U.S. Department of Veterans Affairs

Public Access Author manuscript

J Trauma Dissociation. Author manuscript; available in PMC 2019 April 25.

Published in final edited form as:

J Trauma Dissociation. 2011 ; 12(3): 216–231. doi:10.1080/15299732.2011.551506.

Unreliability and Error in the Military's "Gold Standard" Measure of Sexual Harassment by Education and Gender

Maureen Murdoch, MD, MPH^{1,2}, John B. Pryor, PhD³, Joan M. Griffin, PhD^{1,2}, Diane Cowper Ripley, PhD⁴, Gary D. Gackstetter, PhD, MPH, DVM^{5,6}, Melissa A. Polusny, PhD^{1,2}, and James S. Hodges, PhD²

¹Minneapolis VA Medical Center, Minneapolis, MN

²University of Minnesota, Minneapolis, MN

³Illinois State University-Normal, Normal, IL

⁴HSR&D/RR&D Rehabilitation Outcomes Research Center, Florida/South Georgia Veterans Health System, Gainesville, FL

⁵Analytic Services, Inc. (ANSER), Arlington, Virginia, Bethesda, Maryland

⁶Uniformed Services University of the Health Sciences, Bethesda, Maryland

Abstract

The Department of Defense's "gold standard" sexual harassment measure, the Sexual Harassment Core (SHCore), is based on an earlier measure that was primarily developed in college women. Furthermore, the SHCore requires a reading grade level of 9.1. This may be higher than some troops' reading abilities and could generate unreliable estimates of their sexual harassment experiences. Results from 108 male and 96 female soldiers showed that the SHCore's temporal stability and alternative forms reliability was significantly worse in soldiers without college experience compared to soldiers with college experience and in men compared to women. For men without college experience, almost 80% of the temporal variance in SHCore scores was attributable to error. A plain-language version of the SHCore had mixed effects on temporal stability, depending on education and gender. The SHCore may be particularly ill suited for evaluating population trends of sexual harassment in military men without college experience.

Keywords

Military Personnel; Sexual Harassment; Satisficing; Reliability; Questionnaires; Test Construction; plain language; educational attainment

Corresponding Author: Maureen Murdoch, MD, MPH, Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center (111-0), One Veterans Drive, Minneapolis, MN 55417, Tele: 612-467-2679 Fax: 612-725-2118, Maureen.Murdoch@med.va.gov

Contributing Author Contact Information John Barron Pryor, PhD, Department of Psychology, Illinois State University, Normal, IL 61790-4620, T: (309) 438-5191, pryor@ilstu.edu

Joan M Griffin, Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center, One Veterans Drive (152-CCDOR), Minneapolis, MN 55417, T: (612) 467-4232 F: 612-725-2118, Joan.Griffin@va.gov

Diane Cowper Ripley, PhD, RORC/Research Service (151-B), VA Medical Center 1601 SW Archer Rd., Gainesville, FL 32608, T: (352) 271-4540, Diane.Cowper2@va.gov

Gary Dean Gackstetter, PhD, MPH, DVM, Colonel, USAF (Retired), Analytic Services, Inc. (ANSER), 2900 South Quincy St., Suite 800, Rm 4076, Arlington, VA 22206, T: (571) 303-2087, gary.gackstetter@anser.org

Sexual harassment, reported by up to 74–80% of active duty men and women (Bastian, Lancaster, & Reyst, 1996; Rosen & Martin, 1998), has been associated with symptoms of depression, anxiety, and PTSD (e.g., Kang, Dalager, Mahan, & Ishii, 2005); diminished job satisfaction and performance; and even poorer non-work interpersonal relationships (Pryor, 1995). Progress in reducing military sexual harassment has primarily been gauged by periodically surveying troops and comparing their reports of sexual harassment to those of previously surveyed troops. However, earlier sexual harassment surveys yielded widely differing prevalence estimates across service branches, primarily because each branch defined and measured sexual harassment differently. Consequently, in 1998 the Deputy Assistant Secretary of Defense for Equal Opportunity mandated development of a standardized sexual harassment instrument that was to be “used in all [Department of Defense] and Service-wide surveys” (Lipari & Lancaster, 2003, pg. 97). The military’s chosen measure, the Sexual Harassment Core Measure (SHCore, Stark, Chernyshenko, Lancaster, Drasgow, & Fitzgerald, 2002), is a 12 item derivative of the Sexual Experiences Questionnaire (SEQ, Fitzgerald et al., 1988).

The SEQ was primarily developed in college students and faculty (Fitzgerald et al., 1988; Gelfand, Fitzgerald, & Drasgow, 1995), and most of its psychometric evaluation has been conducted in women (Fitzgerald, Gelfand, & Drasgow, 1995; Gelfand, et al., 1995). In the Armed Forces, however, men outnumber women by a ratio of 6:1. Yet, most psychometric evaluation of the Department of Defense’s SEQ version has either excluded men or, when included, evaluated only a subset of the SEQ items they responded to (Fitzgerald, Magley, Drasgow, & Waldo, 1999). The SEQ/SHCore’s performance in military men therefore remains uncertain.

Furthermore, perhaps because the SEQ was developed in college populations, it and the SHCore require relatively high reading grade levels of 9.8 and 9.1, respectively, based on the Flesch-Kincaide readability formula (Kincaid, Fishburne, Rogers & Chisson, 1975). Since military recruits on average have an 11th grade reading proficiency—which is higher than the US general population—most military members would likely find the SHCore reasonably straightforward and easy to read. However, a substantial fraction with below-average reading skills might struggle to comprehend it. For example approximately 10% of military troops did not complete high school and about a third have cognitive aptitude scores between the 30th and 50th percentiles of national norms (Department of Defense, 2004). Consequently, many military training materials are targeted to the 5th-6th grade reading level, which is the level literacy experts recommend for printed materials to ensure comprehension (Doak, Doak, & Root, 1996).

As with training materials, survey instruments need to perform adequately at all levels of reading comprehension. Otherwise, common results include low survey returns and more satisficed responses. Low response rates have been previously reported in military studies using the SHCore (Bastian, et al., 1996), but no study has evaluated possible satisficing. “Satisficed” responses are the opposite of “optimized” responses and occur when survey respondents do not carefully and accurately interpret questions, do not fully develop an accurate summary judgment in response to questions, or do not transcribe their judgments

correctly (Krosnick, 1991). They are responses that, on the face of it, appear valid but do not actually convey accurate information. Examples of extreme satisficing include endorsing the same response to every item regardless of item content or randomly endorsing various responses for different items.

Satisficing has also been linked to poor motivation (Krosnick et al., 2002; Lam & Bengo, 2003). Military men, who generally face lower risks for sexual harassment than military women (Bastian, et al., 1996; Culbertson, Rosenfeld, & Newell, 1993; Lipari & Lancaster, 2003), could be less motivated than women to respond accurately to sexual harassment questions. Furthermore, some critics have questioned the various SEQ versions' applicability and saliency to men, arguing that several types of experiences important to them, such as challenges to one's masculinity, are not captured (Berdahl, Magley, & Waldo, 1996; Waldo, Berdahl, & Fitzgerald, 1998).

Whether arising from comprehension problems or low motivation, satisficed responses may increase both random and systematic measurement error (Krosnick, 1991; Krosnick et al., 2002). This in turn could cause investigators to over or under estimate the prevalence of military sexual harassment in selected subgroups and hinder appropriate allocation of resources to those subgroups (e.g., to prevent sexual harassment's occurrence or its sequelae). In addition to possible differences in sexual harassment reporting by gender, for example, satisficing could plausibly account for some of the inverse association described between years of formal education and reporting sexual harassment (Antecol & Cobb-Clark, 2006). Clarifying the SHCore's performance in groups at higher risk of satisficing is therefore important.

Since the SHCore was derived from a scale primarily developed in female college populations, we hypothesized that soldiers without college experience would be more likely to satisfice on the SHCore compared to soldiers with college experience, and military men would be more likely to satisfice compared to military women. We also hypothesized that a plain language version of the SHCore, written at a lower reading grade level, would diminish satisficing. Simpler text tends to improve reading comprehension across all reading abilities, even among those with very high reading proficiency (Doak, et al., 1996), and it can sometimes counter low motivation problems by reducing cognitive burden (Krosnick, 1991).

As is common in this field, we tested these hypotheses by examining various aspects of test reliability (Krosnick et al., 2002). Test reliability is the fraction of total variance in test scores accounted for by true subject differences divided by the total test score variance, where total test score variance is the sum of variance from true subject differences and error (Guyatt, Walter, & Norman, 1987). Since satisficing increases error (that is, increases the denominator), reliability is diminished (Krosnick et al., 2002). Thus, if satisficing related to reading comprehension were present, then soldiers without college experience should have lower internal consistency, temporal stability, alternative forms reliability, and case concordance (kappa) on the SHCore compared to soldiers with college experience. If men were less motivated than women to optimize their responses to sexual harassment questions, these same parameters should likewise be lower in men than in women. By extension, any differences between those with and without college experience and between men and

women should be reduced when using a more simply written SHCore version. As a further check of our hypotheses, we used a measure of problem drinking, written at the fourth grade reading level, to see if satisficing and unreliable self-report were specific to the SHCore or were a more general characteristic of our sample. If satisficing were specific to the SHCore, then we expected the drinking measure's internal consistency and test-retest reliability to not differ by education or gender.

Methods

Setting and Sample

The Minneapolis VA Medical Center's Human Studies Subcommittee reviewed and approved the study protocol, as did the relevant Army installation's Commanding Officer and Division Equal Opportunity office. The study's primary goal was to examine the psychometric performances of the Department of Defense's sexual harassment measure, its plain language revision, and several other measures not discussed here.

In October, 1999, a convenience sample of 108 active duty enlisted men and 96 active duty enlisted women at the same installation and for the most part in their first duty tour (mean service length 3.1 years, *SD* 2.7), completed a consent document and 21-page questionnaire on site (Time 1). On the basis of their availability one month later, 86 participants (42% of the original sample) were invited to complete the questionnaire a second time (Time 2). On both occasions participants were offered food while completing the survey, as well as a \$10 gift certificate. The refusal rate for each test administration was less than 1%.

Measures

Sexual harassment.—Participants completed the Department of Defense's original, 26-item SEQ version (Fitzgerald, et al., 1999) from which the SHCore was derived and a 31-item, plain language version we created by unbundling several double- and triple-barreled items and by simplifying each item's language. We counterbalanced the two SEQ versions' order of appearance in the questionnaire. Both SEQ versions were pre-fielded to focus groups of active duty troops at the same Army post. While officers in the focus groups tended to prefer the original version because it more clearly described unwelcome behaviors, junior enlisted soldiers tended to favor the plain language version because it was more understandable. Junior enlisted personnel found the original SEQ's preamble confusing and wordy, and they appeared to have difficulty meaningfully distinguishing among the 5 original response categories, which were "Never," "Once or twice," "Sometimes," "Often," or "Very often." The inability to meaningfully distinguish between response options independently undermines measurement reliability (Hawthorne, Mouthaan, Forbes, & Novaco, 2006). Therefore, before fielding the two SEQ versions to our sample of 204 soldiers, we simplified the preamble to read, "At work, have any of these happened to you in the last year? Someone. . .," followed by the list of SEQ behaviors. Response options were reduced to "Never," "Once," or "More than once." Other SEQ versions have used these response options.

Since the Department of Defense now uses the SHCore exclusively, we present findings only for the 12-item SHCore and its 16-item, plain-language counterpart (SHCore-PL, Appendix). The SHCore-PL had a Flesch-Kincaide reading grade level of 2.8¹.

Problem drinking.—We used the TWEAK’s “hold” version (Russell et al., 1994) to assess problem drinking. The TWEAK is a five-item test developed originally to screen for risky drinking during pregnancy, but also has excellent psychometric properties in men. It includes three of the CAGE (Mayfield, McLeod, & Hall, 1974) alcohol screening questions and also asks about respondents’ alcohol tolerance (i.e., how much can they “hold”?) and blackouts. The TWEAK has a 3.8 reading grade level.

Analysis

Scores for the SHCore and SHCore-PL were calculated by summing respondents’ answers to the relevant items. The TWEAK was scored according to its published algorithm (Russell et al., 1994). Analyses were conducted for respondents overall and for subgroups defined by education (with versus without some college experience) and gender. All reported *p* values are two tailed.

Reliability.—We followed Murphy and Davidshofer’s (1998, pg. 143) conventions for interpreting reliability coefficients: .95 indicates that measurement error has “virtually no effect” on test score variance; .90 reflects high to moderate reliability; .75, moderate to low reliability; .65, low reliability; and .50, “true scores and error have equal effects on test scores.” Reliabilities below .50 indicate that error accounts for more test score variance than do true score differences. For all measures, we used Cronbach’s alpha to assess internal consistency and Pearson’s correlation coefficient to assess one month test-retest reliability; we also report the mixed model intraclass correlation (ICC) between tests and retests. For the sexual harassment measures, we used Pearson’s correlation coefficient to assess immediate- and one-month alternative forms reliability, with the SHCore and SHCore-PL serving as each other’s alternative form. Defined subgroups’ internal consistency and test-retest and alternative forms reliabilities were compared by applying Fisher’s *z* transformation to each subgroup’s correlation and doing a *z* test on the transformed scale.

Case concordance (kappa).—Not all behaviors on the two SHCore versions necessarily meet legal thresholds for sexual harassment. However, we followed common research practice and classified participants as sexual harassment cases if they responded “once” or “more than once” to any test item. Without assigning gold standard status to either test version, we used kappa (κ) to compare case agreement overall and for subgroups defined by education and gender. We used *z* tests to compare κ across subgroups.

¹The 12-item SHCore had correlation .96 with the Department of Defense’s original SEQ version. The 16-item SHCore-PL had correlation .96 with the 31-item, plain-language SEQ version. Analyses of the Department of Defense’s full SEQ version and its plain language counterpart are available on request and essentially mirror results reported here.

Results

Sociodemographic characteristics

Education category was missing for 2 participants, and gender, for 6 participants. The sample's mean age was 23.5 years ($SD = 5$). On average, those with college experience were 3.0 years older than those without college experience ($p < .001$), and women were 1.4 years older than men ($p = .04$). Forty-seven percent of men and 50% of women reported some college experience ($\chi^2 = 0.06$, $df = 1$, $p = .81$). Remaining sociodemographic characteristics are presented in Table 1. Except for age and paygrade, sociodemographic characteristics did not differ significantly by college experience or gender.

Reliability

SHCore.—Table 2 shows reliability estimates and case prevalence for the SHCore. Contrary to expectations, internal consistency did not differ significantly across subgroups. However, compared to those with some college experience, soldiers without college experience demonstrated significantly poorer temporal stability ($z = 2.23$, $p = .01$) and immediate alternative forms reliability ($z = 2.07$, $p = .02$).

Error accounted for more of the variance in men's SHCore responses over time than did true score differences. Compared to women, men demonstrated substantially poorer temporal stability ($z = -2.15$, $p = .02$), immediate alternative forms reliability ($z = -2.76$, $p = .003$), and one month alternative forms reliability when the Time 2 SHCore was compared to the Time 1 SHCore-PL ($z = -3.04$, $p = .001$).

SHCore-PL.—The SHCore-PL's internal consistency was similar to the SHCore's and did not differ across subgroups (Table 2). However, in contrast to the SHCore, the SHCore-PL's temporal stability did not differ significantly for soldiers with and without college experience ($z = 1.09$, $p = .17$). Although men's temporal stability was somewhat improved using the SHCore-PL, their responses remained significantly less stable than women's ($z_r = -2.27$, $p_r = .01$; $z_{ICC} = -2.33$, $p_{ICC} = .01$), and error continued to account for half of their test scores' variance.

Kappa.—Case agreement did not differ by college experience, but men demonstrated substantially poorer case agreement than women ($z = 5.89$, $p < .001$).

Post hoc, exploratory analysis

To further explore the striking gender difference in reliability, we created 4 substrata defined by college experience and gender and then compared reliability parameters across them (Table 3). Because within-stratum numbers are small, there were few statistically significant results. Even so, men without college experience had strikingly poorer SHCore test-retest reliabilities and ICCs than any other group. For men, the SHCore-PL's temporal stability relative to the SHCore's depended on college experience: for men without college experience, the SHCore-PL's temporal stability was qualitatively better; for men with college experience it was qualitatively worse. For women, the SHCore-PL yielded slightly better reliability parameters than the SHCore regardless of college experience.

TWEAK

As is often seen in brief screening tests designed to maximize criterion-related validity, the TWEAK's internal consistency was low (.38-.46); however, it did not differ significantly by education ($z = 0.16$, $p = .44$) or gender ($z = 0.44$, $p = .33$). One month test-retest reliabilities and ICCs were moderate (.77-.86) and also did not differ significantly by education ($z_r = 1.00$, $p_r = .16$; $z_{ICC} = 1.06$, $p_{ICC} = .14$) or gender ($z_r = 0.14$, $p_r = .44$; $z_{ICC} = 0.40$, $p_{ICC} = .35$).

Discussion

The military's "gold standard" for assessing sexual harassment performed well in the population for which it was designed: women and individuals with college experience. Most US troops, however, are males without college experience. Among them, we identified substantial reliability problems. Particularly for men without college experience, more variance in test scores over time was explained by error than by true test score differences. Consequently, the SHCore may not be the ideal instrument for monitoring changes in sexual harassment for most US troops. Claims that sexual harassment prevalence has been declining in the Armed Forces based on responses to the SHCore (Lipari & Lancaster, 2003) should be viewed cautiously.

Since most of the SEQ's psychometric development has excluded men, it is perhaps not surprising that the instrument does not perform as well in them as in women. However, emerging data using other sexual harassment measures suggest that men sexually harassed in the military are at least as adversely impacted by such experiences as women, if not more so (Murdoch, Pryor, Polusny, Wall, Cowper Ripley, & Gackstetter, in review; Vogt, Pless, King, & King, 2005). Inaccurate accounting of men's experiences, especially those resulting in undercounts, could lead policy makers and health care workers to minimize the threats sexual harassment might pose for men and divert resources to other, less urgent problems. Therefore, as with women, men's military sexual harassment experiences need to be quantified as accurately as possible.

Writing tests to a population's average reading level may overshoot the reading abilities of many, which in turn could lead to satisfied or inaccurate responses. Even though military recruits on average read at the 11th grade level and the SHCore reads at the 9th grade level, our focus groups showed that several participants had problems understanding it. In the present sample, the SHCore's test-retest reliability and alternative forms reliability were also substantially poorer in soldiers without college experience compared to those with college experience. Since our sample actually had more college experience than US troops at large, it seems likely that these problems would be amplified in the larger military population. Likewise, since the US general population tends to have lower literacy skills than the military, comprehension problems might also affect civilian SEQ versions.

While we created the SHCore-PL in hopes of reducing cognitive burden for soldiers with lower reading aptitudes, results using the SHCore-PL were mixed. It performed very well in women, regardless of college experience, and improved test performance for men without college experience as expected; however, it actually degraded performance for men with

college experience. Schuman and Presser (1996, pg. 6) speculated that when asking survey questions, well educated individuals might be particularly sensitive to “precise verbal distinctions.” As suggested by our focus groups, by simplifying the SHCore we may have discarded nuances important to college educated men’s responses. Furthermore, despite the SHCore-PL’s better performance in men without college experience, error still accounted for almost as much of the variance in their SHCore-PL scores as was accounted for by true test-score differences. In contrast, the TWEAK, written at the fourth grade level, showed no evidence of unreliability by education or gender. This suggests that the problems we identified in the SHCore/SHCore-PL were specific to those measures and not a general tendency of our sample to satisfice.

Although provocative, the present study had several limitations. First, findings were based on a convenience sample. While our enrollment method has been used in other military studies and is thought to generate reasonably representative samples from any given military facility (Rosen & Martin, 1998; Wolfe, Brown, & Kelley, 1993), findings may not generalize to soldiers stationed at other locations or to other military branches. Additionally, our participants were more likely to have college experience than the military at large.

Second, actual changes in sexual harassment events could have given an appearance of temporal unreliability in soldiers’ responses between the first and second tests. Reactivity, where answering questions induces changes in respondents’ perceptions (Carmines & Zeller, 1979), and priming effects, which trigger test takers to remember additional events after their first sitting, could also lead to spuriously low test-retest reliability estimates. However, all of these effects should have resulted in soldiers reporting more events on the second test compared to the first one. Among respondents with discordant test results, half to four fifths indicated no sexual harassment experiences on retesting but some during the original assessment.

Third, satisficing can be measured directly only by comparing participants’ answers to their actual experiences. Since this is usually infeasible, most investigators rely on indirect evidence, such as reliability measures, to infer the presence of satisficing (Krosnick et al., 2002). However, satisficing is not the only reason why tests prove to be unreliable.

Fourth and most importantly, we do not know if the unreliability we documented in men and soldiers without college experience resulted in biased or unbiased test error. As long as very large sample sizes were used, only biased error would meaningfully influence sexual harassment prevalence or effect-size estimates. However, satisficing that leads to biased error (e.g., systematic under or over counting) cannot be overcome with larger sample sizes. Furthermore, data collected from smaller samples would continue to be at risk for misinterpretation. While it seems inefficient and undesirable to knowingly employ a survey instrument that underperforms in most US troops, our data are too preliminary to recommend abandoning the SHCore. We do, however, encourage replication in larger, more representative samples. Comprehensive clinical and cognitive interviews might also clarify whether the SHCore systematically over or underestimates men’s sexual harassment experiences. If systematic bias were identified, perhaps the Armed Forces could consider creating a military-specific measure of sexual harassment that followed more usual survey

development (i.e., item generation from key informants and focus groups, psychometric analysis of the entire item pool, and cognitive interviews and reliability testing of the resulting instrument).

The military independently represents a large segment of the United States population; however, these results carry implications beyond the Armed Forces and even beyond the field of sexual harassment. Specifically, our findings highlight problems that can arise when measures developed in one population are applied to a new population without adequate pretesting. Feedback, in the form of focus groups or cognitive interviews, must be elicited from members of the new population to ensure that the measure remains readable, applicable, and consistently and correctly interpreted. When feedback cannot be obtained, post hoc reliability assessments across selected subgroups may help identify potential problems. Unfortunately, although Cronbach's alpha is easily obtained and popular in social science research, correlated errors tend to inflate it (Green, 2003). Thus, it may not be terribly useful for identifying problems related to satisficing. In the present study, Cronbach's alpha seemed a less sensitive indicator of potential satisficing than test-retest and alternative forms reliability. We encourage researchers to consider risks for satisficing (e.g., by attending to reading level and saliency) when developing new survey tools and when applying preexisting measures to new populations. These concerns are particularly acute when the topic is socially important and carries policy implications.

Acknowledgements:

Funding for this study was provided by the Department of Veterans Affairs (VA) Health Services Research and Development (HSR&D) Service (IIR 96-014). Drs. Murdoch, Griffin, and Polusny are core investigators for the Center for Chronic Disease Outcomes Research (CCDOR). CCDOR is a VA HSR&D Service Center of Excellence. We thank two anonymous reviewers for their helpful critiques. The views, opinions, and findings are ours and do not necessarily reflect the views of the Departments of Veterans Affairs or Defense.

Appendix

Item Content of the SHCore and its Plain Language Counterpart (SHCore-PL)

SHCore Items		Corresponding Plain Language Items (SHCore-PL)	
Item		Item'	
1	Told sexual stories or jokes that were offensive to you?	1	Told sexual stories or jokes that were offensive to you?
2	Made unwelcome attempts to draw you into a discussion of sexual matters (for example, attempted to discuss or comment on your sex life)?	2'	Tried to talk "sexy" with you when you didn't want to?
3	Made offensive remarks about your appearance, body, or sexual activities?	3a'	Made crude comments about your body?
		3b'	Made crude comments about how you looked?
4	Made gestures or used body language of a sexual nature which embarrassed or offended you?	4a'	Made crude sexual gestures that made you blush?
		4b'	Acted out sex acts in front of you?
5	Made unwanted attempts to establish a romantic sexual relationship with you despite your efforts to discourage it?	5a'	Kept trying to go out with me, even after I said I didn't want to?

SHCore Items		Corresponding Plain Language Items (SHCore-PL)	
		5b'	Kept trying to have a sexual romance with me, even though I said I wasn't interested?
6	Continued to ask you for dates, drinks, dinner, etc., even though you said "No"?	6' ^b	Kept asking you to go out with him/her even after you said you didn't want to?
7	Made you feel like you were being bribed with some sort of reward or special treatment to engage in sexual behavior?	7'	Hinted that you would be treated extra nice if you had sex with him or her?
8	Made you feel threatened with some sort of retaliation for not being sexually cooperative (for example, by mentioning an upcoming review)?	8'	Said you'd get a bad review if you didn't have sex with him/her?
9	Touched you in a way that made you feel uncomfortable?	9'	Touched you in a sexual way?
10	Made unwanted attempts to stroke, fondle, or kiss you?	10a'	Tried to stroke you when you didn't want him/her to?
		10b'	Tried to fondle you when you didn't want him/her to?
		10c'	Tried to kiss you when you didn't want him/her to?
11	Treated you badly for refusing to have sex?	11'	Treated you badly after you said you wouldn't have sex with him/her?
12	Implied faster promotions or better treatment if you were sexually cooperative?	12'	Offered you a promotion in exchange for sex?

^bThis item was omitted from the SHCore-PL because it is nearly identical to item 5a'.

References

- Antecol H, & Cobb-Clark D (2006). The sexual harassment of female active-duty personnel: Effects on job satisfaction and intentions to remain in the military. *Journal of Economic Behavior & Organization*, 61, 55.
- Bastian L, Lancaster A, & Reyst H (1996). Department of Defense 1995 Sexual Harassment Survey (Report No. 96-014). Arlington, VA: Defense Manpower Data Center.
- Berdahl J, Magley V, & Waldo C (1996). The sexual harassment of men? Exploring the concept with theory and data. *Psychology of Women Quarterly*, 20, 527-547.
- Carmines E, & Zeller R (1979). Reliability and validity assessment (Vol. 17). Newbury Park, CA: Sage Publications.
- Culbertson A, Rosenfeld P, & Newell C (1993). Sexual harassment in the active-duty Navy Findings from the 1991 Navy-wide survey. San Diego: Navy Personnel Research and Development Center TR-94-2.
- Department of Defense. (2004). Office of the Undersecretary of Defense, Personnel and Readiness: Population representation in the military services: Fiscal year 2002.
- Doak C, Doak L, & Root J (1996). Teaching patients with low literacy skills. Philadelphia: Lippincott Publishers.
- Fitzgerald L, Gelfand M, & Drasgow F (1995). Measuring sexual harassment: Theoretical and psychometric advances. *Basic and Applied Social Psychology*, 17, 425-445.
- Fitzgerald L, Magley V, Drasgow F, & Waldo C (1999). Measuring sexual harassment in the military: The Sexual Experiences Questionnaire (SEQ-DoD). *Military Psychology*, 11, 243-263.
- Fitzgerald L, Shullman S, Bailey N, Richards M, Swecker J, Gold Y, Omerod M, Weitzman L (1988). The incidence and dimensions of sexual harassment in academia and the work place. *Journal of Vocational Behavior*, 32, 152-175.
- Gelfand M, Fitzgerald L, & Drasgow F (1995). The structure of sexual harassment: a confirmatory analysis across cultures and settings. *Journal of Vocational Behavior*, 47, 164-177.

- Green SB (2003). A coefficient alpha for test-retest data. *Psychological Methods*, 8, 88–101. [PubMed: 12741675]
- Guyatt GH, Walter S, & Norman G (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Disease*, 40, 171–178.
- Hawthorne G, Mouthaan J, Forbes D, & Novoco RW (2006). Response categories and anger measurement: Do fewer categories result in poorer measurement? Development of the DAR5. *Social Psychiatry and Psychiatric Epidemiology*, 41, 164–172. [PubMed: 16362166]
- Kincaid J, Fishburne R, Rogers R, & Chisson B (1975). Derivation of new readability formulas (Automated Readability Index, Fog count, and Flesch Reading Ease Formula) for Navy enlisted personnel Research Branch Report 8–75 Millington, TN: Naval Technical Training, U. S. Naval Air Station Memphis, TN.,
- Kang H, Dalager N, Mahan C, & Ishii E (2005). The role of sexual assault on the risk of PTSD among Gulf War veterans. *Annals of Epidemiology*, 15, 191–195. [PubMed: 15723763]
- Krosnick JA (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Krosnick JA, Holbrook AL, Berent MK, Carson RT, Hanemann W, Kopp RJ, Mitchell RC, Presser S, Rudd PA, Smith V, Moody WR, Green MC, Conway M (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66, 371–403.
- Lam TC, & Bengo P (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24, 65–80.
- Lipari R, & Lancaster A (2003). Armed Forces 2002 Sexual Harassment Survey (No. 2003–026). Arlington, VA: Defense Manpower Data Center, Department of Defense.
- Magley VJ, Waldo CR, Drasgow F, & Fitzgerald LF (1999). The impact of sexual harassment on military personnel: is it the same for men and women? *Military Psychology*, 11, 283–302.
- Mayfield D, McLeod G, & Hall P (1974). The CAGE questionnaire: Validation of an alcoholism screening instrument. *American Journal of Psychiatry*, 131, 1121–1123. [PubMed: 4416585]
- Murphy KR, & Davidshofer CO (1998). Using and interpreting information about test reliability in Murphy KR & Davidshofer CO (Eds.), *Psychological testing: Principles and applications* (4th ed., pp. 127–145). Upper Saddle River, New Jersey: Prentice Hall, inc.
- Pryor J (1995). The psychosocial impact of sexual harassment on women in the U.S. military. *Basic and Applied Social Psychology*, 17, 581–603.
- Rosen L, & Martin L (1998). Childhood maltreatment history as a risk factor for sexual harassment among US Army soldiers. *Violence & Victims*, 13, 269–286. [PubMed: 9836414]
- Russell M, Martier S, Sokol R, Mudar P, Bottoms S, Jacobson S, Jacobsen J (1994). Screening for pregnancy risk-drinking. *Alcoholism: Clinical and Experimental Research*, 18, 1156–1161.
- Schuman H, & Presser S (1996). Questions and answers in attitude surveys: Experiments on question form, wording, and context. Thousand Oaks, CA: Sage Publications, Inc.
- Stark S, Chernyshenko OS, Lancaster AR, Drasgow F, & Fitzgerald LF (2002). Toward standardized measurement of sexual harassment: Shortening the SEQ-DoD using item response theory. *Military Psychology*, 14, 49–72.
- Vogt D, Pless AP, King LA, & King DW (2005). Deployment stressors, gender, and mental health outcomes among Gulf War I veterans. *Journal of Traumatic Stress*, 18, 115–127. [PubMed: 16281203]
- Waldo C, Berdahl J, & Fitzgerald L (1998). Are men sexually harassed? If so, by whom? *Law and Human Behavior*, 22, 59–79. [PubMed: 9487791]
- Wolfe J, Brown P, & Kelley J (1993). Reassessing war stress: Exposure and the Persian Gulf War. *Journal of Social Issues*, 49, 15–31.

Table 1

Sociodemographic Characteristics Overall and by Subgroup, Reported as Percentage (%) With the Characteristic

Characteristic	Overall N = 204	College Experience		Gender	
		None n = 105	Some n = 97	Male n = 108	Female n = 90
Male Gender	54.5	55.9	53.1	100	0
Race/Ethnicity:					
White	50.0	52.4	48.5	54.6	46.7
African American or Black	28.9	31.4	26.8	25.9	31.1
Hispanic	9.3	8.6	10.3	9.3	10.0
Other	12.2	8.6	14.4	11.1	12.2
Highest Educational Achievement:					
Less than 12 years	0.5	1.0	0	0.9	0
GED or other high school equivalency	4.0	7.6	0	1.9	6.7
High School Diploma	47.0	91.4	0	50.0	43.3
Less than 2 years of college credit	35.9	0	74.2	33.3	38.9
2–4 years of college credit	11.1	0	22.6	12.1	10.0
At least some graduate school credit	1.5	0	3.1	1.9	1.1
Paygrade:					
E1-E2	6.6	6.6	5.2	4.7	8.9
E3	32.8	37.1	28.9	28.7	37.8
E4	52.0	53.3	50.5	58.3	44.4
E5	8.6	1.9	15.5 ^{***}	8.3	8.9

Note: Education category was missing for 2 participants; gender was missing for 6 participants.

 $p < .001$ for those without college experience compared to those with college experience.

Table 2
SHCore and SHCore-PL Reliability Estimates and Sexual Harassment Case Prevalence (%) Overall and by Subgroup

Sample	Reliability Estimates										Sexual Harassment Case Prevalence		Case Agreement
	SHCore						SHCore-PL				SHCore	SHCore-PL	
	Internal Consistency	1 Month		Alternative Forms ^{a,b}		Internal Consistency	1 Month Test-Retest ^a						
		α	r	ICC	r		α	r	ICC	κ			
Overall (N = 204)	.85	.70	.70	.89/.72/.70		.87	.73	.73	.73	59.8	65.2	.80	
By Education													
Without College (n = 105)	.85	.57 [*]	.56 [*]	.86 [*] /.68/.61		.88	.69	.67	.67	57.1	62.9	.80	
With College (n = 97)	.85	.81	.81	.92/.79/.79		.86	.79	.76	.76	63.9	69.1	.79	
By Gender													
Male (n = 108)	.85	.40 [†]	.39 [†]	.83 [†] /.45/.37 [†]		.86	.49 [†]	.46 [†]	.46 [†]	48.1 [†]	55.5 [†]	.71 [†]	
Female (n = 90)	.83	.72	.72	.92/.72/.79		.86	.78	.77	.77	75.6	77.7	.94	

Note: Education category was missing for 2 participants; gender was missing for 6 participants. ICC= Intraclass Correlation Coefficient.

^aResults are based on 86 respondents (39 without college education, 50 with college education; 39 men, 47 women.

^bCorrelations are presented in this order: SHCore with SHCore-PL, both taken at Time 1 with SHCore taken at Time 1 with SHCore-PL 1 month later/SHCore-PL taken at Time 1 with SHCore 1 month later.

^{*} p .05 for those without college experience compared to those with college experience.

[†] p < .05 for men compared to women.

Table 3
SHCore and SHCore-PL Test-Retest Reliability and Case Agreement after Stratification by Education and Gender

Stratum	Reliability Estimates				Case Agreement ^b	
	1 Month Test-Retest		Alternative Forms ^a			
	SHCore	SHCore-PL				
	<i>r</i>	ICC	<i>r</i>	ICC	<i>r</i>	κ
Men						
Without College Experience (n = 19)	.24 [*]	.21 [*]	.53	.52	.80/.48/.25	.76
With College Experience (n = 20)	.74	.74	.49	.40	.87/.49/.66	.65
Women						
Without College Experience (n = 17)	.64	.63	.78	.67	.91/.62/.78	.90 [†]
With College Experience (n = 26)	.77	.77	.81	.81	.93/.77/.81	1.00

Note: Education category was missing for 2 participants; gender was missing for 6 participants. ICC= Intraclass Correlation Coefficient.

^aCorrelations are presented in this order: SHCore with SHCore-PL, both at Time 1/SHCore at Time 1 with SHCore-PL 1 month later/SHCore-PL at Time 1 with SHCore 1 month later.

^bResults are based on the full sample of 57 men without college experience, 51 men with college experience, 45 women without college experience, and 45 women with college experience.

^{*} *p* .05 comparing men with and without college experience.

[†] *p* < .05 comparing women with and without college experience.