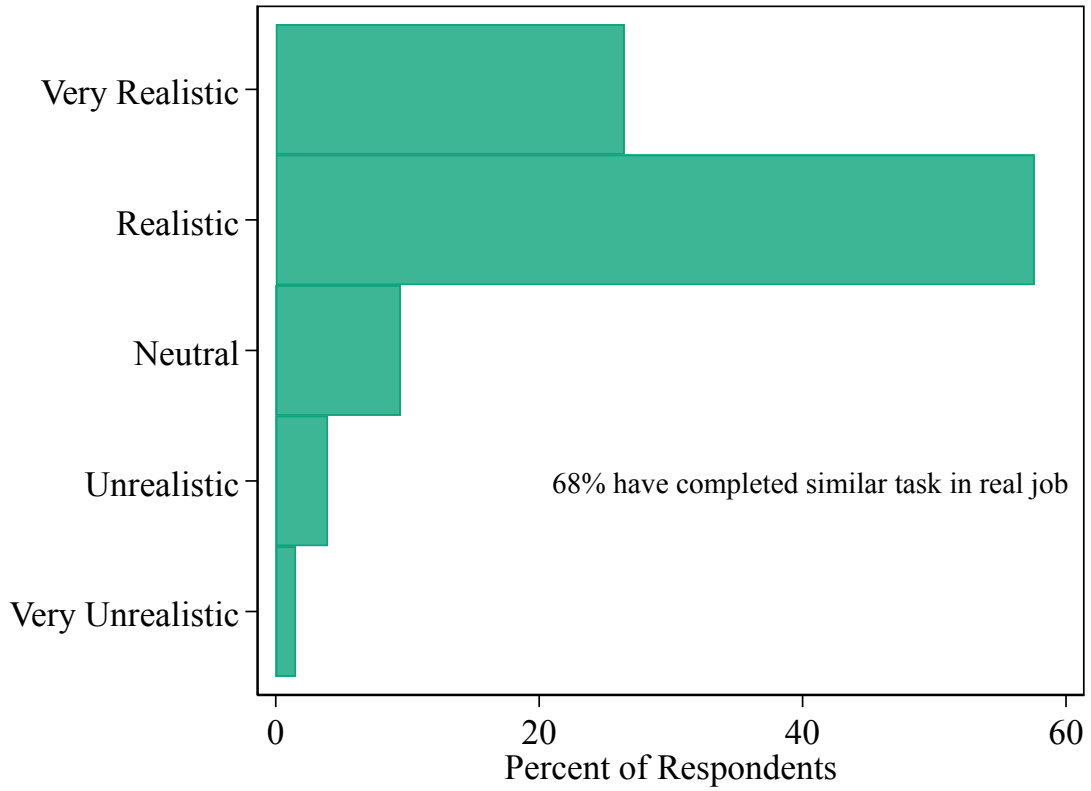
(b) No Clear Heterogeneity in Benefits of ChatGPT by Relative Writing Skills



*Note:* Panel (A) shows self-reported time allocation on three separate task components, separately in the treatment and control group on the post-treatment task. Panel (B) shows that ChatGPT does not have greater benefits for treated respondents whose writing skills are poor relative to their other skills. Specifically, in the left-hand column, respondents are sorted by their self-assessed skill rankings: at the start of the survey, they rank their communication, problem-solving, and creativity skills. Those who rank communication 1st are defined as “Good” communicators, similarly for those who rank it 2nd and 3rd. In the right column, participants are sorted according to their grades: evaluators give each essay a separate overall, writing quality, content quality, and originality score. We define people as “Good” communicators if their average writing score exceeds their overall score, (about 30% of respondents), “Medium” if the two are equal (50%), and “Bad” if overall exceeds writing (20%). In the top row, the outcome is how much the participant is willing to pay to access ChatGPT in their job (elicited hypothetically). In the bottom row, the outcome is the participant’s grade gain between the first and second task (everywhere we restrict to treated participants). Flat slopes within each of the



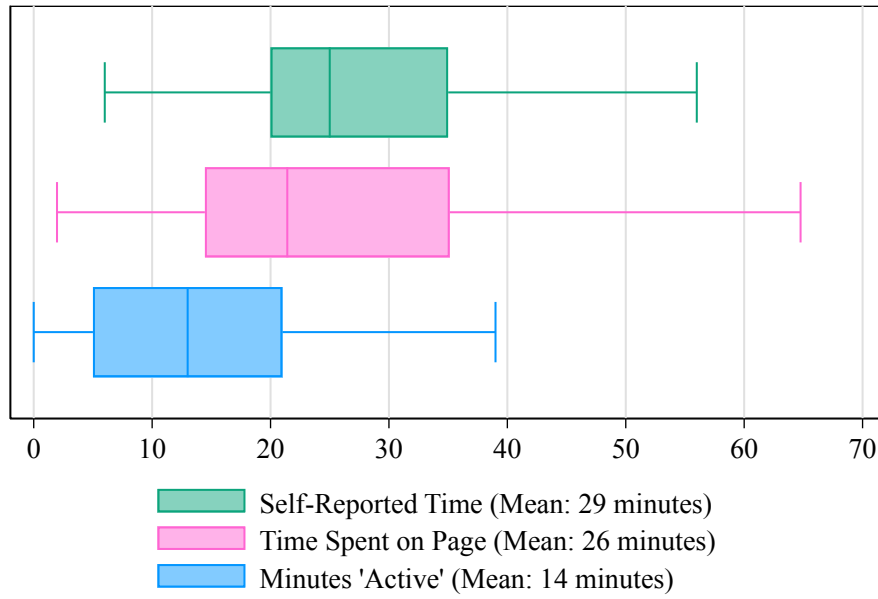
Figure S.6: Task Realism Scores



*Note:* After the first task, participants are asked to rate how realistically the task imitates real tasks performed by people in their occupation, on a 1-5 scale. This is the histogram of responses to that question. Additionally, participants are asked if they have performed a similar task in their job before; 68% say yes.

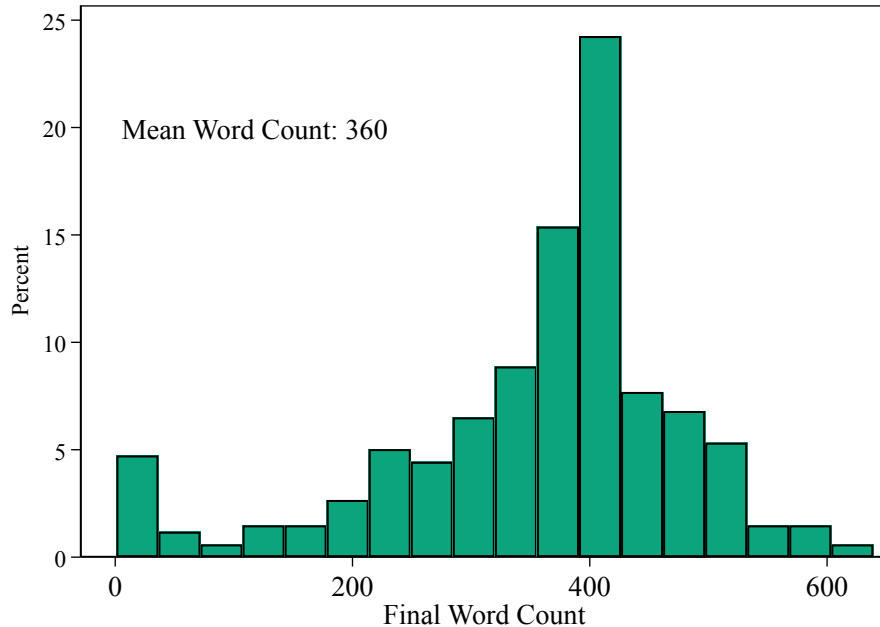
Figure S.7: Task Compliance (Linear+Convex Incentive Groups, First Task)

(a) Distribution of Time Taken (3 Measures)



excludes outside values

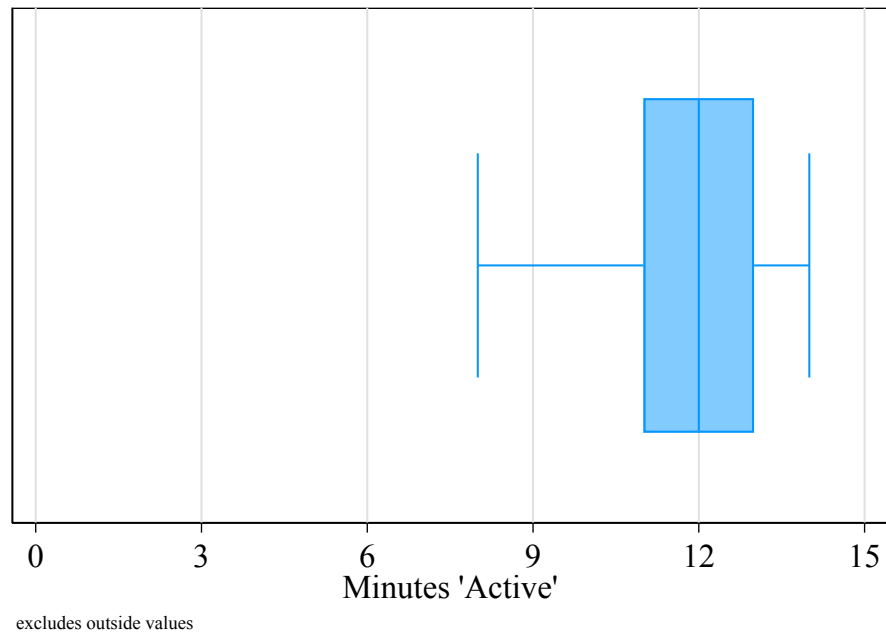
(b) Distribution of Final Word Count



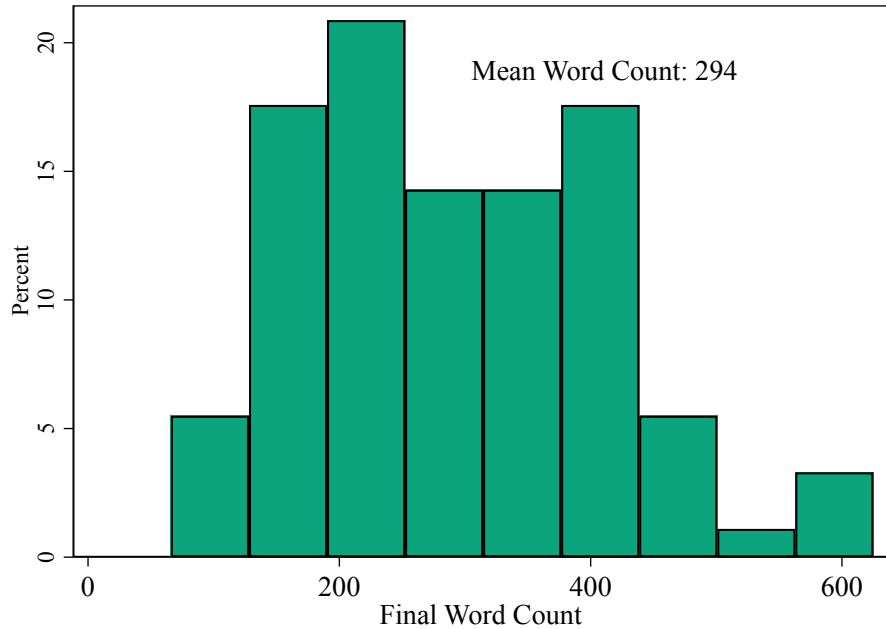
*Note:* This figure shows that respondents (in the “linear” and “convex” incentive groups) comply with our instructions and spend significant time and effort completing the first task. Panel (a) shows the distribution of time taken according to 3 measures: participants’ self-reports of their total time spent, their time spent on the survey page, and “minutes active.” The final measure is calculated by counting the number of minutes in which the respondent adds or modifies at least 3 words in the task text box. Minutes active can be less than total self-reported time because respondents are taking time to read the prompt, because they are spending time brainstorming (literally) outside of the box, or because they are overreporting their time taken and are in fact off-task for some of the time they’re on the page. Panel (b) shows the distribution of final word counts on the first task.

Figure S.8: Task Compliance (Exact-Time Group, First Task)

(a) Distribution of Time Taken (Objective “Time Active” Measure)

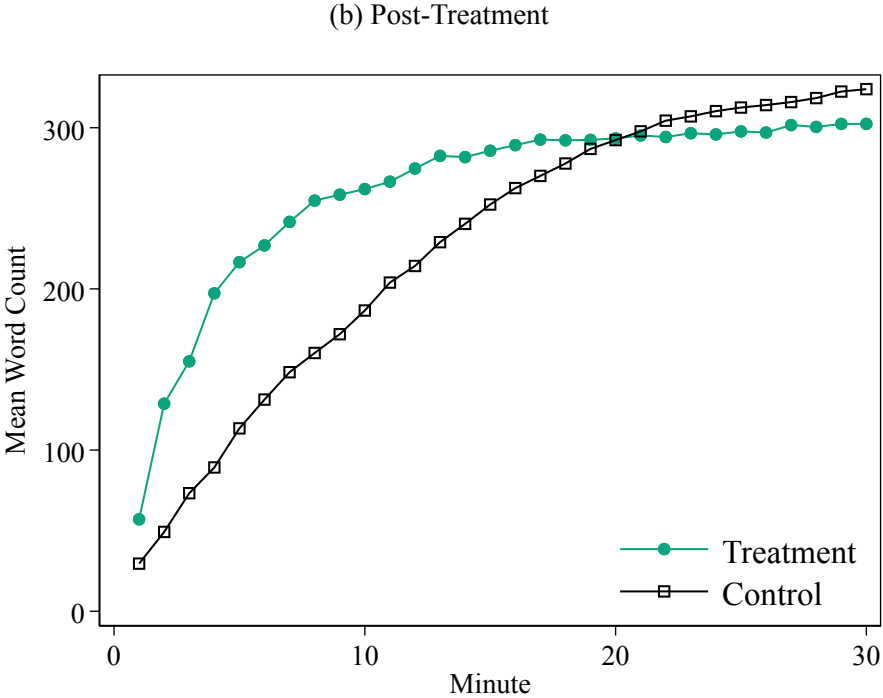
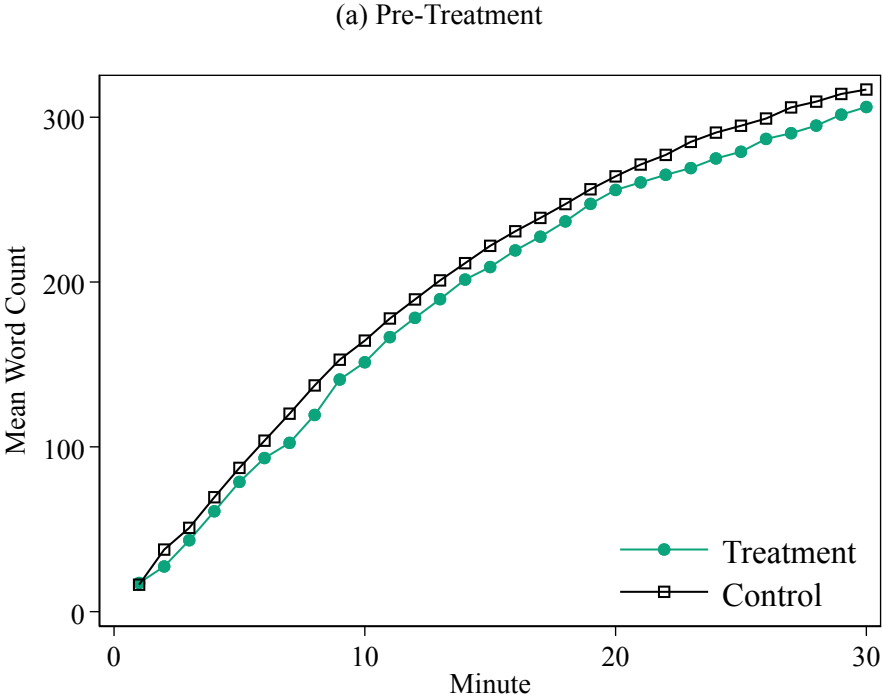


(b) Distribution of Final Word Count



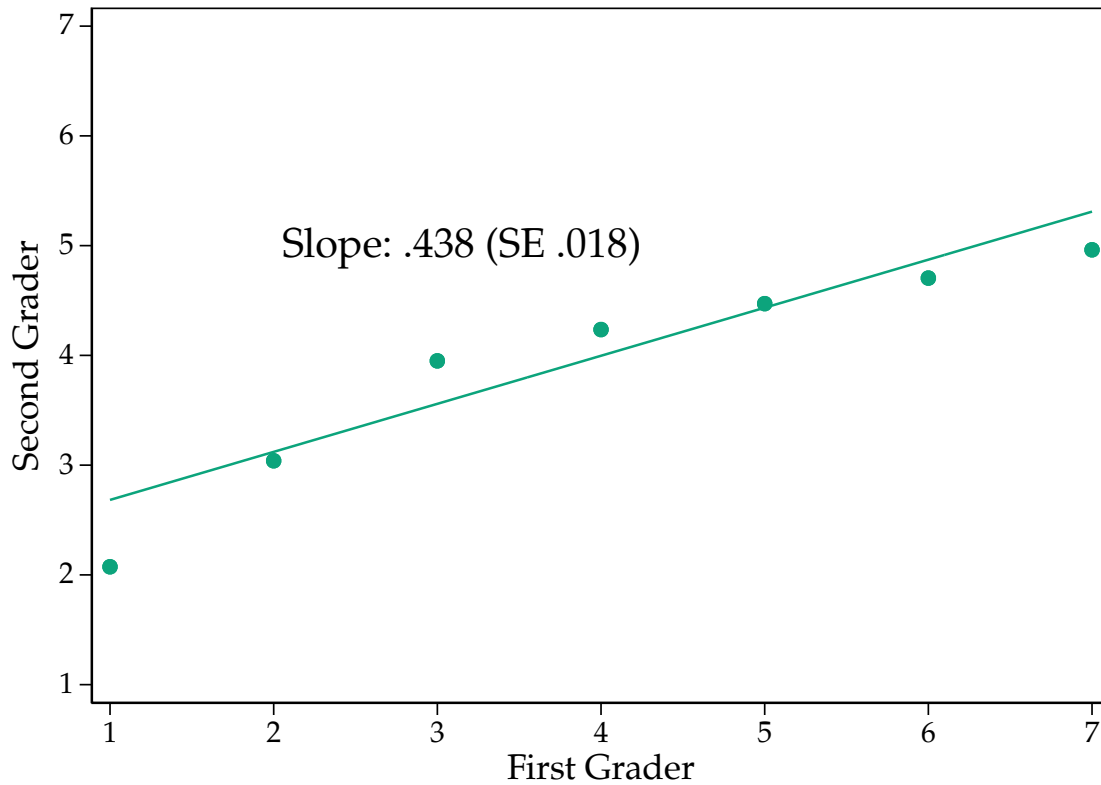
*Note:* This figure shows that respondents (in the “exact time” group, who are required to spend exactly 15 minutes on the task) comply with our instructions and spend around 15 minutes completing the first task. Panel (a) shows the distribution of time taken according to our objective “minutes active” measure, which is calculated by counting the number of minutes in which the respondent adds or modifies at least 3 words in the task text box. Minutes active can be less than total self-reported time because respondents are taking time to read the prompt, because they are spending time brainstorming (literally) outside of the box, or because they are overreporting their time taken and are in fact inactive for some of the time they’re on the page. Panel (b) shows the distribution of final word counts on the first task.

Figure S.9: Word Count Evolution in Treatment vs Control (Linear+Convex Arms)



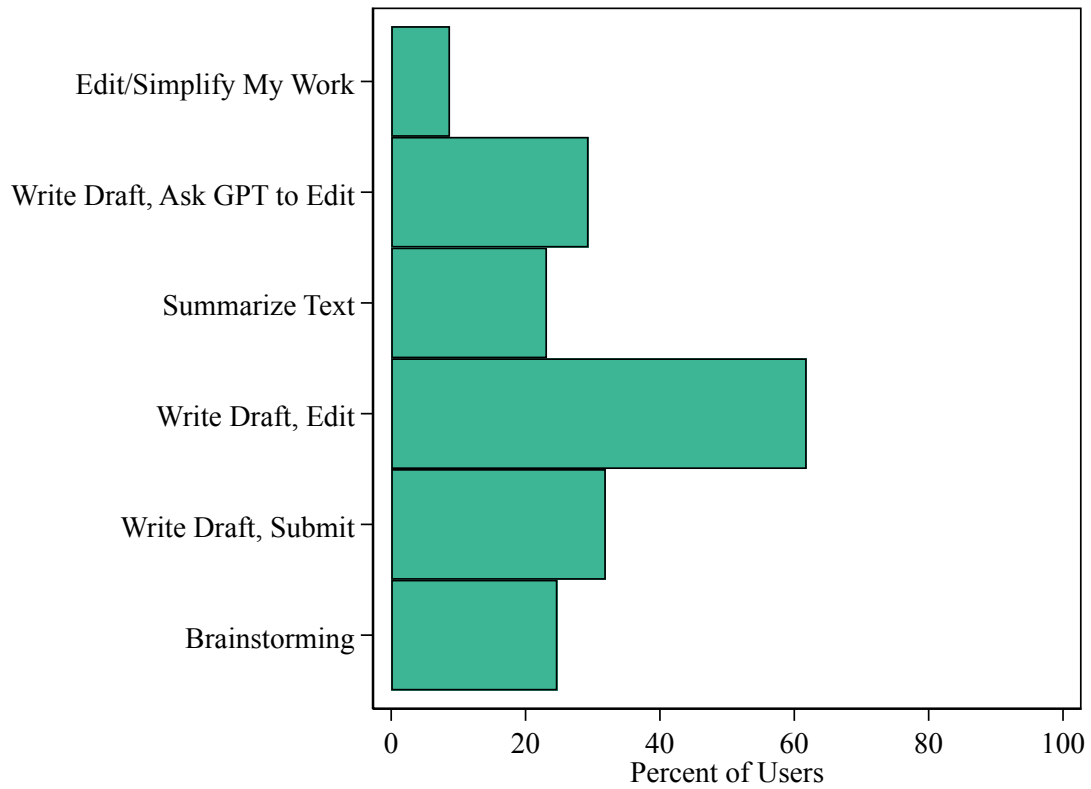
Note: This figure uses our snapshots of the respondent’s essay box each minute to plot mean word counts in the treatment versus control groups, before and after treatment, on the tasks. We restrict to the linear and convex incentive arms and include already-submitted essays at each minute (i.e., when a respondent submits their essay, their word count stays at its final value in every subsequent minute).

Figure S.10: Within-Essay Cross-Evaluator Grade Correlation



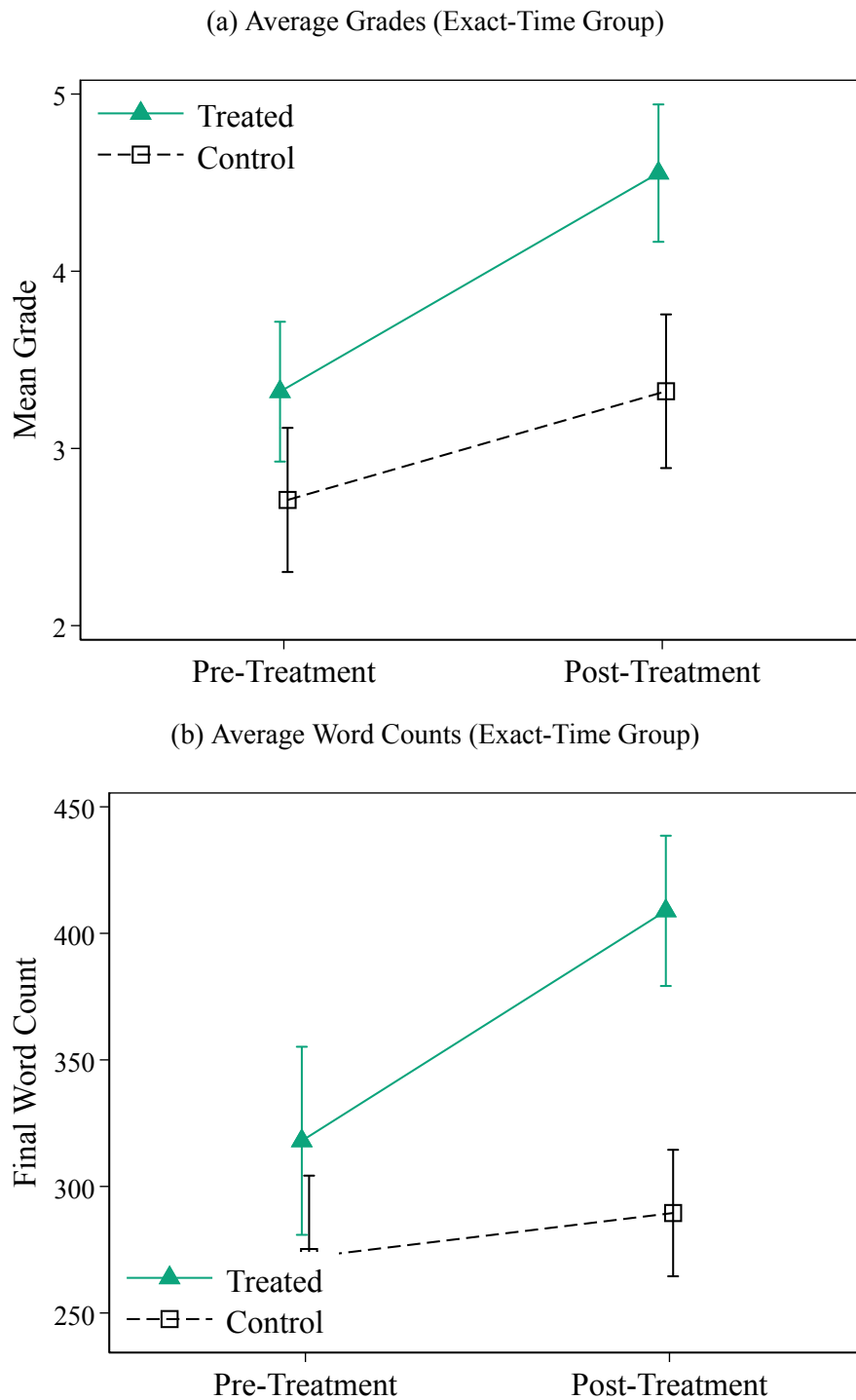
*Note:* This graph displays a binned scatterplot of the cross-grader within-essay grade correlation in our dataset. Specifically, we create a dataset where an observation is a unique grader1-grader2-essay pair, and then regress grader1score on grader2score. This plot represents the correlation after dropping graders who are totally uncorrelated with other graders seeing the same essays, defined as a correlation with those other graders of  $<0.1$ . The corresponding version of this slope when *not* dropping those graders is 0.34.

Figure S.11: How People Use ChatGPT



*Note:* Respondents who report using ChatGPT on the second task (both treatment and control group) are asked how they used it from a set of options. They can tick multiple. This figure shows how many respondents selected each option.

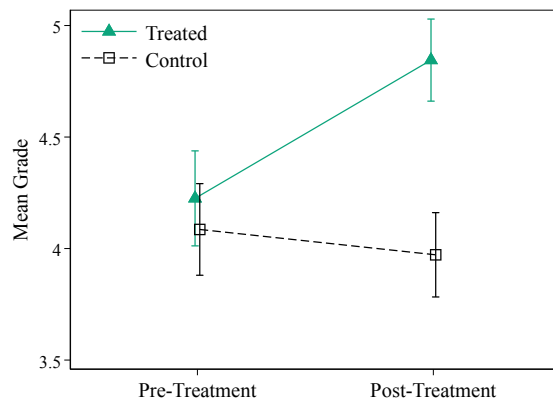
Figure S.12: Grade and Wordcount Effects for Exact-Time Group



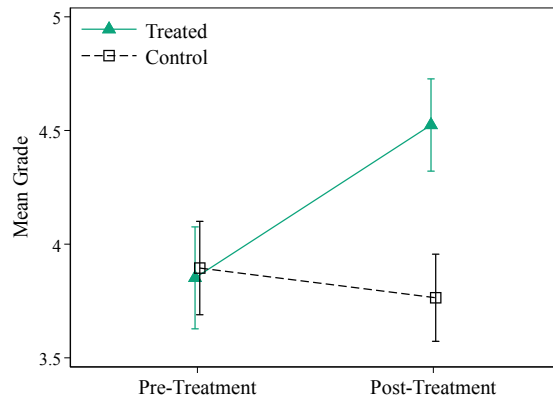
*Note:* This figure shows average grades and average word counts, before and after treatment, separately by treatment and control in our “exact time” group (where participants are required to spend exactly 15 minutes on each task). There is almost zero selective attrition in this incentive arm so we think the pre-treatment outcome imbalance is just due to random chance (the number of observations in this incentive arm is only 94). Our estimates will still identify the causal effects of ChatGPT under the standard difference-in-differences assumption of parallel trends in the absence of treatment, which we think is plausible.

Figure S.13: Grade Effects for Subgrades

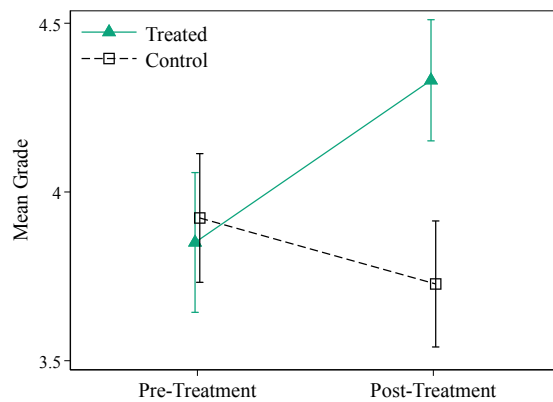
(a) Writing Quality Grades (Linear+Convex Groups)



(b) Content Quality Grades (Linear+Convex Groups)



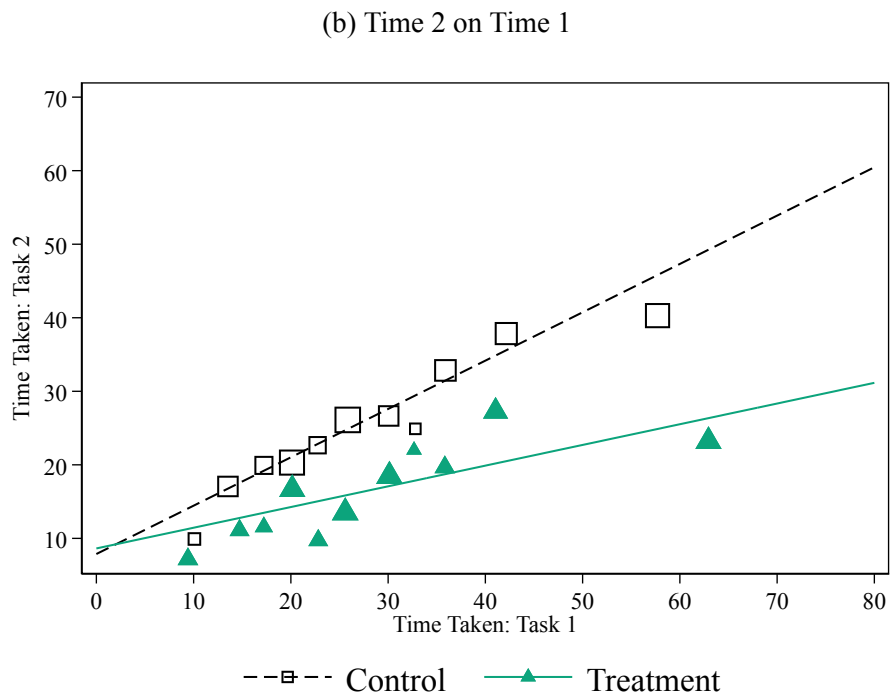
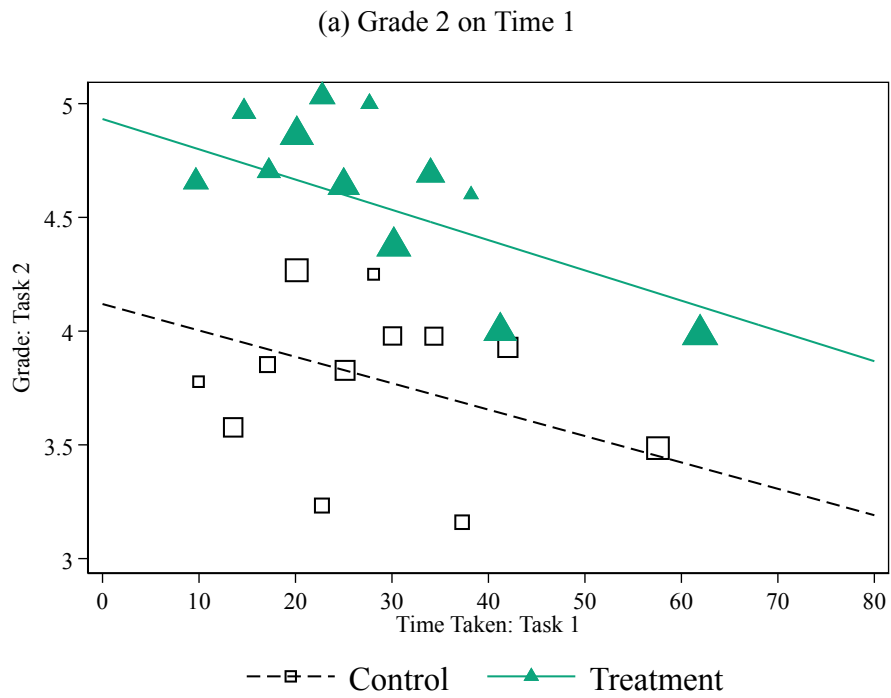
(c) Originality/Creativity Grades (Linear+Convex Groups)



*Note:* This figure shows mean subgrades for “writing quality,” “content quality,” and “originality/creativity,” before and after treatment, separately by treatment and control in our linear and convex incentive groups. Evaluators separately give an overall grade and subgrades for writing quality, content quality, and originality.



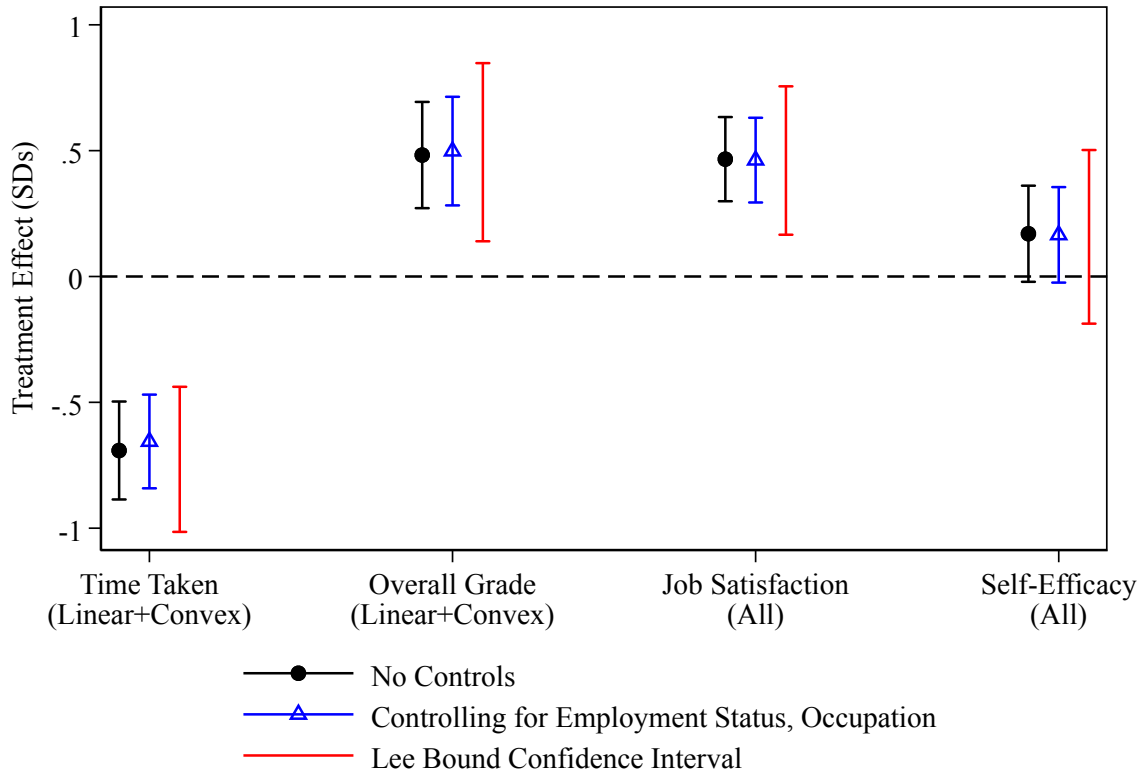
Figure S.14: Alternative Versions of Inequality Figures



*Note:* This figure complements Figure 2 by giving the other two cells in the 2x2 quadrant of [grade 2, time 2] on [grade 1, time 1] separately by treatment and control. “1” and “2” denote pre-treatment and post-treatment tasks, respectively.

### D.3 Robustness Checks

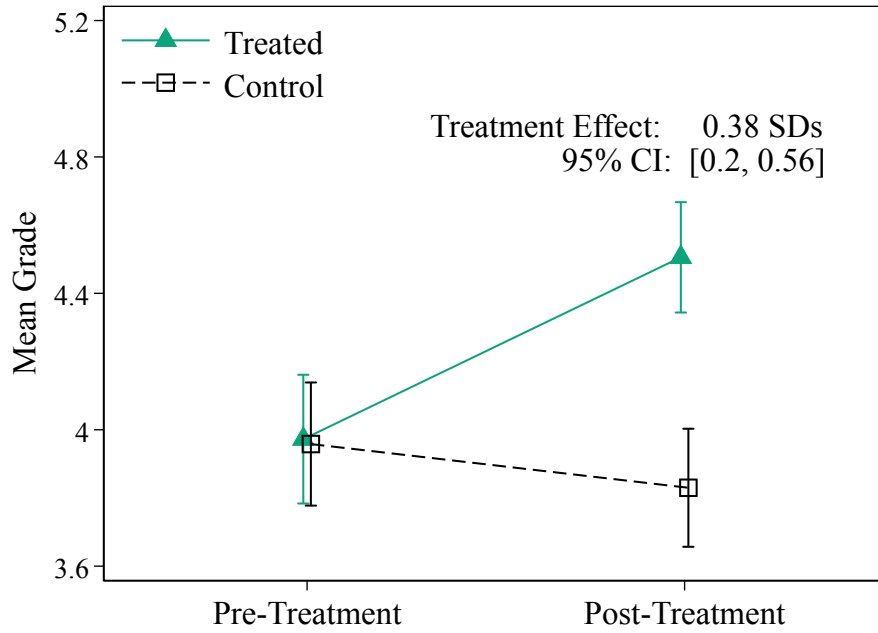
Figure S.15: Robustness of Main Results to Selective Attrition



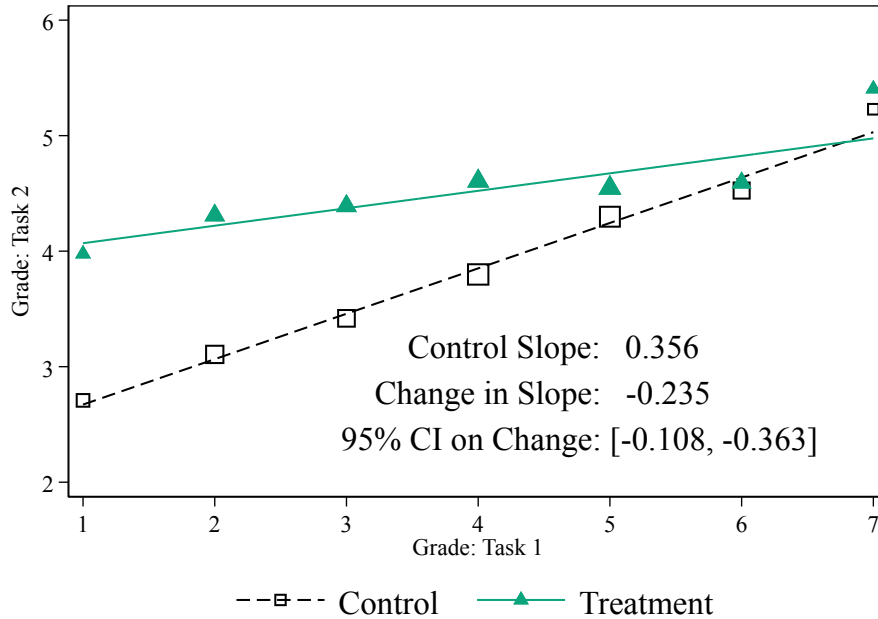
*Note:* This figure shows robustness of our main treatment effects to two corrections for selective attrition: Lee bounds (24), and controlling for employment status (the only variable that exhibits a significant treatment-control difference in our balance tests; our main estimates already control for occupation fixed effects). Error bars are 95% CIs. The main coefficients correspond to the ones printed in Panels A and B of Figure 1 and Panels A and B of Figure 3; they are slightly different in this figure because, to simplify the implementation of Lee bounds, we estimate them at the participant level (overall grade is averaged across all graders) with no controls (as opposed to at the participant-grader level with grader fixed effects and occupation/incentive-arm fixed effects).

Figure S.16: Grade Results, Not Excluding Bad Graders

(a) Main Treatment Effect on Grades

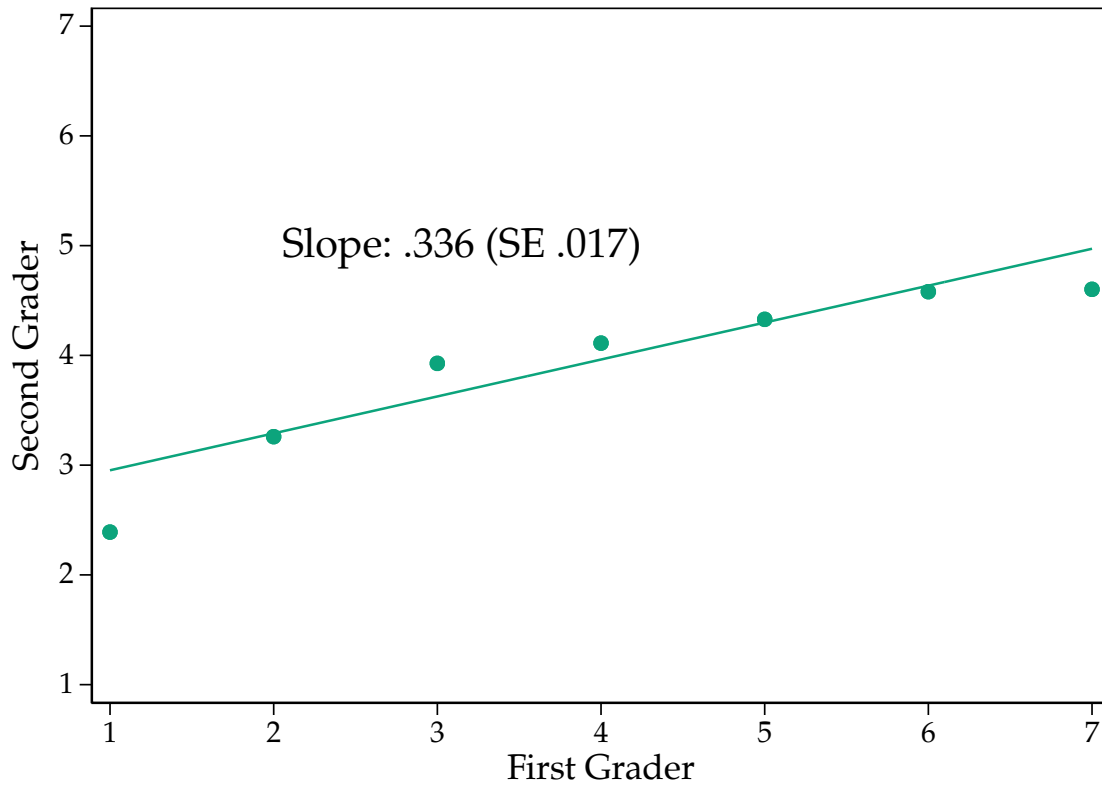


(b) Inequality Graph



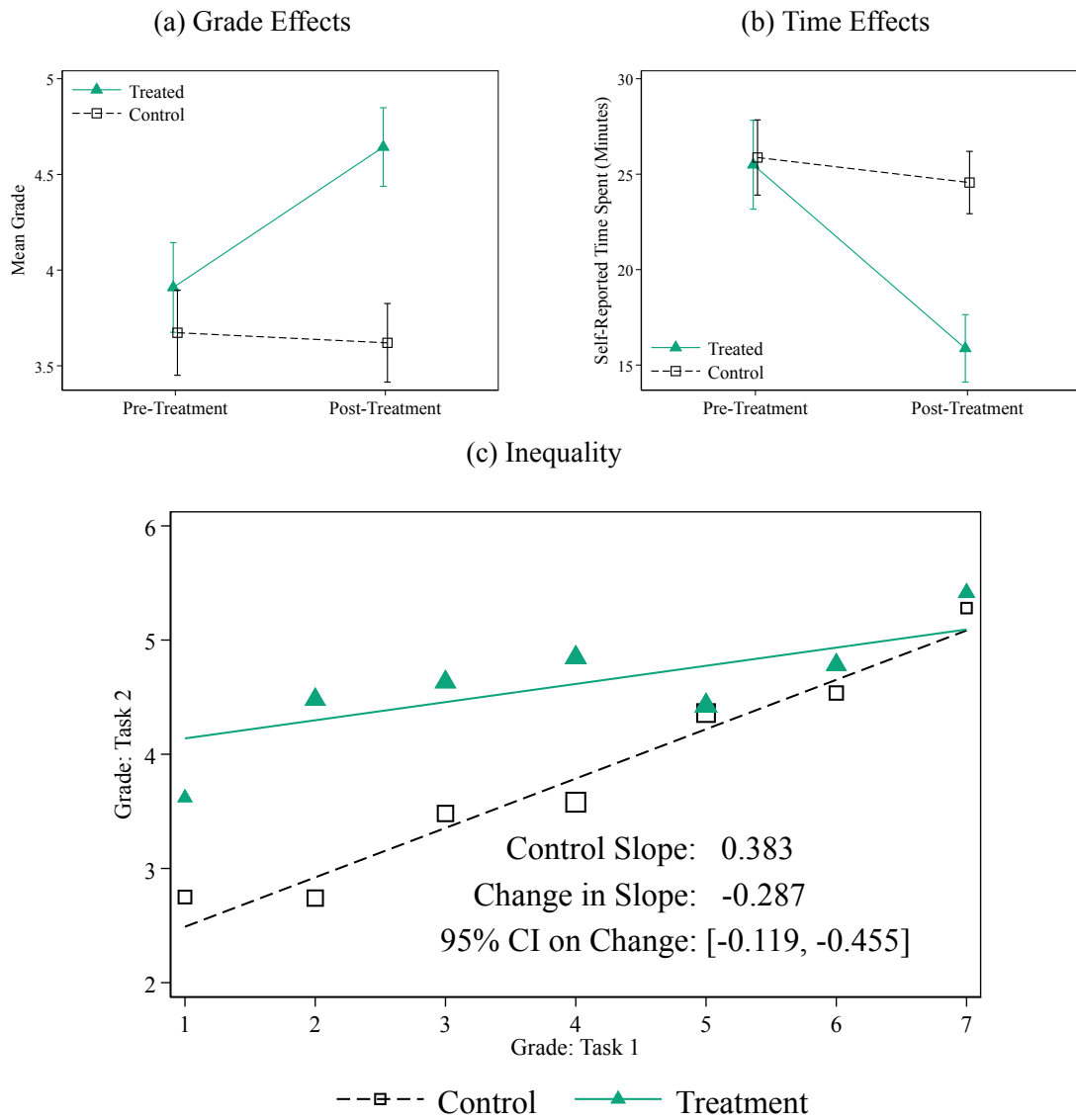
*Note:* This figure replicates our main grading results without excluding “bad” graders (graders whose grades have a correlation  $<0.1$  with the grades of other graders seeing the same essays, whom we drop from our main analyses).

Figure S.17: Within-Essay Cross-Evaluator Grade Correlation, Not Excluding Bad Graders



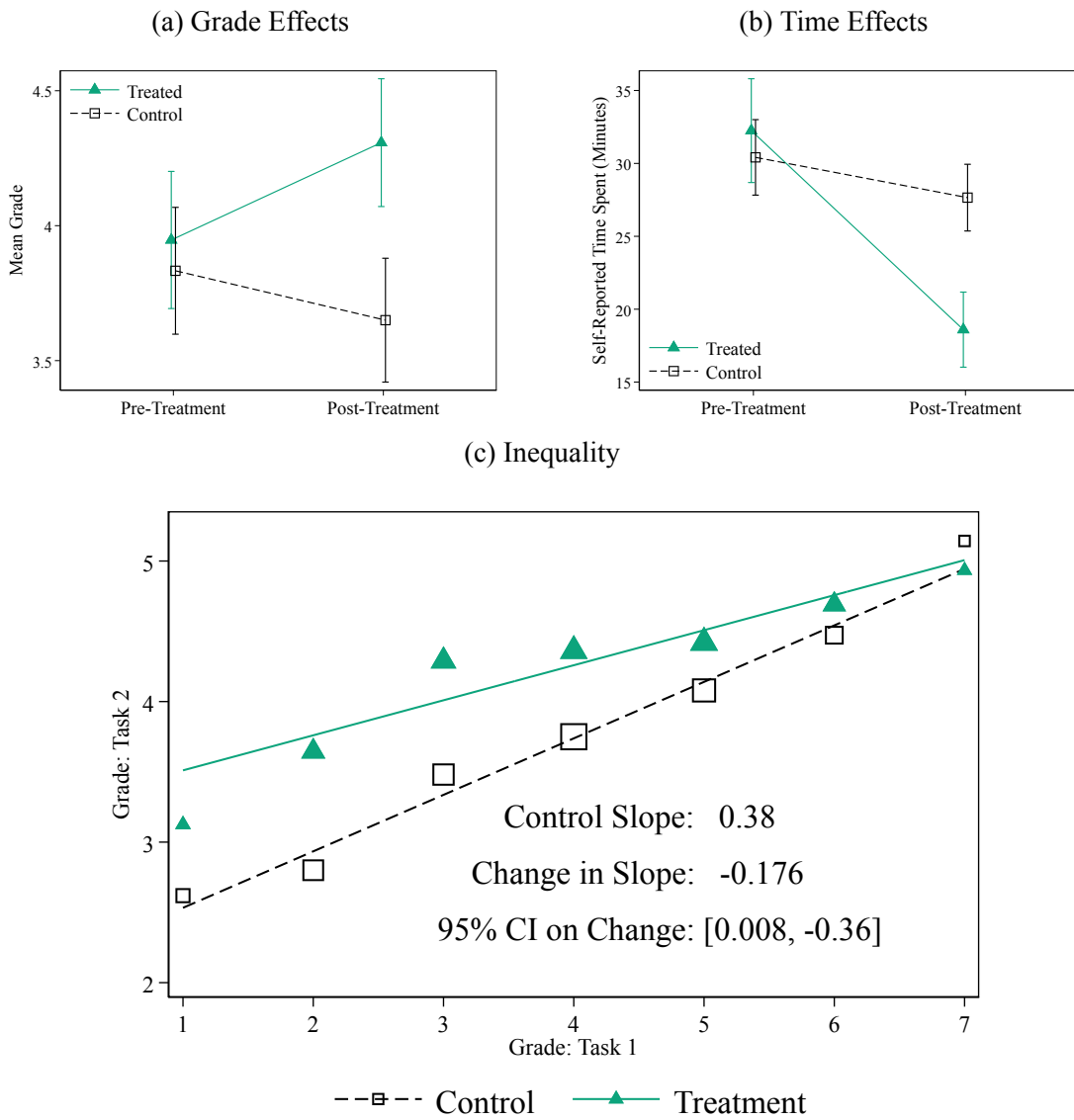
*Note:* This replicates Figure S.10, but without dropping “bad” graders first.

Figure S.18: Main Results for Experienced Taskers



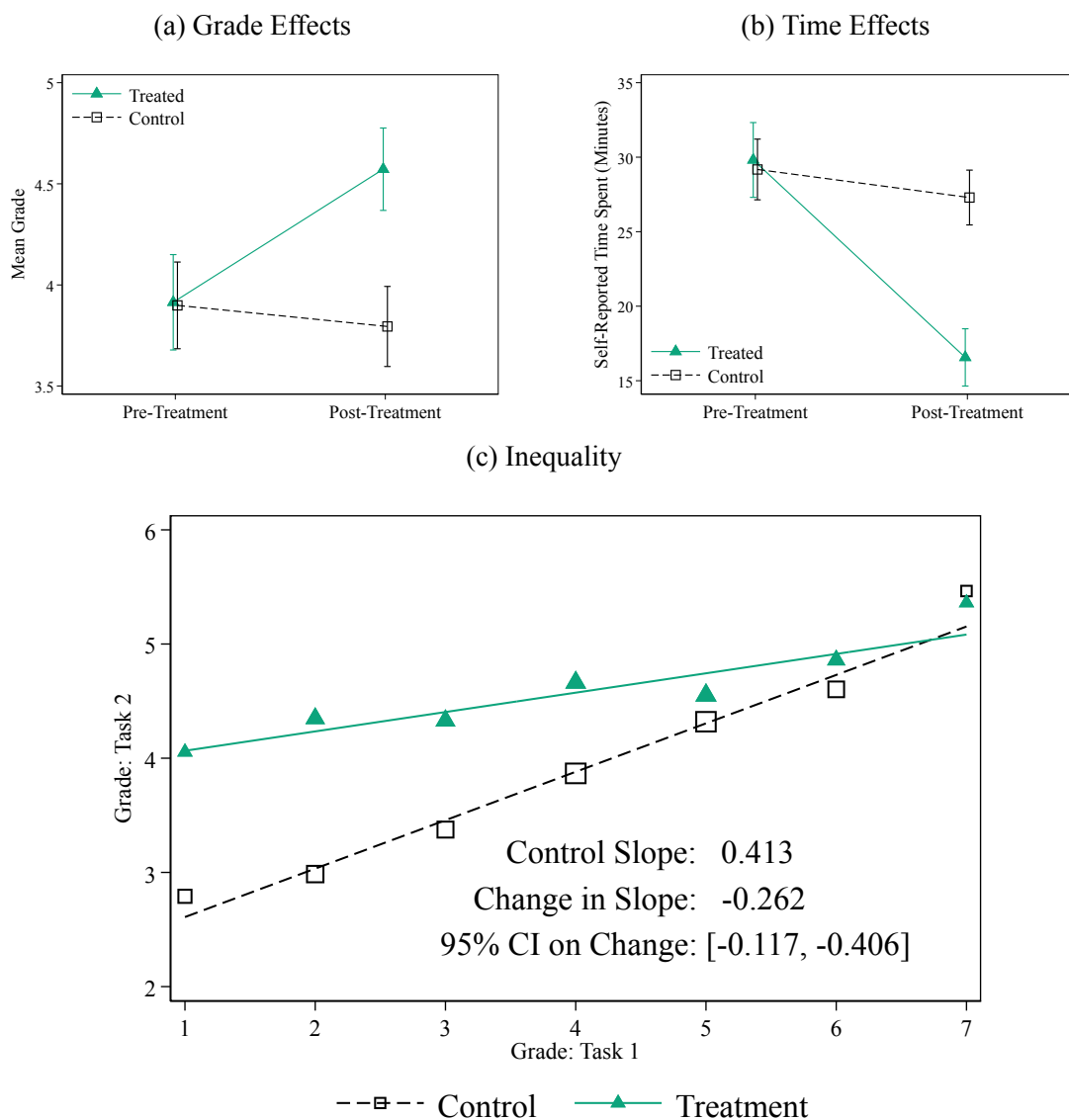
*Note:* This figure replicates our main treatment effects, restricting to people who said after the pre-treatment task that they had completed similar tasks in their real job before.

Figure S.19: Main Results Excluding Managers



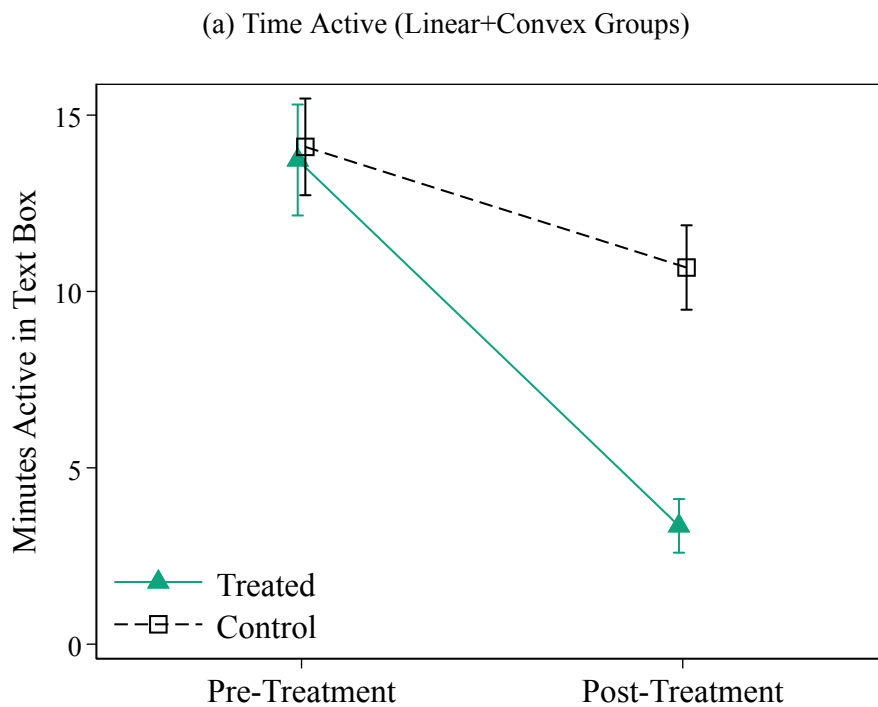
*Note:* This figure replicates our main treatment effects, restricting to people who said after the pre-treatment task that they had completed similar tasks in their real job before.

Figure S.20: Main Results Excluding Data Analysts



*Note:* This figure replicates our main treatment effects, restricting to people who said after the pre-treatment task that they had completed similar tasks in their real job before.

Figure S.21: Robustness of Time Results to Objective “Time Active” Measure



*Note:* This figure shows “time active” (number of minutes in which at least 3 words in the text box are added or modified), before and after treatment, separately by treatment and control in our linear and convex incentive groups.

## E Supplementary Tables

Table S.1: IV Versions of Main Treatment Effects I

	(1) Time (OLS)	(2) Time (IV)	(3) Grades (OLS)	(4) Grades (IV)
<b>Treatment</b>	-10.842*** (1.382)	-16.507*** (1.986)	0.651*** (0.145)	0.949*** (0.185)
<b>Nb. obs</b>	346	346	1,135	1,135

Standard errors in parentheses

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



Table S.2: IV Versions of Main Treatment Effects II

	(1) Job Satisfaction (OLS)	(2) Job Satisfaction (IV)	(3) Self-Efficacy (OLS)	(4) Self-Efficacy (IV)
<b>Treatment</b>	1.164*** (0.211)	1.928*** (0.297)	0.409** (0.204)	0.805*** (0.280)
<b>Nb. obs</b>	449	449	449	449

Standard errors in parentheses

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Table S.3: Pure ChatGPT Grades

Occupation	# Unique Essays	# Grades	Mean Grade	SD Grade	# Unique Graders	Mean Human Grade
HR professional	6	39	4.41	1.90	11	5.00
Consultant	10	51	4.73	1.43	19	4.81
Data analyst	10	60	4.85	1.44	19	4.14
Grant writer	9	58	4.47	1.29	20	4.36
Manager	24	154	4.95	1.55	45	5.00
Marketer	8	46	5.78	1.23	11	5.13
Total	67	408	4.88	1.53	122	4.76

*Note:* For each occupation, the table shows the number of unique ChatGPT essays that were graded, the number of grades that were given, the mean grade, the standard deviation of grades, the number of unique graders, and the mean human grade. The mean human grade is defined as the mean grade of individuals in the treatment group who stated they used ChatGPT in the second task.

## Bibliography

- (21) Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina, “Data Quality of Platforms and Panels for Online Behavioral Research,” *Behavior Research Methods*, 2021, pp. 1–20.
- (24) Lee, David S, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *The Review of Economic Studies*, 2009, 76 (3), 1071–1102.