

Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness

JOHN H. KRANTZ, JODY BALLARD, and JODY SCHER
Hanover College, Hanover, Indiana

The Internet provides a new method to obtain subjects. Arguments supporting the use of the Internet as an experimental medium include the ease of data collection, limited resources needed, and ability to gain a sample more representative of the population. However, it has not been demonstrated that the Internet provides the conditions necessary to yield valid data. This paper compares results from experiments conducted over the Internet and in a laboratory. Correlating the means from the Internet samples to the laboratory samples on important interactions yielded values near 1.0, indicating that the two data sets are driven by the same psychological variables. Further evidence of the validity of these Web studies comes from regression analyses. However, more studies of this sort are needed to further identify what types of experiments may be validly conducted over the Internet.

The development of the Internet, in particular the World-Wide Web (WWW), provides a new medium for psychological research (Reips, 1996; Welch & Krantz, 1996). The WWW is particularly intriguing because of its ability to present integrated images and sounds that can be used as stimuli. There are a number of arguments for conducting research over the WWW, ranging from ease of implementation and the relatively limited resources that are required (Schmidt, MacDonald, & Hoffmann, 1996) to extending the external validity of studies by having a more representative sample (Reips, 1996). However, arguments can also be made that there are several threats to the internal validity of WWW experiments. By conducting an experiment over the WWW, the researcher loses that all-important control over the experimental environment and equipment. For example, some subjects may be in a quiet room whereas others work in a noisy computer laboratory (Reips, 1996) and, in the psychoacoustical experiments reported by Welch and Krantz (1996), the data collection and interpretation were hampered by the fact that many computers accessing the WWW do not have adequate audio attachments to play the sound files that were used as the stimuli in the study. Even when the computers do have the necessary equipment, the quality of the sound reproduction varies considerably across computers. These variations in the environment and equipment potentially contaminate the data.

The authors would like to express their deep gratitude to the two anonymous reviewers and the editor for their many helpful comments, which greatly improved the quality of this paper. We would also like to thank the many subjects both from the laboratory and World-Wide Web studies for sharing their perceptions. Correspondence should be addressed to J. H. Krantz, Hanover College, P.O. Box 108, Hanover, IN 47243-0108.

Thus, it is necessary, as with all new experimental techniques, to establish the validity of the WWW as an experimental medium. One approach is to conduct the same experiment both in the laboratory under well-controlled conditions and over the WWW. If the same psychological variables are driving the results of both data sets, the trends in the data should be very similar, leading to a high correlation between the results of the two studies. The present paper reports a comparison of two laboratory studies on variables influencing the perception of attractiveness of females with the same studies ported to the WWW.

Background Issues on the Research Questions

The psychological issue under investigation is what features of the female figure determine attractiveness (Fallon & Rozin, 1985; Singh, 1993; Wiggins, Wiggins, & Conger, 1968). One reason for interest in this topic is the hypothesis that the use of very thin models by the media may play a role in the general dissatisfaction women have with their bodies and the increasing number of cases of eating disorders among women (Mable, Balance, & Galgan, 1986; B. Silverstein, Perdue, Peterson, & Kelly, 1986).

Although different studies have examined many different issues related to female attractiveness, the samples of nearly all of the studies have been drawn from college-age students. These samples differ in many ways from the general population and may limit generalization. For example, the age range of most college samples is limited to individuals in their late teens to early twenties. As they age, people may change in what they find attractive in female figures. In the hope of examining some of these issues, two WWW studies were developed to match two laboratory studies. One set of studies, the front-view studies, investigated weight and hip-to-bust proportion with figures facing front. The other set of studies, the side-view

Table 1
Subject Characteristics of Laboratory Samples

| | Frontal View | | Side View | |
|----------------|--------------|-----|-----------|-----|
| | No. | % | No. | % |
| Total <i>N</i> | 112 | | 121 | |
| Female | 56 | 50 | 65 | 54 |
| Male | 56 | 50 | 56 | 46 |
| Age | | | | |
| 18–22 | 112 | 100 | 121 | 100 |
| Race | | | | |
| White | 112 | 100 | 121 | 100 |
| Continent | | | | |
| N. America | 112 | 100 | 121 | 100 |

studies, investigated weight, breast size, and buttock size in figures viewed from the side.

METHOD

Subjects

In the laboratory studies, subjects were obtained from introductory psychology classes and from upper-level psychology classes and were recruited as volunteers (Table 1). The subjects from introductory psychology classes received extra credit for their participation, and the other subjects participated solely on a volunteer basis. WWW subjects were obtained on the basis of using the Web pages that collected the data (Table 2). All participation was voluntary, and only findings regarding those subjects who completed an entire experiment were included in data analyses. The only two places the experiments were posted were the Hanover College psychology department home page (<http://psych.hanover.edu/>), where the experiments were housed, and the page of on-line research maintained by the American Psychological Society (<http://psych.hanover.edu/APS/exponnet.html>). From there the experiments may have been linked to by other pages and by search engines. However, these links were not sought out. During the last 3 months of the data collection reported here, the front-view experiment collected a mean of 7.1 complete subjects a day. This number represented 17% of all who reached the experiment's home page and 29% of all who began the experiment. The side-view experiment collected a mean of 2.7 complete subjects a day. This number represented 10% of all who reached its home page and 17% of all who began the experiment. The differences in the rate of data collection and rate of completion can probably be ascribed to the different lengths of the two experiments. The front-view experiment required half as many trials as the side-view experiment.

Stimuli

The figures for the front-view studies were adapted from Fallon and Rozin (1985). The nine images differing in weight were scanned into a computer where they could be modified. First, all figures were adjusted so that the bust and hip width were equal. Then, two new sets of figures were created. One set was adjusted so that the figure had a bust width 1.2 times greater than the hip width at each weight value. Another set had the hip width 1.2 times greater than the bust width. Thus, there were 27 figures: nine weight values \times the three proportions: hips $>$ bust, hips = bust, and hips $<$ bust.

The figures for the side-view studies were adapted from Wiggins et al. (1968). One of the figures was scanned to allow for the images to be created and modified. First the figures were adjusted for weight to approximately match the first six weight values used in Fallon and Rozin (1985). Then, for each weight value, three levels of breast size and buttock size were added. The six levels of weight

\times three levels of breast size \times three levels of buttock size yielded 54 total figures.

Procedures

All studies employed a magnitude estimation procedure with a standard (Engen, 1971). The standard figure was given an arbitrary attractiveness value of 200. In this method, subjects rated the attractiveness of each figure by assigning a value relative to the standard; for example, a figure perceived to be half as attractive as the standard would be assigned a value of 100. The value of 200 was chosen to avoid suggesting scales of 1 to 10 or 1 to 100, so that subjects might feel freer in assigning their attractiveness ratings. As part of the informed consent procedure, subjects completed a personal information form giving their age, race, country of birth, and gender.

In the laboratory, subjects were run in groups viewing images presented with overhead projectors. In the front-view study, the standard was presented immediately prior to each figure to be judged. For the side-view study, both the standard and the figure to be judged were projected simultaneously. Figures were presented in a random order. To conduct the studies on the Web, each figure was placed on its own Web page with the standard immediately to the left of the figure to be judged. The standard value of 200 was adjacent to the standard, and instructions were presented on each page as well as on the informed consent page. Subjects entered their judgments in the single form field on the page. Subjects clicked on a button to submit the data for that trial, and then the page with the next image to be judged was delivered to the browser. Figures were randomized, with the order rotated across blocks of subjects. Therefore, the Web experiments very nearly approximated the way that the laboratory studies were conducted.

Data Analyses

Data were range corrected, which adjusts each subject's data to a 1 (for the minimum rating) to 100 (for the maximum rating) range. Thus, the range of judgments was standardized across subjects (L. D. Silverstein, Krantz, Gomer, Yeh, & Monty, 1990).

In all statistical comparisons, α was set at .05. The laboratory front-view study range-corrected data were submitted to a 9 (weight) \times 3 (proportion) \times 2 (gender) analysis of variance (ANOVA)

Table 2
Self-Reported Subject Characteristics of World-Wide Web Samples

| | Frontal View | | Side View | |
|----------------|--------------|----|-----------|----|
| | No. | % | No. | % |
| Total <i>N</i> | 428 | | 145 | |
| Female | 191 | 44 | 66 | 44 |
| Male | 237 | 56 | 79 | 56 |
| Age | | | | |
| 18–22 | 101 | 24 | 39 | 27 |
| 23–30 | 141 | 33 | 54 | 37 |
| 31–50 | 162 | 37 | 43 | 30 |
| >50 | 24 | 6 | 9 | 6 |
| Race | | | | |
| Nonwhite | 34 | 8 | 17 | 11 |
| White | 382 | 89 | 125 | 84 |
| Continent | | | | |
| Africa | 2 | 1 | 0 | 0 |
| Asia | 9 | 2 | 4 | 3 |
| Australia | 13 | 3 | 3 | 2 |
| Europe | 31 | 7 | 12 | 8 |
| N. America | 368 | 86 | 123 | 82 |
| S. America | 5 | 1 | 1 | 1 |

Note—Since the subject characteristics of race and country of origin were not used in the current studies, subjects were included if they did not fill in these questions.

with repeated measures on weight and proportion. For the WWW front-view study, an additional between-subjects factor of age was also included in the ANOVA. Age was broken into the four groups indicated in Table 2. The laboratory side-view study range-corrected data were submitted to a 6 (weight) \times 3 (breast size) \times 3 (buttock size) \times 2 (gender) ANOVA, with repeated measures on all factors except gender. Again, for the WWW side-view figures, the ANOVA had the additional between-subjects factor of age. All reported differences between means were determined to be significant using Tukey's HSD post hoc comparison.

RESULTS

Front-View Studies

The main effects of weight and proportion were significant [$F(8,880) = 276, MS_e = 720, p < .01$; $F(2,220) = 28.33, MS_e = 283, p < .01$, respectively]. The two-way interactions of proportion \times weight [$F(16,1760) = 15.35, MS_e = 232, p < .01$], gender \times proportion [$F(2,220) = 10.73, MS_e = 283, p < .01$], and the three-way interaction of weight \times proportion \times gender [$F(16,1760) = 2.23, MS_e = 232, p < .03$] were also significant. No other main effect or interactions were significant. The same main effects and interactions were also significant in the WWW version of the study: weight [$F(8,3360) = 353, MS_e = 877, p < .01$], proportion [$F(2,840) = 28.38, MS_e = 390, p < .01$], proportion \times weight [$F(16,6720) =$

$24.67, MS_e = 247, p < .01$], proportion \times gender [$F(2,840) = 9.00, MS_e = 390, p < .01$], and weight \times proportion \times gender [$F(16,6720) = 1.99, MS_e = 247, p < .02$]. Although the degrees of freedom for the error terms in WWW data were much larger, it must be remembered that the explained variance in the WWW studies was partialled many more ways because of the additional variable of age and all the additional interactions that resulted from this additional variable. Power ratings for the independent variables were very consistent across the two experiments, indicating that increasing the number of subjects in the WWW version of the experiment did not inflate the significance levels.

The three-way interaction of weight \times proportion \times gender is shown in Figure 1 for both data sets. (This interaction is shown in the figure because the main effects and other interactions can be seen fairly clearly in this presentation of the data and also because using one figure saves space.) As can be seen from the figure, both studies yielded the same general data trends, including an overall preference for figures at the 30 weight value, significantly lower ratings for the hips < bust figures near the most preferred weights, and a preference by males for these hips < bust figures over other proportions when they were thin (weight values of 10 and 20). Since interpretation of the data is not the focus of the

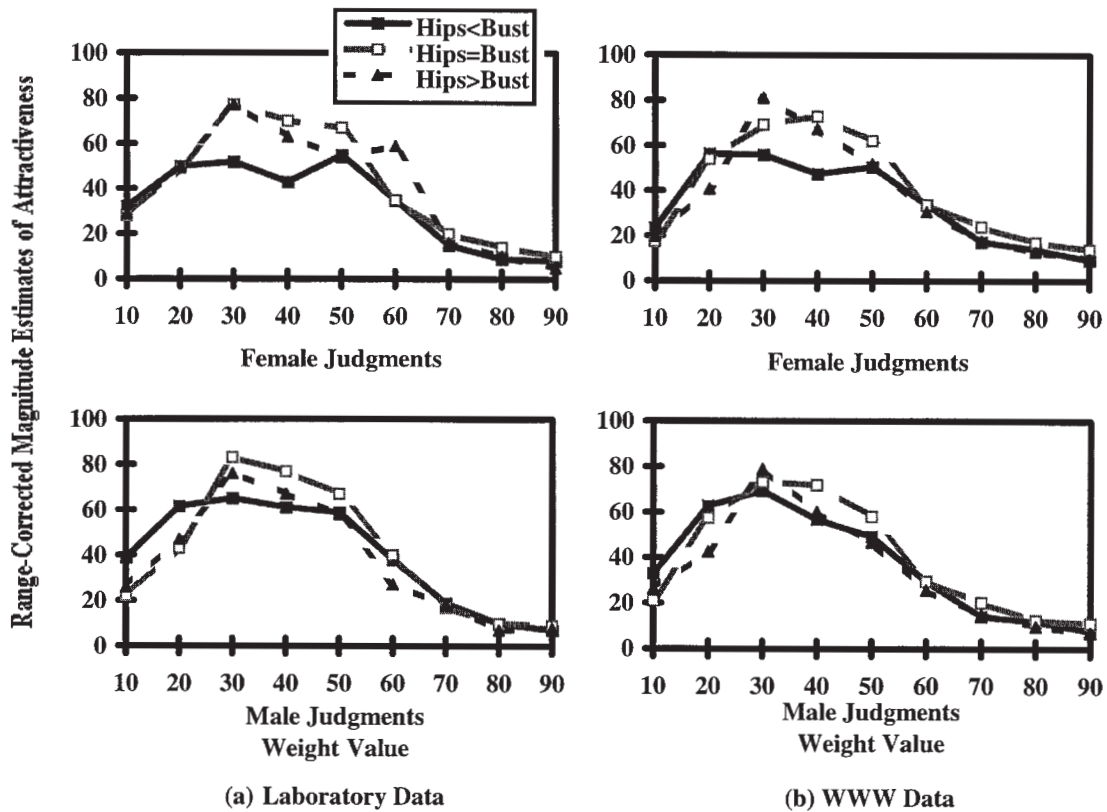


Figure 1. Front-view data for weight \times proportion \times gender interaction. (a) Data from the laboratory study. (b) Data from the World-Wide Web (WWW) study.

paper, further discussion of these data will not be presented here.

A procedure often used in attempting to determine the validity of a psychological test—correlating the untested measures to established measures (Shaughnessy & Zechmeister, 1994)—was adapted in order to examine the validity of the Web data. The use of magnitude estimation is well established as a method of collecting a wide range of psychological judgments in the laboratory (Engen, 1971). In this case, because different subjects were used, the means from the weight \times proportion \times gender interaction from the two studies were entered into a Pearson's correlation. This interaction was chosen because the means for the main effects and other interactions were simple averages of the means in the three-way interaction, because each study was primarily a within-subjects design, and because the number of male and female subjects were nearly the same in both studies. Thus, if the data from this interaction fit the criteria of validity established here, so would all lower order effects because each three-way interaction would have the same values that would be used in determining the values for the lower order effects. If the means were the result of the same psychological variables operating in each study, the correlation of the two data sets should be near 1.00. The resulting correlation was significant and extremely high [$r(52) = .96, p < .001$]. However, we argue that in this case the data should be more than highly correlated. For both the laboratory and WWW data sets, each subject's judgments were restricted to the same range of values through the range-correction procedure. Thus, the means should have been more than moving in the same direction, which is what the positive correlation indicates; they should have been the same or very similar if driven by the same psychological variables. Thus, a regression analysis of these data should yield a slope of near 1.00 and an intercept of near 0. A least squares regression using the WWW data set as the predicted variable yielded the following equation:

$$\text{WWW mean} = 0.91 (\text{laboratory mean}) + 1.85$$

The standard error of the slope was 0.04, and the standard error of the intercept was 1.70. Thus, although the slope was significantly different from 1, it was very close to 1, and the intercept was not significantly different from 0. The fact that the slope was slightly different from 1 may have been due to the difference in the age ranges and other person variables between the two subject groups. However, both the regression and the correlational analyses showed that the WWW data set is valid.

Side-View Study

In the laboratory side-view study, the main effects of weight [$F(5,595) = 273, MS_e = 1,281, p < .01$], breast size [$F(2,238) = 18.73, MS_e = 649, p < .01$], and buttock size [$F(2,238) = 29.05, MS_e = 523, p < .01$] were significant. The following interactions were also significant: weight \times gender [$F(5,595) = 3.21, MS_e = 1,281,$

$p < .01$], breast size \times gender [$F(2,238) = 33.01, MS_e = 649, p < .01$], weight \times breast size [$F(10,1190) = 29.07, MS_e = 226, p < .01$], and weight \times breast size \times gender [$F(10,1190) = 7.09, MS_e = 226, p < .05$]. The variable of buttock size proved to be of very low power. Even when a main effect or interaction involving buttock size was significant, the post hoc Tukey's HSD usually did not reveal any significant differences because of the low power of this variable. Thus, for the sake of brevity, buttock size will not be discussed here. In the WWW version of the study, the following main effects and interactions from the list above were significant: weight [$F(5,685) = 129, MS_e = 1,664, p < .01$], breast size [$F(2,274) = 15.58, MS_e = 674, p < .01$], breast size \times gender [$F(2,274) = 5.50, MS_e = 674, p < .05$], weight \times breast size [$F(10,1370) = 24.03, MS_e = 280, p < .01$], and weight \times breast size \times gender [$F(10,1370) = 4.02, MS_e = 280, p < .001$]. The only difference was that the weight \times gender interaction was not significant for the WWW data set. The interaction of weight \times breast size \times gender for both data sets is shown in Figure 2 and, as in the front-view studies, the other main effects and interactions are fairly clear from this figure. Again, many superficial similarities in the data are seen, including a preference by males for large breasts in many of the thinner figures, and women tending to prefer moderate breast sizes except perhaps at the very thinnest weight value in the WWW study. This interaction was submitted to the same correlation and regression analyses. As in the front-view studies, the correlation was near 1 [$r(34) = .95, p < .001$]. The regression equation also demonstrated the great validity of the WWW data set:

$$\text{WWW mean} = 0.99(\text{laboratory mean}) + 2.72$$

The standard error of the slope was 0.06 and the standard error of intercept was 2.21. Thus, the slope was not significantly different from 1 and the intercept was not significantly different from 0.

DISCUSSION

In both studies, the WWW data were comparable to the laboratory data sets. This comparability was found despite the great potential for variations in experimental environment and equipment that was discussed in the introduction. The comparability in the data also occurred despite the differences in the experimental procedures that are a necessary consequence of the use of the WWW. For example, in the front-view laboratory study, the standard and comparison figures were presented sequentially, whereas in the Web study, the figures were presented simultaneously, and the experiments in the laboratory were experimenter paced, whereas the experiments over the Web were subject paced. These findings attest to the robustness of the psychological phenomena under investigation.

Moreover, these findings suggest that some examination of the subject variable of age, which in the present research is possible only in WWW data sets, may be fruit-

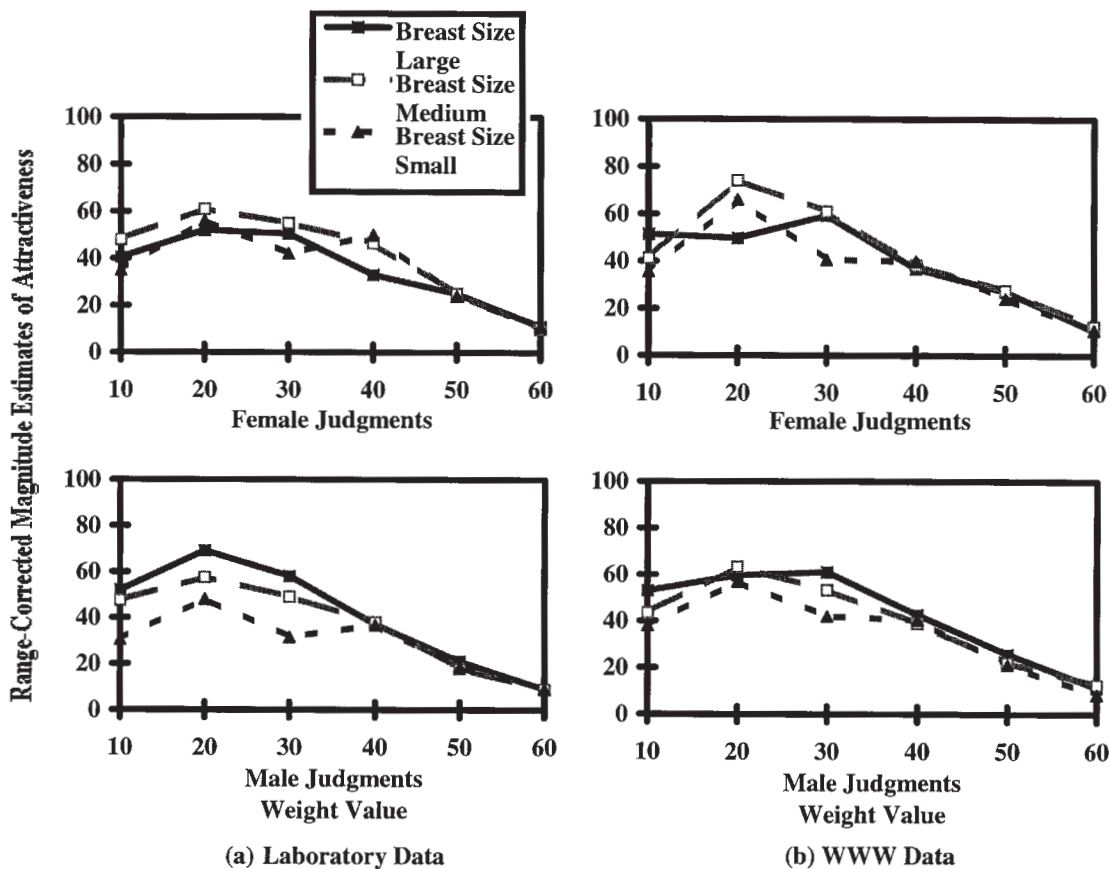


Figure 2. Side-view data for the weight \times breast size \times gender interaction. (a) Data from the laboratory study. (b) Data from the World-Wide Web (WWW) study.

ful. For example, the interaction of age \times weight was significant in the front-view WWW study [$F(24,3360) = 1.97, MS_e = 877, p < .005$]. This interaction is mainly seen in that the oldest group ($n = 24$) gave the lowest ratings for figures at the 10 and 20 weight values compared with the other age groups, which did not significantly differ in their ratings of these figures. The older subjects also gave their highest ratings for the figures with a 40 weight value, whereas all other age groups preferred the figures with a 30 weight value. One potentially interesting question raised by this finding is whether the interaction was due to a cohort difference, in which standards of beauty are different for different generations, or due to a true aging difference in that older people tend to prefer heavier figures perhaps as a result of the tendency of many people to gain weight as they age. Hopefully, the number of subjects differing in racial heritage and country of birth will reach a level where these variables can successfully be examined.

The rate of data collection was generally slow but did increase steadily since the experiments were posted. It is almost certain that the rate of data collection would be faster if fewer trials were used. Comparison of the rate between the two Web experiments indicated that the shorter front-view experiment had more subjects begin

the experiment (25/day vs. 16/day) and a greater percentage complete the experiment (29% of those that began vs. 17% of those that began) relative to the side-view experiment. More actively posting links to the experiments on many different sites might have led to a greater data collection rate. However, limiting the postings of the experiment to standard psychological sites helps give potential subjects additional assurance that this is a psychological experiment and that the data will be used in accordance with appropriate psychological purposes. A recent highlighting of the American Psychological Society on-line research page by the Planet Science Web site has greatly increased the rate of data collection for the experiments, suggesting that it is a useful page on which to post experiments.

It will require many such studies to truly determine what type of research can validly be conducted over the WWW. However, some initial observations are possible at this point. First, several features of this study are relatively unusual and may have contributed to the success of this endeavor. First, the figures were black-and-white drawings, and thus the simplest of monitors can render these images with high fidelity, and changes in resolution will change only the size of the figures. If the figures had been in color or gray shaded, variations in mon-

itor calibration and the video card could have caused a greater variation in the stimulus appearance and quality and could have significantly affected the data. In addition, these data were extremely robust and the manipulations quite powerful; the main trends in the data are visible when examining the data from only a few subjects. Thus, studies employing more sophisticated images and subtle manipulations may not run as successfully on the WWW. A comparative rating of attractiveness of the type made in the present studies is relatively simple and may not have been greatly effected by the variations in the environment. More complex judgments may be more influenced by the variations that will occur over the WWW. Finally, no attempt was made to alter the subjects' cognitive or affective state or attitudes. Such manipulations probably would be ineffective over the Web since the variations in the environment would interact with these manipulations.

REFERENCES

- ENGEN, T. (1971). Psychophysics II: Scaling methods. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlossberg's Experimental psychology* (3rd ed., pp. 47-86). New York: Holt, Rinehart & Winston.
- FALLON, A., & ROZIN, P. (1985). Sex differences in perceptions of desirable body shape. *Journal of Abnormal Psychology, 94*, 102-105.
- MABLE, H., BALANCE, W., & GALGAN, R. (1986). Body image distortion and satisfaction in university students. *Perceptual & Motor Skills, 63*, 907-911.
- REIPS, U. (1996, October). *Experimenting in the World-Wide Web*. Paper presented at the 1996 Society for Computers in Psychology Conference, Chicago.
- SCHMIDT, W. C., MACDONALD, J., & HOFFMANN, R. (1996, October). *Setting up your own World-Wide Web server*. Paper presented at the 1996 Society for Computers in Psychology Conference, Chicago.
- SHAUGHNESSY, J. J., & ZECHMEISTER, E. B. (1994). *Research methods in psychology* (3rd ed.). New York: McGraw-Hill.
- SILVERSTEIN, L. D., KRANTZ, J. H., GOMER, F. E., YEH, Y., & MONTY, R. W. (1990). The effects of spatial sampling and luminance quantization on the image quality of color matrix displays. *Journal of the Optical Society of America A, 7*, 1955-1968.
- SILVERSTEIN, B., PERDUE, L., PETERSON, B., & KELLY, E. (1986). Some correlates of the thin standard of bodily attractiveness for women. *International Journal of Eating Disorders, 5*, 145-155.
- SINGH, D. (1993). Adaptive significance of female physical attractiveness: Role of waist-to-hip ratio. *Journal of Personality & Social Psychology, 65*, 293-307.
- WELCH, N., & KRANTZ, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, & Computers, 28*, 192-196.
- WIGGINS, J., WIGGINS, N., & CONGER, J. (1968). Correlates of heterosexual preference. *Journal of Personality & Social Psychology, 10*, 82-90.

(Manuscript received November 18, 1996;
revision accepted for publication February 5, 1997.)