DOCUMENT RESUME

ED 222 843 CG 016 317

AUTHOR Richert, Alphons J.; Hoyenga, Katharine I.
TITLE Measurement of Individual Differences in Sex-Role Stereotyping.
PUB DATE May 82
NOTE 19p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (54th, Minneapolis, MN, May 6-8, 1982). Best copy available.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; College Students; Higher Education; *Individual Differences; Interpersonal Relationship; Personality Traits; Responses; *Sex Role; *Sex Stereotypes; Social Cognition; *Test Construction; Test Reliability
IDENTIFIERS *Bayes Theorem

ABSTRACT

It has been suggested that a procedure based on the logic of Bayesian probabilities would make it possible to assess individual differences in stereotyping. Given the possible advantages of using the McCauley and Stitt (1978) procedures to measure individual differences, three groups of college students were tested to see if they would use Bayes rule appropriately in responding to the measure. Results indicated that the basic assumption underlying the use of the McCauley and Stitt (1978) procedures was being met and that subjects used Bayes appropriately. The correlations of the estimated percentage of trait given sex with both base rate (the overall frequency of the trait in the population) and with representativeness (the frequency of a particular sex showing a specific trait) were all large and significant. The findings suggest that a measure of sex-role stereotyping based on estimations of conditional probabilities is viable. (JAC)

Measurement of Individual Differences in Sex-role Stereotyping

Alphons J. Richert and Katharine I. Hoyenga

Western Illinois University

Measurement of Individual Differences in Sex-role Stereotyping

Several assumptions and procedures appear to characterize many of the current attempts to measure sex-role stereotyping. First, many current measures such as Spence's PAQ (Personal Attributes Questionnaire; Spence, Helmreich & Stapp, 1975) or the BSRI (Bem's Sex Role Inventory; Bem, 1974) attempt to assess the extent to which the respondent behaves--or says he/she behaves--in sex-role stereotypic ways. Except for Broverman's work (Broverman, Broverman, Clarkson, Rosenkrantz & Vogel, 1970), research on sex roles does not typically address the extent to which people view and respond toward others in sex-role stereotypic ways. Even Broverman's work was concerned with the stereotyping exhibited by certain groups or types of people (e.g. clinicians) rather than the degree to which one individual sees others in stereotypic ways.

A second assumption involves the development of the scales. Current measures have typically been formed out of item pools developed either on the basis of differential endorsement by males and females (the M-F scale from the CPI) or the differential mean ratings of how "typical" or "desirable" a particular trait or item is for males and females (BSRI or PAQ). This means that when an individual responds to stereotype measures developed in these ways, he/she is indicating degree of agreement with an average or group perception of a certain "typical" or "desirable" sex-role rather than revealing the nature and extent of his/her personal stereotype.

Finally, current measurement procedures all appear to have taken an a priori stance on the relationship between masculinity and femininity. Some, such as the M-F scales on the MMPI or the CPI, are constructed on the assumption that masculinity-femininity is a bipolar trait. Others, such as the BSRI (Bem, 1974), assume that there is an orthogonal relationship between masculinity and femininity. And Spence's PAQ scale assumes that there are two types of masculinity-femininity: bipolar and orthogonal.

McCauley and Stitt (1978; McCauley, Stitt & Segal, 1980) argue that measures of stereotyping based upon the assumptions and procedures sketched above cannot adequately measure individual differences in the degree to which stereotyping occurs. McCauley and Stitt (1978) define a stereotype as any trait that is seen as more probable for the target group than for the world in general. Working from this definition, they point out that "stereotypes exist in indidivuals, but, as noted above, the checklist can measure only a kind of group-average stereotype" (McCauley & Stitt, 1978, p. 929). They go on to suggest that a procedure which is based on the logic of Bayesian probabilities would make it possible to assess individual differences in stereotyping. The degree to which a person sees a trait or group of traits as uniquely characteristic of a target group would be measured. The degree of uniqueness of a trait in a target group for an individual is, of course, quite different from a count of the number of times a person agrees or disagrees with statements that a group has decided are characteristic of a target group. The measure of sex-role stereotyping presented here is based on this logic and uses the McCauley and Stitt (1978) procedures.

When a stereotype is defined and measured in the ways proposed by McCauley and Stitt (1978), the measure is no longer a measure of image of self or self-reported behavior. Rather the procedure assesses the degree to which the respondent perceives or attributes unique, distinguishing characteristics to others in a specific target group, in this case women. The procedures, then, tap cognitive products or processing in the subject rather than self-image. Clearly when such an approach to measuring sex-role stereotyping is used, no a priori position regarding the relationship between masculinity and femininity is implied. In fact, the relationship may vary from person to person and from trait to trait, since it is not dictated by the measuring technique.

Finally, since the measure of sex-role stereotyping presented here is based upon Bayesian probabilities, it is possible that this measure may be more closely related to actual behavior toward women than self-image measures of sex-role stereotyping would be. Mischel (1973) has argued that expectancies are one of several critical person variables involved in the tailoring of a person's behavior to specific situations. He suggests that such expectancies are typically conditional probabilities, or, in other words, that they are Bayesian probabilities. Therefore, a procedure which directly assesses a subject's probabilities about the uniqueness of women, as does a sex-role stereotype measure based on the McCauley and Stitt (1978) procedures, should predict behavior toward women.

Given these differences and possible advantages in using the McCauley and Stitt (1978) procedures to measure individual differences in sex-role stereotyping, several specific questions were framed regarding the use of the procedure in this context. First, would subjects use Bayes rule appropriately in responding to the measure? Unless this condition is met, the arguments for the uniqueness of the procedure are pointless. Second, can a reliable measure be developed using these techniques? Third, how would the stereotype of females developed on the basis of the new procedure compare to the stereotypes derived from existing measures? Fourth, would the new measure show evidence for validity as a measure of sex-role stereotyping? The study presented here addresses the first three of these questions.

Development of the New Measure

The 60 items from the Bem Sex-Role Inventory were selected for use as an original item pool because they are well researched, they do not impose bipolarity upon the subjects, and they include supposedly "sex neutral" items as a control. Moreover, the use of items from a well-known measure allowed for assessing the overlap between a group stereotype and the stereotype developed on the basis of individual differences in perceived uniqueness of women.

Bem's (1974) 60 items were divided into 6 sets of 10 items in sequential order from Bem's scale. This insured that each of the 6 preliminary item sets would have an approximately equal number of masculine, feminine, and neutral items based upon Bem's (1974) ratings. Each adjective was then used as the basis for three questions: "What percent of all people are . . . ?", "What percent of women are . . . ?", and "Of the people who are . . . what percent are women?". A final question asked the subject for an estimate of the percent of women in the world's population.

Each item was scored following McCauley and Stitt (1978) by computing the likelihood ratio for that item. The likelihood ratio is computed by dividing the percentage of women showing the trait by the percentage of all people showing the trait. Although it was not used in the present study, a person's overall score on the measure is conceptualized as the average of the item scores, or the average likelihood ratio.

Use of the likelihood ratio as a scoring procedure allows for identification of those traits that the subject sees as uniquely descriptive of women, and, when averaged across a group of homogeneous items, provides a direct measure of the extremity of the person's stereotyping on that dimension. It is important to recognize that in this procedure items are not simply assigned a fixed weight based on the likelihood ratio. Rather, the likelihood ratio is calculated from the appropriate percentage estimates for every item for each subject, thus avoiding the problem of mere endorsement of a group perception.

The task was presented to subjects as a task in estimation of actual percentages, and accuracy of estimates was emphasized in the instructions. Two sets of instructions were used. The difference involved the instruction for the third estimate, namely "Of the people who are . . . what percent are women?"

Initially this estimate was introduced by the example "What percent of all Chevrolets in the world are in America?", and in the instrument the items were worded "Percent of . . . people who are women." Subjects given these instructions frequently commented that what was wanted in this third estimate was unclear, especially in not being sufficiently distinct from the second estimate "Percent of women who are . . . ". For this reason, the example for the third estimate was changed to read "Of all the Chevrolets that have been manufactured, what percent of them are still in America?" The form of the items was correspondingly changed to "Of all the people in the world who are . . . what percent are women?"

Subjects in this initial study were all undergraduate volunteers from introductory psychology classes who received class points for their participation. Approximately 15 subjects responded to each set of 10 items under each set of instructions. The actual sample sizes are shown in Table 1. The sample receiving the original instructions consisted of 42 males and 53 females across the six subsets of items while for the sample receiving the revised instructions the figures were 49 males and 36 females.

The data were analyzed in two distinct ways. First, to provide information on the use of Bayes Rule in responding to the measure, the average across subjects of each of the three percentage estimates for each item was found. The average estimated percentage of women in the world was also found, and the predicted percentage of trait given sex was computed from the other averages for each item. Correlations were then computed among these variables for each sample, as well as across samples, using the 60 items as the sample size since the data had already been collapsed across subjects.

Second, a cumulative homogeneity analysis (Fiske, 1971) was performed on the data from each of the six subsets of items from each sample separately to provide information on the psychometric quality of the new measures. The homogeneity analyses were performed twice, once on all items in each subset and once on a selected group of items from each subset. Items were selected for inclusion in the second analysis on the basis of average likelihood ratio. Items with likelihood ratios of 1.20 or more in both samples or .80 or less in either sample were chosen. Items with more extreme likelihood ratios such as these are more uniquely characteristic of women in the judgment of the subjects than items with likelihood ratios closer to 1.00. Items selected in this way should be a homogeneous group because they share the quality of being judged as uniquely characteristic of women. The scoring of items with likelihood ratios of .80 or less was reversed by taking the inverse of the observed likelihood ratio. This conversion was necessary to make item scoring comparable among items before analyzing for internal consistency.

Results

The correlations among the various estimated percentages are relevant to the question of whether the subjects used Bayes Rule in responding to the measures. The correlations were all substantial and significant at the .01 level in both samples. The correlations of percent trait given sex with base rate (percent trait) were .62 and .65 and the correlations of percent trait given sex with representativeness (percent sex given trait) were .89 and .92. The correlation between the direct estimate of percent trait given sex and that same percentage as calculated from the other estimates using Bayes Rule was .92 in both samples. These findings are summarized in Table 2.

Correlations of each of the estimates across the samples were also computed as shown in Table 3. The correlations for percent trait (.71), percent trait given sex (.81), and percent sex given trait (.86) were all large and significant at the .01 level. The correlation of average likelihood ratios (item scores) across the two samples was also significant but smaller being only .61. The smaller size of this correlation coupled with the change in the instructions between the two samples prompted use of the conservative strategy of not combining the two samples in performing the homogeneity analyses on each of the six subsets of items.

The results of the cumulative homogeneity analyses on all items in each of the six subsets are shown in Table 4, and the comparable results on the subsets of items selected as uniquely descriptive of women are given in Table 5. The average intercorrelation among items (r_{ii}) gives an indication of the commonality among items based on an estimate of the amount of true score variance in the typical item (Nunnally, 1978). Although there are the expected sampling variations from item set to item set, since both items and subjects vary across sets, r_{ii} ranged from .01 to .29 under the original instructions and from .11 to .57 under the revised instructions in the unselected sets of 10 items. In the sets of selected items, r_{ii} ranged from .00 to .29 under the original instructions and .03 to .46 under the revised instructions. The average of r_{ii} across groups of items was .12 for the original and .35 for the revised instructions in the unselected item sets. For the sets of selected items the average r_{ii}'s were .14 for the original and .18 for the revised instructions. The revised instructions showed greater internal consistency on both the unselected and selected item sets. Item selection produced a small increase in r_{ii} under the original instructions but a decrease under the revised instructions.

The average intercorrelation among persons (r_{pp}) gives an indication of the commonality in subjects' perceptions of the items. In other words, it reflects the consistency with which subjects sort items (Fiske, 1971). As with r_{ii}, there are wide sampling variations in r_{pp}. The average of the r_{pp}'s in the unselected item groups was .23 and .21 for the original and revised instructions, respectively. For the selected item sets these averages were .23 and .16.

Of the 32 items selected as uniquely descriptive of women on the basis of likelihood ratios, 11 were items that are listed as feminine in Bem's (1974) scale. The remaining 21 items were listed as either neutral or masculine in Bem's scale. Twelve items that were sex-role neutral on Bem's scale showed likelihood ratios greater than 1.19 in the present data. Among the nine male items selected by likelihood ratio, 7 showed likelihood ratios of .80 or less. That is, these items were rated as uniquely atypical for women. Two male items from the Bem scale, however, showed likelihood ratios greater than 1.19, indicating traits uniquely typical of women.

Discussion

The data suggest that the basic assumption underlying the use of the McCauley and Stitt (1978) procedure is being met and subjects do use Bayes Rule appropriately in their responses. The correlations of the estimated percentage of trait given sex with both base rate, the overall frequency of the trait in the population, and with representativeness, the frequency of a particular sex showing a specific trait, were all large and significant in both samples. As McCauley and Stitt (1978) point out, the logic of Bayes Rule demands that both types of information be taken into account in arriving at a conditional probability, since by Bayes Rule $P(A/B) = (P(A)*P(B/A))/P(B)$. The correlation of the subject's estimated percentage of trait given sex and the percent of trait given sex calculated using Bayes Rule was also large and significant in both samples, lending further support to the conclusion that subjects used Bayes Rule properly in responding to the items.

The results of the cumulative homogeneity analyses indicate that the items, when scored by the likelihood ratio method, have a sufficient degree of commonality to produce a reliable measure. The observed average intercorrelations among items suggest that between 12 and 35% of the variance in the items is common variance. Given a test of 30 items from this pool of items, these figures mean thar r_{tt} would range from .81 to .95. The logic of test construction used here suggests that this common variance should tap a dimension of perceived uniqueness of women. Conclusions in the area must, of course, await validational evidence.

If Bem's (1975) items constitute a heterogeneous pool with regard to the uniqueness of women, as seems likely given the inclusion of masculine and neutral items in that pool as well as the evidence for the multifactorial nature of those items (Waters, Waters & Pincus, 1977; Feather, 1978; Bohannon & Mills, 1979; Kimlicka, Wakefield & Friedman, 1980), then selection of items with extreme likelihood ratios should increase the commonality among selected items above that in the total pool. For this reason, greater internal consistency was expected in groups of selected items. A very small increase was seen in the sample obtained under the original instructions, but under the revised instructions a rather substantial decrease appeared. This may have occurred simply because selection of items on the basis of extreme likelihood ratios plus the reversal of scoring on items with low likelihood ratios, which was done only in the selected samples, necessarily limits the variability in item scores. Such limitation of variability would tend to attenuate the intercorrelation among items.

It is difficult to assess the effect of the change in instructions on the reliability of the measures despite the fact that r_{ii} was higher under the revised instructions. The reason for the uncertainty is that the increase in r_{ii} could be artifactual inflation resulting from sampling bias. In the data gathered under the original instructions 56% of the subjects were females while in the sample gathered under the revised instructions 58% of the subjects were males. It is possible that young males have a less differentiated view of women than females have of themselves as a gender. The lack of differentiation in the preponderantly male sample could have increased intercorrelations among items, apart from any effect of the instructions.

The average intercorrelation among persons reflects how consistently people sort items, or the extent to which subjects share common judgments regarding the extent to which a particular trait (item) is uniquely characteristic of women. Overall the r_{pp}'s observed in this work suggest that there is reasonable commonality although far from complete agreement regarding the uniqueness to women of various items. This evidence for commonality in perception of the items reinforces the use of the likelihood ratio as a scoring procedure, since that ratio acts as an item weight giving greater importance in an individual's score to items that he/she rates as more extreme. Scores are more meaningfully comparable across subjects if there is evidence for commonality in interpretation of items as reflected in r_{pp}. The fact that r_{ii} and r_{pp} are similar in size is also encouraging as it suggests that it may be possible to discriminate among many levels of stereotyping using a limited number of clearly distinct items, as implied in the logic of the cumulative homogeneity model (Fiske, 1971).

Finally, it seems clear that a stereotype of women based upon identification of traits that are uniquely descriptive of women, in the sense of occurring with greater or lesser than base rate frequency among women, is not likely to be the same as one developed on the basis of ratings of what is appropriate for women, although a strong correlation is probable. Of the 21 items selected for having likelihood ratios of 1.20 or greater, eleven were more desirable for women according to Bem's judges. Nine items that were sex-role neutral for Bem and two that were judged more desirable for men in her work were seen as uniquely typical of women in the present study. Of the 11 items seen as uniquely atypical of women, 7 were masculine by Bem's ratings, but four were sex-role neutral.

In summary, the preliminary data presented here suggest that a measure of sex-role stereotyping based on estimation of conditional probabilities is viable and offers a significant alternative to group based methods of measurement of sex-role stereotyping. This technique permits assessment of individual differences in the attribution of unique traits to women. Furthermore, the technique shows good promise of reliability and of providing a view of the feminine stereotype that is different from that produced with at least one existing measure, the BSRI in its original form. The next steps are to compare this new measure with the PAQ, and to assess the behavioral validity of this measure of stereotyping.

References

Bem, S. The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 1974, 42, 155-162.

Bohannon, W., & Mills, C. Psychometric properties and underlying assumptions of two measures of masculinity/femininity. Psychological Reports, 1979, 44, 431-450.

Broverman, I. K., Broverman, D. M., Clarkson, R. E., Rosenkrantz, P. S., & Vogel, S. R. Sex-role stereotypes and clinical judgment of mental health. Journal of Consulting and Clinical Psychology, 1970, 34, 1-7.

Feather, T. Factor structure of the Bem Sex-Role Inventory. Australian Journal of Psychology, 1978, 30, 241-254.

Fiske, D. Measuring the concepts of personality. Chicago: Aldine, 1971.

Kimlicka, T., Wakefield, J., & Friedman, A. Comparison of factors from the Bem Sex-Role Inventory for male and female college students. Psychological Reports, 1980, 46, 1011-1017.

McCauley, C., & Stitt, C. An individual and quantitative measures of stereotypes. Journal of Personality and Social Psychology, 1978, 36, 929-940.

McCauley, C., Stitt, C., & Segal, C. Stereotyping: From prejudice to prediction. Psychological Bulletin, 1980, 87, 195-208.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. Psychological Review, 1973, 80, 252-283.

Nunnally, J. Psychometric Theory (2nd Ed.). New York: McGraw Hill, 1978.

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and person, sex-role attributes and their relation to self-esteem and conceptions of masculinity and femininity. Journal of Personality and Social Psychology, 1975, 32, 29-39.

Waters, C., Waters, L., & Pencus, S. Factor analysis of masculine and feminine sex-typed items from the Bem sex-role inventory. Psychological Reports, 1977, 40, 567-570.

Table 1

Samples by Size, Sex and Instruction

for 6 Subsets of Items

	Item Set	Old Instructions			New Instructions		
		M	F	Total	M	F	Total
Adapt	1	5	13	18	10	5	15
Affec	2	4	7	11	4	11	15
MDE	3	6	9	15	8	2	10
Relia	4	6	9	15	7	7	14
Self-Rel	5	11	8	19	11	5	16
Warm	6	10	7	17	9	6	15
		42	53	95	49	36	85

May be that some of the differences are due not only to instructions but to differing sex ratios in each condition.

Table 2

Correlations of estimated percent of trait given sex with base rate, representativeness, and calculated percentage of trait given sex across sixty items for two samples

	Sample 1 (Original Instructions)	Sample 2 (Revised Instructions)
%T w %T/S (Base Rate)	.62*	.65*
%T/S w %S/T (Representativeness)	.89*	.92*
Calculated %T/S w %T/S	.92*	.92*

$^{*}p \leq .01$

Table 3

Correlations of Estimates Across Samples

with Revision of Instructions

%T	.71*
%T/S	.81*
%S/T	.86*
LR	.61*

*$p \leq .01$

Table 4

Average intercorrelations among items and persons under two instructions in 6 sets of 10 unselected items

Item Set	Original Instructions		Revised Instructions	
	r_{ii}	r_{pp}	r_{ii}	r_{pp}
1	.29	.19	.57	.29
2	.14	.12	.34	.28
3	.13	.05	.23	.11
4	.11	.54	.23	.34
5	.07	.29	.23	.15
6	.01	.11	.11	.06
Averages	.12	.23	.35	.21

Table 5

Average Intercorrelations among items and persons under two instructions in 6 sets of selected items

Item Set	# of Items	Original Instructions r_{ii}	Original Instructions r_{pp}	Revised Instructions r_{ii}	Revised Instructions r_{pp}
1	6	.14	.16	.46	.29
2	4	.00	.13	.07	.08
3	5	.29	.00	.27	.00
4	6	.10	.67	.09	.35
5	7	.08	.22	.13	.14
6	4	.23	.09	.03	.07
Averages		.14	.23	.18	.16