DOCUMENT RESUME

ED 128 463 TM 005 656

AUTHOR Shaffer, W. Michael
TITLE The Use of Item-Favorability Data as Evidence of Sex Bias in Interest Inventories.
PUB DATE [Apr 76]
NOTE 13p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Interest Tests; *Item Analysis; Junior High Schools; Response Style (Tests); Sex Differences; *Sex Discrimination; Sex Stereotypes; Statistical Analysis; *Test Bias
IDENTIFIERS Ohio Vocational Interest Survey

ABSTRACT

Item data from two scales of the Ohio Vocational Interest Survey were used to investigate the relationship between item-favorability and sex bias. Item-response data, item sex group interaction data, and item-total score correlational data were examined. It was found that item-favorability was not a suitable criterion for the identification of existing sex bias. Of the three procedures, the item-total score correlation technique appeared to be the most sensitive to the existence of possible sex bias in the interest scales. (Author)

The Use of Item-Favorability Data as Evidence of Sex Bias in Interest Inventories

W. Michael Shaffer

The Psychological Corporation

A paper presented at the annual convention of the National Council on Measurement in Education, San Francisco, April 20, 1976.

The Use of Item-Favorability Data as Evidence of

Sex Bias in Interest Inventories

Objectives of the Inquiry

The purpose of this investigation was to study the appropriateness of using item-favorability data as evidence of sex bias in interest inventories. The large differences which are evident between the mean scores of males and females on certain interest scales may be the direct result of differences existing in the response styles of men and women or they may be indicative of existing sex bias (National Institute of Education, 1974). The literature, however, suggests that the results of interest inventories merely reflect basic differences in the interests of men and women (Anastasi, 1958, 1968; Noeth, Roth & Prediger, 1975). It is understandable, therefore, that a recommendation made by the National Institute of Education that the items within a given interest scale should be balanced by sex with respect to favorability has been the subject of considerable controversy (NIE, 1974). It is hoped that the results of this investigation will add new insight into the issues related to use of item-favorability data as evidence of sex bias in interest inventories.

Instrumentation

The data in this investigation were obtained from the administration of two scales, Machine Work and Clerical Work, from the Ohio Vocational Interest Survey. The two scales were selected because the pattern of scores generally obtained by males and females reflect quite dramatically

two widely held sex-role stereotypes. The two OVIS interest scales contain eleven activity statements which were selected to represent typical job activities performed by workers in a specified group of occupations[1]. A five-option Likert response format is used with each item. The response options and corresponding scoring weights are: dislike very much, 1; dislike, 2; indifferent, 3; like, 4; and like very much, 5. The range of possible scores is 11 to 55. The higher the score the greater the person's preference for the job activities associated with the particular cluster of occupations.. As indicated above, the two sex groups perform quite differently on the two scales. For the research sample, the mean score for the males on Machine Work was 13.1 score points higher than the mean score for the females. For Clerical Work, the mean score for females was 9.2 score points higher than the mean score for males.

Methodology

Sample: The subjects used in this study were participants in the national standardization of the Ohio Vocational Interest Survey (OVIS). The sample consisted of 10,225 eighth-grade students from 39 school systems throughout the country[2].

1 Each OVIS scale is based on one or more of the 114 homogeneous worker-trait groups defined by the U.S. Department of Labor in the Dictionary of Occupational Titles, Third Edition.

2 A more detailed description of the sample may be found in Chapter 4 in the OVIS Manual for Interpreting.

Analyses

Three different procedures were used to study the item performance data collected during the OVIS standardization program. The first analysis was an examination of item-favorability. The chi-square statistic was used to test for significant dirrerences between the two sex groups. For each item, a fourfold contigency table (Sex X Favorability) was used to obtain the cell frequencies. For the purpose of the study, a "like very much" response and a "like" response were designated as favorable responses. The remaining three response categories were designated as unfavorable responses.

The second analysis was an examination of the item data for evidence of significant item-sex group interactions. A plotting procedure suggested by Ecternacht (1974) was used for this analysis. The assumption underlying this procedure is that the presence of item-sex group interactions can be used as evidence of possible sex bias.

The final analysis was an examination of the item-total score correlations for the two sex groups. Each student's score on a given item (the item-response weight) was correlated with the student's total score (the sum of the eleven item weights). This is a standard statistical procedure for evaluating item homogeneity and construct validity for psychological scales. Sex differences in the item-total score correlations would suggest that the items in a given scale are functioning in a biased manner with respect to the stated construct. The significance of the difference between the male and female correlation coefficients was tested using the Fisher z transformation (McNemar, 1962).

Results

The item-favorability data are reported in Table 1. It should be noted that all twenty-two chi square values were significant at the .01 level. Thus, with respect to these two sets of items, males and females did not view the items with the same degree of favorability. Furthermore, all eleven activities in the Machine Work scale were found to be favored more by males than females. In the Clerical Work scale, the reverse pattern was found. All eleven activities in this scale were favored more by females than males. Thus, in terms of the NIE Guidelines for Assessment of Sex Bias and Sex Fairness in Career Interest Inventories (1974), the two scales could be considered potentially sex biased.

Figures 1 and 2 provide a graphical presentation of the item-sex group interaction analysis for Machine Work and Clerical Work respectively. The solid, straight line in each graph represents a hypothetical normal distribution of the cummulative item-delta differences ($\Delta_M - \Delta_F$). The item deltas reflect the proportion of favorable responses. The broken, curved lines represent the confidence bands for the .05 level of significance. The points, which are plotted on normal probability graph paper, represent the item-delta differences for males and females for the respective items. As shown in the two graphs, all of the plotted points fell within the designated confidence bands. Thus, the departure from normality was not found to be statistically significant at the .05 level. This, suggests that the item-sex group interaction was not significant. Item-sex group interaction could

be concluded only in the case where the condition of normality had been rejected. The absence of a significant item-sex group interaction, therefore, suggests an absence of sex bias in the performance of the two groups on the respective interest scales, assuming that a constant bias is not inherent in the items themselves.

Tables 2 and 3 contain item-total score correlational data for a random sample of 1000 males and 1000 females drawn from the larger sample of 10,225 eighth-grade students. For the Machine Work scale, the range of the correlation coefficients was .60 to .76 and .56 to .77 for the males and females respectively. The median values were .72 for the males and .73 for the females. The pattern in the Clerical Work scale was similar. The range of coefficients was .50 to .75 for the males and .50 to .82 for the females. The median values were .68 and .71 for males and females respectively.

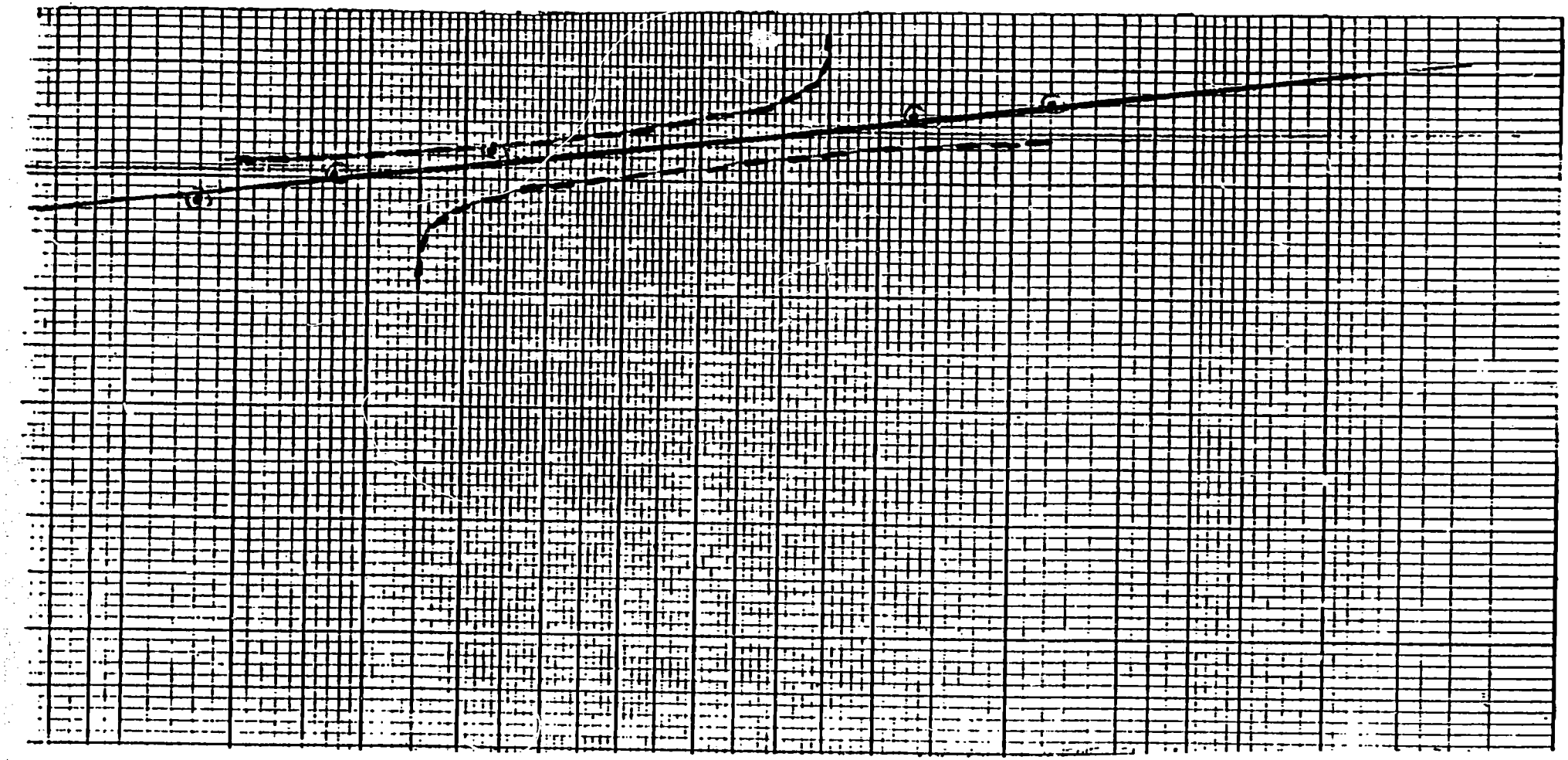
As shown in Table 2, the differences between the 11 pairs of male and female correlation coefficients were not statistically significant at the .05 level. Therefore, the eleven Machine Work items were found to contain the same statistical properties for both sex groups with respect to stated construct. In Table 3, however significant differences were found for five of the eleven pairs of correlations. This suggests that these five items may be functioning differently for males and females with respect to the construct identified as Clerical Work. However, the median correlation coefficients showed no significant difference overall. Thus, further analysis appears to be necessary to determine if the small, but significant differences have practical implications.

Conclusions and Implications

Although the item-favorability data suggested possible sex bias in the two interest scales, consistent support for this position was not found in the other item data. The item-sex group interaction analysis failed to provide any evidence of existing item-sex group interaction. The item-total score correlation analysis did detect a possible source of sex bias in the Clerical Work scale but no evidence of bias was found in the Machine Work scale. These findings raise serious questions regarding the appropriateness of using item favorability data as evidence of possible sex bias. While additional investigations will be necessary to confirm these findings, the conclusion reached at this time is that item favorability data by themselves do not appear to be valid indicators of sex bias in interest inventories.

These findings help to point out the need for a workable operational definition of sex bias as the term applies to interest measurement. The existing definitions, while politically expedient, tend to be a curious mixture of emotional subjectivity and psychometric objectivity. As a result, the available guidelines lack the specificity which would make them useful criteria for assessing sex bias in interest inventories.

gure 1. A Plot of the Item-Sex Group Interaction Data for Scale 2, Machine Work.



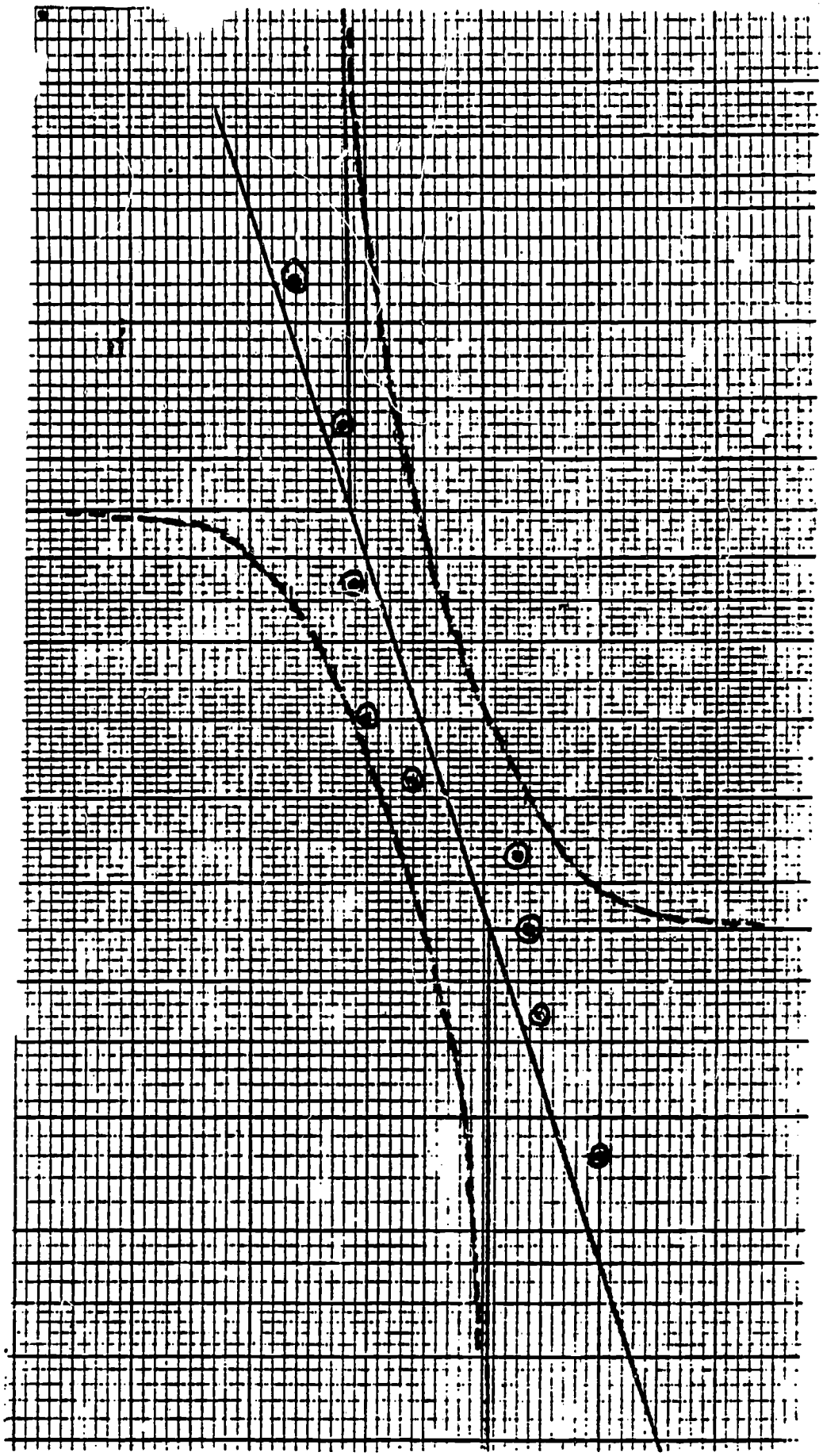


Figure 2. A Plot of the Item-Sex Group Interaction Data for Scale 5, Clerical Work.

Table 1. Summary of Item Favorability Data for the Machine Work and the Clerical Work Scales on the Ohio Vocational Interest Survey.

Manual Work				Clerical Work			
Item #	% Male Favored Responses	% Female Favored Responses	Chi-Square	Item #	% Male Favored Responses	% Female Favored Responses	Chi-Square
9	22	1	1053.9*	1	14	27	266.1*
23	37	2	1970.9*	31	6	47	2224.9*
98	23	1	991.9*	45	10	23	315.2*
109	24	2	1000.0*	55	8	44	1735.1*
130	29	1	1510.1*	100	7	31	964.6*
157	23	1	1116.7*	112	8	42	1588.2*
219	27	2	1235.9*	133	8	44	1735.1*
239	23	1	1079.1*	164	6	37	1468.6*
243	18	1	810.5*	200	9	23	375.0*
268	22	1	1610.3*	202	22	30	85.1*
256	32	1	1053.9*	261	8	41	1516.6*

* Statistically significant at or beyond the .05 level

Table 2. Comparison of the Item-Total Score Correlations for Scale 2, Machine Work.

Item No.	Males		Females		Zm-Zf	Sign Level
	r	≠	r	≠		
9	.60	.693	.56	.633	.058	N.S.
23	.61	.709	.60	.693	.016	N.S.
98	.72	.908	.69	.848	.060	N.S.
109	.74	.973	.75	.950	.023	N.S.
130	.74	.950	.77	1.020	.070	N.S.
157	.71	.887	.74	.950	.063	N.S.
219	.76	.996	.75	.973	.023	N.S.
239	.71	.887	.73	.929	.042	N.S.
243	.72	.908	.73	.929	.021	N.S.
256	.76	.996	.74	.950	.046	N.S.
268	.72	.908	.71	.887	.021	N.S.
Median	.72	.908	.73	.929	.021	N.S.

Table 3. Comparison of the Item-Total Score Correlations for Scale 5, Clerical Work.

Item No.	Males		Females		Zm-Zf	Sign Level
	r	z	r	z		
1	.50	.549	.50	.549	.000	N.S.
31	.54	.604	.57	.648	.044	N.S.
45	.57	.648	.62	.725	.077	N.S.
55	.69	.848	.75	.973	.125	.05
100	.68	.829	.75	.973	.144	.05
112	.75	.973	.82	1.157	.184	.05
133	.73	.929	.78	1.045	.116	.05
164	.72	.908	.71	.887	.021	N.S.
200	.66	.793	.64	.758	.035	N.S.
202	.64	.758	.64	.758	.000	N.S.
261	.70	.867	.76	.996	.129	.05
Median	.68	.829	.71	.887	.058	N.S.