

GROUP GENDER COMPOSITION EFFECTS IN GENDER-ROLE ATTITUDE MEASUREMENT: TWO FAILURES TO REPLICATE

Joel M. Savell

Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave., Alexandria, VA 22333

KEY WORDS: Gender, context, measurement, group

Some years ago an experiment was reported (Shomer & Centers, 1970) indicating that male college students respond differently to gender-role attitude questionnaires depending on the gender composition of the group in which the questionnaire is administered. The study found--in addition to the usual effect of respondent gender (women more egalitarian on this topic than men)--that the egalitarianism scores of men (but not women) were higher in mixed-gender than in single-gender groups.¹ If the finding is reliable, it suggests either (a) that the presence of women in the room completing the questionnaire at the same time raises the level of gender-role egalitarianism expressed by the men, (b) that the presence of other men in the room completing the questionnaire at the same time lowers the level of gender-role egalitarianism expressed by the men, or (c) both.² Either way, the results raise the question of potential bias in the measuring of a particular variable (gender-role attitude) in the circumstance of other people being present--a circumstance that varies somewhat with the data collection mode but is apparently not uncommon when questionnaires are administered in an organizational setting.³

The primary purpose of the experiments described here was to test the applicability to soldiers of the effect reported by Shomer and Centers (1970) by replicating their experiment some fifteen years later in a military population.⁴

EXPERIMENT I

Method

At a relatively large Army post in the southern part of the United States, 51 male and 56 female soldiers undergoing basic combat training (Replication 1) and 51 male and 59 female soldiers receiving standard follow-on technical training (Replication 2) completed a questionnaire under one of two (same-gender-vs-mixed-gender) group-gender-composition conditions. In each replication, subjects were first assembled at the post testing center and from there were taken to one of four rooms that were being used for the

experiment. In the case of the men, half were taken to a room (room 1) to which only men had been assigned, while the rest were taken to one of two rooms (rooms 2 and 3) to each of which an approximately equal number of women had also been assigned. The case of the women was exactly the reverse: Half were taken to a room (room 4) to which only women had been assigned, while half were taken to one of the two rooms (rooms 2 and 3) to each of which, as indicated above, an approximately equal number of men had also been assigned. In each replication, therefore, there were four experimental locations: one (room 1) in which the questionnaire was administered only to men, a second (room 4) in which the questionnaire was administered only to women, and two others (rooms 2 and 3) in each of which the questionnaire was administered to an approximately equal number of men and women. Each replication was thus a 2x2 (subject gender x group composition) between-subjects design, with subjects randomly assigned within gender to the various conditions. In the all-male group the questionnaire was administered by an enlisted man, and in the all-female group it was administered by an enlisted woman. In one of the two mixed-gender groups it was administered by an enlisted man, and in the other it was administered by an enlisted woman.

The primary dependent variable was the score obtained on a seven-item scale (Cronbach alpha=.76) that had been constructed to measure attitude concerning the role of women in the Army. Other dependent variables consisted of measures of such things as attitude regarding noncommissioned officers (NCOs).

Results

As indicated above, the variable of proctor gender was not combined factorially with the other two variables but was instead held constant (in the single-gender condition) or varied systematically (in the mixed-gender condition). The effect of this variable in the latter condition was examined by testing the proctor-gender x subject-gender interaction. In neither replication was the effect significant (both $p > .05$) and the data from the two mixed-gen-

der conditions were combined. One question asked subjects whether the soldiers in the room were "all or mostly male", "all or mostly female", or "about half and half". Virtually all subjects answered this question correctly.

In each replication, the subject's scale score was computed by summing his/her score on the seven items and computing the mean. The theoretical range was 0 to 6, with higher scores indicating greater gender-role egalitarianism. In replication 1, mean scores for the men and women were 1.8 and 2.4 respectively; and the difference was statistically significant ($F=23.78$, $df=1/103$, $p<.001$). The subject-gender x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all $ps>.05$). In replication 2, mean scores for the men and women were 1.7 and 2.4 respectively; and the difference was again statistically significant ($F=36.70$, $df=1/106$, $p<.001$). As before, the subject-gender x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all $ps>.05$). The results of this first experiment thus failed to detect the phenomenon reported by Shomer and Centers (1970).

EXPERIMENT II

Method

Two years later, at the same post at which the first experiment was conducted, a second effort was made to replicate the Shomer and Centers (1970) results. Serving as subjects in the first session (Replication 1) were 100 male and 91 female soldiers who were receiving phase 2 of their general basic training (individual technical training), and serving as subjects in the second session (Replication 2) were 102 male and 105 female soldiers who were also receiving this training. At each session, subjects assembled in an auditorium where they were randomly assigned by gender to experimental conditions in accordance with the requirements of a 2x4x2 between-subjects design and then taken to one of five rooms which were being used for the experiment. Independent variables were (a) gender of the subject, (b) percentage of soldiers in the subject's room who were of the subject's own gender (100%, 75%, 50%, and 25%), and (c) gender of the two proctors who administered the questionnaire in the subject's room (both male or both female). For each replication the proportion of male and female subjects in the various rooms was as follows: Room 1 (100% male), Room 2 (75% male and 25% female), Room 3 (50% male and 50% female), Room 4 (25% male and 75% female), and Room 5 (100% female). The number of subjects in each room

was approximately 20; and at each level of group-gender-composition, one of the two rooms was proctored by two male soldiers and the other was proctored by two female soldiers. Altogether there were four male and four female proctor pairs, and for the second session (Replication 2) the variable of proctor identity was systematically rotated across conditions.

Subjects were asked about the relative proportion of male and female soldiers who were in the room with them (all or mostly male, more than half male, about half male and half female, more than half female, and all or mostly female); and, as before, virtually everyone gave the correct answer.

Again, the primary dependent variable was the score obtained on the gender-role attitude scale, with other variables consisting of measures of such things as expressed willingness to volunteer for combat, attitude toward Army life, and plans for staying in or leaving the Army. A number of measures were included which sought to tap subjects' feelings of pressure (if any) to respond differently from the way they would ordinarily respond. For example, questions were included that asked what subjects thought about the value of surveys like the present one, the extent to which they thought subjects would respond honestly, and whether there were any questions they had considered not answering honestly. There was also a measure of social desirability. Finally, in an effort to identify soldiers with a situational basis for being concerned about possible monitoring of their responses, we included a question that asked: "Just now, while filling out this questionnaire, did you happen to be sitting next to someone of the opposite gender?" (Yes/No).

Results

During the two years following the first experiment, the method of scoring the gender-role attitude scale was revised (primarily by dropping one of the items and by developing a set of weights to be applied to the individual items), and scores were now able to range up to 31. The replication variable did not interact with any of the other variables in the design (all $ps>.05$), and the data for the two replications were combined. With respect to gender-role attitude, mean scores for men and women were 16.4 and 19.7 respectively, and the difference was statistically significant ($F=56.8$, $df=1/382$, $p<.001$). The subject-gender x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all $ps>.05$). Examination of the other variables revealed no obvious patterns. The results of this second experiment thus failed to detect the phenomenon

reported by Shomer and Centers (1970).

DISCUSSION

Why did the pattern reported by Shomer and Centers (1970) fail to replicate in the experiments described here? One possibility (which is consistent with some of the discussion by the authors of that report) is that gender was not as salient a factor in these experiments as they were in the earlier one. Responses to relevant questionnaire items confirmed that subjects were at least aware of the fact that there were or were not members of the opposite gender in their rooms; but most soldiers were in uniform when they came to the sessions, and it is possible that being dressed alike and in the common uniform served to reduce the attention that subjects paid to this fact.

A second possibility is that the effect observed by Shomer and Centers (1970), while reliable at the time, is no longer reliable because of historical/societal changes that have taken place with respect to attitudes toward women. Given, however, the controversial nature of some of the topics asked about in the questionnaire (e.g., use of women in combat roles) and the fact that male and female soldiers are generally recognized as differing on this question, it seems unlikely that historical/societal factors provide a sufficient explanation.

A third possibility is that subjects in the recent experiments suffered from what Rosenberg (1965) called "evaluation apprehension" and, as a result, were rendered impervious to some of the forces usually operating in a situation of this sort. Efforts were made in the experiments to reduce the unnaturalness of what the subjects were being asked to do, but just how successful these efforts were is not known.

A fourth possibility, which requires some assumptions about the different populations represented in these studies, is that the kinds of people (soldiers? non-college young people?) who took part in the later experiments are for some reason less susceptible to group-gender-composition effects than the kinds of people (non-soldiers? college students?) who took part in the original study. We have no data on this, one way or the other.

A fifth possibility focusses on the different measures of gender-role attitude used in the studies and suggests that the measures used in the recent experiments were less affected by normative considerations than the measures used in the original study. Among the items used in the recent experiments, however, are several that are similar in character to those used in the original study; and none of these items (even

when examined singly) showed group-composition effects.

A sixth possibility is simply that there was no effect to replicate--i.e., that the original effect was not reliable. In this connection it should be noted that (a) the effect reported by Shomer and Centers (1970) was different in pattern from the one these investigators originally predicted and (b) the statistical reliability of the reported effect was not great ($.05 > p > .01$).

Looking back over the results of these two experiments, it is difficult to avoid concluding that--at least for soldiers (or soldiers in uniform)--the effect reported by Shomer and Centers (1970) is not very reliable and, if it exists at all, is of limited scope. (We have not yet tried this experiment with soldiers in combat units although we have used soldiers who were undergoing basic training, and many of these soldiers were expecting eventually to be assigned to combat units.) For the time being (and until additional research on the topic suggests otherwise) there seems little reason to be concerned that surveys of soldier attitude regarding the role of women in the Army will be invalidated by the kind of group-composition factors discussed here.

REFERENCES

- Rosenberg, M.J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1, 28-42.
- Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys: Experiments in question wording, form, and context. New York: Academic Press.
- Shomer, R. W., & Centers, R. (1970). Differences in attitudinal responses under conditions of implicitly manipulated group salience. Journal of Personality and Social Psychology, 15, 125-132.

FOOTNOTES

¹This finding appears consistent with the usual interpretation of observed response patterns where black and white interviewers have interviewed black and white respondents on certain race-related topics (cf Schuman & Presser, 1981). It is not known, however, whether respondents use group membership information about an interviewer in the same way they use such information when it pertains to the other people in the room completing the questionnaire the same time.

²The question could also of course be raised, mutatis mutandis, with respect to women respondents.

³With more and more women entering jobs and workgroups that previously were all male, one could speculate (again assuming the finding to be reliable) that over time men would report more and more gender-role egalitarianism--not (or not necessarily) because they were becoming more egalitarian in their gender-role attitudes but because more and more often they were responding to these questions in situations that elicit the expression of such

egalitarianism.

⁴The replication is not exact. Shomer and Centers (1970) used several types of mixed-gender groups but were unable to demonstrate empirically that these within-mixed-gender-group distinctions were valid. What seemed important, therefore (particularly given the methodological focus of the present research and the presumably infrequent occurrence in the Army of some of these types), was to try to replicate the effect (single-gender/mixed-gender) that was found to be statistically reliable in the original study.