The effect of graduate-level higher education on women's likability

Aryana Far, Jonah Grossman, Derek Yao, Jun Park

Abstract

Purpose: In this study, we investigated the influence of graduate-level education level on women's likability by potential partners in the context of heterosexual online dating.

Methods: We ran two experiments. First, we ran a field experiment on the dating app Tinder, successfully creating 6 fictitious female profiles with varying educations. For each fictitious profile, we collected 300 observations (N = 1800) corresponding to Tinder users that our accounts swiped right on, gathering the outcomes of resulting likes or messages. Profile bans posed a grave challenge. Second, we ran a Tinder simulation survey, presenting participants with two profiles of varying education levels and collecting outcomes of stated attractiveness rating and stated liking behavior. For each of the 100 survey respondents, we collected 2 observations (N = 200).

Hypothesis: Due to different existing claims in literature, we maintained a two-tailed hypothesis that there would be some significant effect of graduate-level education on women's likability.

Results: In the field experiment, simple treatment-outcome contrasts revealed a significant but small positive treatment effect. In the simulated experiment, models with and without covariates consistently indicated small statistically significant positive treatment effects.

Conclusion: While these results are interesting, we suggest further research that combines methodologies of both our experiments, leverages the benefits of both approaches, and subsequently weights results to provide a more robust treatment effect estimate.

Keywords: online dating, dating preferences, gender dynamics, education level, social status, relationship intentions, mate selection, partner preferences, dating behavior, gender roles, socioeconomic factors, digital interactions, assortative mating, relationship initiation, social desirability

Research Question

The objective of this study was to answer the following question: What is the effect of graduate-level higher education on women's overall likability in American heterosexual online dating contexts, both real or simulated? By examining this question from both naturalistic and artificial experimental lenses, we aimed to contribute to a better understanding of modern mating dynamics.

Background

In recent decades, the method of finding romantic partners has evolved significantly. By 2009, around 22% of heterosexual couples met online (Rosenfeld & Thomas, 2012). Furthermore, 33% of marriages initiated between 2005 and 2012 in the US sprung from online dating (Cacioppo et al., 2012). This shift is expected to grow further, aided by the rise of mobile dating apps like Tinder (Ranzini & Lutz, 2016), which is the primary vehicle of our study.

Amidst this transformation in how individuals connect romantically, understanding the nuances of mate selection criteria, particularly regarding factors such as education, becomes increasingly pertinent. Existing findings on the impact of higher education on women's attractiveness in heterosexual dating contexts vary considerably. One observational study suggests that highly educated women tend to match up with equally highly educated men, which is in line with the theory of educational assortative mating (Eika et al., 2017). Conversely, another observational study suggests a departure from traditional assortative mating patterns, with highly educated women more likely to partner with less educated men rather than remaining single (De Hauw et al., 2017). Conflicting results also emerge from field experiments: one posits that higher education diminishes a woman's attractiveness due to perceived intimidation (Egebark et al., 2021), while another rejects this notion, asserting that educational attainment does not influence women's appeal (Neyt et al., 2019). In light of the conflicting findings, continued research efforts are necessary to elucidate the influence of higher education on women's likability in heterosexual dating contexts.

Hypothesis

Due to a diverse synthesis of literature on this topic, our hypothesis is two-tailed, positing that a graduate-level degree has some significant effect on a woman's overall likability. Firstly, we consider the possibility that in a contemporary society, a woman's educational attainment may positively influence her perceived likability due to associated traits such as intelligence, ambition, and potential for success. Secondly, we hypothesize that the presence of a graduate degree could negatively affect likability due to ingrained social biases, intimidation, or perceived compatibility issues.

Two Approaches, Two Experiments

Conducting both real and simulated Tinder experiments was necessary for comprehensive investigation. Hence, we present two distinct sections, each corresponding to one experiment. This approach ensures a thorough examination of the phenomena under study, providing insights into both real-world dynamics and simulated scenarios.

First, an Explanation: What is Tinder?

Tinder, a popular online dating application, boasts a monthly user of more than 75 million all across the globe, with on average 1.5 million users arranging meetups each week (Business of Apps). When creating a Tinder account, users are required to provide basic information such as their name, age, gender, and romantic preferences, along with at least three photos. Additional details like education, occupation, and interests can be included optionally. Upon completing profile setup, users can adjust their preferences regarding the distance radius and age range of profiles they wish to view. Subsequently, users can begin swiping through profiles that match their criteria, swiping right to indicate interest and swiping left to pass. When two users both swipe right on each other's profiles, a match is made, allowing either party to initiate a conversation via chat message.

Experiment A

Overview

Experiment A was a real Tinder field experiment. This experiment allowed us to observe thousands of actual interactions in a naturalistic environment, although it certainly presented various challenges. Firstly, throughout this experiment, there was a constant threat of our fictitious accounts being flagged by Tinder's excellent cybersecurity engineers. Secondly, controlling the attractiveness of fictitious profiles was difficult due to subjective picture selection and availability. Thirdly, maintaining consistent liking behaviors across all accounts was challenging, as miscounting and losing track could have occurred easily. Fourthly, obtaining covariate information from participants was infeasible due to the large number observations and privacy issues.

Comparison of Potential Outcomes

In Experiment A, we compared the following potential outcomes. The untreated potential outcome, or $Y_i(0)$, was whether the subject found the fictitious woman in question likable given that she had only a high school diploma. The treated potential outcome, or $Y_i(1)$, was whether the subject found the fictitious woman in question likable given that she had a graduate degree. But, what do we mean by "likable"? In this real-world Tinder context, we define likability as the *actual* degree of appeal that other users perceive in a profile of interest. A Tinder user's likability might be influenced by many factors including physical appearance, profile presentation, perceived personality, and social status. Thus, we measured actual likability by two metrics: whether a participant actually "swiped right on," or "liked," the fictitious woman in question and whether a participant actually initiated a messaging conversation with the fictitious woman in question.

Fictitious Profile Creation

We successfully created 6 fictitious female Tinder profiles, using unique phone numbers and emails for verification during account creation. All profiles were created and maintained mainly in the Bay Area, but also some other populous parts of the United States, such as New York and Los Angeles. For profile names, we randomly selected 6 generic American girl names. All profiles were configured to represent 25 year old heterosexual females. For profile pictures, three stock images featuring the same model in various environments and poses were selected. The images were chosen based on specific criteria, including portrait-style composition, minimal distractions, clear image quality, and at least one image showcasing a smile. Additionally, the images were broadly sampled to encompass a diverse range of races while maintaining a consistent, albeit subjective, level of attractiveness. See Appendix A for profile details.

Participant Recruitment

After profile creation – which had a success rate of about 20% due to Tinder's strong cybersecurity efforts to prevent fictitious profiles – we set our eligibility criteria as such: males between the ages of 24 and 32 in a 100 miles radius. Profiles that fit these criteria were presented to us through the complex, yet consistent, Tinder algorithm. The Tinder algorithm broadly works by considering factors such as user activity, profile information, and preferences to determine which profiles are shown to a user. While the exact workings of the algorithm are proprietary to Tinder and subject to change, it generally aims to optimize matches based on compatibility and engagement.

We implemented covert recruiting of participants such that every like we sent corresponded with a participant we covertly recruited. Consistency was paramount across all fictitious profiles so we aimed to maintain uniform liking behavior across accounts. Our liking pattern was as such. 50 "right, right, left" swipes until reaching the maximum daily limit of 100 likes. One day on, one day off liking for 5 days, resulting in 300 likes sent. Finally, a 3-day account rest period for matches to accumulate. Each profile thus yielded 300 observations, leaving us with a total of 1800 observations.

Unfortunately, due to the time constraints and privacy issues, it was both impractical and unethical to collect covariate information for all observations, which led to two specific concerns: (1) the possibility of undocumented instances of duplicate participants, and (2) undocumented timewise inconsistencies in match accumulations. Firstly, the presence of undocumented duplicate instances hampers our ability to effectively incorporate fixed effects across participants who engage with more than one of our fictitious accounts. Secondly, there exists a discrepancy in the time available for match accumulations, with likes sent on the first day of our liking window having five additional days to accumulate matches compared to those sent on the fifth day. However, we found assurance in the consistent and spaced-out liking pattern employed, the uniformity of the Tinder algorithm across our accounts, and the high level of user activity on the platform, which collectively mitigated the impact of these concerns.

Randomization Process

3 of our fictitious profiles presented a high school level education, corresponding to 900 observations in the control group. 3 of our fictitious profiles were randomly chosen to present a graduate level degree – either Master's or PhD – , which corresponded to 900 observations in the treatment group.

We utilized a random selection process to assign profiles to either the control group (high school level education) or the treatment group (graduate-level degree). This random assignment method ensures that each profile had an equal chance of being assigned to either group, without any systematic bias or preconceived notions influencing the selection process.

Treatment

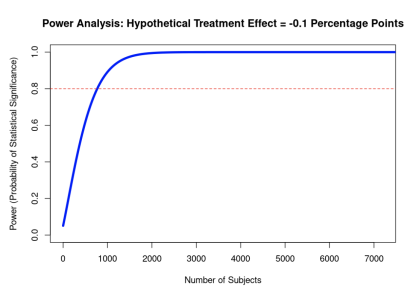
Since participation recruitment was a covert process, so was treatment assignment, so there were no issues of compliance or spillover. Treated participants were presented with one of our graduate-level educated fictitious profiles, and untreated participants were presented with one of our high school educated fictitious profiles.

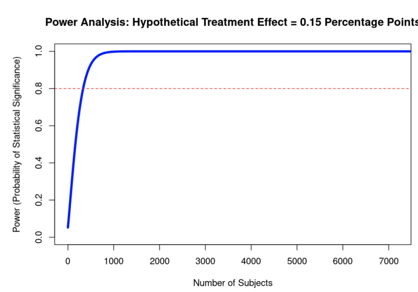
For our manipulation check, we ensured that participants assigned to both the treatment and control groups were effectively subjected to their respective conditions of being exposed to a fictitious profile with a graduate level degree or with high school education. To achieve this, we designed our fictitious profiles such that we strategically included or strongly hinted at information regarding the education level in multiple appropriate profile sections.

For the treated fictitious profiles, we clearly stated their education level (either Master's or PhD) in both the bio and the dedicated education section. However, we deliberately omitted the exact field of study for treated profiles to avoid introducing additional covariates and to allow for greater generalizability of findings.

Conversely, for the untreated profiles, we refrained from directly stating a high school level of education. Instead, we portrayed their profession as plausible blue-collar occupations such as waitress, hostess, and similar roles. This information was presented in the bio and in the dedicated career section without specifying a particular company.

Power



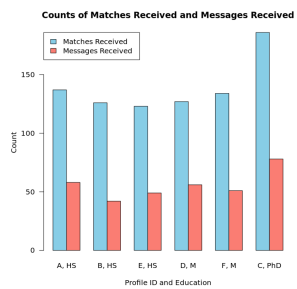


The two power analysis graphs provided depict curves representing the relationship between the number of subjects in a study and the power of the study. Power is defined as the probability of detecting an effect if there is one. The conventional threshold for sufficient power is often set at 0.8, marked by the horizontal dashed line on the graphs. The first graph shows the power analysis for a hypothetical treatment effect of -0.1 percentage points. As the number of subjects increases, the power of the study rapidly increases, reaching the threshold at around 1,000 subjects. After this point, additional subjects slightly increase the power, which approaches 1.0 but with diminishing returns. The second graph represents a hypothetical treatment effect of 0.15. Here, the power reaches the threshold of 0.8 with fewer subjects, at around 750. This suggests that detecting a treatment effect of 0.15 percentage points requires fewer subjects than detecting an effect of -0.1 percentage points, given the same study conditions. Similar to the first graph, the power in this case also approaches but does not quite reach 1.0, even as the number of subjects extends toward 7,000.

Overall, it is clear that the magnitude of the effect size has a considerable impact on the power of the study. Given our hypotheses, our study is well-powered at 1,800 observations. However, it is important to consider the variability of likability in online dating contexts as well as other study limitations when interpreting these power analysis results.

Data

The outcomes of matches and messages for each profile are demonstrated in this aggregate plot, which was only visualized after our modeling process.



We converted the summarized counts of likes and messages mentioned earlier into binary data, aligning with our representation of potential outcomes, denoted as $Y_i(z)$. The treatment condition of having a graduate degree is integrated into the operational framework of the models via a binary treatment indicator. This indicator mirrors the behavior of z, where z = 0 corresponds to the control condition (high school diploma only), and z = 1 corresponds to the treatment condition (graduate degree, including Master's or PhD).

Models:

In our analysis, we incorporated the likability outcomes of both likes and messages. For each of these two outcomes, we linearly regressed them on our binary indicator of treatment. For each model, we also ran regressions on the categorical profile variable to demonstrate profile-specific effects on likability.

Comparison of Models with Liked Outcome

	Dependent variable:	
	liked_back	
	(1)	(2)
treated	0.068***	
	(0.023)	
profile_typeB_high school		-0.037
		(0.040)
profile_typeC_phd		0.163***
		(0.040)
profile_typeD_masters		-0.033
		(0.040)
profile_typeE_high school		-0.047
		(0.040)
profile_typeF_masters		-0.010
		(0.040)
Constant	0.429***	0.457***
	(0.017)	(0.029)
Observations	1,800	1,800
R^2	0.005	0.021
Adjusted R^2	0.004	0.018
Residual Std. Error	0.498 (df = 1798)	0.494 (df = 1794)
F Statistic	8.344*** (df = 1; 1798)	7.662*** (df = 5; 1794)
Note:		*p<0.1; **p<0.05; ***p<0.01

Comparison of Models with Messaged Outcome

	Dependent variable:	
	messaged	
	(1)	(2)
treated	0.040**	
	(0.018)	
profile_typeB_high school		-0.053*
		(0.032)
profile_typeC_phd		0.067**
		(0.032)
profile_typeD_masters		-0.007
		(0.032)
profile_typeE_high school		-0.030
		(0.032)
profile_typeF_masters		-0.023
		(0.032)
Constant	0.166***	0.193***
	(0.013)	(0.022)
Observations	1,800	1,800
R^2	0.003	0.009
Adjusted R^2	0.002	0.007
Residual Std. Error	0.388 (df = 1798)	0.388 (df = 1794)
F Statistic	4.772** (df = 1; 1798)	3.361*** (df = 5; 1794)
Note:		*p<0.1; **p<0.05; ***p<0.01

For the simple regression of "liked_back" on "treated," the treatment coefficient was 0.068, significant at the 1% level, with a standard error of 0.023. Similarly, for the regression of "messaged" on “treated,” the treatment coefficient was 0.040, significant at the 5% level, with a standard error of 0.018. These regression results indicate that being presented with a fictitious profile featuring a graduate-level education has a small positive significant effect on likability.

Next, we examined profile-specific effects on likability. In the regression of “liked_back” on the profile-specific effects, each profile corresponded with a specific education level, indicated in the variable name. For example, profile_typeC_phd was a profile with a Ph.D. education level. The coefficient for profile_typeC_phd was 0.163, statistically significant at 1%, with a standard error of 0.04. This suggests this particular Ph.D. educated profile was associated with a higher likelihood of being liked compared to the other profiles. Other coefficients for different profiles (profile_typeB_high school, profile_typeD_masters, profile_typeE_high school, profile_typeF_masters) were not statistically significant. In the regression of “messaged” on profile-specific effects, we see the coefficient for profile_typeC_phd is 0.067 and statistically significant at 5%, with a standard error of 0.032. This suggests that profiles with a Ph.D. education level are more likely to receive messages compared to profiles with other education levels, but it is important to note the 95% confidence interval of the coefficient being between 0.00428 and 0.12972. The only other statistically significant coefficient in this regression was for profile_typeB_high school, at -0.053, statistically significant at the 1% level; however, its standard error of 0.032 makes its 95% confidence interval overlap with 0.

Overall, these models demonstrate an overall positive effect of the treatment. However, our small number of profiles and inability to realistically duplicate the same profiles with different educations prevented us from fully attributing this positive effect to the treatment alone. Future research could shed further light on the complexities of mate selection in digital contexts. For a more comprehensive picture, we ran Experiment B.

Experiment B

Overview

Experiment B was a Tinder simulation survey. The Tinder simulation survey allowed us to observe stated interactions in a simulated environment. Although stated interactions often differ from actual interactions, this design enabled the collection of additional covariate information such as participants' education level, race, and age. Despite a reduced number of observations compared to our naturalistic experiment, the ability to regulate potential confounding variables empowered us to conduct an insightful analysis through an alternative perspective.

Comparison of Potential Outcomes

In Experiment B, we compared the following potential outcomes. The untreated potential outcome, or $Y_i(0)$, was whether the subject found the fictitious woman in question likable given that she had only a high school diploma. The treated potential outcome, or $Y_i(1)$, was whether the subject found the fictitious woman in question likable given that she had a graduate degree.

In this simulated Tinder context, we define likability as the *stated* degree of appeal – rather than *actual* degree of appeal, which Experiment A measured – that other users perceive in a profile of interest. We measured stated likability by one main metric and one secondary metric: whether a participant stated that they would “swipe right on,” or “like,” the fictitious woman in question and, secondarily, a participant’s stated attractiveness rating for the fictitious woman.

Survey Creation

Using Pollfish, an online market research platform, we were able to record 100 survey responses with each surveyor reacting to two simulated Tinder profiles, one woman with a high school degree and the other with a graduate-level degree. Similar to Experiment A, the age for these women was displayed as 25 to ensure realism for the possibility of having a graduate degree and to correspond to an age range where individuals typically begin contemplating serious relationships. We selected profile images according to what we thought were of similar attractiveness so that we could somewhat control respondents' personal preferences for physical attractiveness. See Appendix B for the images used and information provided for each profile. In terms of survey eligibility criteria, we set filters to ensure that respondents would satisfy our ideal demographic of heterosexual 18-34 aged males. We also recorded respondents' education level and race as covariates.

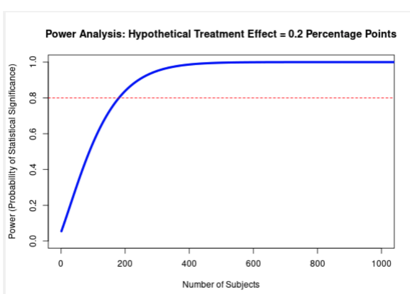
Randomization Process

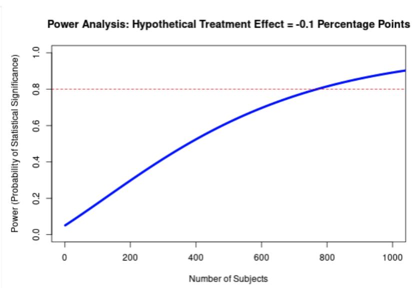
We created two versions of our survey. Both surveys were the same other than swapped images of the women. Again, see Appendix B for exact simulated profile presentation. Randomizing survey versions was a straightforward process through Pollfish, with an even 50/50 split of respondents per survey.

Treatment

In each survey, respondents reacted to two profiles: one untreated (high school degree), and one treated (graduate-level degree). We used different images for treatment and non-treatment, but randomized which woman was treated across surveys to avoid bias (two ways images could be randomized, thus two survey versions). Using images of subjectively similar physical attractiveness helped us decrease the influence of personal preference in determining attractiveness. We collected two observations for every respondent. With 100 survey respondents, our final dataset featured 200 observations.

Power



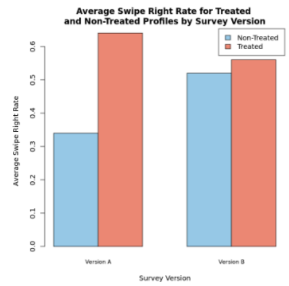


The first graph shows the power analysis for a hypothetical positive treatment effect of 0.2 percentage points. As the number of subjects increases, the power of the study rapidly increases, reaching the conventional threshold of 0.8 at around 181 subjects. After this point, additional subjects slightly increase the power, which approaches 1.0 but with diminishing returns. The second graph represents a hypothetical negative treatment effect of -0.1 percentage points. Here, the power reaches the threshold of 0.8 with many more subjects, approximately at around 770. This suggests that detecting a treatment effect of 0.2 percentage points requires many more subjects than detecting an effect of -0.1 percentage points, given the same study conditions.

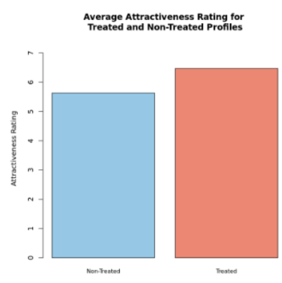
Similar to the power analysis done in Experiment A, the difference between the two hypothetical treatment effects (-0.1 and 0.2) shows that the magnitude of the effect size has a considerable impact on power. If the actual effect of graduate education on likability is smaller than the effect sizes we hypothesized, even a well-powered study would not detect it. Although these analyses confirmed the power of our study study, it is still crucial to consider study limitations in our interpretation.

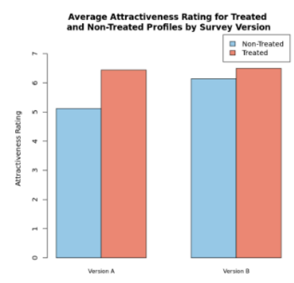
Data

After collecting the 200 observations from 100 survey responses (two observations per respondent) and deciding on our subsequent modeling methodology, we explored our data to identify any patterns. The charts below visualize different views of our collected data focused around swipe right (like) rates and attractiveness ratings. Our bar charts show data aggregated by treatment, survey version, and education. We chose to not include other variable aggregates like age or race because they did not hint at any significant patterns we could further investigate.

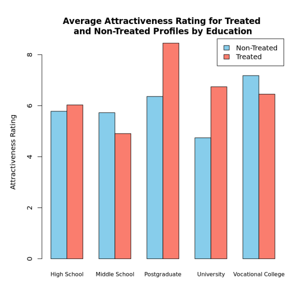


The above charts show us how the average swipe right rates differ between non-treated and treated profiles. On the left, the swipe right rates increase from the untreated profiles (0.43) to the treated profiles (0.60). However, to account for any differences in appearance between the two profile images, we aggregated the data by survey version. In Version A, untreated profiles had a 0.34 swipe right rate and treated profiles had a 0.64 swipe right rate. For Version B, the difference in swipe right rates wasn't as big, but there still is an increase (0.52 for untreated, 0.56 for treated). Given this increase from non-treated to treated, we looked at how treatment influenced attractiveness ratings on a scale of 1-10.





In the chart on the left, non-treated profiles received a lower average attractiveness rating (5.63) than the treated profiles' average attractiveness rating (6.47). To investigate whether it was the difference in physical appearance or treatment that was what led to these observed differences, we also split the untreated and treated by survey version. In Version A, non-treated profiles received a 5.12 average rating and treated profiles with 6.44. In Version B, there was a smaller difference, but still an increase from untreated to treated (6.14 to 6.50).



This bar chart shows us how average attractiveness ratings varied by non-treated and treated profiles, now aggregated by the different education levels recorded by the survey respondents. We can see that for postgraduate and university educated respondents, the treated profiles were ranked significantly higher than untreated profiles (high school educated respondents did show a better liking towards treated profiles, but the difference wasn't as large). For middle school-educated and vocational college respondents, they preferred the non-treated profiles over the treated profiles (maybe signifying that at lower education levels, education is not a major factor when selecting partners, but it is for higher educated individuals).

Models

Comparison of Linear Regression Models for SwipeRight

	Dependent variable:			
	SwipeRight			
	(1)	(2)	(3)	(4)
Treated	0.170**	0.170**	0.170**	0.170**
	(0.070)	(0.070)	(0.070)	(0.070)
SurveyVersion		0.050		0.011
		(0.070)		(0.077)
Educationmiddle_school			0.077	0.072
			(0.123)	(0.128)
Educationpostgraduate			0.168	0.162
			(0.123)	(0.131)
Educationuniversity			0.017	0.016
			(0.086)	(0.086)
Educationvocational_technical_college			0.122	0.120
			(0.123)	(0.124)
Constant	0.430***	0.405***	0.384***	0.380***
	(0.049)	(0.061)	(0.071)	(0.076)
Observations	200	200	200	200
R^2	0.029	0.031	0.042	0.042
Adjusted R^2	0.024	0.022	0.018	0.013
Residual Std. Error	0.495 (df = 198)	0.496 (df = 197)	0.497 (df = 194)	0.498 (df = 193)
F Statistic	5.898** (df = 1; 198)	3.196** (df = 2; 197)	1.714 (df = 5; 194)	1.424 (df = 6; 193)
Note:				*p<0.1; **p<0.05; ***p<0.01

We started with a baseline linear regression of "SwipeRight" on "Treated." Here, we can see that "Treated" was statistically significant in impacting "SwipeRight" and had an estimate of 0.170. This value represents the increase in the odds of "SwipeRight" for when a profile is treated. The constant (which was significant throughout all four models) represents the baseline probability that a profile will be swiped right. The next model we considered was one with the "SurveyVersion" covariate. Including the "SurveyVersion" helped us understand whether the actual image of the profile was influencing "SwipeRight" outcomes. However, "SurveyVersion" was not statistically significant, telling us any difference in image-specific effects did not pose significant issues. In the next model, we use the categorical "Education" (highest education achieved by respondent) variable as a covariate. Our results showed no significant effect of respondent education on the "SwipeRight" outcome. In our fourth model, we combined all of the aforementioned covariates. As the previous models implied, only "Treated" was a statistically significant variable in predicting "SwipeRight" (this was true across all four models).

While our models do provide evidence of the treatment having a significant impact on profiles' swipe right rates, we were not successful in identifying other significant predictors in this experiment. We collected information about respondents' age, race, and education, but none proved to have a statistically significant relationship with swipe right rates.

We also tested these same regressions on the "AttractivenessRating" outcome, which represented the respondents attractiveness rating from 1 to 10 for a given simulated profile.

Comparison of Linear Regression Models for AttraactivenessRating

	Dependent variable:			
	AttractivenessRating			
	(1)	(2)	(3)	(4)
Treated	0.840**	0.840**	0.840**	0.840**
	(0.402)	(0.401)	(0.397)	(0.397)
SurveyVersion		0.540		0.372
		(0.401)		(0.434)
Educationmiddle_school			-0.588	-0.764
			(0.693)	(0.724)
Educationpostgraduate			1.503**	1.293*
			(0.693)	(0.736)
Educationuniversity			-0.163	-0.184
			(0.485)	(0.486)
Educationvocational_technical_college			0.912	0.837
			(0.693)	(0.699)
Constant	5.630***	5.360***	5.486***	5.358***
	(0.284)	(0.347)	(0.403)	(0.430)
Observations	200	200	200	200
R^2	0.022	0.030	0.066	0.070
Adjusted R^2	0.017	0.021	0.042	0.041
Residual Std. Error	2.843 (df = 198)	2.837 (df = 197)	2.805 (df = 194)	2.807 (df = 193)
F Statistic	4.365** (df = 1; 198)	3.097** (df = 2; 197)	2.760** (df = 5; 194)	2.419** (df = 6; 193)
Note:				*p<0.1; **p<0.05; ***p<0.01

Although "AttractivenessRating" wasn't the main variable of interest for this experiment, it was a good indicator of whether the respondent would swipe right. Starting with a simple treatment-outcome contrast, we can see that "Treated" had a coefficient of 0.840 with a p-value < 0.05. This means that for treated profiles, given attractiveness ratings increased by 0.840 rating points on average. The constant is also significant at a 5.630 estimate, meaning that all simulated profiles were given a baseline average attractiveness rating of 5.630. In the second model for this outcome, we use "SurveyVersion" as a covariates. Similar to our first set of models, 'SurveyVersion' was not statistically significant. In the third model, we use "Education" to test whether respondents' own education levels influenced their ratings. Only postgraduate education was significant with a 1.530 coefficient. This means that for postgraduate educated respondents, their profile attractiveness ratings were higher by 1.530 on average, than those without. In the fourth model, we included both covariates and found that "Treated" and "EducationPostGraduate" were statistically significant predictors of "AttractivenessRating," although the standard error of "EducationPostGraduate" indicates an overlap with 0 in the 95% confidence interval for the coefficient.

Overall, our linear regression models regressing attractiveness on the treatment, plus some covariates, support the idea that profiles treated with graduate degrees were more likable in terms of attractiveness ratings. The only difference is that these regressions found that survey respondents' education level was statistically significant, specifically only for those with a postgraduate degree. To test whether respondents with higher education scored treated profiles greater than those with lower levels, we ran a linear regression using an 'Education' interaction term with 'Treated.'

Linear Regression Model with Education Interaction

	Dependent variable:
	AttractivenessRating
Treated:Educationhigh_school	0.401
	(0.569)
Treated:Educationmiddle_school	-0.721
	(0.890)
Treated:Educationpostgraduate	2.825***
	(0.890)
Treated:Educationuniversity	1.113**
	(0.550)
Treated:Educationvocational_technical_college	0.825
	(0.890)
Constant	5.630***
	(0.280)
Observations	200
R^2	0.070
Adjusted R^2	0.046
Residual Std. Error	2.800 (df = 194)
F Statistic	2.912** (df = 5; 194)
Note:	*p<0.1; **p<0.05; ***p<0.01

Looking at the results above using 'Education' as an interaction term, we now see that 'Treated:EducationUniversity' and 'Treated:EducationPostGraduate' are statistically significant. This tells us that for treated profiles, respondents with university and postgraduate educations were more likely to give higher average attractiveness ratings than respondents with lower level educations—this aligns with the bar chart we created earlier depicting average attractiveness ratings split by respondent education levels. However, while this idea that higher educated respondents care more about their potential partners' education levels may be true for our dataset, this does not guarantee it will always hold true. The standard errors (values in the parentheses) are very high, indicating that there is a high level of uncertainty with our estimates.

In summary, the models explored in Experiment B highlight that profiles treated with graduate degrees did receive more swipe rights and higher average attractiveness ratings than untreated profiles. While we were able to explore how other covariates influenced these metrics, we were only successful in identifying a relationship with the education level of the survey respondents and the treatment (higher educated respondents were paying more attention to the profiles' education). However, we take this finding with caution because of high standard errors; hopefully, in future experiments, we can work with larger datasets to reduce this error and also identify new relationships with more certainty.

Appendix A - real Tinder profile details

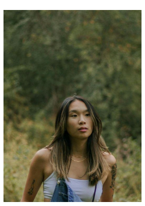
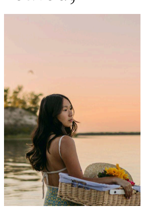
Profile A – Katrina
Bio: Waitress who loves chill matcha chats and photography!
Treatment: Not treated, high school

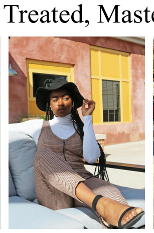
Profile B – Naomi
Bio: Currently a hostess at a local restaurant! I love to read and I love all animals!
Treatment: Not treated, high school

Profile C – Jasmine
Bio: Stanford PhD geek with endless hobbies – let's chat and geek out together!
Treatment: Treated, PhD

Profile D – Aisha
Bio: Just earned my Master's at UC Berkeley – ready to conquer the world with a sprinkle of Berkeley spirit! Let's celebrate together!
Treatment: Treated, Master's

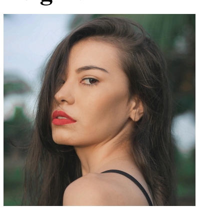
Profile E – Samantha
Bio: Spreading joy, one dish at a time! Let's savor the flavors of life together!
Treatment: Not treated, high school

Profile F – Bella
Bio: Freshly minted with my Masters from UC Berkeley, armed with ambition and a hint of Berkeley's zest for life! Join me in celebrating our victories and conquering new horizons together!
Treatment: Treated, Master's

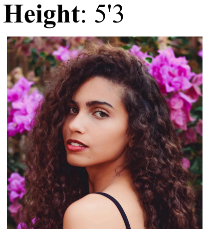
Appendix B - simulated Tinder survey details

Treatment Survey Profiles

Woman A – Isabella
Looking for: a life partner
Highest education obtained: high school diploma
Height: 5'3



Woman B – Sofia
Looking for: a life partner
Highest education obtained: graduate school degree
Height: 5'3



Control Survey Profiles

Woman A – Isabella
Looking for: a life partner
Highest education obtained: high school diploma
Height: 5'3

Woman B – Sofia
Looking for: a life partner
Highest education obtained: graduate school degree
Height: 5'3

Conclusion

In conclusion, our investigation into the impact of graduate-level education on the perceived likability of heterosexual women, utilizing both field experimentation and simulation studies, has yielded intriguing results. Both the initial Tinder field experiment and our subsequent simulation study revealed a statistically significant positive effect of graduate-level education on the likability of heterosexual women.

Despite the promise of these findings, we advocate for a cautious interpretation of its external validity and emphasize the need for continued research in this domain. The simulated nature of our study warrants a cautious approach as well as the limitations that we faced in our field experiment.

Future work can explore the area of meta-analysis by combining the strengths of both approaches of a field experiment and a simulation study. By leveraging the “naturalness” of a field experiment alongside the covariate-controlling capability of a simulation, such an approach could yield a more robust understanding of the average treatment effect of graduate-level education.

Our findings underscore the importance of interdisciplinary approaches in investigating the complex interplay between education and social perception. Through our work here we hope that our findings can point towards a more nuanced understanding of these dynamics, ultimately informing both academic discourse and practical interventions in the realm of social psychology.