Glasses for Intelligence Impression
Final Deliverable
Cohort B Team 7

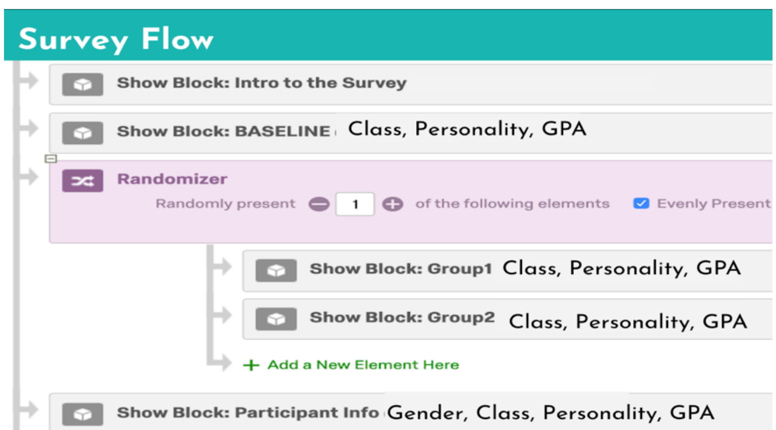
Kunpeng Huang, Yoki Liu, Lyufan Pan, Yunlei Zhou, Jiayuan Zou, Sherry Zuo

Causal Question of Interest

In this experiment, we study the effect of whether wearing glasses affects people's perception of intelligence. We use wearing glasses as the treatment and "GPA" with different grade levels as the outcome variables in our experiment. The result of the experiment shows that just by wearing a pair of glasses, you can make yourself look smarter in others perception.

Experiment Design and Why Choose the Experiment Design

In this experiment, we employed the A/B test on the IP address level. The reason we used this method instead of other alternatives is that we can do random assignments on the individual level but unable to do it on time or region level. We used Qualtrics to design the questionnaire and sent out the survey link to the participants. We employed simple randomization utilizing the Randomizer function of the Qualtrics website. By using this function, participants will be randomly assigned to the treatment or control group with a 50% possibility at the first moment they open the website. As a result, we achieved not only the random assignment but also the double-blindness that neither the experimenters nor the participants know who is receiving the treatment. With double-blind, we had eliminated subjective biases and personal preferences that may appear in the consciousness of experimenters and participants.



The questionnaire consisted of three parts. The first part was used as a baseline that all the participants will see the same photo of a male without glasses, and will be asked questions based on the photo (Photo A). Questions included the first impression of the person's class, personality, and GPA. The reason we used a baseline is that people's judgment on GPA might be very different. By adding a baseline, we create a benchmark for evaluating the effect. Besides, we want to use this as a balance check to our experiment, to check the randomization of dividing the treatment and control groups, and see if their standards for considering people's GPA are the same. So, for the first picture shown to both groups, we expected to see that both of the groups' responses the same, and this is also a prerequisite that must be satisfied to further the experiment.

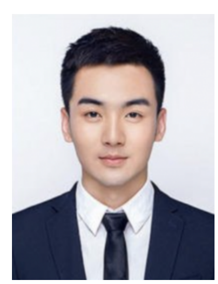


Photo A

In the second part of the questionnaire, participants in the control group saw a male without glasses (Photo B-1) and participants in the treatment group saw the same person but with glasses (Photo B-2). The same questions were asked as in the first part. The two photos in the second part are identical except wearing glasses or not and it is similar to the photo in the first part on clothing but not facial features. By doing so, we assure the excludability that only the treatment, wearing glasses, should affect the result.


Photo B-1


Photo B-2

In the last part of the questionnaire, we asked survey participants' personal information including gender, class, personality, GPA, and whether they can guess the purpose of the experiment. The reason why we put this part at the end of the experiment is to avoid the potential influence on the participants' first impression by letting them aware of our testing goal.

Before finalizing the questionnaire, we did a pilot experiment to help discover potential problems. From the feedback of the participants, we found several weakness: First, asking people about the GPA right below the photo will make the participant aware of the purpose of the experiment and enhance the influence of their judgment; second, some of the questions were offensive such as the relationship question; last, some participants were taking too serious on the questions and took a long time to examine the photos which make their judgment no longer based on their first impression. To solve these problems, we redesigned the questionnaire including moving the GPA question backward, removing the relationship question, and adding a timer on the top of the website with a limitation of 60 seconds for each part of the questionnaire.

Balance check / Randomization Check

Since the p-value is 0.9251 which is larger than 0.05, the R-square is -0.0048823 which is too small, so we fail to reject the null hypothesis, and we have a 95% confidence that the glasses treatment does not affect the baseline person's grade took before the experiment started. So these results indicate that the validity of the study's conclusion is high.

```
Call:
lm(formula = grade_baseline ~ Treatment, data = analysis)

Residuals:
     Min       1Q   Median       3Q      Max
-6.62136 -0.62136  0.37864  0.39216  1.39216

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.607843   0.101792 74.7392   <2e-16 ***
TreatmentTRUE 0.013516   0.143606  0.0941   0.9251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.028 on 203 degrees of freedom
Multiple R-squared:  4.3636e-05,	Adjusted R-squared:  -0.0048823
F-statistic: 0.0088585 on 1 and 203 DF,  p-value: 0.92511
```

Statistics about the Data

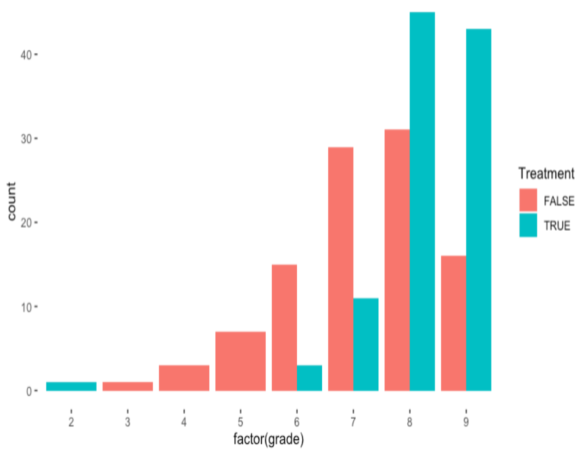
From the survey, we collected a raw dataset with 228 rows and 43 columns. During the data cleaning, we removed unused variables like date, unfinished observations, and duplicate observations that some participants have done the survey more than once. We also converted some variables into a numeric type. For all personality variable, we converted "logical" to 1 and "emotional" to 0; for gender, 1 for "male" and 0 for "female"; for grade, 9 for "A(3.75-4.0)", 8 for "A-(3.5-3.75)", 7 for "B+(3.25-3.5)", 6 for "B(3.0-3.25)", 5 for "B-(2.75-3.0)", 4 for "C+(2.5-

2.75)”, 3 for “C(2.25-2.5)”, 2 for “C-(2.0-2.25)”, 1 for “Below C-(<2.0)”. Variables with suffix “_baseline” are the question results in part one of Person A, variables without suffix are the question results in part two of Person B, and variables with suffix "_p" are the results of the personal questions on part three of participants.

After the data cleaning, there were **205 observations and 15 variables** with 103 people in the treatment group and 102 people in the control group, which showed a good 50/50 split.

```
 [1] "Duration..in.seconds." "RecordedDate"          "LocationLatitude"      "LocationLongitude"
 [5] "class_baseline"        "personality_baseline"  "grade_baseline"        "class"
 [9] "gender_p"              "class_p"               "personality_p"         "grade_p"
[13] "personality"           "grade"                 "Treatment"
```

We also visualized the participant’s location on the world map, and we found out that people took this survey at different times randomly. Thus, we think it shows that the probability of interference is low.




```
Call:
lm(formula = grade ~ Treatment, data = analysis)

Residuals:
     Min       1Q   Median       3Q      Max 
-6.19417 -0.20588 -0.19417  0.80583  1.79412 

Coefficients:
              Estimate Std. Error t value  Pr(>|t|)    
(Intercept)    7.20588    0.11520  62.551 < 2.2e-16 ***
TreatmentTRUE  0.98829    0.16252   6.081 5.838e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.1635 on 203 degrees of freedom
Multiple R-squared:  0.15409,    Adjusted R-squared:  0.14992 
F-statistic: 36.978 on 1 and 203 DF,  p-value: 5.8378e-09
```

From the bar chart, we could find out that the treatment group participants gave higher grades than the control group participants. The ATE of treatment on grade is 0.98829, and the 95% confidence interval is [0.6697508, 1.3068292], and since the p-value of Treatment is less than 0.05, so we think this causal effect is statistically significant, we could conclude that wearing glasses have a positive impact on people’s impression of intelligence.

Treatment Effects

1) Fixed Effect Check

In this part, we would like to check the relationship between grade and other variables (variables except treatment), to observe if these variables also affect our initial outcome of grade. Since our experiment goal is not to observe the effects of these variables, we would like to include them as fixed effects in our final regression model if we find these variables are significantly related to the outcome grade. To do so, we would run several regression models of grade on each of these variables, and see the results of them. In our final model, we would add all of these significant variables as fixed effect to eliminate their effects on the outcome.

Grade_baseline

Since the p-value (9.889e-08) is less than 0.05, we would reject the null hypothesis and have a 95% confidence that the baseline affects the grade, that's why we want to do baseline during our survey. Thus, it would be a fixed effect on our results.

Personality_p

With the t-test, the p-value is 0.056235 which is larger than 0.05, so we fail to reject the null hypothesis and don't have 95% confidence that the personality of each subject affects the grade. Thus, it's not a fixed effect on our results.

Gender_p

With the t-test, the p-value is 0.030423 which is less than 0.05, so we reject the null hypothesis and we have a 95% confidence that the gender of each subject affects the grade. Thus, it's a fixed effect on our results.

Class_p

While comparing the distributions of the class of person A (baseline), person B, and the subject itself, we find that although most of the subjects are graduates, their judgment on person B is still almost evenly distributed. Hence, we don't believe that it's a fixed effect.

2) Causal Effect Model with Fixed Effect

Based on our fixed effect check, we would like to include the fixed effects of grade_baseline and gender_p in our final regression to find the effect of treatment for letter grade impression. The p-value is less than 0.05, so the treatment effect on glasses with fixed effects is statistically significant. Here, we could find that after considering fixed effects and compared with the original causal effect model, this model has a smaller standard error, a larger adjusted R-squared, so it's more precise. Thus, we believe that the causal effect model with the fixed effect would be better.

```
Call:
   felm(formula = grade ~ Treatment | grade_baseline + gender_p,      data = analysis)

Residuals:
     Min       1Q   Median       3Q      Max
-2.86147 -0.57468  0.13280  0.55207  1.85032

Coefficients:
              Estimate Std. Error t value  Pr(>|t|)
TreatmentTRUE  1.00573    0.13825   7.275 8.977e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.95269 on 189 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared(full model): 0.33205   Adjusted R-squared: 0.30731
Multiple R-squared(proj model): 0.21877   Adjusted R-squared: 0.18983
F-statistic(full model):13.422 on 7 and 189 DF, p-value: 5.028e-14
F-statistic(proj model): 52.925 on 1 and 189 DF, p-value: 8.9765e-12
```

Limitations of the Study

1) Sample Selection Bias: The participants' characteristics are not representative. We sent our survey links through the social networking platforms, so most of the participates are our friends, rather than random participants.

2) Interference: We supposed to run our experiment on an individual level without interference of each other. Although the map shows we have partially achieved non-interference, we could not make sure that participants were not communicated with the rest of the participants when they answered questions from our experiment. Such as some participants come from the MSBA program at Boston University, they may interfere with others when they open the survey link at the same time although they are randomly assigned to different groups.

3) Questionnaire Design Limitation: We supposed that the experiment's questions should not interacted with other questions. However, we found that the question about people's personality has a positive correlation with the guessed grade, and adjusted R2 would be increased after adding personality in the regression. In other words, people have stereotypes that people with logical personalities would have higher GPA. Thus, showing both the question about personality and GPA on the same page may influence participants' judgments, which means that we did not fulfill our assumption.

	Dependent variable:	
	grade	
	(1)	(2)
Treatment	1.006***	0.663***
	(0.138)	(0.130)
personality		1.112***
		(0.147)
Observations	197	197
R2	0.332	0.488
Adjusted R2	0.307	0.466
Residual Std. Error	0.953 (df = 189)	0.836 (df = 188)
Note:	*p<0.1; **p<0.05; ***p<0.01	

4) Photo Representation Limitation: Additionally, since we would like to keep the statistical power within limited count of samples, our selection of photos include only one race and gender (i.e. Photo A & Photo B), which is not representative of our research purpose as it did not evaluate the treatment effect on other races and genders. To reach a more comprehensive result, more samples should be tested based on a diverse selection of photos.

5) Evaluation Scale Limitation: We used GPA as the scale to evaluate the intelligence in this experiment. However, GPA cannot perfectly represent one's intelligence.

Conclusion

Based on the experiment, wearing glasses has a positive impact on people's impression and judgment on others' intelligence. Thus, we conclude that wearing a glass can make people think you are more intelligent.

Open sources on https://github.com/Sherry-Zuo/Glasses-for-Intelligence-Impression