

New interventions for sexual harassment prevention

Olle Folke, Torbjørn Hanson, Åshild Johnsen, Andreas Kotsadam, Johanna Rickne

February 24, 2023

Summary: We study an intervention against sexual harassment that provides information on attitudes and job performance. We randomize the intervention across small rooms with military recruits in the boot camp of the Norwegian military. Outcome variables include sexual harassment prevalence and attitudes toward voicing disagreement with harassment when it occurs. An extended analysis studies attrition from the training to understand if our interventions can prevent occupational exit among women.

Contribution: Sexual harassment is a severe and prevalent workplace hazard. It is discrimination by law and a form of violence against women. Previous evaluations of the most common prevention policies have shown a lack of efficiency or even counterproductive results. We suggest a new prevention method and test its efficiency with randomized control trials. This produces new actionable knowledge for sexual harassment prevention.

Setting

The experimental setting is the boot camp training for the Navy and Air force in the Norwegian armed forces. Recruits from all over the country participate in an 8-week training program at a military base. Each person is assigned to a *troop* (a military unit consisting of around 90 individuals) depending on when they arrive to the camp. Within the troop, we randomly assign recruits to rooms. A room typically consists of six or ten people. In the rooms people do a lot of tasks together, such as room cleaning, and previous data shows that people spend most of the time together with people from their room. Survey data from March 2020 shows that 95% of recruits reported spending “some” or “most” of their spare time with their roommates.

The military exemplifies the type of male-dominated occupation and workplace environment where women experience the highest rates of sexual harassment (e.g., USMSPB 1994; Fitzgerald et al. 1997; Folke and Rickne 2022). In Norway, women have had the right to serve and receive military training in since 1985, female conscription was introduced and enforced from 2016 for all women born after 1997, and women currently make up about 15% of the military workforce. Importantly, men and women share rooms in both the Navy and the Air force.

Our data collection in 2022 among 780 recruits shows that women self-report more sexual harassment than men among boot camp participants (see Tables 1 and 2), and we therefore design our

interventions to target harassment against women while leaving harassment against men for future research. This data collection was done with a behavioral survey instrument consisting of 15 items, based on the SEQ-DoD and SEQ-DoDs and listed in Table 1 (Fitzgerald et al. 1999; Stark et al. 2002), but with minor adaptations to a Norwegian context (see Rones et al. 2018 and Fasting and Kjøber 2018 for details).

Defining sexual harassment

We employ a broad definition of sexual (or “sex-based”) harassment as behaviors that humiliate or demean people based on their sex or gender (Berdahl 2007). We study two out of the three categories of conduct typically included under this broader umbrella. *Unwanted sexual attention* is unwelcome verbal or physical sexual advances. *Gender harassment* is verbal and nonverbal behaviors that convey hostility, objectification, or second-class- status based on gender. We do not study *sexual coercion*, which often comes from superiors rather than same-level colleagues and constitutes the third and least common category of sexual harassment.

Theory of change

Workplace sexual harassment is an outcome of contextual factors. People hold latent proclivities to harass which manifest as perpetration under certain environmental conditions. Male-dominated social environments like the military have several features that makes sexual harassment more prevalent against the female minority. We draw on theories about these features to suggest new interventions for harassment prevention and to discuss which mechanisms a treatment effect might be operating through.

Information nudges. Women in male-dominated jobs such as the military are often viewed as less competent than their male counterparts (Eagly and Mladinic 1994). One reason is that societal views about women’s skills do not conform with views about what it takes to be a good soldier. Sex-role spillover theory describes how an occupation dominated by one gender takes on traits associated with that gender as desirable for performing the job (Gutek and Morasch 1982; Gutek and Cohen 1987). In this way, male-dominated jobs take on traits like aggression, dominance, and status competition. People with these traits are considered—rightfully or not—to be more competent in performing the work. “Having what it takes” for the job becomes conflated with being a “real” man (Ely and Meyerson 2010) and women, who are perceived as nurturing, communal, and physically weak, become viewed as less capable.

The view that women lack competence in the military profession likely helps *underpin gender harassment* against women soldiers. Attitudes that some people (women in this case) are not suitable for a certain job or task (here the military) are common within this category of sex-based harassment. Data from the boot camp context in 2022 shows that 28% of women and 6% of men had experienced this type of behavior at least once during their training (see Table 1).

There is little evidence that women soldiers are actually worse at their job. Extensive assessments of military recruits within the Norwegian Armed Forces show no gender differences across various characteristics including the ability to lead, responsibility, teamwork, military skills, overall judgment, and oversight (Køber et al. 2019). This opens for an information nudge consisting of information that women are not, in fact, worse than men when it comes to military competence.

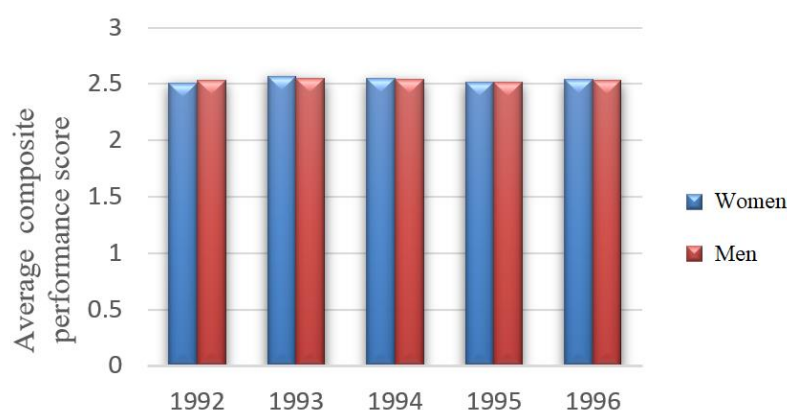


Figure 1. Gender gaps in military performance by cohort.

Source: Køber and Hanson (2019); see also Køber (2016; 2017).

Theory also links *unwanted sexual attention* to behavioral processes in male-dominated work environments. One mechanism is women’s increased visibility in the everyday work environment (building on Kanter 1977), which creates sexualization as the female minority becomes viewed as a member of their gender first, and a professional colleague second. The incorporation of masculine traits into the organizational culture is another factor. These traits include men’s superiority to (and dominance over) women, and encourage “masculinity contests” that link men’s status in the group to proving their masculinity—for example by aggressively propositioning or badgering female colleagues in a sexual manner (discussed by, e.g., Berdahl et al. 2018; Glick et al. 2018).

Most men might not approve of a work culture influenced by aggression and masculinity contests that tolerates or even encourages unwanted sexual advances against women. But individual men might perceive themselves as an exception in this regard and “go along to get along.” Recent economics research has shown that correcting these types of erroneous beliefs about the progressive nature of

other men's attitudes may be a forceful tool for cultural and behavioral change (Bursztyn et al. 2020). In the spirit of this work, we devise an information nudge about other recruits' views on unwanted sexual attention.

In the boot camp survey data from 2022, we found that "unwanted jokes of a sexual nature" were the most common form of unwanted sexual attention experienced by female recruits (21% had experienced "that someone has told sexual jokes or stories that upset you/ you found offensive/ made you feel uncomfortable", see Table 1). A separate question asked recruits whether this behavior "can be called sexual harassment." More recruits agreed than disagreed with this statement, and only 5% disagreed strongly. Based on this data, we formulate the second information nudge to reveal other recruits' negative attitudes toward unwanted jokes of a sexual nature.

We deliver both of our two information treatments at the same time and expect them to make attitudes less tolerant for sex-based harassment. We formulate this first key treatment effect as a first testable hypothesis:

- H1: The information treatments reduce tolerant attitudes toward sex-based harassment.

We expect this attitudinal shift to affect harassment prevalence via the processes described above and predict that:

- H2. The information treatments reduce sex-based harassment.

Victim and bystander behaviors. The information nudges might affect victims' and bystanders' attitude to sexual harassment. For victims, a change in the attitudes to harassment in the person's social circle might affect how they cope with harassment experiences. A specific type of coping behavior of interest is so-called *advocacy seeking*, which consist of voicing grievances, either by confronting the harasser directly or reporting their behavior as a formal complaint (Cortina and Magley 2003, Cortina and Wasti 2006). For bystanders, changes in the attitudes about harassment in their social circle might affect their attitudes to intervening when they observe a harassment event, or their actual intervention behaviors.

A specific mechanism whereby our intervention might affect victims' and bystanders' behaviors is via the development of *legal consciousness* about sexual harassment (e.g., Felstiner et al. 1980–1981; Blackstone et al. 2009). The concept of *legal consciousness* describes a mental process whereby a person recognizes the wrongfulness of a certain type of behavior. Legal consciousness of sexual harassment has three components: recognizing a situation as wrongful, recognizing the victim as a victim, and recognizing the perpetrator as a perpetrator. If our information nudges shift attitudes about one of these components—the harassment situation—they might, as a knock-on effect, also affect

attitudes about the other two—the victim and the perpetrator. Via this mechanism, victims and perpetrators might become more likely to actively voice their grievance by advocacy seeking or bystander interventions, rather than staying silent (for a theory on the silence around sexual harassment, see Herschovis et al. 2019).

- H3. The information treatments increase the probability that harassment victims cope by advocacy seeking or social coping after experiencing sex-based harassment.
- H4. The information treatments increase positive attitudes toward bystander intervention in sex-based harassment events.
- H5. The information treatments increase bystander intervention behaviors in sex-based harassment events.

Empirical challenges

There are four major empirical challenges in this experiment. One consists of allocating and surveying recruits who leave boot camp before completion and thereby remove themselves from the endline survey. Attrition among women who leave due to a bad work environment is particularly problematic since it risks biasing any potential treatment effects toward zero. Attrition might also bias results in the opposite direction, however. Women in treated rooms might become less likely to leave if the information treatments reduce harassment (H1 and H2), empower women (H3), makes bystanders more helpful in harassment events (H4 and H5), or changes some other cultural aspect that drives harassment in the room. We will check whether attrition and missing outcomes are correlated with treatment. If there are statistically significant differences in attrition or non-response between treatment and control (controlling for the strata variables), we will follow the correction proposed by Lee (2009). We also provide a shortened version of the endline survey to individuals that quit the training and we will analyze answers from that survey when available.

A second difficulty is that our treatments might change people's perceptions of harassment in ways that create bias in the prevalence measurements (and possibly in other measurements, as well). The treatments might affect the likelihood to recall a harassing event because the person now finds that event more problematic, which makes it stick in their memory. This same process might also make it more likely that a person finds a certain situation more fitting to the description in the sexual harassment survey questions of being "unwanted". These mental processes might make treated respondents more likely to check the boxes in the Sexual Experiences Questionnaire and raise our incidence rate, counteracting a potential negative treatment effect. This empirical challenge is not necessarily restricted to the harassment prevalence measurement, but might also affect our measurements of victims' and bystanders' behaviors and attitudes. One indication that the treatments

might not have a large effect on recall is that they did not correlate with military recruits' self-reported prevalence of sexual harassment in the year *before* entering training, as reported in their enrollment survey.

A third challenge is the fact that most sexual harassment reported by boot camp recruits is not coming from members in their own room. In our 2022 data most harassment—86% of women's experiences—come from other recruits in the camp who do not reside in the same room. In fact, 61 percent comes from someone in the same troop but not in the same room. One difficulty stemming from this fact is that treated men might harass women less outside their room, which will result in a lower overall level of harassment but no variation across treated and un-treated rooms.

A fourth challenge is the possibility of information spillovers between rooms. If recruits discuss the information with each other across rooms, any differences in attitudes or behaviors between treated and un-treated rooms will be attenuated.

Specification of treatments

We use data previously cited above, collected at the same military camp, to specify the two information treatments already pre-viewed above. The first treatment targets attitudes toward the most common category of sex-based harassment, namely gender harassment (see Table 2). The second targets unwanted sexual attention, the second-highest category. Within this category we design a treatment that describes attitudes toward the highest-prevalence harassment behavior, namely exposure to unwanted jokes of a sexual nature (Table 1).

Treatment 1. Gender Harassment

The first treatment draws on the statistics provided by Kjøber et al. (2019, replicated in Figure 1) on women's and men's military performance. We formulate our treatment as:

T1. The Norwegian Armed Forces uses the statement of service to assess recruits at the end of their service. The service statement measures many characteristics, such as ability to lead, responsibility, teamwork, military skills, overall judgment, and oversight.

Which statement do you think is correct with respect to the statement of service and gender?

1. *Men generally receive a better grade than women*
2. *Men and women generally receive an equal grade*
3. *Men generally receive a lower grade than women in general*

After they have answered the question, the treated individuals get to see Figure 1 and a text box that says that *"Answer: When comparing the service statements of women and men over years, there are generally no gender differences: Men and women generally receive an equal grade."* The control group only answers the question but do not get any information.

Treatment 2. Unwanted Sexual Attention

The second treatment uses information collected in 2020 on attitudes about "telling sexualized jokes can be called harassment". On this question, 24% agreed or strongly agreed with the statement and 15% disagreed or strongly disagreed. Only 5% strongly disagreed. Based on this data we formulate a treatment consisting of two parts, one providing information about other recruits' attitudes and a second part aiming to imprint this information by asking about whether the respondent "was surprised" about it:

T2a. In 2022, we asked recruits in the Madla camp at the end of recruit school whether they agreed or disagreed with the statement "Telling sexualized jokes can be called harassment". What do you think that they answered?

- a) More recruits agreed than disagreed that telling sexual jokes can be called harassment*
- b) More recruits disagreed than agreed that telling sexual jokes can be called harassment*

After answering this question, an information box appears with the following text: *"Answer: More recruits agreed than disagreed that telling sexual jokes can be seen as harassment. Only 5 percent fully disagreed with this statement."*

T2b. Are you surprised that only a small minority fully disagreed with this statement?

- a) No*
- b) Yes*

The control group only answers the question in T2a but do not get the answer and are not asked T2b.

Data collection and randomization

Before starting their military training, recruits go through checks and final screenings at a military camp outside of Stavanger. Soldiers do not know each other at this point. Recruits take a survey as part of this enrollment procedure and another survey at completion of boot camp 8 weeks later. Our pre-treatment and post-treatment data collections form part of these two surveys.

We randomize the two information nudges at the room level. The randomization was done by the troop leader, who was asked to divide the rooms into two groups and give everyone in one of the groups a sticky note to put on their key card. The troop leader was told that the two groups would be given different questionnaires and when the groups arrived to take the survey they were seated at different tables. Those with a sticky note used a different QR code to access the survey and got the treatments.

Empirical strategy

The empirical strategy consists of balance tests, treatment validation, hypothesis testing, and exploratory analysis. This analysis will use different specifications and samples. To test effects of the Treatments on fixed traits (balance tests), harassment attitudes (treatment validation) and endline experiences, attitudes, and behaviors (hypothesis tests), we use the main specification:

$$Y_{i2r} = \beta Treated_{i1r} + \chi X_{i1} + \delta Troop_{i1} + \varepsilon_{ir} \quad (1)$$

where i indexes individuals, r is room, and t is time (t_2 then implies endline). β gives the effects of the treatments. X_{i1} is a vector of controls (further discussed below). To make the models fully saturated, we partition the covariate space and add control variables as indicator variables rather than using their multi-valued codings (Athey and Imbens, 2017). If cells are too small, with less than 5 percent of the observations, adjacent cells are combined. When using interaction terms and in tests of balance we will retain the continuous coding of the variables. If we have missing values on explanatory variables we will code the variables as zero and include dummy variables controlling for missing status so that we do not lose observations. Troop fixed effects are included in all analyses since randomization is conducted within troops. We cluster the standard errors at the room level in all estimations.

We will present results from estimations without other control variables than the troops and with the full set of controls. Our main estimation will however be one with optimal controls being chosen from the total list of controls using a post-double LASSO selection approach of Belloni et al (2014). The LASSO selection approach selects those variables that are correlated with both treatment and the outcomes which may improve precision in the estimates and it also helps to correct for imbalances across groups.

Independent variables

The following variables will be used as control variables:

- **Attitude towards living in mixed gender rooms:** Based on the question: "To what degree do you prefer to live in a room where everyone has the same gender as you?" The answer categories are on a five-point scale from Strongly prefer that everyone is of the same gender to Strongly prefer a mixed gender room. We will create up to four dummy variables but ensuring that at least 5 percent of the individuals are in each variable (if not, we combine the categories).
- **High social status:** A dummy for answering high on the question: "In Norway there are groups that are more or less at the bottom and groups that are more or less at the bottom of the societal ladder. Below you see a scale that goes from bottom to top. If you think about the family you grow up in, where would you place it. Scale (1=bottom to 10=top). We will transform this into a binary

variable by splitting it in a way that retains the ordering while it minimizes the difference in number of observations between the two categories.

- **Female:** A dummy variable equal to one if the respondent is female.
- **High GPA:** Self-reported grades from high school at baseline, equal to 1 if grades are above median and zero otherwise.
- **Immigrant background:** A dummy equal to one if the individual or any of their parents are born abroad.
- **Mother and Father employed** (2 variables): Based on the question: “Are your parents working?” Original: 1= Yes, both, 2=My mother is in work, my father is not, 3=My father is in work, my mother is not, 4=No, neither of them is in work. Recode: We recode into two variables: Mother employed (1/2=1, 3/4 = 0) and Father employed (1 and 3=1, 2 and 4=0)
- **Mother and Father with high education** (2 variables): Based on the question: “Do your parents have higher education (university/college)?”. Original: 1= Yes, both have higher education, 2=My mother has higher education, my father has not, 3= My father has higher education, my mother has not, 4=No, neither of them have higher education. We recode the answers into two variables: Mother with high education (1/2=1, 3/4= 0) and Father with high education (1 and 3=1, 2 and 4=0)
- **Plan higher education:** Based on the question: “Do you plan to take higher education?” Original: 1=Yes, 2=Don’t know, 3=No. We recode the answers such that 2 and 3 equal 0.
- **Plan continue serving:** A binary variable equal to one for those who are considering to continue in the armed forces (answer 1 “Yes” or 2 “Maybe) and zero otherwise (answer 3)
- **Suited for service:** A binary variable equal to one for the ones fully agreeing (1) with the statement “I believe I have what it takes to carry out service in the Armed Forces”, otherwise 0 (answer 2 to 4).
- **Motivated effort:** A binary variable equal to one for those who respond that they are strongly motivated (9 or 10) on the question for those who respond that they are strongly motivated on the question (“How motivated are you to make an effort in the Armed Forces?”. Those answering 1 to 8 are coded as zero.
- **Consider working in Armed Forces:** We ask recruits whether they would consider working in the armed forces at a later stage in your life, and code all positive answers (4 to 6) as yes and the other answers (1 to 3) as zero.
- **Fixed effects for the number of women in the room.**
- **Fixed effects for combinations of room size and the number of women.** We include fixed effects for all combinations of the number of individuals and the number of women in the room for which there is both a control room and a treated room. For the combinations that there is not at least

one room of each type, we will group the most similar rooms, according to the scheme shown in Table 3.

The following variables will always be included as control variables in the analysis of treatment effects, i.e., they will not be subject to the LASSO selection procedure:

- ***Troop fixed effects.***
- ***Fixed effects for initial room size.*** Four categories with pooling to the most common sizes such that 1 and 2 get value 2; 3 and 4 get value 4; 5 and 6 get value 6; and 7 to 11 get value 10.
- ***Fixed effects for four categories for the share of women in the room:*** 0; 0.01—0.49; 0.50; >0.5. For some analyses, the category with zero women will naturally disappear.

The troop fixed effects have to be included since randomization was within troops so they act as strata variables. The fixed effects for room size and share of women in the room are conceptually seen as important for the probability of harassment and make the results easier to interpret.

Balance and attrition

To test for balance, we will regress treatment on a set of independent variables measured at baseline, both individually and together, while controlling for the troop fixed effects. More specifically, we will test balance for the following variables:

Attitude towards living in mixed gender rooms, High social status, Female, High GPA, Immigrant background, Mother and Father employed, Mother and Father with high education, Plan higher education, Plan continue serving, Suited for service, Motivated effort, and Consider working in Armed Forces, Share of women in the room (continuous).

With many variables tested, some of them are likely to be different and we will conduct an F-test of whether this set of control variables jointly predict treatment status.

We will not be able to reach all the respondents initially sampled since many people quit during boot camp. We will check whether attrition and missing outcomes are correlated with treatment. If there are statistically significant differences in attrition or non-response between treatment and control (controlling for the troop fixed effects), we will follow the correction proposed by Lee (2009). We also provide a shortened version of the endline survey to individuals that quit the training and we will analyze answers from that survey when available. These data will likely have a low response rate and they do not include all our main measures, so they will be used only for descriptive purposes.

Treatment validation

Attitudes are mainly measured with three questions. One asks for agreement (from 1 for “disagree completely” to 7 for “agree completely”) with the statement that “Telling sex-related jokes can be called harassment”. The other two measure agreement (from 1 for “disagree completely” to 5 for “agree completely”) for the items “A higher share of women in the defense forces lowers a country’s defense ability”, and “Women are equally capable as men to do military service”.¹

- **Index of harassment attitudes.** We combine the three questions to an index in three steps. We standardizing each variable to have a mean of zero and a standard deviation of 1. Then, we weigh them together with a 50% weight on the statement about sexual jokes and 25% weights on the two statements about women’s relative capabilities. The third step standardizes this weighted average.
- **Harassment attitudes.** We test each standardized variable separately.

We tested if the treatment impacted attitudes at baseline and we see in Table 4 that there were clear short term effects.

Main outcome variables

- **Harassment attitudes.** We use the index of harassment attitudes at endline as the main outcome variable to test hypothesis 1. The three separate attitude questions are used to detect if an overall impact stems from one of the items (jokes or women’s capabilities) more than the other. We will also test if there are effects on perceiving men to be better troop leaders at endline (a dummy equal to one if they answer “Men” on the question “Which sex do you think is best at leading troops” and zero if they answer “Women” or “Equally good”).

Sexual harassment prevalence. These outcomes will rely on the Sexual Experiences Questionnaire (Table 1) combined with a follow-up question about perpetrator identity. This second question identifies whether the perpetrator was a roommate of the victim and is asked for all harassment items jointly, and separately for the two harassment behaviors closest to our treatment content.

- **Prevalence of specific treated items.** As the main outcome variable, we set a dummy variable to 1 if a woman self-reported, from a person residing in the same room as them, either unwanted sexual jokes (SEQ item 5), gender harassment directed at her competence (SEQ item 1, or on a separate item on a bullying survey instrument, namely “negative comments about

¹ In Norwegian: Å fortelle seksualiserte vitser kan kalles trakassering; Økt kvinneandel i Forsvaret svekker forsvarsevnen; Kvinner er like egnet som menn til å avtjene førstegangstjenesten.

your qualifications or competence”²). Conditional on finding a treatment effect on this outcome, it is also interesting to explore if that effect is coming from one type of behavior rather than the other.

- *Prevalence of any gender harassment or unwanted sexual attention.* We set a dummy variable to 1 if a woman self-reported, from a person residing in the same room as her, any sex-based harassment behavior in Table 1 plus the item on women’s competence from the bullying inventory.³
- *Third-party observations of harassment.* We ask whether recruits’ have observed sexual harassment against a woman in their own room, and perpetrated by someone else in their own room.⁴ We create a dummy variable for having observed (any) sex-based harassment against a woman roommate from another roommate. Notably, this question on third-party observations is asked after the sexual experiences questionnaire and refers to that questionnaire as a whole. This means that it also includes three behavioral items for sexual coercion (see footnote 4). This variable will not include the item on negative comments about competence, because we have no third-party measurement of that behavior.
- *Harassment from perpetrators inside or outside the room.* An exploratory analysis widens the perpetrator group for the three harassment outcomes above to also include fellow recruits who live in other rooms than the victim.

Victim coping behaviors. For people who marked any behavior on the Sexual Experiences Questionnaire, we ask two follow-up questions about how they coped with this experience. One question asks if the person “pressed charges or notified someone about one or more of the incidents?” and the other if they “Talked about what happened with someone you trust or with a friend to get support?”⁵

² In Norwegian: Negative kommentarer om dine kvalifikasjoner og kompetanse

³ This variable will be measured with some error due to the (small) possibility that sexual coercion may have been perpetrated by a peer. There are three sexual coercion items in the survey, namely: “Unwanted requests for sexual services in exchange for promises of rewards, for instance gifts, better grades, a permanent position on the team or similarly?”; “that someone has treated you badly because you didn’t want to have sex, for example by giving you a low grade, preventing you from opportunities, or by other means oppose you?”; and “that someone has taken advantage of their position to pressure you into unwanted sexual activities?”. In all, less than 0.5 percent reported sexual coercion in the 2022 data.

⁴ In Norwegian: Tenk på de spørsmålene vi nettopp stilte deg om ulike typer trakassering. Har du observert at noen av de kvinnene du delte rom med ble utsatt for slike handlinger? Hvem gjorde det? (flere valg er mulig)

⁵ In Norwegian: Har du varslet eller anmeldt ett eller flere av tilfellene over?; Snakket du om det du skjedde med noen du stoler på eller en venn for å få støtte?

- **Advocacy seeking** will be a dummy for answering yes to the question about pressing charges or notifying someone.⁶
- **Social coping** will be a dummy for answering “yes” to the question about seeking support by talking to trusted people or friends.
- **Victim coping** will be a dummy for answering yes to either the advocacy seeking or social coping question and will be part of our set of main outcome variables.

Bystander attitudes and behaviors. Our first main variable for bystanding will measure self-reported bystanding behavior in response to harassment situations observed either inside or outside the room.

- **Self-reported bystanding** will be a dummy variable that takes the value 1 for individuals who witnessed someone being harassed inside or outside their own room, using the same question discussed above for the prevalence outcomes, and who also self-reported that they “intervened or did something”⁷ in response to what happened.
- **Third-party observed bystanding** will be measured as a binary indicator for third-party observations. As a follow-up question for the sexual experiences questionnaire we ask, as stated above, if the recruit observed any of the listed behaviors happening to a woman in their room. We also have a question of whether they saw it happening to a woman that was not in their room. After the questions about who did it, we ask if “Anyone who witnessed the situation did anything or intervened”⁸. One of the answer categories for these questions is “Fellow soldiers from your room”. We let the dummy variable take the value 1 if this answer is checked for any of the questions.
- **Descriptive norms for bystanding.** We ask recruits to tell us, on a scale from 0 to 100%, “what fraction of their fellow soldiers they think would confront a fellow soldier who sexually harasses someone”.⁹ This variable will be standardized to have mean zero and standard deviation one.
- **Injunctive norms for bystanding.** We measure how recruits’ think that others ought to behave by asking them to rate the statement “It is important to confront the perpetrator when you observe that colleagues/co-students are targets of unwanted sexual attention”¹⁰, on a scale

⁶ If an individual has talked to a superior about the incident, which is how the question is commonly phrased in the literature, this would be captured by the Norwegian word “varslet”.

⁷ In Norwegian: “Sa du selv i fra, eller gjorde noe?”

⁸ In Norwegian: “Var det noen som var vitne til situasjonen som sa i fra eller gjorde noe?”

⁹ In Norwegian: “Hvor stor andel av dine medsoldater tror du vil konfrontere en medsoldat som utøver seksuell trakassering?”

¹⁰ In Norwegian: Det er viktig å si ifra når man observerer at kolleger/medelever blir utsatt for uønsket seksuell oppmerksomhet

from (1) Agree completely to (5) Disagree completely. We create a dummy variable equal to one if the respondents answer “Agree completely” and zero for all other answers.

Our second main measure for bystanding is intended to measure attitudes toward bystanding and we do this with the help of a vignette to which we now turn.

Vignette analysis

As described in the section on empirical challenges, our measurements of harassment prevalence, victim and bystander behaviors may be affected by higher likelihoods to recall harassing events and think of them as problematic. This can bias our potential treatment effects toward zero. We attempt to circumvent this measurement problem by a vignette. This vignette describes harassment situation (giving everyone the same “recall”) and notes how the behavior is unwanted by the target (giving everyone the same value judgement on part of the victim). We then ask about attitudes and hypothetical behaviors. The vignette reads:

The following situation occurred in a room like yours during boot camp training:

After the training session is over, the recruits return to their rooms. A male recruit puts his female roommate in an uncomfortable situation by shouting to her "Can you pull up your sweater, please, I want to see your tits".

What do you think about the following statements concerning the described scenario? [with each item having a scale from (1) Disagree strongly to (5) Agree strongly.

- *This situation constitutes sexual harassment*
- *The woman should just let the situation go*
- *The woman should confront the man and clearly state that this is not okay*
- *If there are others present in the room, they should clearly tell the man that this is not okay*
- *Imagine that the man did not give up, and despite the woman trying to fend him off, he managed to touch her anyway. Given that reporting the male recruit will have clear consequences for him, do you think she should make a report?*
- *If there were others present in the room who did not react to the situation, they would be co-responsible*

We standardize each of the likert scales to have a mean of zero and a standard deviation of one. We then use each one as the outcome variable in our main regression equation (1). The answer to the last question after the vignette, which we call **Co-responsible bystanders**, will be our main measure of **bystanding attitudes**.

Power

We expect to have an endline sample of at least 650 individuals in at least 90 rooms. The treatment is clustered at the room level and we assume an intraclass correlation coefficient at the room level of 0.05 (in 2022 all endline intraclass correlations for similar outcomes were below 0.05). We use two-

sided tests of all hypotheses. At the conventional level of significance of 0.05 and a power of 0.8, our sample size would allow for a minimum detectable effect of 0.26 standard deviations. For some of the outcomes, the sample is restricted to women but we do not expect to have many fewer rooms. The calculations do not take into account the potential gains in precision from including the covariates in the estimation.

We will also adjust the p-values for the fact that we are testing the impact on seven outcomes. We use the false discovery rate method developed by Benjamini and Hochberg (1995), which implies that the m p-values of the i hypotheses are ordered from low to high and that the critical value of the p-value is then $p(i) = \alpha \cdot i/m$. To illustrate, with 5 main hypotheses and a significance level (α) of 0.05, the critical p-value would be 0.01 for the one with the lowest p-value ($0.05 \cdot 1/5$, which is the same as a Bonferroni correction). For the second hypothesis, the critical p-value is 0.02 ($0.05 \cdot 2/5$) and for the fifth it is 0.05 ($0.05 \cdot 5/5$). The minimum detectable effect for the most statistically significant effect after accounting for multiple hypothesis testing is 0.3 standard deviations.

Our five main outcomes variables and corresponding estimation samples are the following:

Harassment attitudes. The main sample includes all soldiers answering at endline.

Sexual harassment prevalence of specific treated items from a person in the room. The main sample for this test is restricted to women.

Victim coping. The main sample for this test is restricted to women.

Self-reported bystanding. The main sample includes all soldiers answering at endline .

Co-responsible bystanders. The main sample includes all soldiers answering at endline.

Exploratory analyses of room gender composition and masculinity

Room composition is randomized at baseline (see the Appendix for the randomization procedure). This gives us random variation in the share of women in the rooms and other aspects of the room composition. There are many possible effects of these room compositions and we will explore the effects of the female share and the effects of being randomized to rooms with more (self-described) masculine men in particular. These exploratory analyses are less detailed and it is difficult to know how to analyze the data beforehand. We may have the possibility of adding data from previous waves of data collection and/or to collect data on sexual harassment again in September 2023. If so, we will also include these datasets in our analysis for this study.

Room gender-composition. Theory on sexual harassment leads us to expect that a larger share of men in a work group will increase women's risk of harassment by men in that group. One theoretical

basis for this expectation is the theory on women's increased visibility and sexualization in such situations (Kanter 1977). Another basis comes from the so-called *contact hypothesis*, according to which women's risk of harassment grows linearly with the proportion of the workday spent interacting with men for simple "mechanical" reasons. The theory assumes that men have a latent proclivity to sexually harass women that is largely constant across socio-economic categories such as education level, age, and social class (e.g., Pryor 1987; Pryor et al. 1993; Pina et al. 2009). Under this assumption, each interaction that a woman has with a man at work entails a uniform risk of harassment. If sexual harassment against women is a function of the proportion of men in the immediate work environment along the lines stated above, a higher share of men should increase prevalence.

A room's gender composition might affect women's and men's attitudes toward voicing grievances against harassment, i.e., affect victims' coping behaviors and bystanders' attitudes toward intervention. A larger share of women in a room might increase advocacy seeking among women by strengthening their self-confidence and personal empowerment (Stoddard et al. 2020, Born et al. 2022). Improving these aspects of women's sense of self may lead them to more forcefully voice objections to their own and others' harassment. A larger share of women may also shift men's behavior toward advocacy seeking on women's behalf by increasing men's support for gender equality and lowering masculinity (Dahl et al. 2021). According to previous research, sexist attitudes are commonly thought to underpin a proclivity to sexually harass (e.g., De Judicibus and McCabe 2001).

We will look at the baseline answers among the men in the room and create the following variable:

- **Masculinity.** A standardized variable for the difference in self-ratings on two scales from 1 (Not really) to 5 (Entirely) for the statements "I am masculine" and "I am feminine".

If there are effects of the information treatment on the baseline answers we will also conduct analyses restricting the sample to the control group. The reason for this is that the answers of the men in the room are measured after the information treatments and we believe the information treatments to affect harassment also via other channels.

Exploratory analysis on career ambitions

One important goal of sexual harassment prevention is to equalize people's opportunities to develop their skills and talents in occupations of their choice. A better work environment for the female gender minority in the military might raise perceptions of occupational fit and future career ambitions within that sector. These impacts are further removed from our treatments, of course, but will be subject to exploratory analysis. We create an *index of military career and ambitions* at endline with the following variables:

- **Plan continue serving:** A binary variable equal to one for those who are considering to continue in the armed forces (answer 1 “Yes” or 2 “Maybe”) and zero otherwise (answer 3)
- **Suited for service:** A binary variable equal to one for the ones fully agreeing (1) with the statement “I believe I have what it takes to carry out service in the Armed Forces”, otherwise 0 (answer 2 to 4).
- **Motivated effort:** A binary variable equal to one for those who respond that they are strongly motivated (9 or 10) on the question for those who respond that they are strongly motivated on the question (“How motivated are you to make an effort in the Armed Forces?”. Those answering 1 to 8 are coded as zero.
- **Consider working in Armed Forces:** We ask recruits whether they would consider working in the armed forces at a later stage in your life, and code all positive answers (4 to 6) as yes and the other answers (1 to 3) as zero.

We combine these measures into an index ranging from 0-4. In the main specification, we will restrict the sample to women. We will then explore effects on the questions separately and for other samples. We will also explore effects on a variable equal to 1 for women who are planning to attend the Norwegian Military Academy (Answer 4 on q165, which asks about recruits’ plans for the future).

To understand potential negative impacts on career ambitions, either via attrition from the military training or measured by the attitudes above, we use two additional questions that capture recruits’ satisfaction with their room allocation. We define these outcomes as follows:

- **Room satisfaction** is measured with a standardized likert-scale variable from 1 (Very unhappy) to 7 (Very happy) for the statement “All in all, how happy were you with the room you stayed in?”¹¹
- **Room social environment.** A weighted index for three ratings from 1 (Disagree entirely) to 7 (Agree entirely) with the statements “In my room, everybody collaborates effectively to solve tasks such as cleaning the room”; “If someone in the room faces challenges linked with the training, others assist them”; “I consider at least one of the persons I share a room with to be a friend”.¹²

IRB

This study was reviewed and approved by the IRB officer at the Frisch center. Informed consent for storing and combining the data is given by the participants.

¹¹ In Norwegian: Alt i alt, hvor godt har du trivdes på rommet hvor du bodde?

¹² På rommet mitt jobber alle godt sammen for å løse felles oppgaver som romvask; Hvis noen på rommet har utfordringer knyttet til tjenesten så bistår de andre på rommet; Jeg vil kalle minst en av personene jeg deler rom med for en venn.

Registration

The pre-analysis plan is archived before any endline data is collected. We archive it at the registry for randomized controlled trials in economics held by The American Economic Association: <https://www.socialscienceregistry.org/> on February 24 2023. We collect endline data on February 26 2023.

Tables

Table 1. Survey instrument for sexual harassment with prevalence in the 2022 data.

Gender harassment		Women	Men
1	Offensive or unpleasant comments about your gender or sexual orientation, for example, that you, because of your gender or sexual orientation, are not suited to do certain types of tasks or should not have the opportunity to perform certain tasks?	0.28	0.06
2	that someone has treated you poorly, ignored, or insulted you because of your gender or sexual orientation?	0.15	0.04
3	that someone has shown condescending behavior or has made demeaning or unpleasant comments or throwaway remarks to you because of your gender or sexual orientation?	0.11	0.02
4	to receive or been shown unsolicited content (such as sms, email, letter or message on social media) related to your gender or sexual orientation?	(not in 2022 data)	
Unwanted sexual attention			
5	that someone has told sexual jokes or stories that upset you/ you found offensive/ made you feel uncomfortable?	0.21	0.04
6	sexually charged stares or looks that upset you/ you found offensive/ made you feel uncomfortable?	0.19	0.05
7	that someone has tried to establish a romantic or sexual relationship with you, despite your discouragement?	0.07	0.04
8	comments or throwaway remarks about your appearance or body that upset you/ you found offensive/ made you feel uncomfortable?	0.11	0.05
9	Bothersome or intrusive invitations for dates, dinners or the like, despite the fact that you had said no to this?	0.04	0.01
10	unwanted sexual attention (glances, throwaway remarks, jokes and teasing) about your body, your clothes, your private life, your sexual orientation or the like?	0.11	0.03
11	that someone has taken or shared sexually suggestive photos/videos of you that upset you/ you found offensive/ made you feel uncomfortable?	0.02	0.00
12	Unwanted physical contact that upset you/ you found offensive/ made you feel uncomfortable, such as touching, petting, hugging or kissing against your will?	0.11	0.06
13	that someone has had sex/intercourse with you without your consent?	0.00	0.00
14	to be forced into sexual acts?	0.00	0.00
15	that someone has tried to have sex with you without your consent or against your will, but without success?	0.01	0.01

Note: The data are collected from the same camp in 2022 at the end of boot camp.

Table 2. Prevalence of sexual harassment during base camp.

	N	Gender Harassment	Unwanted Sexual Attention	Gender Harassment or Unwanted Sexual Attention
Men	420	0.08	0.18	0.21
Women	273	0.32	0.37	0.43
Total	693	0.17	0.26	0.30

Note: The data are collected from the same camp in 2022 at the end of boot camp. The outcomes are equal to one if a recruit answers yes to at least one of the items included under each of the four categories of sexual harassment (see Table 1).

Table 3. Grouping scheme for combinations of room size and number of women in room.

# recruits in room	# women in room	# of rooms in control group	# of rooms in treatment group	Grouping for the fixed effect
5	2	3	1	1
5	3	4	2	2
6	2	2	4	3
6	3	3	8	4
6	4	1	.	4
6	5	1	.	5
8	4	2	.	5
8	5	.	1	5
9	3	.	1	6
9	4	2	1	7
9	5	3	3	8
10	3	.	1	6
10	7	.	1	9
10	5	1	.	9
11	5	.	1	9

Table 4. Treatment effects at baseline.

VARIABLES	(1) Index	(2) Jokes are SH	(3) Capable for service	(4) Defense capacity
Treated	0.448*** (0.061)	0.515*** (0.057)	0.120* (0.065)	0.173*** (0.062)
Observations	948	948	949	949
R-squared	0.111	0.111	0.052	0.052
Troop FE	YES	YES	YES	YES
Roomsize FE	YES	YES	YES	YES
Female share FE	YES	YES	YES	YES
Mean for control group	-0.223	-0.261	-0.061	-0.081

Data is based on all recruits answering the survey at baseline in 2023. The index refers to *Index of harassment attitudes* as described in the text. Columns 2-4 show the effects on the individual *Harassment attitudes*. Robust standard errors, clustered at the room level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

References

- Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.
- Berdahl, J. L. (2007). Harassment based on sex: Protecting social status in the context of gender hierarchy. *The Academy of Management Review*, 32(2), 641–658.
- Berdahl, J. L., Cooper, M., Glick, P., Livingston, R. W., & Williams, J. C. (2018). Work as a masculinity contest. *Journal of Social Issues*, 74(3), 422–448.
- Blackstone, A., Uggen, C., & McLaughlin, H. (2009). Legal consciousness and responses to sexual harassment. *Law & society review*, 43(3), 631–668.
- Born, A., Ranehill, E., & Sandberg, A. (2022). Gender and willingness to lead: Does the gender composition of teams matter?. *Review of Economics and Statistics*, 104(2), 259–275.
- Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, 110(10), 2997–3029.
- Cortina, L. M., and Magley, V. J. (2003). Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. *Journal of Occupational Health Psychology*, 8(4), 247–265.
- Cortina, L. M., & Wasti, S. A. (2005). Profiles in coping: Responses to sexual harassment across persons, organizations, and cultures. *Journal of Applied Psychology*, 90(1), 182.
- Dahl, G. B., Kotsadam, A., & Rooth, D. O. (2021). Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams. *The Quarterly Journal of Economics*, 136(2), 987–1030.
- De Judicibus, M., & McCabe, M. P. (2001). Blaming the target of sexual harassment: Impact of gender role, sexist attitudes, and work role. *Sex roles*, 44, 401–417.
- Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European review of social psychology*, 5(1), 1–35.
- Ely, R. J., & Meyerson, D. E. (2010). An organizational approach to undoing gender: The unlikely case of offshore oil platforms. *Research in organizational behavior*, 30, 3–34.
- Fasting, Kari, and Petter Kristian Kjøber. "Anbefalt metode for kartlegging av mobbing og seksuell trakassering i Forsvaret." (2019). Eksternnotat 19/00805. Available here (in Norwegian): <https://ffi-publikasjoner.archive.knowledgebase.net/bitstream/handle/20.500.12242/2820/19-00805.pdf?sequence=1&isAllowed=y>
- Felstiner, William L. F., Richard L. Abel, and Austin Sarat. 1980–81. "The Emergence and Transformation of Disputes: Naming, Blaming, Claiming . . ." *Law & Society Review* 15:631–54.
- Fitzgerald, L. F., Magley, V. J., Drasgow, F., & Waldo, C. R. (1999). Measuring sexual harassment in the military: The sexual experiences questionnaire (SEQ-DoD). *Military Psychology*, 11(2), 243–264.
- Fitzgerald, L. F., Drasgow, F., Hulin, C. L., Gelfand, M. J., & Magley, V. J. (1997) Antecedents and consequences of sexual harassment in organizations: a test of an integrated model. *Journal of Applied psychology* 82(4): 578–589.
- Folke, O., and Rickne, J. (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics*, 137(4), 2163–2212.
- Glick, P., Berdahl, J. L., & Alonso, N. M. (2018). Development and validation of the masculinity contest culture scale. *Journal of Social Issues*, 74(3), 449–476
- Gutek, B. A., & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harassment of women at work. *Journal of Social issues*, 38(4): 55–74.
- Gutek, B. A., & Cohen, A. G. (1987). Sex ratios, sex role spillover, and sex at work: A comparison of men's and women's experiences. *Human Relations*, 40(2), 97–115.
- Herscovis, M. S., Vranjes, I., Berdahl, J. L., & Cortina, L. M. (2021). See no evil, hear no evil, speak no evil: Theorizing network silence around sexual harassment. *Journal of Applied Psychology*.
- Kanter, Rosabeth Moss (1977) *Men and Women of the Corporation*. New York: Basic Books.

- Køber, Petter Kristian (2016): Fra sesjonsplikt til verneplikt for kvinner – motivasjon og seleksjon på sesjon del 1 for årskullene 1992–1997. FFI-rapport 16/00014.
- Køber, Petter Kristian (2017): Friskere, raskere, sterkere? – en kvantitativ analyse av medisinske og fysiske krav til førstegangstjeneste. FFI-rapport 17/01691.
- Køber, Petter Kristian and Torbjørn Hanson, 2019. "Rekruttering fra førstegangstjeneste - en statistisk analyse" FFI-RAPPORT 19/01036.
- Lee, D. S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76: 1071–1102
- Pina, Afroditi, Theresa A. Gannon, and Benjamin Saunders (2009). An Overview of the Literature on Sexual Harassment: Perpetrator, Theory, and Treatment Issues, " *Aggression and Violent Behavior*, 14, 126–138.
- Pryor, John B. (1987) Sexual Harassment Proclivities in Men, *Sex Roles*, 17, 269–290.
- Pryor, John B., Christine M. LaVite, and Lynnette Stoller (1993). A Social Psychological Analysis of Sexual Harassment: The Person/Situation Interaction, *Journal of Vocational Behavior*, 42, 68–83.
- Rones, Nina, Petter Kristian Køber, Kari Fasting, and Trond Svela Sand. "Metodeutvikling for kartlegging av mobbing og seksuell trakassering i Forsvaret." (2018). Notat 17/16767. Available here (in Norwegian): <https://ffi-publikasjoner.archive.knowledgearc.net/bitstream/handle/20.500.12242/2210/17-16767.pdf?sequence=1>
- Stark, S., Chernychenko, O. S., Lancaster, A. R., Drasgow, F., & Fitzgerald, L. F. (2002). Toward standardized measurement of sexual harassment: Shortening the SEQ-DoD using Item Response Theory. *Military Psychology*, 14(1), 49–72
- Stoddard, O., Karpowitz, C., & Preece, J. (2020). Strength in numbers: A field experiment in gender, influence, and group dynamics. Working Paper
- U.S Merit Systems Protection Board (USMSPB), *Sexual Harassment in the Federal Workplace: Trends, Progress, Continuing Challenges*, 1994.

Appendix: Randomization procedures

The researchers had several meetings with the officers in charge of allocating soldiers randomly to rooms in advance of bootcamp, as well as throughout the allocation process, and all officers displayed a strong understanding of what was meant by randomization and how to proceed in different instances to follow the principles of randomization. Soldiers are randomized to rooms in the following steps:

1. Soldiers arrive to the camp from Sunday to Wednesday, early morning to late evening. They are divided into troops as they arrive, up to 92 individuals in each troop.
2. When a troop is complete (up to 92 individuals), all recruits are given a key card. The key card uniquely identifies each individual within the troop:
 - a. Women are given key cards with even numbers; males are given key cards with odd numbers.
 - b. The key cards decks are shuffled beforehand and handed out to soldiers within a troop, women getting key cards from the even number deck, men getting cards from the odd numbers deck.
3. The researchers have provided the officers in charge with randomization protocol using these key card numbers to allocate soldiers to rooms:
 - a. These protocols are handed out to the officers in charge of each troop beforehand.
 - b. The protocol contains two columns, one for males (odd numbers) and one for females (even numbers): the sequence of the numbers was randomized with a computer.
 - c. The officer then allocates the males with key card number according to the first x numbers for males in the randomization protocol to the first room, and the first y number of females in the randomization protocol to the first room.
 - i. The rooms vary in size, and the army has a rule that there should be at least two women in each room.
 - ii. The share of females within a troop varies and is not known in advance. The officers are allocating people adjusting to the number of people, of each gender, and according to room size. The officers are instructed to allocate people following the randomization protocol to ensure that individuals are randomly placed in rooms.
4. The information treatment:
 - a. The troop leaders are asked to divide the rooms for their troop into two groups as similar as possible with respect to the size of the rooms:
 - i. the soldiers in the treatment group are given key cards with a yellow sticky note to put on their key card.
 - b. The troop leader was told that the two groups would be given slightly different questionnaires.
 - c. When the groups arrived to take the survey they were seated at different tables. Those with a sticky note used a different QR code to access the survey and got the treatments.