Michael Powers, Jessica Vincent, Cathy Zhou

W241.1 Final Project

April 23, 2018

**Examining the short term effects of sexually degrading music on self-image**

[**1. Background**](#_5tled2lt8o2l) **2**

[**2. Research Question and Hypothesis**](#_ot6s986wpchd) **2**

[**3. Methods and Experimental Design**](#_4v24aklwfhnu) **3**

[3.1 Administration](#_klopstivzkpm) 4

[3.2 Participants and Randomization](#_dz2546hb6207) 6

[3.3 Pilot](#_jao0rz7ytxu8) 7

[3.4 Placebo Design](#_nmmnicar2dyp) 7

[3.5 Treatment Variables](#_q35uuio0v83v) 7

[3.6 Dependent Variables](#_ervwo2ckleut) 8

[3.7 Controls](#_e2nlvb2j2dfq) 8

[3.8 Experimental Design Diagram](#_yxu55ejd4h2) 9

[**4. Results and Analysis**](#_5texep6biyq2) **10**

[4.1 Distribution of Outcomes](#_xcsplne3xrzz) 10

[4.2 Placebo vs. Control](#_qyku55yhae88) 11

[4.3 Final Results](#_izap5i5s2so6) 12

[**5. Challenges, Confounds and Concerns**](#_rxch6rd0org7) **13**

[5.1 Distribution of Randomization](#_r9w7v94ll6mr) 13

[5.2 Defining Control versus Treatment](#_7lrmez12cqe4) 14

[5.3 Difference in Environment](#_rb09uprei11l) 14

[5.4 Sampling Bias](#_bejs0suthozy) 14

[5.5 Observation Effect](#_kt0q149dwzoc) 14

[5.6 Does Not Measure Long Term Effect](#_5svlqrbyvncx) 15

[**6. Opportunities for further study**](#_e25f1lvfeowu) **15**

[6.1 Measure Greater Short-Term Exposure](#_m2ixp7k2p182) 15

[6.2 Measure Long Term Effects of Continued Exposure](#_tzs1kchlgxn9) 15

[6.3 Explore Different Genres and Songs](#_usd883nnyq63) 15

[6.4 Measure Effects on Attitude Toward Others](#_896c8xsg6eiw) 16

[**7. Conclusion**](#_eq2kl08ungb1) **16**

[**8. References**](#_n559rfgdensg) **16**

## 1. Background

In light of the recent and pervasive sexual misconduct and assault allegations across Hollywood, sports franchises, government, and more, many people in our country are looking for answers around what is fueling this national epidemic. You don’t have to look very hard to see that these broad themes of toxic masculinity and objectification of women are extremely prevalent in our society. They are commonly observed in movies and tv shows, advertising, and music. Social cognitive theory demonstrates that people “learn not only by direct experience but also by exposure to modeled and rewarded behavior.”1 In this way, we are reinforcing objectification and power-dynamics through many channels we engage with. One such channel we choose to explore further is popular music.

Several studies have looked to observe the impact of modern popular music on teenage sexual behavior, most commonly the age at which teenagers first have sex. But what effect does this music have on adults? It has been demonstrated that “more than one third of popular songs contain explicit or strongly implied references to sexual activity” with “degrading references” constituting two thirds of these sexual references.”1 Degrading references are those in which lyrics refer to women as objects, contain violent sexual language, infer non-consent, etc. Given the associated magnitude therefore with which people are exposed to such language and ideals, propagated by artists who receive positive status and reward in society, we looked to broaden the existing research to measure effects of such music on the adult psyche as well. More specifically, how does such music make women and men feel about themselves?

## 2. Research Question and Hypothesis

*What impact does short-term exposure to sexually degrading music have on short-term self-image for male and female adults, compared to non-sexually degrading music?*

Adults were defined as individuals 18 years or older. Sexually degrading music was defined as songs with at least 5 unique references that represent women as sex objects, use violent language, depict men as sexually voracious, or reference sex as casual and meaningless. For the purposes of this study, we focused only on male artists singing about women, though different configurations could be explored in the future.

Since ran a two-tailed test for robustness, our formal hypotheses were as follows:

H0: self-image non-degrading music = self-image degrading music

Ha: self-image non-degrading music self-image degrading music

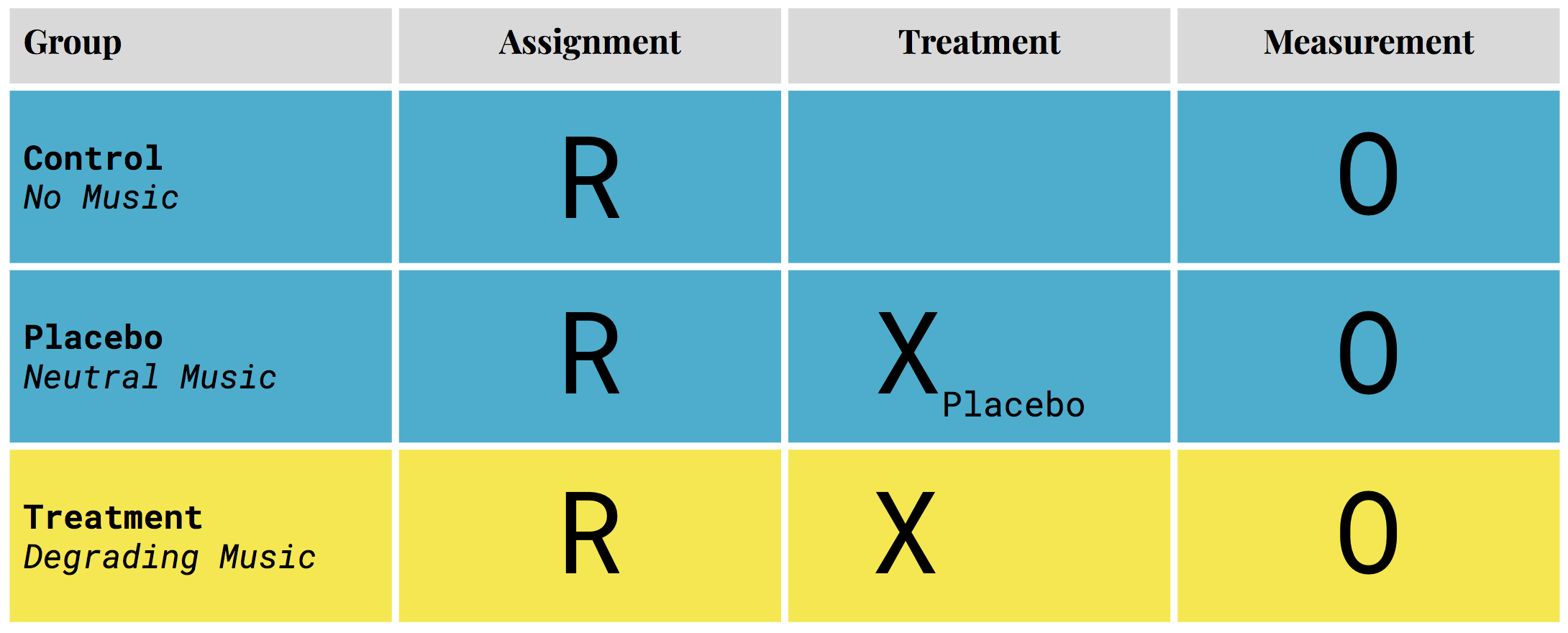
We informally hypothesized that the self-image score of those who did not listen to degrading music would be greater than those who did. We additionally hypothesized that since this experiment focused on lyrics degrading to women (sung by men), that this directional hypothesis would be especially true for female participants.

When speculating about the directional alternative hypothesis for men, we were inconclusive. On one hand, perhaps such lyrics could make men feel more powerful and therefore more confident, but on the other hand, perhaps degrading toxic masculinity could also make men feel worse about themselves and the cultural expectations of them. Therefore, we simply conjectured that ATEwomen ATEmen prior to running the experiment.

## 3. Methods and Experimental Design

Summary: In this experiment, participants were told that they would be asked to listen to two 30-second clips of a random song and would then be given a short quiz to test their recollection of the song lyrics before answering some questions about themselves. They were then asked to record their age and gender, and anyone that selected under 18 years of age was exited from the survey. Participants were then randomly assigned to control, placebo, or treatment using the Qualtrics randomization feature. Participants in the control group received a message that they were randomly selected to skip the music clips, participants in the placebo group were presented with two 30-second clips of a non-degrading song, and participants in the treatment group were presented with two 30-second clips of a song containing sexually degrading lyrics. If a participant was in the placebo or treatment group, they were asked whether they had heard the song previously and if they liked the song. All groups were then asked to answer 10 questions related to self-image and esteem (adopted from the Rosenberg self-esteem scale) using a 7- point scale.

In general, we executed an RXO, post-test, between subjects experimental design.



The following provides additional details and insights into this set up.

### 3.1 Administration

This experiment was conducted using Qualtrics. All background context and instructions were provided to the subject as written text on the screen. This setup isolated all human elements to only the song and the lyrics. We speculated that the presence of an in-person researcher might cause the participant to behave or think differently than if they were alone. They might also feel compelled to answer the survey questions more or less strongly due to the perceived preferences of the researcher/ interviewer. Even if the instructions were administered by a voice recording, it could still influence the participant. Furthermore, the gender of the researcher might also have a confounding effect. Therefore, we decided to avoid these potential confounds by having the experiment be self-read and navigated.

The first page of the Qualtrics form stated:

*Welcome!*

*This study contains two 30-second audio clips from a random song. After you listen to each clip, you will be given a short quiz that will test your ability to recall the song lyrics. Once you complete the quiz, you will be asked to provide some information on yourself. Please note, once you move to a next page you will not be able to move back to listen to clips or edit responses.*

*This study will take under 5 minutes and all responses are anonymous. Thank you for your participation!*

After advancing off the intro page, participants were asked their age and gender. Participants that selected less than 18 years-old were exited from the form. After clicking the next button on this page, that Qualtrics randomizer tool assigned participants to either the control, placebo, or treatment group.

Participants assigned to the control group next saw a message informing them that they were randomly selected to skip the audio clips and to continue to the next page.

Participants in the placebo or treatment groups were advanced to a new screen with an embedded play button and reminder instructions to listen to the entire 30-second clip and that there would be a quiz on the lyrics. The next button to advance the screen only appeared once the clip had been played in its entirety to ensure all participants listened to the clips for an equal amount of time and listened to the entire treatment. There was no back button.

After the first 30-second clip, participants clicked next and answered two multiple choice fill-in-the-blank lyrics quiz questions. They then clicked next and were presented with a second 30-second clip from the same song. After that they were presented with a second two-question lyrics quiz.

The clips of the treatment song were selected to highlight the objectifying and degrading sections of the music. We utilized clips rather than playing the entire song so that participants would focus on the degrading sections without fighting a waning attention span. We selected only one song to keep the overall time to complete the survey under 5 minutes, since we expected that survey abandonment rates would greatly increase with survey length, especially since we did not incentivize completion.

The lyrics quiz was implemented to both emphasize treatment as well as mask the true intention of our experiment. From our own experience and feedback from others, we knew that sometimes people can enjoy a song without having truly listened to the meaning of the lyrics. Additionally, due to known findings such as the Hawthorne effect in which individuals modify an aspect of their behavior in response to their awareness of being observed, we were concerned that being too transparent with our research question could cause participants to modify their responses. Therefore, the lyrics quiz was designed to incentivize participants to truly listen to the lyrics and reduce bias in the answers to the self-esteem questions.

After two clips and quizzes, participants in the placebo and treatment groups advanced to a page that asked them to select if they had heard the song before (yes - many times, yes - at least once, maybe, no), and how much they liked the song they just heard (like a great deal, like somewhat, neither like nor dislike, dislike somewhat, dislike a great deal). We speculated that emotional response to a song regardless of lyrics may be influenced by these factors, so these questions were included to be used as controls in our analysis.

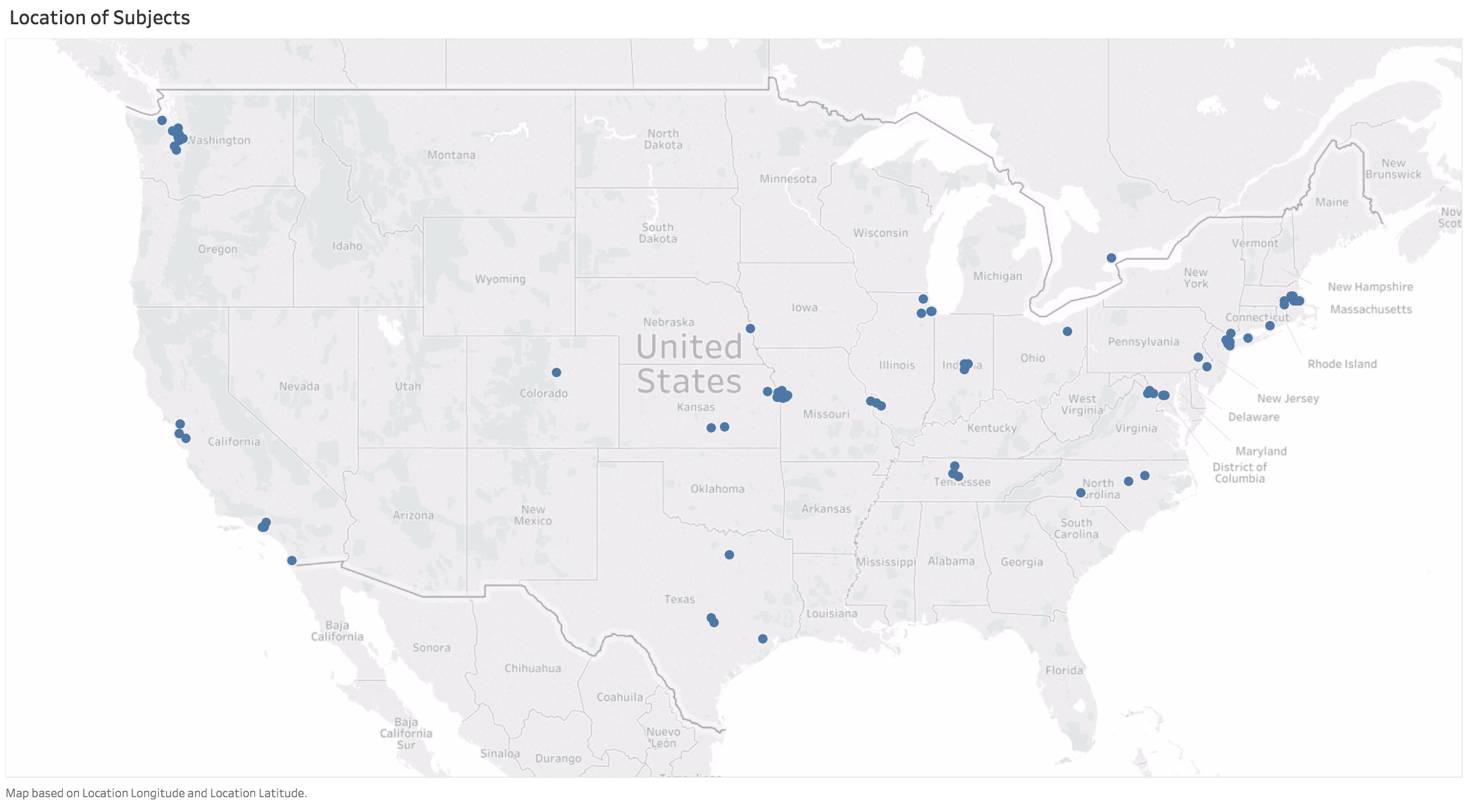
Lastly, participants in placebo and treatment reached a screen with 10 self-image questions. Participants in the control group were advanced directly to this screen after clicking next on the page where they were informed they would skip the audio clips. The 10 questions were adopted from the Rosenberg Self-Esteem scale, so that the self-esteem outcome metric would be psychologically sound.2 Participants answered these questions using a 7-point Likert Scale, and the direction of the questions varied so that participants could not simply select the same answer to all 10 questions without raising suspicion that they had not truly read the questions (e.g. On the whole, I am satisfied with myself vs. All in all, I am inclined to feel that I am a failure).

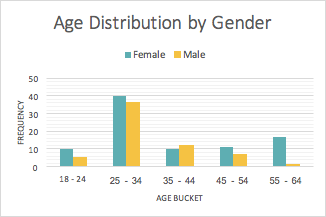
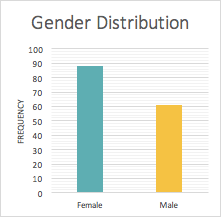
While the official Rosenberg Self-Esteem Scale uses only 4-point Likert Scale options for response (strongly disagree, disagree, agree, strongly agree), we expanded this to a 7-point scale (strongly disagree, disagree, slightly disagree, neither agree nor disagree, slightly agree, agree, strongly agree) to provide a neutral option and increase the distribution of scores so that we could analyze differences with more precision.

### 3.2 Participants and Randomization

We recruited participants by reaching out directly to friends, family, and coworkers, as well as by posting requests for participation on our personal social media accounts. We aimed to keep the survey completion time under 5 minutes total, since we were not monetarily incentivizing completion and knew that attention spans are relatively short. Since we expected that the treatment effect could differ for males and females, we aimed to get a relatively even distribution of male and female respondents and monitored completion progress accordingly, though we did end with more female participants.

Our participants ended up having the following gender, age, and geographic distributions.





We used Qualtrics’ built in randomization feature to randomize assignment to the control, placebo, or treatment group after the participants advanced past the gender and age screen. Furthermore, we used this feature to randomize the order in which the self-esteem questions appeared, in case the order in which they were presented effected scores.

### 3.3 Pilot

Prior to officially conducting our experiment, we ran a pilot study using friends and family as our respondents to ensure that the Qualtrics form and randomization tool were displaying and working correctly, and to collect any initial feedback.

We assigned pilot participants only to placebo or treatment at this point in time, and collected 10 responses. We had an even distribution of participants between placebo and treatment, indicating that the randomization tool was working as expected. Initial feedback from participants was that the length of time to complete was appropriate, that the lyrics quiz and songs were fun, and that they had been mostly unaware of the true research question.

### 3.4 Placebo Design

After the successful pilot, we opted to make only one change to our experiment. We had originally planned only two groups - one that listened to a “neutral” song and one that listened to a degrading song. We conducted the pilot with these two.

However, after some discussion, we speculated that our choice of “neutral” song could also be influencing responses. Since music is not “binary,” even non-degrading songs could have other emotional triggers that could influence outcome. We therefore opted to create a placebo design by also including a control group that listened to no music. This group was included to help ensure that our neutral song did not have an effect on self-image as compared to listening to no music. With control now defined as no music, we thought of the group listening to the neutral song as our placebo group. This group acts as a placebo because participants were subjected to the same conditions and screens as the treatment group, some of which the control was not, but with a “benign” treatment song.

If we think of a listening to a degrading song as taking an experimental drug, then listening to the neutral placebo song was like taking a sugar pill, and being in the control group was like taking no pill at all.

### 3.5 Treatment Variables

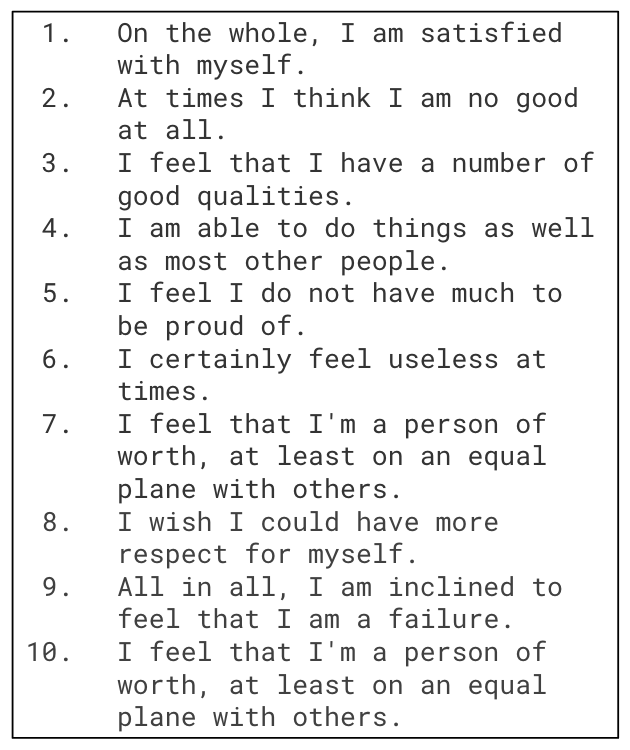
Participants in the control group did not listen to any music.

Participants in the placebo group listened to two 30-second clips from Blue by Eiffel 65. This song was selected because the lyrics are both non-degrading and relatively nonemotional. Selecting a “neutral” song is a hard task when it comes to music.

Participants in the treatment group listened to two 30-second clips from Talk Dirty by Jason Derulo, featuring 2 Chainz. The first 30-second clip was sung by Jason Derulo and featured objectifying language, and the third 30-second clip was sung by 2 Chainz and featured objectifying and explicit language. While there are more explicitly degrading songs with even more profane and shocking lyrics, we selected Talk Dirty for this round of experimentation because it was a more currently popular song that participants would therefore not be surprised to hear, and so that we could examine the effects of a slightly more “subtle” form of objectification since we suspect this is the most common form of degrading lyrics in mainstream music.

### 3.6 Dependent Variables

The Rosenberg Self-Esteem Scale generates an overall esteem score by summing the answers to each of the 10 self esteem questions below, while adjusting for the direction of the question. All questions were answered using a 7-point Likert Scale (strongly disagree, disagree, slightly disagree, neither agree nor disagree, slightly agree, agree, strongly agree), but for some questions “strongly agree” was the most positive answer while for others it was the most negative. We always assigned numbers to responses such that the most positive response was rated “1” and the most negative was rated “7”. Overall esteem scores therefore ranged from 10-70, with lower scores indicating higher self-esteem.



### 3.7 Controls

We asked all participants their age since we conjectured that age might also have an influence on overall self-esteem (research suggests that younger people are sometimes less self-assured than older people, since age typically brings more successes with life-experience).

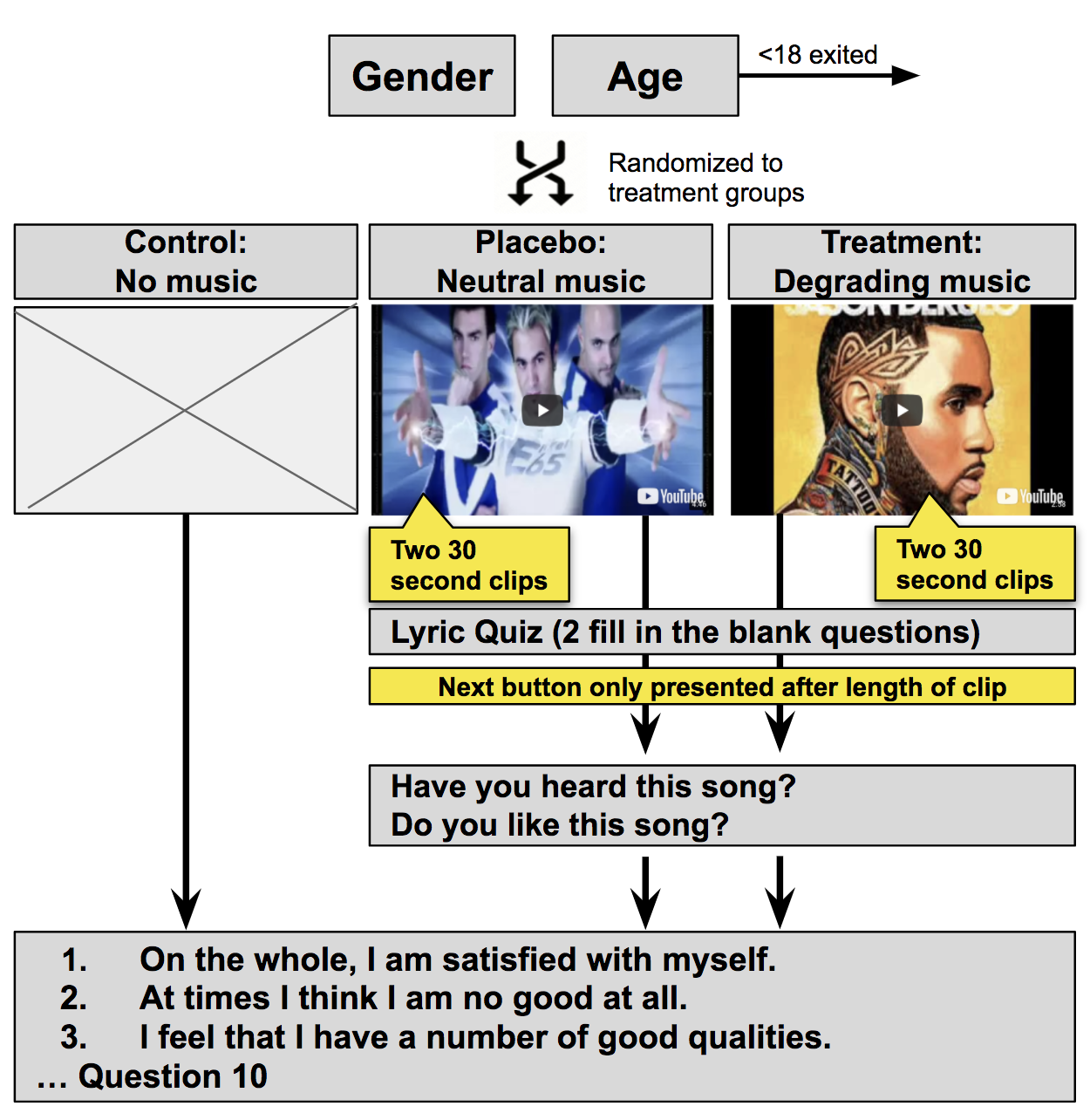
We asked all participants their gender since we hypothesized that music that was sung by men and was degrading toward women might have a more negative effect on women’s self-esteem than men’s.

We asked all participants to rate their level of familiarity with the song since we conjectured that high familiarity might cause participants to become “immune” to the lyrics themselves.

We asked all participants to rate their enjoyment of the song since we conjectured that factor might strongly influence the outcome (we had heard people say things like “I know this song is pretty bad, but I love the beat”).

Based on these conjectures and hypotheses, we included these factors as controls in our analyses.

### 3.8 Experimental Design Diagram

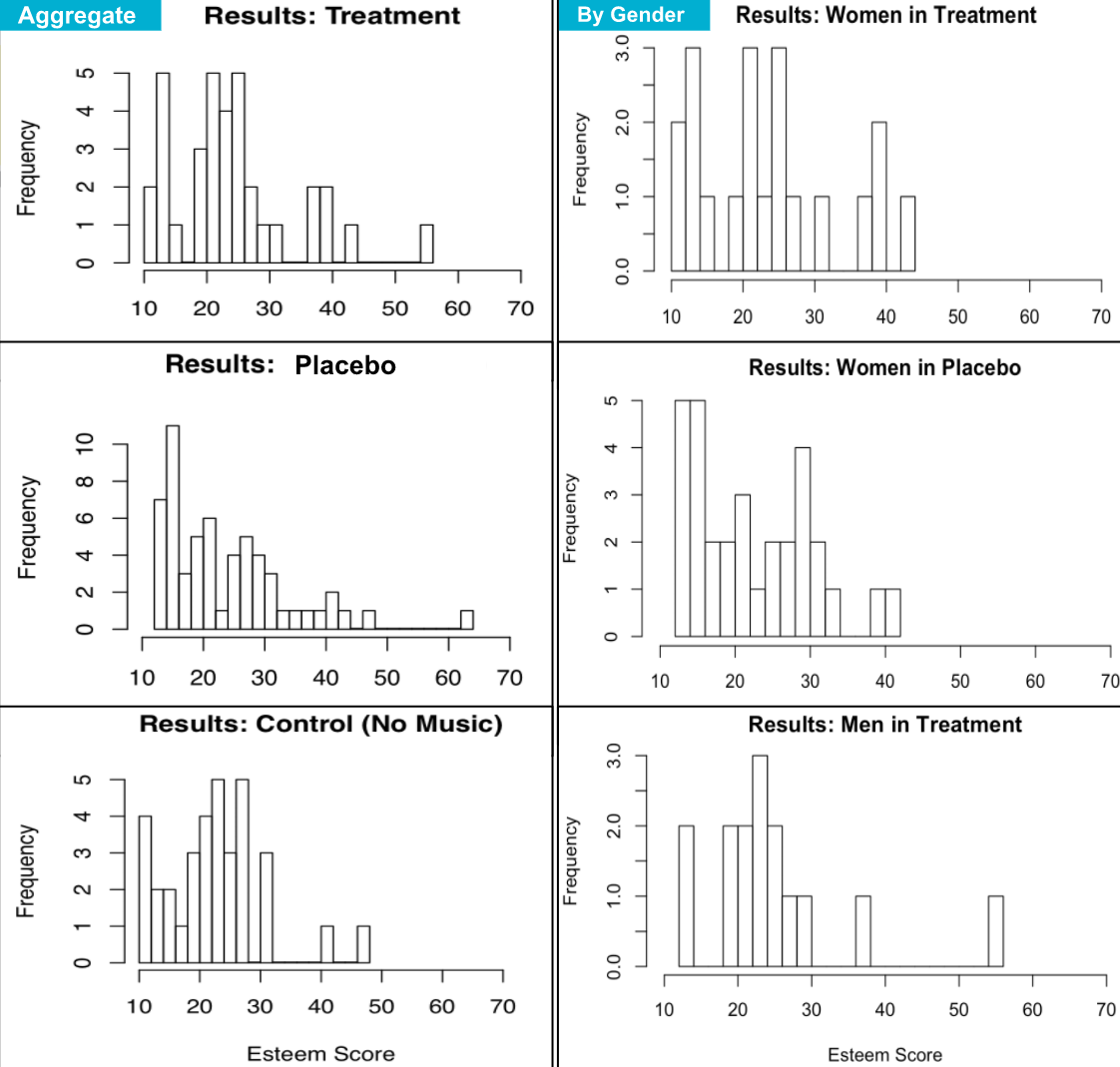


## 4. Results and Analysis

## 

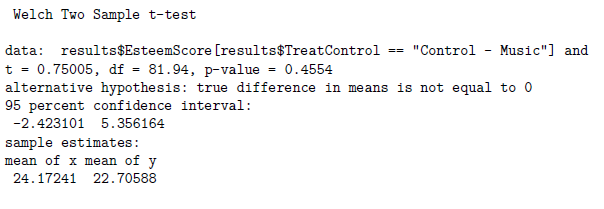
The experiment concluded with a total of 160 subjects. We did not have complete results for 33 of these subjects because 25 did not finish, and 8 were not allowed to advance because they were under 18. The results of our analysis below are based on the 127 complete observations.

### 4.1 Distribution of Outcomes



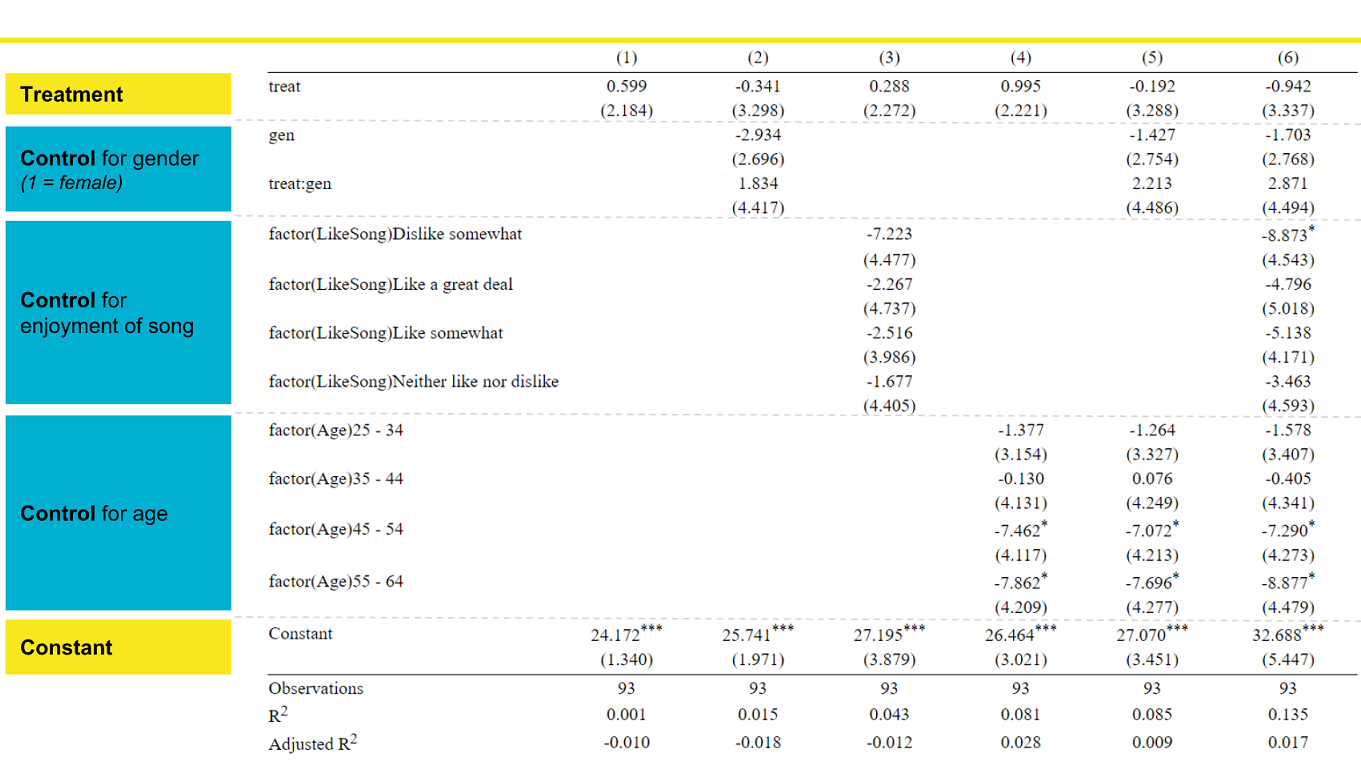
The outcome, or dependent variable, is the esteem score which was based on the Rosenberg Self-Esteem Scale. This score can range from 10, which signifies the most esteem, to 70, which signifies the least esteem. We started to see normal distributions forming around the center of esteem score 23. These distributions also had another spike around the lowest possible score of 10. This additional spike represents people with overall high self esteem. The charts above seem to indicate that the treatment group has a larger right tail compared to the control and placebo groups. These right tail observations represent people with lower self esteem. This tail appears to be even larger in the female observations. We anticipate these distributions would become more refined if we were able to obtain more observations.

### 4.2 Placebo vs. Control



We previously mentioned that in addition to the treatment group, we included a control group with no music and a placebo group with neutral music. We included a control group with no music so we could attempt to ensure that the placebo song did not have an effect on the outcome esteem score. A two sample t-test indicates the average esteem score of the placebo group was 24.2, and the average esteem score of the control group was 22.7. The difference in these scores is not statistically significant (p-value = 0.455). This resulted in us concluding the song used in the placebo group did not have an effect on the outcome esteem score. However, we decided to focus our analysis on the treatment and placebo groups, since these groups experienced the most similar conditions (i.e. listen to music and take lyrics comprehension quiz).

### 4.3 Final Results

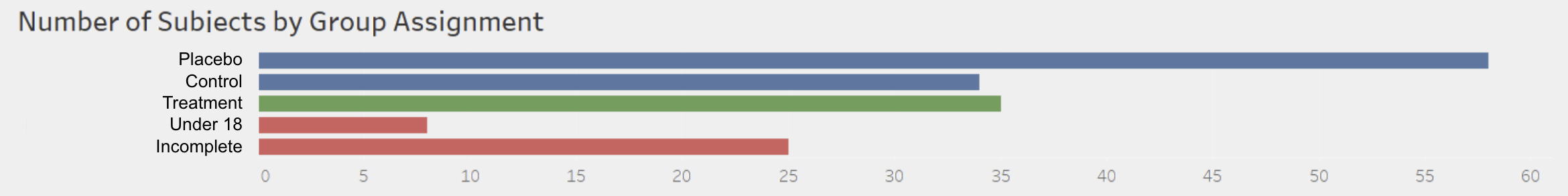


The models above reflect the experimental results of the treatment and placebo groups. The effect of our treatment is not statistically significant in any of the models. However, we see coefficients that matched the directions of our hypotheses. The [gen] variable, which is coded 1 for female and 0 for male, has a negative coefficient which indicates that women had a baseline lower score (i.e. higher esteem) than men. The coefficient of the interaction term [treat:gen] is positive, which suggests our treatment raised scores (i.e. lowered esteem) more in women than men. We anticipated this outcome because the treatment song was sexually degrading towards women. We also noticed that older subjects disliked our song choice more and had higher esteem, compared to younger subjects.

## 5. Challenges, Confounds and Concerns

The following were observed challenges, or factors that could possibly introduce confounding variables or unobserved heterogeneity that we must minimize or be aware of.

### 5.1 Distribution of Randomization



We used Qualtrics’ randomization tool to assign participants to treatment, placebo, and control. Given the nature of randomization, we would expect roughly ⅓ of participants in each of these buckets, and we did in fact see treatment distribution in line with these expectations in our pilot study. However, we saw a much greater number of participants assigned to the placebo group than the treatment or control in our experiment.

Uneven treatment groups caused us to view attrition in more detail to ensure that the 25 abandoned survey didn’t show any potential bias (e.g. people in treatment and control abandoning at higher rates than placebo for an unknown reason that could skew results). Based on our further analysis, of the 25 subjects started the experiment and did not complete it, 19 stopped on the first page after age and gender. These participants would not yet have been assigned to a treatment group. Beyond that, 3 started the treatment song but never completed esteem questions and 3 started the placebo song but never completed esteem questions. Since it seems that attrition is equal among our different groups, we are not too concerned about its impact on our results.

We note that any time randomization does not work as expected, we must be concerned that there could be an underlying reason for the unequal distribution that might impact the results (e.g. if the randomization formula is based on time of day, and the time of day also impacts outcome). However, based on our analysis of the survey flow and underlying randomization formula. we do not have reason to believe that this distribution negatively impacted the validity of our results.

### 5.2 Defining Control versus Treatment

Unfortunately, music is not binary. It isn’t simple to divide it into degrading and not degrading, since it is comprised of the intricacies of the human language and emotional impacts of a given melody. Therefore, defining a placebo versus treatment was not as simple as simply changing one variable. A song that is defined as sexually degrading may differ greatly in many other aspects from a song that is not sexually degrading, making it challenging to pinpoint the exact cause of the treatment effect. To minimize this, we implemented both a control and placebo group so we could compare the effects of the placebo song to the group that did not listen to music. We did not see a significant difference between the control and placebo group outcomes.

As an interesting note, we did explore the idea of selecting songs that have both clean and explicit versions since this would provide the greatest amount of similarity between the placebo and treatment (since the songs contain the exact same melody, singer, and majority of the lyrics). However, we decided against this approach since we cannot control for whether the participant is familiar with both versions prior to the experiment. It is possible that a participant listening to a clean version in the control group might know the explicit version, introducing a confounding effect.

### 5.3 Difference in Environment

Since this experiment was administered online, we could not control the environment in which participants listened to the songs. It is possible that listening in silence versus with background noise, in public versus in private, at a loud volume versus a soft one, with headphones versus through a speaker, or any other such difference could affect the outcome. However, we hope these differences were minimized by the principles of randomization.

### 5.4 Sampling Bias

By selecting individuals through social media, we are perhaps introducing a form of sampling bias. For example, it is possible that soliciting participation this way over represented millennials who may listen to sexually derogatory music more frequently anyway, or that our aquantieses, by virtue of sharing things in common with us, also share more in common with each other than a traditional random sample.

### 5.5 Observation Effect

By seeing very pointed questions about self-esteem immediately after listening to music, it may have become readily apparent to the participants that they were being evaluated on how the music affected their mood. This knowledge may have caused them to overthink their answers, under or over report their feelings, etc. (e.g. perhaps they really like the given song and didn’t want to be complicit in demonstrating a negative effect even if they did recognize concerning mood changes). We hoped to minimize some of this effect by masking the true intention of the experiment by including the lyrics quiz. We received feedback from pilot participants that they were generally aware that we were looking at the impact of music on mood, but they did not pinpoint the role that the sexually degrading lyrics played.

### 5.6 Does Not Measure Long Term Effect

Given the nature and timeline of this experiment, it was not feasible to measure the longer-term or persistent effects of listening to derogatory music. Given this, even though this experiment failed to find any significant results, it does not mean that there are not long-term repercussions that build over time, or with more frequent exposure to degrading lyrics.

## 6. Opportunities for further study

While we did not find statistically significant results this time, we did find general trends that were in line with our hypotheses that we think could warrant further study. Given some of the limitations described above, we propose the following considerations.

### 6.1 Measure Greater Short-Term Exposure

As is, the experiment only measured the immediate impact of listening to 1 minute of lyrical content. While the results were not significant, it would be meaningful to explore whether a greater amount of short-term exposure has an effect. For example, we could have participants listen to 5 degrading songs in a row and then measure self-esteem. This format might necessitate some incentive for participants to complete the survey since it will greatly impact the time to completion, and we know that people have relatively short attention spans.

### 6.2 Measure Long Term Effects of Continued Exposure

As is, the experiment only measured the immediate impact of listening to 1 minute of lyrical content. It would be meaningful to additionally explore the impact of longer-term exposure to degrading music, as we suspect this would be more impactful and significant. For example, we could have treatment participants listen to a degrading song every day for a given period of time and measure self-esteem before and after for a within subjects design. This format would be more costly in terms of time and participant recruitment, and would likely introduce more issues with non-compliance that we would have to plan and adjust for.

### 6.3 Explore Different Genres and Songs

As is, the experiment only measures the impact of 1 degrading song. This leaves a lot of room to test some variations in future experiments. For example, we could test the effects of degrading lyrics among different genres, or the effects of more strongly and shockingly degrading songs versus more subtly objectifying lyrics. These variations could be applied to either short-term or long-term experiments.

### 6.4 Measure Effects on Attitude Toward Others

As is, the experiment looked to identify any measurable difference in self-esteem caused by degrading lyrics. However, it did not consider the secondary impact of how this self-esteem impacted a participant’s attitude towards others. One idea to explore is to develop a secondary part of the experiment with a mechanism that allows us to measure the participant’s attitudes towards other women as well.

## 7. Conclusion

Despite the fact that our sexually degrading treatment song did not appear to have a statistically significant effect on self-esteem, we feel enthusiastic about the results of our experiment because the direction of the treatment effect matched our hypotheses. Specifically, the treatment seemed to lower esteem, and this effect appeared to be larger in females, which was anticipated because the treatment song was degrading towards women.

We view this experiment as a large pilot study. We hope it will be used as a guide for a larger study with more subjects and a more refined experimental design. We believe the negative effect of sexually explicit music on self-esteem is real, and this experiment is one step in helping prove that.

## 8. References

1. [Summary of previous research examples](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3008595/)
2. [Summary of Rosenberg Self-Esteem Scale](http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Self_Measures_for_Self-Esteem_ROSENBERG_SELF-ESTEEM.pdf)
3. [Experiment in-class presentation](https://docs.google.com/presentation/d/106KPRWHGBqr9ATFMEkiYB3OUBH1WBNdZfrdXP_Xlb78/edit#slide=id.g36f6fd2b17_0_94)
4. [Survey Preview](https://berkeley.qualtrics.com/jfe/preview/SV_eXW4vPppOJ0PaYJ?Q_SurveyVersionID=current&Q_CHL=preview)
5. [Treatment Song](https://www.youtube.com/watch?v=P18bCciV_7c&feature=youtu.be&t=65)
6. [Placebo Song](https://youtu.be/zA52uNzx7Y4)