Effects of Sexual Misconduct Training on University Campuses

Early draft, please do not circulate, post or cite *

Carlos Contreras†
Mala Htun‡
Francesca R. Jensenius§
Melanie Sayuri Sonntag¶
Justine Tinkler∥

April 25, 2018

Abstract
What are the effects of sexual misconduct training on college campuses? As the #metoo movement grows, public and private organizations around the world are exploring training and other policy mechanisms to combat sexual assault and harassment. This study examines the case of a diverse university in the Southwest that introduced mandatory, in-person training on sexual misconduct, including harassment and assault, for all students. Drawing on two sets of studies, one with a quasi-experimental and one with an experimental design, we explore the effects of the training on attitudes towards rape myths, gender stereotypes, and expressed willingness to report of episodes of violence. We find that students (particularly men) who had undergone training were less likely to endorse common rape myths and were more likely to say that they thought campus authorities would believe them if they reported an incident. However, the training also made students more likely to express traditional gender stereotypes and less likely to say they would report incidents of sexual assault. The change in stated likelihood to report was driven primarily by changes among women respondents. Concluding with reflections on which aspects of the training should be changed to avoid such counterproductive effects, this paper has broad implications for the design of institutional responses to sexual misconduct.

*The Office of the Vice President for Research at UNM provided seed funding for this research project.
†PhD student, University of New Mexico.
‡Professor of Political Science, University of New Mexico.
§Senior Research Fellow, Norwegian Institute of International Affairs.
¶PhD student, University of New Mexico.
∥Associate Professor of Sociology, University of Georgia.
1 Introduction

In response to federal mandates and public pressure, universities around the country require that incoming and current students participate in specialized training to change climates and reduce sexual and gender violence and harassment on campus. The Campus SaVE Act, part of The Violence Against Women Reauthorization Act of 2013, ordered universities and other institutions of higher education to offer primary prevention and awareness training about various forms of misconduct. Many public officials, including judges, look to the existence of training programs, as well as the broader framework of policies intended to reduce sexual misconduct, as evidence of compliance with these federal mandates including Title IX of the Education Amendments of 1972 and Title VII of the Civil Rights Act of 196.

Mandatory training and other mechanisms to prevent misconduct and discrimination were adopted by universities and corporations to comply with law, not in response to evidence of their efficacy (Dobbin, 2009; Edelman, 2016). In fact, there is mixed evidence of whether sexual misconduct training achieves its intended effects of changing climates and preventing violence. Some studies conducted in university contexts show that training produces positive effects on the cultural beliefs conducive to violence known as “rape myths,” as well as on bystander awareness and intentions to engage in bystander interventions (Banyard, Moynihan and Crossman, 2009; Gidycz, Orchowski and Berkowitz, 2011; Storer, Casey and Herrenkohl, 2016; Vladutiu, Martin and Macy, 2011). However, other studies of misconduct training have found that it may activate traditional gender stereotypes, incite backlash against women, and trigger resistance (Bingham and Scherer, 2001; Tinkler, 2012, 2013). Studies of different types of training, such as training to promote diversity or sensitivity in the workplace, has also found that it may be counterproductive to the goal of promoting more women and minorities in leadership (Dobbin, Schrage and Kalev, 2015).

It is imperative to develop strategies that work, since studies show that sexual assault
and harassment are widespread in university settings. Campus surveys show that between 11 and 25 percent of students have experienced non-consensual sexual contact, and some 10 percent have experienced intimate partner violence (Wood et al., 2017). Results vary depending on the scope, questions, and methods of the surveys, and may underestimate rates of victimization for vulnerable populations such as LGBTQ and Native American students (de Heer and Jones, 2017). At one university with a well-designed survey and high response rate, 28% of women students and 38% of gender non-conforming students, compared to 13% of men, reported experiencing sexual assault since enrollment (Mellins et al., 2017). When it comes to sexual harassment, graduate students are particularly vulnerable: in a large multi-campus survey, some 10% of women graduate students reported suffering harassment by a faculty member (Cantalupo and Kidder, 2018).

In this paper, we analyze the effects of mandatory, in-person sexual misconduct training on student attitudes at a large public university. We are concerned to explore the impact of the training on self-reported attitudes toward sexual and gender violence, including rape myths, gender stereotypes, and willingness to report experiences of misconduct. We draw on two sets of studies, one with a quasi-experimental and one with an experimental design, conducted in 2017.\(^1\) We find that students (particularly men) who had undergone training adhered less to some rape myths and were more likely to say that they thought campus authorities would believe them if they reported an incident of sexual assault. However, the training also made students more likely to express traditional gender stereotypes and less likely to say they would report incidents of assault. The change in stated likelihood to report was driven primarily by changes among women respondents.

Our results conform to the findings of other studies showing that training on sexual harassment and assault can produce mixed effects. We interpret these findings as the result of defensive reactions to gender status beliefs and norms of interactions, as well as a reaction

\(^1\)The study was approved by the UNM Office of the Institutional Review Board.
to the legalism and mandatory nature of the training. In future studies, we plan to test whether modifying certain aspects of the training, including its framing as well as its delivery, can change its effects on gender beliefs and on willingness to report assault.

2 Background

In the years following the Civil Rights Act of 1964, organizations across the United States adopted policies and programs to prevent discrimination and promote equal opportunity. Over time, in response to feminist arguments that sexual harassment constituted a form of employment discrimination, understandings of equal opportunity came to include protection from harassment (Dobbin, Kalev and Roberson, 2013; MacKinnon, 1979). The evolution of jurisprudence and federal legislation related to Title IX of the Educational Amendments Act similarly came to require that institutions of higher education take action to prevent and punish sexual misconduct, as well as to protect victims, to insure equal opportunities.

Is the giant equal opportunity infrastructure – which includes harassment training, civil rights and sexual misconduct grievance procedures, performance evaluations, diversity training – symbolic or effective? Much legislation adopted as part of the “rights revolution” to promote equality can be seen as “window dressing”: lawmakers and officials adopt it to look good on the outside but have little interest in, or resources dedicated to, enforcement (Brinks and Botero, 2014; Htun and Jensenius, 2018; Levitsky and Murillo, 2009). Another possibility is that organizational mechanisms introduced as part of the rights revolution are actually enforced, but may not be effective in solving problems of exclusion, misconduct, and discrimination. Especially when such mechanisms are adopted quickly to comply with federal mandates, they may suffer from poor design, a lack of pre-testing, and fail to incorporate evidence-based methods. Edelman (2016) shows that courts are likely to absolve employers of responsibility for harassment or hostile working environments if they merely
have programs in place to prevent harassment and receive complaints, regardless of their efficacy.

Many studies show that programs intended to prevent and reduce discrimination and other mistreatment of disadvantaged groups are not only ineffective but can be counterproductive. Why? Many equal opportunity programs, such as those involving job tests, performance evaluations, grievance procedures, and diversity training, are seen as externally imposed, involve limiting individual discretion, and correcting people’s beliefs. But self-determination theory shows that people resist external controls and imposition. Managers and employees tend to rebel against initiatives that micromanage them, which may end up activating, rather than reducing, social biases (Dobbin and Kalev, 2016; Dobbin, Schrage and Kalev, 2015). In particular, when the locus of motivation for diversity training programs is perceived to be imposed from the outside, such as from a legal mandate, it is less effective than when the locus of motivation is internal to the organizational culture and people participate in training on a voluntary basis (Kalev and Dobbin, 2015). Furthermore, Kaiser et al. (2013) find that the mere existence of diversity-promotion programs can offer an illusion of fairness which makes dominant groups less able to recognize discrimination and undermines the legitimacy of claims of bias.

Tinkler’s work on sexual harassment training identifies a different reason for a lack of efficacy. She shows that some men and women perceive sexual harassment policies as threats to gendered interaction norms (Tinkler, 2008, 2012, 2013). People’s investments in existing gender norms affects their responses to policies designed to change relations between men and women. In other work, Tinkler and others shows that men and women perceive sexual harassment policies as a threat to their respective status positions (Tinkler, Li and Mollborn, 2007; Tinkler, Gremillion and Arthurs, 2015). In particular, men perceive potential losses to material and interactional power, and women perceive the policies as disempowering in its paternalism. For both of these reasons, individuals receiving training may react against
its messages.

These perspectives suggest several reasons why mandatory, in-person sexual misconduct training may produce mixed or counterproductive effects. First, people may perceive the training as an external imposition and rebel against it as a threat to their autonomy and self-determination. Rebellion against mandates from the outside may be particularly acute when it comes to misconduct training. Since the 2013 Campus Sexual Violence Elimination Act (Campus SaVE Act), all private and public universities that receive federal funding must train students on sexual misconduct policies and reporting procedures. To the extent that the training builds on a legalistic frame and identifies the locus of motivation as a federal mandate, it may incite a negative reaction.

Second, the training by definition activates gender beliefs. It aims to combat sexual and gender violence, as well as harassment, on campuses. As such, the training runs up against the underlying reasons for sexual misconduct, which include power asymmetries between men and women, deeply held status beliefs about their roles and characteristics, and long-standing cultural beliefs that legitimize male sexual entitlement and women’s submission and shame, known in the literature as “rape myths.” These are all highly gendered issues. Though men are also victims of harassment and assault, the majority of the perpetrators tend to be men and the victims tend to be women. The training may trigger defensive resistance simply by bringing up the issue, regardless of how open or egalitarian or inclusive the training’s content is.

Third, the training attempts to produce a change in sexual practices. In response to federal mandates, universities have adopted “affirmative consent” policies, which require both parties to seek and receive affirmative, conscious, and verbal consent at each stage of a sexual interaction. This clear communication about sexual desires involves a shift in norms of heterosexual interaction, which historically prescribed men to be assertive and women to be deferential (Connell and Connell, 2005; Hamilton and Armstrong, 2009; Lorber, 1994).
Not surprisingly, affirmative consent policies are controversial. Critics in the media have blasted them for "ruining good sex," turning sexual encounters into legal contracts between parties, and giving alleged victims more rights than the accused.\(^2\)

Finally, the training introduces new categories of legitimate and illegitimate behavior and shows results of studies on their prevalence. In doing so, it categorizes a variety of behaviors as sexual misconduct, which includes a broad spectrum of non-consensual activities ranging from verbal harassment to assault and rape.

Our study explores the effects of the training on student attitudes, but does not test or identify the underlying mechanisms that explain these effects. In future work, we hope to vary the content of the training, which will allow us to pinpoint mechanisms.

3 Title IX Reforms at the University of New Mexico

In October of 2016, the University of New Mexico (UNM) entered into a three-year agreement with the U.S. Department of Justice (DOJ) to improve its compliance with Title IX of the Educational Amendments Act. The agreement resulted from a DOJ investigation into allegations of Title IX violations at UNM, which culminated in a “findings” letter of April 2016. According to the findings, UNM was “out of compliance in several respects.” It did not have adequate policies on misconduct, procedures to report and investigate, leadership commitment, and a hostile climate. In the 30-page agreement, UNM committed to five principal pillars of action, including revision of policies, revision of reporting and investigative procedures, annual climate surveys, annual sexual misconduct training of employees, and in-person sexual misconduct training of 27,000 students within the first year of the agreement. The in-person training for students, called the Grey Area training, was developed and implemented by staff in the university’s student advocacy and response center, called Lobo.

\(^2\)See contributions by Dalmia 2014; Friedersdorf 2014; Grasgreen 2013, for example.
Respect.

The training involves an approximately one hour lecture, followed by a small group discussion which lasts about 30 minutes. The topics covered in the lecture include definitions and examples of sexual misconduct, university policies, reporting procedures, data on the prevalence of assault and harassment, Title IX and federal law, bystander interventions, and the concept of affirmative consent and an illustrative video. Small group discussions engaged participants in a conversation about the features of “good hookups,” “bad hookups,” “alcohol,” and “rape,” with the goal of emphasizing that consent is part of the former two experiences, but not part of the latter two.

According to the terms of the agreement with the DOJ, the in-person training was required for all enrolled students. Incoming students, whether first-year or transfer students, participated in the training as part of their new student orientation programs. Continuing students were sent multiple notices and reminders informing them that they needed to sign up for the training, under penalty of a registration hold. Students were instructed to go to the Dean of Students website, where they selected and registered for a training session on a particular date. To avoid triggering reactions in students who may have experienced violence in the past, the website included “escape” buttons. Students were able to opt out of the training, and advocates were on hand during the sessions in order to consult with students in the event of adverse reactions. Between March and December of 2017, Lobo Respect trained approximately 22,000 enrolled students.

Though other universities similarly require training of students, the vast majority of these programs involve training of incoming students only, and many programs include only an on-line component. To our knowledge, UNM’s experience is unique in involving mandatory, universal in-person training. It was a massive task for trainers and advocates to organize and deliver these sessions within a relatively short period of time.
4 Study Overview

The DOJ agreement provided an opportunity to study student reactions to mandatory sexual misconduct training. We employed experimental methods to conduct two studies on the effects of the Grey Area training on undergraduate students’ understandings of sexual assault and harassment, adherence to rape myths, beliefs about gender, willingness to report assault, and beliefs about the consequences of reporting. Although the research project is still ongoing, we completed two studies during the summer and fall of 2017 – a fully randomized field experiment and a quasi-experiment. Since our goal was to measure effects of the Grey Area training, we were interested in comparing how students responded to our survey before and after the training. One option would have been to compare the responses of the same students before and after going through the training (often referred to as a ‘pre-experimental’ research design or ‘pre- and post-test’ design). However, this approach would be vulnerable to learning or test effects: that changes in responses result from the testing itself due to questions raising awareness and triggering learning after the pre-test (see, e.g., Marsden and Torgerson, 2012). Another option was to compare responses from students who had undergone the training to another group of students who had not yet undergone the training, ideally with similar characteristics. This was the strategy we employed by assigning one group to complete the survey measuring our dependent variables immediately before the training began (our control group) and the other group to complete the survey immediately after the training (our treatment group).

4.1 Summer study procedures and sample demographics

The summer study was conducted during two successive sessions of New Student Orientation (NSO). The NSO took place once a week over two days and one night. The Grey Area training was administered during the evening. Students self-selected into which session they
signed up for, and sessions closed as they filled up. Our assumption was that there might be substantial self-selection bias between the students who ended up signing up for an early and a late session (e.g., if more eager students signed up as soon as they could), but that this bias was likely to be small if we selected successive sessions – identical in all respects other than being one week apart.

In this quasi-experimental set-up our “Control group” responded to our survey instrument right before they sat through the Grey Area Training, and the “Treatment group” was a group of students who took the training exactly one week later, and who responded to the survey right after the main part of the Grey Area Training. The survey was distributed before the small group discussions, since the dynamics of each small group discussion would be different. In each of the sessions included in the summer study, there were approximately 300 students, and since the surveys were conducted in the training room itself almost all of the students present responded to the survey.

The survey instrument included questions relating to understandings of sexual harassment and assault, indicators of adherence to rape myths based on the Illinois Rape Myth Acceptance Scale (Payne, Lonsway and Fitzgerald, 1999), a gender beliefs scale developed in Tinkler (2013), questions about willingness to report sexual assault, as well as beliefs about the consequences of reporting. In addition, the survey instrument included questions that gathered information on the demographic characteristics of student respondents such as age, gender identity, racial identity, and whether or not the respondent was a first-generation college student. We provide more information on the instrument below.

We call this a quasi-experimental design since, though the students were not randomly assigned to the control and treatment conditions, their participation in successive new student orientation sessions can be considered as if random, and we should find the groups to be very similar in all respects. This intuition is confirmed in Table 2, where we show that the demographic composition of the survey respondents was similar across the control and
treatment conditions. For the age question, respondents could choose below 18, 18–24 or older than 24. In these groups none said they were older than 24, and the vast majority reported being 18-24. Both control and treatment groups had slightly more female than male respondents, and few respondents who chose other categories (transgender or gender non-conforming). Close to 46% of the students in both conditions said they were first-generation college students. Where we do see some differences is in self-reported ethnic categories. The largest reported ethnic group was Hispanic/Latino, in line with the University of New Mexico’s status as a Hispanic-serving institution, across both groups. The share was somewhat higher in the treatment group, but the difference was not statistically significant. The share of students self-reporting as white was, on the other hand significantly lower in the treatment group, and there was a somewhat higher number of students self-reporting as Native American or American Indian or who checked several of the ethnicity categories.

Table 1: Demographic composition of the control and treatment groups in the summer study

<table>
<thead>
<tr>
<th></th>
<th>Control group (N=339)</th>
<th>Treatment group (N=280)</th>
<th>Difference in means</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent below 18</td>
<td>11.4</td>
<td>12.6</td>
<td>1.2</td>
<td>0.66</td>
</tr>
<tr>
<td>Percent women</td>
<td>52.8</td>
<td>52.3</td>
<td>-0.4</td>
<td>0.92</td>
</tr>
<tr>
<td>First-generation college</td>
<td>45.8</td>
<td>45.5</td>
<td>-0.4</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Self-reported ethnicity

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Control group (N=339)</th>
<th>Treatment group (N=280)</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>39.6</td>
<td>44.8</td>
<td>5.1</td>
<td>0.20</td>
</tr>
<tr>
<td>White</td>
<td>33.4</td>
<td>23.8</td>
<td>-9.6</td>
<td>0.01</td>
</tr>
<tr>
<td>Native American</td>
<td>6.2</td>
<td>9.4</td>
<td>3.2</td>
<td>0.14</td>
</tr>
<tr>
<td>African American</td>
<td>6.5</td>
<td>5.1</td>
<td>-1.4</td>
<td>0.45</td>
</tr>
<tr>
<td>Asian</td>
<td>5.0</td>
<td>5.4</td>
<td>0.5</td>
<td>0.80</td>
</tr>
<tr>
<td>More than one</td>
<td>9.3</td>
<td>11.6</td>
<td>2.3</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Note: The p-values are from two-sample t-tests.

4.2 Fall Study Procedures and Sample Demographics

For the fall study, we were able to employ a fully randomized design. Here, we took the list of students registered for a large training session and randomly assigned them to control and...
treatment conditions in advance. The training took place in two ballrooms of the Student Union Buildings. Each student had to register with their student ID at a registration desk, and was then told to enter either room A or B, where sessions were held by instructors simultaneously. Project staff were present to ensure that the students entered the room they had been assigned to. Students in room B took the survey before the training session (our control group), and students in room A took the survey after the training session (our treatment group). Both sessions started and ended at approximately the same time.

Randomizing the students into a treatment and control ensure good balance on both observable and unobservable characteristics. This means that the control group students should be comparable to the treatment group students before the training. Table 2 compares the demographic composition of the student respondents in the control and treatment groups in the fall study. Here we see both an overall older age – which is not surprising given that these are continuing students, and as a public university UNM has many non-traditional students. We also see an imbalance between the groups on the age variable, as the average age was over 3.3 years higher in the control group. However, this difference is driven by a few older respondents (one was 78 and some were in their 50s). If we compare only the respondents that were younger than 30, the two groups are balanced. Here we also see a somewhat higher share of women in the control group, but the difference is not statistically significant. In this study there were no statistically significant differences in the racial identity composition of the two groups.

Seeing imbalances in a randomized study is not uncommon, particularly with a smaller sample size. What we see here is the trade-off between choosing a fully randomized setup and one with a higher sample size. The literature is not in agreement about whether correcting for imbalances in experimental work (through matching or by introducing control variables in a regression framework) is actually advantageous. It may help to improve balance on observables and to reduce precision in estimates, but it may also worsen balance on
Table 2: Demographic composition of the control and treatment groups in the fall study

<table>
<thead>
<tr>
<th></th>
<th>Control group (N=86)</th>
<th>Treatment group (N=80)</th>
<th>Difference in means</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (all)</td>
<td>27.2</td>
<td>24.0</td>
<td>-3.3</td>
<td>0.02</td>
</tr>
<tr>
<td>Age (below 30)</td>
<td>23.0</td>
<td>22.5</td>
<td>-0.5</td>
<td>0.24</td>
</tr>
<tr>
<td>Percent women</td>
<td>51.9</td>
<td>42.7</td>
<td>-9.2</td>
<td>0.25</td>
</tr>
<tr>
<td>First-generation college</td>
<td>41.7</td>
<td>42.3</td>
<td>0.6</td>
<td>0.93</td>
</tr>
<tr>
<td>Self-reported ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>35.8</td>
<td>35.1</td>
<td>-0.7</td>
<td>0.92</td>
</tr>
<tr>
<td>White</td>
<td>42.0</td>
<td>37.7</td>
<td>-4.3</td>
<td>0.58</td>
</tr>
<tr>
<td>Native American</td>
<td>6.2</td>
<td>10.4</td>
<td>4.2</td>
<td>0.34</td>
</tr>
<tr>
<td>African American</td>
<td>3.7</td>
<td>2.6</td>
<td>-1.1</td>
<td>0.69</td>
</tr>
<tr>
<td>Asian</td>
<td>8.6</td>
<td>3.9</td>
<td>-4.7</td>
<td>0.22</td>
</tr>
<tr>
<td>More than one</td>
<td>3.7</td>
<td>10.4</td>
<td>6.7</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Note: The p-values are from two-sample t-tests.

unobservables and consequently bias the results. In the next section we report differences in responses across the control and treatment conditions without any further control variables. However, to err on the side of caution we also check the robustness of all the patterns by running models with demographic controls.

### 4.3 Outcome Measures

We utilized the Illinois Rape Myth Acceptance Scale (IRMA) developed by Payne, Lonsway and Fitzgerald (1999) to measure rape myth acceptance. The survey instrument included the updated version of the scale developed by McMahon and Farmer (2011), which presents four subscales relating to a respondents’ acceptance of four different rape myths labeled: She asked for it; he didn’t mean to; it wasn’t really rape; and she lied. Each subscale includes five to six statements to which respondents can strongly agree or strongly disagree. For our study, we used one complete subscale and summary indicators of the remaining three subscales (based on work by Theresa Cruz in New Mexico high schools). In the summer study, we used the “it wasn’t really rape” subscale and in the fall study, the “he didn’t mean to” subscale.
We adopted the semantic differential scales used in Tinkler’s research on the effect of sexual harassment policies (Tinkler, 2013; Tinkler, Gremillion and Arthurs, 2015). These items are seven-point scales for various pairs of words measuring participants’ personal opinions about men and women on three dimensions: (1) power: respected/not respected, powerful/powerless, high status/low status, leader/follower; (2) competence: competent/incompetent, knowledgeable/unknowledgeable, capable/incapable; (3) considerateness: considerate/inconsiderate, pleasant/unkleasant, likable/unlikable, cooperative/uncooperative. Students were first asked to mark where they personally thought men fell on each of the polar adjective scales and then they were asked to mark where they thought women fell of each scale.

Finally, questions in the survey measured students respondent’s willingness to report sexual misconduct to campus authorities. These questions were adapted from previous campus climate surveys. These questions included: “How likely would you be to report the assault to a campus authority?”, “How likely would you be to encourage a friend to report the assault to a campus authority?”, and “How likely do you think that campus authorities will believe you?” On a four-point scale, the answers ranged from very likely (1) to not at all likely (4).

Though we do not discuss these results in this paper, the survey instrument also presented respondents with a series of statements asking whether they would consider certain behaviors to be sexual harassment or sexual assault. Among others, some statements measuring students perceptions on sexual harassment included “Your teaching assistant implies that she/he will adjust your midterm grade if you go out with her/him for a drink” and “A faculty member tells you that you are very attractive and asks you out.” With regards to measuring students’ perceptions on sexual assault, some of the statements included “A woman grabs the private parts of a man in a bar, while saying it is a joke” and “A man has sex with a woman who is extremely drunk and unable to speak clearly.”
5 Findings

In Figure 1 we show responses to some of the questions about rape myths, responses to the summer study on the left and the fall study on the right. We are showing the questions that were the same across each study, and for each question the top bar shows the responses in the group surveyed before the training (control) and the bottom bar shows the responses in the group surveyed after the training (treatment). Responses were recorded on a six-level Likert scale running from “Totally agree” to “Totally disagree,” where agreeing meant adhering to the rape myth.

A desired result of the training would be to see a shift in the responses over to the right. We see that for several of the questions. Across both studies, a higher share of the respondents “totally disagreed” with the statement “If someone is raped while they are drunk or on drugs, they are at least somewhat responsible for letting things get out of control.” In the summer study we also see a large shift in the responses to the statement “Rape accusations or charges are often used as a way of getting revenge.”

Collapsing the scale into a dichotomous measure of respondents “agreeing” or “disagreeing” we find that in the July study, the changes in the responses to “If raped while drunk,” “changed their mind,” and “If a girl doesn’t say no” are all highly statistically significant. There patterns are stronger among male respondents, where there is a significant shift in all the four rape myths. These patterns are robust to the inclusion of controls.

In the November study, we see some shifts in the direction of disagreeing more too, but these differences are not statistically significant at conventional levels (with or without controls). Among the male respondents there is a significant shift in “If a girl doesn’t say no” in models that include controls (age, first generation college student, and ethnic group).

In Figure 2 we turn to our measures of gender stereotypes. Here, students were asked to place men and women in a space between a more and less desirable trait. In the survey
the traits were ordered differently to avoid respondents becoming too aware of the patterns in their responses. Here we have ordered the pairs so that the less desirable trait is on the left and the more desirable one is on the right. In the figure the average responses for the group interviewed before the training (control) are indicated by light-colored circles, while average responses after the training (treatment) are indicated by dark-colored circles. Where we can see both circles fully, this suggests we have a statistically significant difference in the responses.

For the responses about men (left plot) we can see two main treatment effects: that after training the respondents were more likely to say that men were “Not respected” and also more likely to label men a “Follower.” These changes are stronger among female respondents. We see stronger patterns in the evaluation of women. Here we see that the training makes respondents more likely to label women “Incompetent,” “Not respected,” “Powerless” and
“Incapable.” Here, the change in “Not respected” and “Incapable” is stronger among male respondents, while the change in “Incompetent” and “Powerless” is stronger among the female respondents. This is certainly not an intended effect of the training and suggests that here – as has been found in other context too – simply talking about issues of sexual harassment and assault may serve to polarize gender attitudes and trigger traditional gender stereotypes.

In the fall study the treatment effects were much weaker. However, if we reduce that sample to the younger respondents – those below 25 of age – we see patterns that are similar to the ones in the summer study. This suggests that it is the younger respondents that are more easily affected, perhaps both by the intended messages of the training and the unintended consequences, such as an activation of gender stereotypes.

Figure 2: Responses to questions about gender stereotypes in the July survey

In Figure 3 we turn to the question of reporting. Here we would want to see a shift to the right in the top three questions (“Would you report” and “Would you encourage a friend to
Figure 3: Responses to questions about reporting in the July and November surveys

Here we results are mixed. Across both the July and November, we see a shift to the left on the intention to report variable: After the training a lower share of students said they were likely to report on in incident that happened to them. If we recode the variable to a dichotomous variable for whether respondents were on the “Not likely” or “Likely” side, we see that this change is statistically significant, that it is robust to the inclusion of control variables, and that the patterns primarily is driven by a change among women. In the July study, the training is associated with a 10 percentage point drop in the share of women who said they would report of assault to campus authorities. In the November study the training is associated with an increase in the share of men who said they were likely to report, but an even larger drop among the women (15 percentage points), but (perhaps due to the smaller
sample size) the pattern is not statistically significant at conventional levels.

On the other hand, almost all the students they they would encourage a friend to report and after the training a higher share reported it as likely that campus authorities would believe them if they reported. Across both studies, a larger share of students also report it as unlikely that they would be blamed for what happened.

6 Conclusions

Based on existing literature, we expected that the training at UNM would reduce adherence to rape myths. Finding that this did happen was therefore not surprising. The two somewhat more surprising findings in our analysis concern the reinforcement of traditional gender stereotypes and the reduction in the willingness to report assault.

Though Tinkler’s previous studies had a similar finding about gender stereotypes, we were still surprised to see this result here. The Grey Area training curriculum, and the Lobo Respect trainers, have taken great pains to make the content of the training as gender neutral, inclusive, and egalitarian as possible. This suggests that other parts of the training, or the mere fact of sitting through a training on this type of topic, were sufficient to activate traditional gender stereotypes.

In the qualitative component of this study (still ongoing), one student spoke about how she found the training to be “isolating.” She explained it was isolating, because “it made everyone feel, like the genders, feel polarized. So, when I walked in there, I was having a comfortable conversation with my neighbor who was a man and by the end of it was like we were trying to distance our seats as much as we could from each other.” She explained that the training overall “didn’t necessarily make people feel safer like they knew what they shouldn’t do. It kind of just made them feel more different from the opposite gender [like] now there are more ways that they have to look to not treat them or make them
uncomfortable.” Instead of talking during the small group discussions, she recalled, “no one wanted to participate […] it makes me feel like people were pulled into themselves […] that it was something that made them more uncomfortable and less open.”

What about the decline in intention to report? It could be the case that the control condition artificially increased the likelihood to report assault. To the extent that participants knew they were entering a sexual misconduct training, they were already primed toward a socially desirable answer, which is that they would report assault under any circumstance. Sitting through the training, however, may have caused them to see the complicated nature of many encounters that could be perceived as assault. Sexual encounters involving drugs or alcohol are common on college campuses, for example. The training emphasizes the importance of consent, and states that people under the influence of drugs or alcohol are unable to give consent. The training thus takes what people may have formerly perceived to be normal and acceptable behavior and moves it to the realm of assault.

Our results point to broader challenges with training. It is easy to preach to the converted, but far more difficult to change the minds of people who disagree merely by compelling them to sit through a lecture. In addition, sexual misconduct training has multiple objectives, which may not always be in sync. The training aims not just to provide information about policies and resources but also to trigger changes in social norms. Notions of affirmative consent lay out a new dynamic of sexual interaction that challenges past behavior. Our study suggests that, since people are defensive about long-standing ways of interacting across gender lines, orienting them toward new models of behavior involves a longer-term and multifaceted strategy.

---

3Interview, March 9, 2018.
References


Dobbin, Frank, Daniel Schrage and Alexandra Kalev. 2015. “Rage against the iron cage:


Tinkler, Justine Eatenson, Yan E Li and Stefanie Mollborn. 2007. “Can legal interven-

