

# Swipe Right or Left: Evaluating the Effectiveness of Interactive Introduction Formats on Dating App

Yu-Hsiang Wang, Zhe-Yu Lin, Ming-Hua Tsai, Ching-Hsuan Lin

## Introduction

Online dating has become one of the most popular and competitive ways to meet partners. According to Pew Research Center (2020), nearly 30% of U.S. adults have used a dating app, with younger users (18-34) being the most active\*. Currently, numerous dating apps, including Bumble, Tinder, Hinge, and many others, are fiercely competing in the market. To secure a leading position, success is not solely dependent on branding; enhancing the user experience and significantly improving the matching rate are crucial. Factors such as advanced algorithms, personalized recommendations, and effective communication features play a pivotal role in increasing engagement and retention.

In this study, we aim to explore whether incorporating more interactive features into user profiles can effectively increase the matching rate. Specifically, our experiment will examine interactive features in self-introductions, such as multiple-choice questions—a feature introduced by Hinge—on user engagement and swipe rates. By analyzing user responses and behaviors, we aim to determine whether interactivity boosts engagement, strengthens connections, and enhances dating app effectiveness. Our findings may offer insights for improving matchmaking algorithms and user experience design.

## Research Question

Does adding interactive questions to a profile enhance swipe rates on a dating app?

## Null Hypothesis

$H_0$  : Including an interactive question has no effect on enhancing the swipe rate.

$H_a$  : Including an interactive question has an effect on enhancing the swipe rate.

## Experimental Design

### 1. Goal of Experiment

The experiment explores whether a simple written self-introduction or an interactive multiple-choice format (e.g., a "Truth & Lie" three-option question) is more effective in increasing right swipe rates on dating apps.

## 2. Unit of Randomization

Our unit of randomization is at the user level. After answering the basic questions and selecting their preferred gender, each participant is randomly assigned to either the treatment group or the control group. This method ensures that we can directly compare the average behavioral differences between the two groups of participants. Additionally, this approach helps us avoid spillover effects and better align with real-world application scenarios.

## 3. Treatment and Control group

Both the treatment and control groups will see ten profiles, similar to those found in a real dating app. Each profile in both groups includes a person's image, age, and a short self-introduction. The only difference is that profiles in the treatment group feature an interactive self-introduction, which includes one additional question with three short possible answers.

## 4. Variables

- Outcome variable

The outcome variable in our experiment is the swipe rate, which is calculated as the number of likes divided by the total number of profiles each user has seen (10). It can be represented as:

$$\text{Swipe rate} = \frac{\sum_{i=1}^{10} x_i}{10}$$

where  $x_i$  represents whether the user liked profile  $i$  (1 for like, 0 for dislike).

- Covariates

In our experiment, we include several covariates, such as age group, preferred gender, and prior experience with dating apps. These variables could help us reduce the variability in the swipe rate, leading to more precise estimates of the treatment effect. Additionally, including these covariates enable us to examine whether the effect of the treatment varies across different demographic groups.

## Survey Design

### 1. Questionnaire Distribution

We designed the questionnaire using Boston University's Qualtrics system and distributed it via email to all students in the MSBA Program. Additionally, we invited target respondents aged 21-35 to participate through social media.

### 2. Questionnaire Design

To increase participants' willingness to complete the survey and protect their privacy, we adopted an anonymous questionnaire that does not collect personal information such as names or email addresses. The questionnaire consists of two main sections:

- First Section

Participants were asked to provide basic demographic information, including age, gender, sexual orientation, and prior experience with dating apps. In this section, we used respondents' sexual orientation as the criterion for determining whether they would be shown male or female profiles. Additionally, participants were randomly assigned to either the control group or the experimental group to ensure a balanced sample size between the two groups.

- **Second Section**

Each participant was shown 10 simulated dating app profile pages, featuring profiles of the gender they are interested in. Each profile contained a photo, name, age, and a self-introduction ("About Me"). The profiles included individuals with diverse ethnic backgrounds and physical features to minimize bias.

- **Treatment Group:**

In addition to the above information, the "About Me" section included an interactive multiple-choice question.

- **Control Group:**

Participants in this group saw only the standard profile information, with the "About Me" section containing only a self-introduction.

## **Data Analysis**

### **1. Data cleaning**

Due to the questionnaire design, the output data columns were quite disorganized. To address this, we extracted the necessary columns using the 'interested in' variable (indicating whether a participant is interested in males or females) and the group designation (treatment or control). We then concatenated the data into two distinct groups: the treatment group and the control group. Additionally, we converted the text responses ("like" or "dislike") into labels and calculated each individual's swipe rate for further analysis.

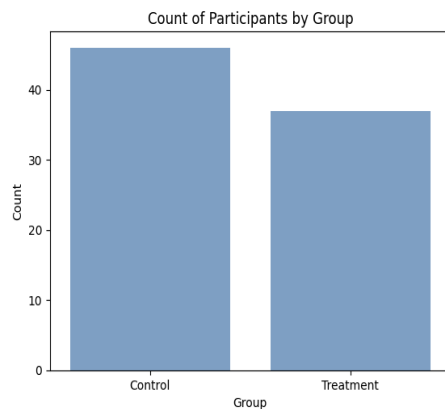
### **2. Data Overview**

Our dataset consists of 83 responses, with 37 in the treatment group and 46 in the control group. The data dictionary is shown in the table below.

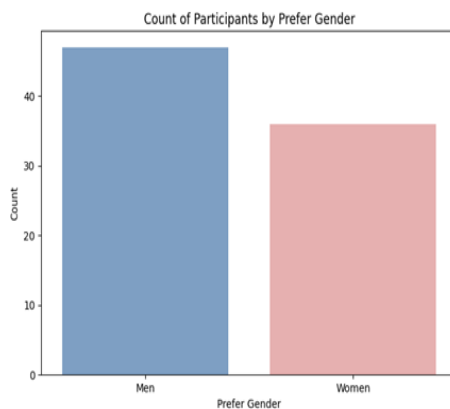
Variable Name	Descriptions
time_spent	Time spent answering the survey (in seconds).
gender	The gender of the participant.: 0 = Male, 1 = Female
interest_in	The participant's preferred gender: 0 = Male, 1 = Female
age_group	Age groups: 0 = 18-22, 1 = 23-27, 2 = 28-32.

experience	Dating app experience: 0 = No, 1 = Yes.
treatment	Profile including interactive self-Introduction: 0=No, 1=Yes
swipe_rate	The swipe rate.

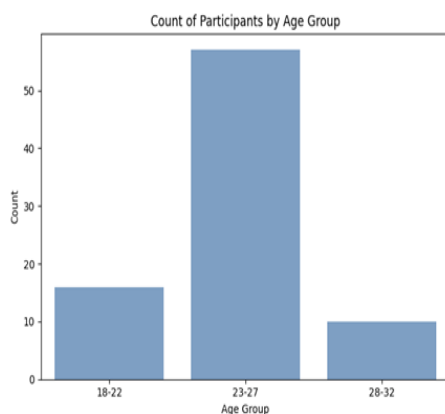
### 3. EDA



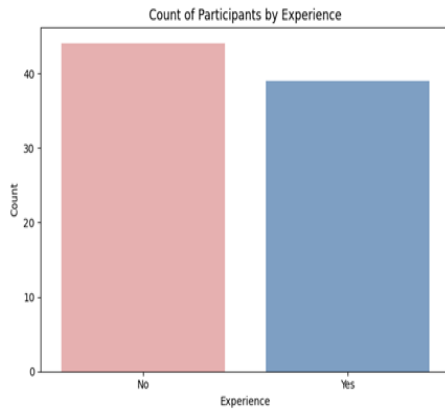
The bar chart shows the number of survey participants in the control and treatment groups. There are 46 samples in the control group and 37 in the treatment group, showing no substantial difference. This suggests that our randomization was likely successful.



This bar chart shows the number of participants by the gender they are interested in. More participants are interested in males than females, with 47 showing interest in males and 36 in females.



This bar chart illustrates the number of participants across different age groups. The results indicate that most participants fall within the 23-27 age range. In contrast, the other two age groups have similar participants, but both are significantly lower than the 23-27 age group.



This bar chart shows the number of participants based on their dating app experience. In our survey, 44 participants have used dating apps before, while 39 have never used them.

#### 4. ATE & Standard Deviation

The calculated Average Treatment Effect (ATE) is 0.02, representing the average difference in Mean Swipe Rate between the treatment and control groups. While the results indicate that the treatment influences Swipe Rate, the impact is not statistically significant. The standard deviation of the ATE is 0.21.

ATE	Std	Cohens'D	Statistical Power	Effect Size
0.02	0.21	0.09	0.07	1790

#### 5. Cohens'D

Cohen's D was calculated to be 0.09, indicating a negligible effect size. This suggests that while interactive self-introduction may have some impact on swipe rate, the effect is minimal.

#### 6. Statistical Power

The statistical power of our experiment is low at 0.07, indicating a limited ability to detect the effect of adding an interactive self-introduction on the swipe rate. One possible reason for this is the limited sample size. To achieve a statistical power of 0.8, we would need to collect at least 1,790 samples.

#### 7. Balance Check

We used OLS regression to perform a balance check on the three covariates—interest\_in, age\_group, and experience—across treatment groups. As shown in the table, none of these covariates is statistically significant, indicating that there is no systematic difference in these characteristics between the treatment and control groups. In other words, random assignment

	interest_in (1)	age_group (2)	experience (3)
Intercept	0.478*** (0.075)	0.935*** (0.085)	0.500*** (0.075)
treatment[T.B]	-0.100 (0.110)	-0.016 (0.123)	-0.068 (0.111)
Observations	83	83	83
R <sup>2</sup>	0.010	0.000	0.005
Adjusted R <sup>2</sup>	-0.002	-0.012	-0.008
Residual Std. Error	0.499 (df=81)	0.562 (df=81)	0.504 (df=81)
F Statistic	0.826 (df=1; 81)	0.017 (df=1; 81)	0.369 (df=1; 81)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

appears to have worked effectively, ensuring that variations in interest\_in, age\_group, or experience are unlikely to confound our subsequent analyses.

## 8. Effect Size

With Cohen's d approximately 0.09, a two-tailed test, and an alpha level of 0.05, each group would need at least 1,789 participants to achieve 80% power in detecting this effect size.

## 9. Proportion Check (Z-test)

The p-value from the Z-test is 0.32, indicating that the test does not show a statistically significant difference.

The p-value for proportions Z test is: 0.3203505658676017

## 10. T-Test

The results of the T-test show that the test statistic is approximately 0.425368, indicating that the difference between the two groups (or conditions) is small relative to the data variability. The high p-value and small Cohen's d also suggest that the difference between the treatment and control groups is not statistically significant. Furthermore, the 95% confidence interval includes 0, reinforcing the conclusion that no significant difference exists.

	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
<b>T-test</b>	0.425638	80.373318	two-sided	0.671509	[-0.07, 0.11]	0.092648	0.249	0.069913

## 11. Regression

We conducted four regression models, each using the swipe rate as the dependent variable. In each model, we gradually added meaningful variables to observe their impact on the swipe rate. The results show that as more variables were included, R<sup>2</sup> steadily increased from an initially low value, indicating that the models gained explanatory power over the swipe rate.

Additionally, it is noteworthy that there was no significant difference between the treatment and control groups.

Dependent variable: swipe_rate				
	(1)	(2)	(3)	(4)
Intercept	0.315*** (0.034)	0.240*** (0.034)	0.190*** (0.046)	0.170*** (0.045)
age_group			0.057* (0.035)	0.028 (0.038)
experience				0.080 (0.049)
interest_in		0.157*** (0.047)	0.150*** (0.048)	0.166*** (0.047)
treatment[T.B]	0.020 (0.047)	0.036 (0.044)	0.036 (0.044)	0.042 (0.044)
Observations	83	83	83	83
R <sup>2</sup>	0.002	0.135	0.157	0.186
Adjusted R <sup>2</sup>	-0.010	0.114	0.125	0.144
Residual Std. Error	0.215 (df=81)	0.201 (df=80)	0.200 (df=79)	0.198 (df=78)
F Statistic	0.181 (df=1; 81)	5.608*** (df=2; 80)	5.239*** (df=3; 79)	5.774*** (df=4; 78)
Note: *p<0.1; **p<0.05; ***p<0.01				

In contrast, introducing the interest\_in variable consistently showed a significant and positive effect. Both findings support the insights derived from our EDA.

12. CATE

This regression analysis aims to examine whether the treatment effect varies based gender interest. The results show that interest\_in has a coefficient of 0.171, meaning that in the control group, for every 1-unit increase in interest\_in, the swipe rate increases by approximately 0.171 on average. Additionally, with a p-value < 0.01, this effect is highly significant, indicating that interest\_in has strong positive impact on the swipe rate.

However, the interaction between treatment and interest\_in is not significant, suggesting that individuals interested in different genders do not experience a distinct treatment effect

Dependent variable: swipe_rate		on
	(1)	
Intercept	0.233*** (0.037)	
interest_in	0.171*** (0.064)	
treatment[T.B]	0.049 (0.048)	
treatment[T.B]:interest_in	-0.032 (0.096)	a
Observations	83	
R <sup>2</sup>	0.137	
Adjusted R <sup>2</sup>	0.104	
Residual Std. Error	0.202 (df=79)	
F Statistic	3.754** (df=3; 79)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

between Groups A and B. Overall, interest\_in emerges as a key explanatory variable. In contrast, treatment and its interaction with interest\_in do not show significant effects.

Challenges & Limitations

1. Challenges

◦ Practical implementation

Achieving realism in our prototype testing was challenging due to time and budget constraints. While we designed a comprehensive UI, we couldn't fully replicate the dynamic interactions, authentic decision-making, and feedback loops of real dating apps. To address this, we integrated the interface into a survey, allowing participants to react to choices within a controlled environment. While this method captures initial reactions and engagement metrics, it lacks the depth of real-world interactions, limiting our ability to assess user behavior in a fully immersive dating experience. Future iterations should incorporate more interactive elements to enhance realism.

◦ Profile diversity

Limited resources restricted the diversity of profiles, reducing representation of racial and ethnic groups. This homogeneity may have influenced participant responses, skewing engagement and

matching rates. As a result, our findings may not fully reflect user behavior in real-world dating apps. Future studies should incorporate a wider range of profiles to improve generalizability.

## **2. Limitations**

- **Sample Bias**

Our survey primarily includes BU MSBA students and social media friends, leading to a homogeneous age range and background that may not represent the broader user group.

- **Limited Sample Size**

Time constraints prevented us from collecting enough data for sufficient statistical power, potentially affecting the detection of treatment effects.

- **Profile Presentation Limitations**

To keep the survey concise, profiles included only an image, age, gender, and self-introduction, unlike real dating apps with more details. This lack of information may have influenced participant decisions.

- **Measurement and Response Bias**

Despite anonymity, participants might have provided socially desirable responses, leading to potential discrepancies from real-world behavior.

## **Conclusion**

Our regression analysis indicates that the coefficient for the treatment variable is positive (0.024), suggesting that users exposed to interactive features exhibit a trend toward increased average swipe rates. However, this effect did not achieve statistical significance, implying that our current sample size or data may not be sufficient to conclusively confirm this impact. As a result, we cannot reject the null hypothesis ( $H_0$ ), meaning that including an interactive question does not significantly affect swipe rate within our study.

Interestingly, the variable "interest\_in" shows a significantly positive coefficient (0.166\*)\*\*, suggesting that users interested in females demonstrate notably higher swipe rates compared to those interested in males. This may indicate that male users generally exhibit higher levels of initiative and actively swipe right more frequently to engage with potential matches.

However, our study faced several limitations that may have influenced the results:

1. **Limited Realism in User Interaction** – Due to time and budget constraints, we could not fully simulate an authentic decision-making environment. While we developed a UI prototype, we relied on survey-based engagement metrics rather than real behavioral tracking tools like Maze. This likely missed key behavioral nuances found in real-world dating apps.
2. **Profile Diversity Constraints** – Our dataset lacked sufficient racial, cultural, and personality diversity, which may have affected participant engagement and responses. A more inclusive range of profiles would better reflect real-world dating app interactions.



## Appendix

1. Figma output: <https://reurl.cc/L5YL9e>
2. Google colab: <https://reurl.cc/EgkLOv>
3. Source of our profile images: <https://www.pexels.com/search/>
4. Survey Link: [https://bostonu.qualtrics.com/jfe/form/SV\\_3yfusZykmvMprSu](https://bostonu.qualtrics.com/jfe/form/SV_3yfusZykmvMprSu)