

1977


Examiner Sex Bias and Wechsler Intelligence Scale for Children Scores

Richard H. Dana
Portland State University

Richard Back

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/rri_facpubs

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Psychiatry and Psychology Commons](#)

Citation Details

Dana, Richard H. and Back, Richard, "Examiner Sex Bias and Wechsler Intelligence Scale for Children Scores" (1977). *Regional Research Institute*. 8.

https://pdxscholar.library.pdx.edu/rri_facpubs/8

This Technical Report is brought to you for free and open access. It has been accepted for inclusion in Regional Research Institute by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

(Extended Report)

Back, R. & Dana, R. H. Examiner sex-bias and Wechsler Intelligence Scale for Children Scores. Journal of Consulting and Clinical Psychology, 1977, 45 (3), 500.

Abstract

The WISC (excluding Mazes and Digit Span) was administered to 64 third-grade children, randomly selected for IQ. Eight examiners, four males and four females, each tested eight children, four boys and four girls. Full Scale IQ scores, as well as Verbal IQ, Comprehension, Similarities, and Vocabulary scores, were significantly affected by sex of examiner. Female examiners elicited higher performance from all examinees, regardless of sex. The implications for such findings, particularly in elementary school education, are discussed.

Examiner Sex-Bias and WISC Scores

The dramatic empirical recognition of experimenter-bias occurred little more than a decade ago. It was noted at that time that male experimenters were more friendly than females (Rosenthal, Persinger, Mulry, Vikan-Kline, & Grothe, 1964a, 1964b). Clinical researchers have been concerned with demonstrating the magnitude and situational relatedness of the sex-bias in assessment. For example, in intelligence testing, female examiners have been credited with eliciting higher test scores on the Stanford-Binet and the WISC (Back, 1975; Cleutat, 1965; Gillingham, 1970; Quereshi, 1968), although one study found equivocal effects (Cleutat & Flick, 1967).

Research has also attended to questions of examiner experience and age of subjects as possible additional sources of influence upon intelligence test scores. Four of five Stanford-Binet studies reviewed did not discern differences related to experience in administration or scoring (Sattler & Theye, 1967). However, testing styles, a byproduct of experience, tended to minimize sex bias present in the first half of a research process (Smith & May, 1967), although test-retest scores were affected by the sheer number of administrations in one unpublished study cited by Sattler and Theye (1967). Finally, younger children appear more responsive to experimenter-bias (Barbe, 1965; Stevenson & Allen, 1964), especially with women experimenters (Stevenson, 1961). The present study was designed to explore in greater detail the examiner-sex child-age sources of bias.

Method

The WISC was administered to 64 third grade children from two middle class elementary schools, 32 from each one, randomly selected from the 110 children who returned signed parental permission slips (return rate of 69%). Males and females were equally represented and their ages ranged from 8 years 4 months to 9 years 10 months, with 8 years 11 months being the mean age. All children were Caucasian and representative of the local population.

Eight undergraduate psychology and sociology majors enrolled in a diagnosis course were examiners. There were five seniors, two juniors, and a sophomore, ages 19 to 32. The examiners were selected by the instructor from a class of twelve after reviewing five WISC protocols each examiner had administered as a course requirement. Before the experiment began, the examiners were given feedback on these five previously administered WISC protocols. Each examiner was then observed upon his/her first experimentally administered WISC. Feedback was provided by the instructor on this initial test (as to scoring and instructions via the WISC manual).

The WISC was administered to each subject (excluding Mazes and Digit Span). Each examiner tested eight children, four from each school, two boys and two girls. The order of administration was random with the restraint that the gender of the subjects had to alternate. If a scheduled child was absent, the examiner randomly selected a child of the appropriate sex from an alternate list. Scheduling was done for the convenience of the examiners. Testing was completed on 13 consecutive school days.

A t -test was conducted to determine if any difference existed between

subjects according to schools. Appropriate main effects, i.e., sex of examiner and subject, and VIQ-PIQ, as well as the nested factor, subtests, and all interactions, were analyzed by a factorial analysis of variance (ANOV). Appropriate post hoc tests were made and individualized error terms were used.

Results

Some scoring errors by the examiners were detected. Only computational errors were corrected, while interpretive errors were allowed to stand. There were no differences in IQ scores between students in the two schools ($t = .547$, $df = 62$). The examiner nested within the sex of examiner effect was not significant ($F = 1.49$, $df = 6/48$).

There was a significant main effect for sex of examiner ($F = 4.98$, $df = 1/60$, $p < .05$). Female examiners obtained an average Full scale IQ of 109.3 compared to that of 102.1 for male examiners. No such effect was present for sex of subject. The interaction between these two factors was not significant. There was also a main effect for VIQ-PIQ ($F = 10.72$, $df = 1/60$, $p < .01$). For all subjects, average PIQ was 107.8, while average VIQ was 102.8. The interaction between female examiners and VIQ was significant ($F = 7.89$, $df = 1/30$, $p < .01$). Female examiners obtained average VIQ of 107.1 while male examiners obtained an average VIQ of 98.6. Neither the interaction between the sex of subject and VIQ-PIQ nor between sex of subject, sex of examiner, and VIQ-PIQ was significant.

The nested factor, subtests within VIQ-PIQ, was significant ($F = 6.78$, $df = 8/480$, $p < .001$), as were the related factors, sex of examiner by subtests

within VIQ-PIQ ($F = 2.64$, $df = 8/480$, $p < .05$), and sex of subject by subtests within VIQ-PIQ ($F = 2.11$, $df = 8/480$, $p < .05$). A post hoc ANOV of the last two factors indicated that the sex of examiner was significant on the Comprehension subtest ($F = 8.98$, $df = 1/60$, $p < .01$), the Similarities subtest ($F = 8.79$, $df = 1/60$, $p < .01$), and on the Vocabulary subtest ($F = 7.19$, $df = 1/60$, $p < .01$).

Examination of these findings indicated that the female examiners obtained significantly higher scores on all three of these subtests. On Comprehension, they got an average score of 10.7 versus 8.5 for the males. On Similarities, the averages were 13.2 (female examiners) and 11.1 (male examiners). Finally, on Vocabulary, female examiners elicited an average score of 11 compared to that of 9.1 for male examiners.

The sex of subject by subtests within PIQ-VIQ was relevant only on the Coding subtest ($F = 8.27$, $df = 1/60$, $p < .01$). Female subjects averaged 12.7 on this subtest, while the males' average was 10.8. The final nested factor, sex of examiner by sex of subject times subtests within VIQ-PIQ, was nonsignificant.

Discussion

A significant female examiner effect was present in this study. Female examiners elicited higher scores on the Comprehension, Vocabulary, and Similarities subtests, which were responsible for higher Verbal and Full Scale IQ scores. Quereshi (1968) accounted for similar results by pointing out that the dominant role played by female adults to children (i.e., originally by mothers, then by elementary school teachers) may significantly influence

the development of verbal proficiency.

The subjects differed by sex in only one area, Coding: females did better than males. This may be due to a combination of maturational lag in the boys and an environment which emphasizes good penmanship and neatness for the girls.

The significant VIQ-PIQ difference for these experimental children with PIQ being greater, may be a population specific characteristic for these rural Arkansans. Another possible reason for this difference could be the .72 inter-correlation between PIQ and VIQ which would permit variability for children at this age. These two factors combine to account for PIQ being greater than VIQ.

These findings suggest that WISC IQ scores should be considered in conjunction with the sex of the examiner. A seven-point, sex-influenced fluctuation could significantly effect the evaluation of the Special Education and the Learning Disabled child. Also, the PIQ-VIQ difference typically existing in these rural children, not only affects Learning Disability diagnoses per se, but is exacerbated by examiner sex-bias. Examiners do need to be familiar with testing population in order to minimize avoidable mistakes.

Two statistical considerations, the standard error of measurement, and the reliability of differences are relevant to this seven-point sex-influenced effect. The standard error of measurement for the Full Scale IQ is 3.19, less than half of this seven-point spread (Manual). The present findings regarding the Vocabulary, Similarities, and the Comprehension subtests suggest that any clinical lore concerning these three subtests should also consider the sex of

examiner. Differences between these subtests and others should not be taken seriously unless there is at least a five-point difference instead of the traditionally accepted three-point difference. The above factors may influence not only the exceptional child's future, but average children as well since these fluctuations in IQ scores are examiner determined. Female examiners can render a "below average" student "average" while male examiners can produce a "below average" score for an "average" student.

Future research should pinpoint relationships between sex of examiner and characteristics of the child including IQ, social class, and age. These data can be utilized to formulate more specific scaled scores for the WISC, i.e., separate sex of examiner scaled scores and subject social class scaled scores. If this is too difficult a task, then the nature of the sex differences for examiners can be discerned and future diagnosis courses could attempt to minimize these biases through appropriate instruction.

References

- Barbe, W. The influence of examiner-child interaction on the measurement of giftedness. Gifted Child Quarterly, 1965, 9, 145-148.
- Cieutat, V. J. Examiner differences with the Stanford-Binet IQ. Perceptual and Motor Skills, 1965, 20, 317-318.
- Cieutat, V. J., & Flick, G. L. Examiner differences among Stanford-Binet items. Psychological Reports, 1967, 21, 613-622.
- Sattler, J., & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 68(5), 347-360.
- Smith, H., & May, T. Individual differences among inexperienced psychological examiners. Psychological Reports, 1967, 20(3), 752-762.
- Stevenson, H. W. Social reinforcement with children as a function of CA, sex of experimenter, and sex of subject. Journal of Abnormal and Social Psychology, 1961, 63, (1), 147-154.
- Stevenson, H. W., & Allen, S. Adult performance as a function of sex of experimenter and sex of subject. Journal of Abnormal and Social Psychology, 1964, 68, (2), 214-216.
- Wechsler, D. Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.