INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format. *

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format. *

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.
EXAMINER DISABILITY, EXAMINER GENDER AND EXAMINEE GENDER AS POTENTIAL SOURCES OF BIAS IN THE ADMINISTRATION OF SELECTED SUBTESTS OF THE WAIS-R
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages √
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received √
16. Other

____________________________________

University Microfilms International
EXAMINER DISABILITY, EXAMINER GENDER AND EXAMINEE GENDER AS POTENTIAL SOURCES OF BIAS IN THE ADMINISTRATION OF SELECTED SUBTESTS OF THE WAIS-R

by

Susan Lynn Voskuil

A Dissertation Submitted to the Faculty of the Department of Rehabilitation In Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY In the Graduate College THE UNIVERSITY OF ARIZONA

1986
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Susan Lynn Voskuil entitled Experimenter Disability, Examiner Gender and Examinee Gender as Potential Sources of Bias in the Administration of Selected Subtests of the WAIS-R.

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director

Date
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his/her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
ACKNOWLEDGEMENTS

I would like to acknowledge the following people for their assistance during graduate school:

Jim Merten – Thank you, my dear husband, for your financial and emotional commitment to me including my goal of completing graduate school. I am thankful for both the uncommon commitment we have to one another’s growth and the unique value system we share.

Jule, Starr, and Sky Merten-Voskuil – Thank you, special children, for experiencing all of graduate school with me. You have made many sacrifices for me, but I feel you have also learned a lot. I think we have all become better persons because of my school experience.

John and Dorothy Voskuil – I want to thank my parents for always believing in my ability to complete school successfully, and for giving me, frequently, the confidence I needed to continue. You are wonderful parents.

Ellen Van Wyhe – Thank you for being a special sister and friend.

Inez Tucker – A special thank you to my committee chairperson, rehabilitation specialist and mentor, Inez Tucker. Without your invaluable knowledge, expertise and
emotional nurturance, I could not have successfully completed graduate school. You were always available and always ready to help with whatever I needed. Your particularly competent grasp of not only many academic topics, particularly rehabilitation principles, but also of the Ph.D. process and the many special needs of students involved in that process (especially older ones with families), is unique and outstanding. You are very dear to me, and Sky and I, in particular, love you a lot.

Amos Sales, Marlene Bence, Bill Downey, Reed Mencke, Bill Thweatt and Hal Arkowitz - Thank you committee members (both past and present) for being part of a smoothly functioning committee. I have appreciated your support and willingness to be a part of my graduate school experience.

Kathy VanDeventer, Anne Wagner, Rhonda Hamacher, Barbara Galehouse, Grace Jasin, Miquela Rivera and Teresa Kinsey - A friend is someone who I know, no matter where she/he is on the planet, has a regard for, and a commitment to, my well being. Thank you all for that commitment, particularly during graduate school. A special thank you to Miquela, Grace and Teresa for being an integral part of my life throughout the internship, dissertation and after.
Marc Linzer and Marla Perry-Simmons - Thank you for all of the hours you helped with my dissertation. You worked diligently and did a great job. I will never forget your special willingness to help me.

Randy Jones - Thank you for your willingness to assist me with my research. Your invaluable aid was of particular help.

Jessie Fryer - Thank you for your professional word processing assistance. I particularly appreciated your flexibility in working with me.

I want to end not with more thank yous but with a brief statement of a part of my value system that is particularly relevant to my graduate school experience:

Success is largely a matter of personal viewpoint, though it is generally impossible for a person to fail permanently, if she/he is determined to succeed. But what is success and who are successful people? From my vantage point, successful people are those who tend to consider all of life, even the parts that are difficult, to be a privilege. Successful people take a chance, work for satisfaction rather than credit, are committed to something or someone larger than themselves, keep their word (remembering who they are and what they are about at all times) and respect the other person.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>x</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>6</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE</td>
<td>8</td>
</tr>
<tr>
<td>Introduction</td>
<td>8</td>
</tr>
<tr>
<td>Examiner Race</td>
<td>10</td>
</tr>
<tr>
<td>Examiner Gender</td>
<td>17</td>
</tr>
<tr>
<td>Examinee Gender</td>
<td>21</td>
</tr>
<tr>
<td>Examiner Expectancies</td>
<td>23</td>
</tr>
<tr>
<td>Examiner Personality Characteristics</td>
<td>27</td>
</tr>
<tr>
<td>Examiner Verbal Behavior</td>
<td>32</td>
</tr>
<tr>
<td>Examiner Nonverbal Behavior</td>
<td>37</td>
</tr>
<tr>
<td>Examiner Administration</td>
<td>40</td>
</tr>
<tr>
<td>Naturally Occurring Examiner Differences in Scoring and Administration</td>
<td>42</td>
</tr>
<tr>
<td>Disability and Related Areas of Examiner and Examinee</td>
<td>44</td>
</tr>
<tr>
<td>3. THE STUDY</td>
<td>58</td>
</tr>
<tr>
<td>Research Question</td>
<td>58</td>
</tr>
<tr>
<td>Null Hypotheses</td>
<td>58</td>
</tr>
<tr>
<td>Sample and Description of Subjects</td>
<td>60</td>
</tr>
<tr>
<td>Research Instrumentation</td>
<td>61</td>
</tr>
<tr>
<td>Data Collection (Procedure)</td>
<td>62</td>
</tr>
<tr>
<td>Design and Variables</td>
<td>64</td>
</tr>
<tr>
<td>Treatment of the Data</td>
<td>66</td>
</tr>
<tr>
<td>Assumptions of the Study</td>
<td>66</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>66</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS--Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. RESULTS AND DISCUSSION.</td>
<td>68</td>
</tr>
<tr>
<td>Introduction</td>
<td>68</td>
</tr>
<tr>
<td>Results</td>
<td>68</td>
</tr>
<tr>
<td>Discussion</td>
<td>75</td>
</tr>
<tr>
<td>Conclusion</td>
<td>80</td>
</tr>
<tr>
<td>5. SUMMARY AND RECOMMENDATIONS</td>
<td>82</td>
</tr>
<tr>
<td>Introduction</td>
<td>82</td>
</tr>
<tr>
<td>General Summary</td>
<td>82</td>
</tr>
<tr>
<td>The Purpose</td>
<td>83</td>
</tr>
<tr>
<td>The Sample</td>
<td>83</td>
</tr>
<tr>
<td>The Procedure</td>
<td>83</td>
</tr>
<tr>
<td>Statistical Treatment</td>
<td>84</td>
</tr>
<tr>
<td>Results</td>
<td>84</td>
</tr>
<tr>
<td>Recommendations for Future Research</td>
<td>85</td>
</tr>
<tr>
<td>APPENDIX A: PROTOCOL AND INSTRUCTIONS FOR VERBAL AND NONVERBAL EXAMINEE BEHAVIORS DURING WAIS-R SUBTEST ADMINISTRATION.</td>
<td>90</td>
</tr>
<tr>
<td>APPENDIX B: PROTOCOL AND INSTRUCTIONS FOR THE USE OF THE WHEELCHAIR CONCERNING ITS MECHANICAL OPERATION.</td>
<td>95</td>
</tr>
<tr>
<td>APPENDIX C: INFORMED CONSENT FORM.</td>
<td>97</td>
</tr>
<tr>
<td>APPENDIX D: EXAMINER AND EXAMINEE DEBRIEFING.</td>
<td>99</td>
</tr>
<tr>
<td>APPENDIX E: TABLE 3 EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER DISABILITY ON ARITHMETIC SUBTEST.</td>
<td>103</td>
</tr>
<tr>
<td>APPENDIX F: TABLE 4 EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER DISABILITY ON PICTURE COMPLETION SUBTEST.</td>
<td>105</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS--Continued

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Table 5 Examination of Cell Means for Independent Variable of Examiner Gender on Arithmetic Subtest</td>
<td>107</td>
</tr>
<tr>
<td>H</td>
<td>Table 6 Examination of Cell Means for Independent Variable of Examiner Gender on Picture Completion Subtest</td>
<td>109</td>
</tr>
<tr>
<td>I</td>
<td>Table 7 Examination of Cell Means for Independent Variable of Examinee Gender on Arithmetic Subtest</td>
<td>111</td>
</tr>
<tr>
<td>J</td>
<td>Table 8 Examination of Cell Means for Independent Variable of Examinee Gender on Picture Completion Subtest</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>115</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Analysis of variance of Arithmetic and Subtest scores.</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>Analysis of variance of Picture Completion subtest scores.</td>
<td>70</td>
</tr>
</tbody>
</table>
ABSTRACT

The purpose of this study was to investigate bias in the areas of examiner disability, and examiner and examinee gender on test scores of selected subtests of the WAIS-R.

The subjects of this study were 101 non-disabled University of Arizona Introductory Psychology students who were randomly assigned to treatment and control groups. The treatment group was administered the Arithmetic and Picture Completion subtests of the WAIS-R by a wheelchair bound examiner (a non-disabled examiner posing as a disabled examiner). The control group was administered the same subtests by a nonwheelchair bound examiner.

The data were analyzed by two three-way ANOVAs. The three-way analysis of variance tested for significant relationships between main effects of the three independent variables - examiner disability, examiner gender and examinee gender, and the scores of the arithmetic and Picture Completion subtests, and also for two-way and three-way interactive effects of the three variables and the subtest scores.

Results indicated that in terms of bias operating to influence the individually administered subtest scores, only examinee gender was found to have a significant effect, with
subtest scores for male examinees significantly higher than those for female examinees, on both Arithmetic and Picture Completion subtests.

Nonsignificant relationships were found between the other two independent variables of examiner disability and examiner gender, and subtest scores. The presence of a visible, physical disability on the part of the examiner, and the gender of the examiner were not seen to be operating as bias, and were not seen to be impacting on subtest scores.

Due to limitations of the study that included instrumentation, interaction of testing and treatment, interaction of selection and treatment, reactive arrangements and history, and due to lack of supportive research in the area of examinee gender as bias influencing test scores, the findings of this study need to be interpreted with caution, and should be considered speculative at best.
CHAPTER 1

INTRODUCTION

Statement of the Problem

Since the experiments Sherman (1927) performed with infants, much research has been completed to illustrate that examiner characteristics can influence and help determine responses given by subjects in a variety of testing situations. Early research involved animal and human conditioning situations, and later emphasis focused on examiner effects on subjects in projective testing. Sanders and Cleveland (1953) discussed the effect of examiner influence and bias in projective testing in their review of a number of studies which investigated that issue. Currently, research has shifted to the standardized procedures and among them, the individual intelligence test. Egeland (1967) indicates that his review of the literature has clearly shown that the interaction of examiner and subject is able to influence the behavior of subjects in a variety of ways that tend to bias the results. Kintz, Delprato, Metee, Persons, and Schappe (1967) advise that results obtained on individually administered intelligence tests must be considered in relation to the specific
examiner of the test if results are to be valid and interpreted correctly, and they even went so far as to suggest that the removal of the examiner from the testing situation might protect the subject's performance from the examiner's influence. Green (1966) also suggests that examiners are a possible source of constant, as well as random, error in intelligence testing.

It is evident that much discussion and research has been devoted to the examination of examiner influenced error in testing. Examiner influenced error is thought of as examiner bias operating to negatively or positively impact on the performance of the person (examinee) taking the test. Examiner bias includes characteristics of the examiner and his/her verbal or nonverbal behavior, in the testing situation, exerting an influence on the performance of the person taking the test, and thereby, influencing the test results or scores. Three other sources of error also can be operating in the administration of a test and these include examinee bias, scoring bias and test environment bias. Examinee bias includes any characteristics of the examinee and his/her, verbal or nonverbal behavior, in the testing situation exerting an influence on the examinee's performance and hence, influencing test scores. Scoring bias includes any error made in the scoring of the test results,
and test environment includes any characteristics of the test environment that positively or negatively impact on the test results.

Although a myriad of error variables have been investigated to determine the significance of their impact on intelligence test scores, examiner bias has not been examined thoroughly in some areas. One area that has been neglected concerns the physical attributes of the examiner. More specifically, the area of disability, and its relationship to the topic of bias has been virtually ignored. A research question not addressed thus far, concerns whether or not the presence of a visible, physical disability on the part of the examiner and the variables of examiner gender and examinee gender would impact on behavior in a testing situation and thereby influence test results. This study addressed the current need for more research in the area of disability and bias. This study was concerned with the aspect of examiner bias as it concerns disability and gender, and the aspect of examinee bias as it concerns gender, and how these impact on test scores of the individually administered intelligence test.

**Significance of the Problem**

Schlenoff (1974) and Grossmann (1978) investigated the subject of disability existing as a viable factor in the testing situation as it concerns the disabled examinee.
Schlenoff (1974) discussed relevant factors to be aware of in the administration of intelligence tests to the physically disabled. He emphasized the need for examiners to consider carefully the appropriateness of the test instrument to the particular examinee, concerning his/her disability, maturation level, training, social competence and education. Schlenoff (1974) also stressed that lack of awareness of any of these factors may result in test scores reflecting the degree of impairment in one of these areas rather than intelligence per se.

Grossman (1978) also investigated the subject of physical disability and intelligence test scores, and his study indicated that significant differences exist between the Arithmetic subtest scores of the WISC-R obtained for the orthopedically disabled and those obtained for a physically normal control sample. These results point to a need on the part of the examiner to be sensitive to any disabling condition present on the part of the examinee and to be aware that biases may exist, due to this awareness, that may tend to influence test results and interject error into the scores.

Thus far, research on the topic of bias has been investigated only with respect to the physical characteristics of race and gender of the examiner and examinee. Disability, although discussed as a potential source of bias as it concerns the examinee, has never been
investigated to determine its potential significance as an important source of examiner bias.

All examiner and examinee variables need to be examined that might negatively or positively impact on test scores resulting in biased results. Sattler and Theye (1967) contend that increased research on the role of procedural, situational and interpersonal variables operating to influence intelligence test scores is needed.

It is known that many problems inherent in intelligence test situations are magnified when testing physically disabled examinees (Schlenoff, 1974). The appropriateness of the test instrument to the physical condition of the examinee is but one factor that must be more carefully scrutinized when administering a test to a disabled individual rather than a nondisabled individual. Awareness of the disabled examinee's medical, social and developmental history in terms of life experiences, and attitudes toward disability, cannot be overlooked as important considerations for the examiner who is administering a test to a disabled examinee. Schlenoff (1974) suggests that "it is necessary that professionals be attuned to the interference that the physical disability can introduce into the testing situation" (p. 363).

This interference is the error variable of bias, and certainly this particular aspect of bias, presence of a disability in the testing situation, represents a vital, and
heretofore, ignored area for significant research in the area of intelligence testing.

The topics of disability and sociocultural responses to disability, are viable variables present in many testing situations. In order to change biased, prejudicial, attitudinal and behavioral patterns toward the disabled, it is necessary to first identify them in the situations in which they occur, determine their significance, and then plan and implement effective treatment aimed at altering them (Sillers, 1976). No efforts have been made thus far, to identify these patterns in the testing arena of the individually administered intelligence test as they concern the disabled examiner and examinee. Specifically, a need does exist to identify whether an examiner with an obvious physical disability would exert a bias on test scores obtained on an individually administered intelligence test. This study attempted to address that need.

**Definition of Terms**

**Examiner:** The individual who is administering the intelligence test.

**Examinee:** The individual who is being administered the intelligence test.

**Disability:** Disability, although applicable to both visible and nonvisible conditions that impair complete functioning of one or more physiological systems, is used in
this study to represent only overt, physically visible, disabling conditions. All disabled examiners were wheelchair bound. A wheelchair bound disability portrayal was utilized as it lends itself to maximum research control over its portrayal, represents an obvious, common physical condition requiring little or no explanation as to what it is, and requires the simplest of equipment.

**Examiner Influenced Error:** Examiner bias operating to negatively or positively impact on the performance (test scores) of the examinee. Examiner bias includes all characteristics of the examiner and his/her verbal and nonverbal behavior and appearance exerting an influence on the results of the being administered.

**Examinee Influenced Error:** Examinee bias operating to negatively or positively impact on the performance of the examinee. Examinee bias includes all characteristics of the examinee and his/her verbal and nonverbal behavior and appearance exerting an influence on the test scores of the examinee.

**Test Results:** Test scores.
CHAPTER 2

REVIEW OF LITERATURE

Introduction

Many education, mental health and rehabilitation professionals place great credence in scores obtained from individually administered intelligence tests, and these scores have been believed to be reliable predictors of scholastic achievement (Miller & Chansky, 1972) and also helpful tools in vocational planning and job development. If these scores are to be utilized appropriately and in the most effective manner, possible sources of error operating in the testing situation must be identified, and their impact, in terms of bias operating negatively or positively to influence test scores, must be taken into consideration when test scores are interpreted and utilized in assessment. If not, the scores cannot be said to legitimately represent the capabilities and potential of the person to whom they are attributed.

The available literature on the subject of examiner bias operating in individually administered intelligence test suggests that several aspects of examiner bias designated as possible sources of error operating in the
testing situation must be identified, and their impact, in terms of bias operating to negatively or positively influence test scores, must be taken into consideration when test scores are interpreted and utilized in assessment.

An examination of all the relevant literature on the subject of examiner bias operating in the individually administered intelligence test situation suggests that several aspects of examiner bias have been designated as possible sources of error. These aspects include race of examiner, gender of examiner, expectations of examiner, prior knowledge and pretest information available to the examiner, personality and emotional characteristics of examiner, rapport established by examiner, scoring and administration expertise of examiner, and a comparison of examiner administered results versus machine administered results. Characteristics of examinees in the testing situation, such as gender of examinee and race of examinee have also been investigated as possible sources of bias as they impact on test scores. In addition, the physical characteristics of the testing environment and the method of administration have also been researched as possible sources of test bias. These four areas of bias comprise the sources of error that can be attributed to negatively or positively influencing the outcome of test scores in testing. Of
particular interest has been the presence or absence of these four sources of bias as they pertain to results obtained on individually administered inventories, especially the intelligence test.

The purpose of this chapter is to examine the literature and research of examiner and examinee bias in individually administered intelligence tests. The review of the literature will focus on several important aspects of bias as categorized in past studies. These aspects include race of examiner and race of examinee, gender of examiner and gender of examinee, expectations of examiner, prior knowledge and pretest information of examiner, personality and emotional characteristics of examiner, rapport established by examiner, scoring and administration expertise of examiner, and a comparison of examiner administered results versus machine administered results.

Examiner Race

Sattler and Gwynne (1982) reviewed much of the available research on the effects of examiner race on the intellectual performance of black children and suggest that the existence of race-of-examiner effects do not negatively affect the performance of black children. Of 27 published reports on the effects of examiner's race on the test scores
of children taking individually administered intelligence
tests, 23 of these found no significant relationship between
the race of the examiner and the examinee's scores. The
tests involved in all of these studies included the WISC,
WISC-R, WPPSI, S-B, Slosson Intelligence Test, Peabody
Picture Vocabulary Test, Draw-A-Man Test, Iowa Test of
Preschool Development and several other tests of cognitive
ability. The grade level of the subjects ranged from
preschool to high school. The four discrepant studies that
disagreed found that black examiners obtained higher scores
than white examiners when testing black children (Forrester
& Klaus, 1964; Moore & Retish, 1974; Ratusnik &
Koenigsknecht, 1977; Savage & Bowers, 1972). Sattler and
Gwynne (1972) present other studies (Abramson, 1969; Buckey
& Pranta, 1972) in which white examiners, as compared to
black examiners, obtained higher scores from both black
children and white children. The majority of the research
however indicates that race of examiner is not a factor in
the intelligence test performance of black or white
children. Sattler and Gwynne (1972) state, "Despite the
plethora of studies conclusively demonstrating that race of
examiner does not influence the intelligence test scores of
black children...the myth is pervasive in the field of
psychology" (p. 206). They go on to suggest that any racial
differences in the examiner-subject relationship are not likely to contribute substantially to the lowered scores of black children and that because many people do not want to believe that lower intelligence scores of black children reflect the true level of their intellectual functioning, they seek ways to attack this belief. Loehlin, Lindsey and Spuhler (1975) support the contention that it is a relatively consistent finding in research that black students, when compared to white students, tend to score lower on standardized intelligence tests, and Wellborn, Reid and Richard (1973) indicate that research has clearly shown that minority groups score lower on intelligence tests than the white groups used for the standardization samples, and to the white groups with which they are compared. That the difference in scores exists, is clear. Why it exists is not clear. The majority of studies, as indicated, refute the contention that the examiner's race is responsible (Caldwell & Knight, 1970; Jacobs & DeGraaf, 1973; LaCrosse, 1964; Samuel, 1977; Terrell & Terrell, 1981; Wellborn, Reed, & Richard, 1973; Uhl, 1974). Graziano, Varca and Levy (1982) report in their review of the literature on the effects of examiner race and validity of intelligence tests that the current literature offers very little support for the hypothesis that examiner's race, per se, has an effect on a
subject's performance on intelligence tests.

Some research, however, does support the contention that the race of the examiner is a source of bias operating in the testing situation that influences subjects' test scores. In addition to those studies mentioned previously, Solkoff (1972) demonstrated that examiner race had a significant effect on the Comprehension and Picture Completion subtests of the WISC, and on Verbal, Performance and Full Scale IQs, with black examiners producing the highest scores. He also found a significant race of subject and race of examiner interaction on the Information subtest, with white subjects achieving the highest scores with a black examiner, and black subjects the lowest scores with a white examiner. Two years later, Solkoff (1974) again investigated the effects of examiner race and found a significant race of subject and race of examiner interaction on the Similarities and Object Assembly subtests of the WISC, with the black subjects achieving their highest scores with a white examiner. Moore and Retish (1974) found that the variable of examiner race clearly influenced test scores of low income preschoolers, with black examiners eliciting statistically significant higher means on the Verbal, Performance and Full Scale IQs from low socioeconomic status
black children than did white examiners. Implications they stressed include the idea that black's perceptions and attitudes toward white professionals and lay people are negative and that these may have been operating to influence the black subject's test performance in the study. In addition, they suggested that white examiners may have maintained perceptions of blacks which were evident in the manner in which they interacted with the black children in the testing, resulting in children less motivated to perform optimally when the examiner was white. France (1973) became more specific and examined the influence of white and black voices on the IQ test performance of white and black children on the Peabody Picture Vocabulary Test and found no examiner racial voice effect on test scores of black children. However, examiner voice, racially, appeared to be an important variable with regard to the test performance of white children.

The research examined thus far concerns examiner race as a variable only for black or white examiners. But what about other minority groups? Less research has been done with other ethnic examiners and subjects, but some does exist. Mishra (1980) investigated the effects of examiner's ethnicity on the intelligence test performance of Anglo and Mexican American children on the WISC and the Raven
Progressive Matrices. The results demonstrate that on the WISC Vocabulary subtest, Mexican American subjects scored significantly higher when the examiner was Mexican American instead of Anglo. In contrast, the performance of Anglo subjects was not significantly affected by examiner ethnicity. The results of other studies (Bordie, 1970; Palomares & Johnson, 1966) concur with the finding that the test performance of Spanish speaking children is affected by examiner variables such as ethnic group membership and bilingualism. Phillips (1966) and Mycue (1968) suggest that white Anglo American examiners arouse anxiety in minority group children resulting in significantly lower test scores for these subjects. Swanson and DeBlassie (1971) found that Mexican American first graders who were given the WISC by an Anglo examiner, with the assistance of a Spanish interpreter, did not perform significantly better than those who were given the WISC entirely in English. On this issue, Gerken (1978) remarks that,

it has been intelligence testing...that has led to the inaccurate classification and placement of minority group children, due to lower test scores, with many of these children...being placed in classes for the EMR. This method of exploring the intelligence of the Mexican American child has been of little value to that child, and instead, (has) been used to reinforce the present hierarchical arrangement within the public schools. One cannot assume that the low scores of the Mexican American child of these (intelligence) tests are indicative of less highly developed intellectual ability when many important variables (including examiner effects), were left uncontrolled (p. 438-439).
Her study, however, found that examiner ethnicity had no significant effect on the test scores of Mexican American children, but that the type of intelligence test administered (verbal as opposed to nonverbal content) and the child's dominant language did. Spanish dominant children obtained significantly lower means than did bilingual children. IQ scores for Mexican American children were higher on nonverbal scales with nonverbal directions, and on nonverbal scales with verbal directions, than on verbal scales with verbal directions. Results support the findings of Sanson and DeBlassie (1971). In another study, Thomas, Hertzig, Dryman, and Fernandez (1971) found that examiner race did not significantly affect the IQ test scores of Puerto Rican working class children but that other examiner characteristics did.

In keeping with the consensus of most researchers, that the effect of examiner race on intelligence test scores seem dubious, but that contradictory findings contribute to lack of clarity on this issue, Solkoff (1972) suggests that many examiner variables such as amount of testing experience, experience in working with minority group children, and testing approach may exert more of an influence on test results than examiner race alone. He additionally suggests that the racial attitudes and
expectations with minority and nonminority teachers could also be influencing test scores. McElwain and Kearney (1970) urge awareness of the visual and auditory cues transmitted by examiners during test administration, as the subject's performance, especially of those of different ethnic backgrounds, may be being influenced by these cues, whether the person taking the intelligence test is a minority or nonminority group member and whether the examiner is black, white or otherwise. Shuey (1966) and Dreger and Miller (1968) report on the need for further investigation regarding the influence of examiner race on the cognitive functioning of subjects taking intelligence tests. As research indicates contradictory findings, Sattler (1970) concurs, and indicates that the evidence supporting slightly negative effects of examiner race on minority group subject's intelligence test scores is only suggested and more research is certainly needed.

Examiner Gender

A second source of examiner bias often discussed in terms of its effects on the results of intelligence test scores is that of the gender of the examiner. Related to this variable is the gender of the subject and the
interaction of gender of examiner and gender of subject. Our discussion will examine both aspects of this variable.

Of the following research reviewed, only one study offered results supporting the contention that the gender of the examiner does not exert an effect on intelligence test scores. In that study (Cieutat & Flick, 1967), the researchers investigated the hypothesis that examiners elicit higher IQs from subjects of the opposite gender on the Stanford-Binet. Their results did not support this hypothesis, and the expected interaction between gender of the examiner and gender of the subject did not occur. The rest of the research suggests strongly that both examiner gender and subject gender exert significant influences on intelligence test scores. Bittner and Shinedling (1968) demonstrated that at the third grade level, male examiners elicited better performances on Piaget's tasks related to the concept of conservation of substance than did female examiners. Pederson, Shinedling and Johnson (1968) suggest that natural biases resulting from sex role conceptions or expectations, manifesting themselves through examiner effects, significantly affect test results with no experimental manipulation. They investigated the effects of examiner gender on the performance of third grade males and females on the Arithmetic subtest of the WISC. Their results
indicate that female subjects did significantly better when tested by female examiners than by male examiners.

A study by Bach and Dana (1977) found that among results of third graders on the WISC, female examiners obtained average Full Scale IQs of 7.2 points above those obtained by male examiners. Results also indicated a female examiner and Verbal IQ interaction, represented by female examiners eliciting verbal IQs 8.5 points higher than the male examiners. Comprehension, Similarities and Vocabulary were the subtests significantly affected by the gender of the examiner. The researchers suggest that the implications of these results include the existence of a significant impact on evaluations for special education and learning disabilities classrooms, and they indicate future research should more clearly delineate gender of examiner variables from subject characteristics, including IQ, social class and age.

In another study, Samuel (1977) found that the gender of the examiner did have an overall significant effect on junior and senior high school student's performance on the WISC, with students achieving higher Full Scale IQs when the examiner was female, especially a black female. Quereshi (1968) examined the influence of examiner gender and subject gender on the scores of children ranging
in age from five to 14, on the WISC, and found that examiner gender seemed to be an important determinant of Vocabulary subtest scores, Verbal IQ, and the Picture Completion Performance subtest. For the Vocabulary subtest and Verbal IQ, regardless of the child's age, those tested by a female examiner obtained significantly higher scores than those examined by a male examiner. On Picture Completion, the reverse was true, with the male examiners eliciting higher scores. Quereshi offered three reasons for the results including: (1) female examiners are more liberal in scoring Vocabulary items than male examiners, (2) female examiners conduct a larger amount of inquiry than male examiners, and (3) there are genuine differences between males and females in providing an optimally stimulating psychological atmosphere for eliciting responses of certain types on certain WISC subtests. Female examiners showed greater cue value for evoking verbal responses while male examiners were more effective in facilitating performance on those tasks requiring nonverbal responses.

DiLorenzo and Nagler (1968) studied test scores of over 1000 children on the Stanford-Binet, and reported that female examiners elicited higher scores, but did caution that the scores they obtained may have been affected by the examiner's expectancies and by communication, during the
study, among the researchers. Cieutat (1965) also looked at the effect of examiner gender on Stanford-Binet scores, and his results concur with those of DiLorenzo and Nagler (1968). Similarly, Ziv (1972) investigated the effects of examiner gender, subject gender and type of reinforcement given to subjects on the Raven Matrix, and found that when reinforcers (regardless of type), were given by male examiners the male's scores increased, and when reinforcement (regardless of type) was given by the female examiners, the female subject's scores increased.

From the research discussed so far, it seems clear that examiner gender, along with gender of the subject, have shown themselves to be potent variables significantly affecting the test scores on individually administered intelligence tests.

**Examinee Gender**

Twelve studies attempted to examine, among other things, the main effects of gender of examinee as an independent variable on the test scores of various individually administered intelligence tests. Of these 12 studies, nine yielded results indicating no significant main effect for gender of examinee on scores. Of these nine, five examined Full Scale WISC scores (Quereshi, 1978;
Thomas, Hertzig & Dryman, 1971; Solkoff, 1974; Schwartz & Flanigan, 1971; Back & Dana, 1977). Of the remaining four studies, Pederson, Shinedling and Johnson (1968) found no significant main effect for gender of examinee on the WISC Arithmetic subtest when administered to nine-year-olds; Mishra (1972) found no significance for gender of examinee on the WISC Verbal Scales; Ziv (1972) found no significant main effect for gender of examinee on the Raven Progressive Matrices as administered to 13-year-olds and Hersh (1971) found examinee gender not significant on the test results of the Stanford-Binet and Peabody Picture Vocabulary Test when administered to 5- to 7-year-olds.

The remaining three of the 12 studies did yield significant results with respect to main effects of gender of examinee on test results. Delprato and Jackson (1975) found that university Introductory Psychology student female examinees performed significantly better on the WAIS Digit Span and Digit Symbol subtests than their male counterparts. Solkoff (1972) found results indicating that male 8- to 11-year-old examinees performed significantly better on the Picture Completion and Object Assembly WISC subtests than female 8- to 11-year-old examinees. He also found that female 8- to 11-year-old examinees performed significantly better on the WISC Coding subtest than their male
counterparts. And the results of Quereshi (1968) indicate that male 5- to 13-year-old examinees performed significantly better on the WISC Comprehension and Vocabulary subtests than did the 5- to 13-year-old female examinees.

From the research discussed so far, it is apparent that the variable of examinee gender, when investigated as a main effect, does present mixed findings with respect to its impact as a source of influence on the test scores of individually administered intelligence tests.

**Examiner Expectancies**

Another variable, that of examiner expectancies resulting from pretest knowledge of, among other things, IQ, personal data and abilities, has been demonstrated to exert quite a significant examiner bias on test score outcome. Of all the studies reviewed, only one (Mishra, 1980), found no examiner expectancy bias occurring. In that study the effects of examiner's prior knowledge of subject's ethnicity and intelligence, on the examiner's scoring of the Stanford-Binet was investigated. The researcher found that examiner prior knowledge of the subject's ethnicity and intelligence did not have a significant effect on test scores.

Another study utilizing the Stanford-Binet (Hersh, 1971) examined the effects of positive and negative
evaluations in referral reports on examiner behavior and intelligence test results, and found that when subjects were referred positively by teachers to examiners, examiners obtained higher IQ scores than subjects referred negatively. Results also indicated that with a positive referral, examiners started the subjects at a significantly higher year level on the test than with a negative referral. Subjects with a positive referral also were rated by the examiner more favorably on a scale measuring factors affecting test performance, and were given more favorable recommendations by the examiners. No significant differences however, were found in the amount of examiner warmth as influenced by a positive or negative referral. Hersh recommends interpreting the process by which the referral reports exert an influence with some caution, but does suggest, on the whole, that information provided by a referral agent does create an examiner expectancy about the intellectual capacity of the subject.

Research on examiner expectancies and the WISC, has provided similar data to that just discussed. A study by Schroeder and Kleinsasser (1972) found that total WISC VIQ scores are significantly affected by the expectancies, as influenced by pretest information about the bright vs. dull intellectual ability of IQ protocols, of examiners. The
results suggest that the bias occurred before the scoring of the tests. Jacobs and DeGraaf (1973) found that when examiners were led to expect high achievement ability of ten-year-old subjects, scores on the WISC were significantly higher than when examiners were led to expect low achievement, especially on the Verbal subtests. They also found that expectancies were most significantly affected when examiners were testing subjects of their own race. In a study by Egeland (1969), examiners received pretest information consisting of supplemental information attached to each test protocol, indicating raw scores on the Information and Arithmetic subtests, and intellectual achievement information, from other sources, about IQ. The results indicate that the pretest information supplied to these examiners significantly affected scoring on Verbal subtests of Comprehension and Similarities, with high expectancy information eliciting significantly higher scores than low expectancy information, and no information at all.

Results of a study by Babad, Mann and Mar-Hayim (1975) indicate similar examiner expectancy biases on the WISC. In their study, it was found that the scoring of the WISC subtests, excluding Arithmetic and Digit Span, were biased by the examiner's expectations of achievement. The mean scores of the high expectancy subjects were higher than
those of the low expectancy subjects, particularly on the Comprehension subtest.

Other studies investigated the subject of examiner expectancy on WAIS scores. Dickstein and Ayers (1973) found that the manipulation of motivation conditions through explicit examiner's expectancy of good performance significantly improved performance on the PIQ of the WAIS. Sattler and Winget (1970) examined the effect of examiner expectancy and subject's intellectual level on examiner's scoring, testing procedures and test observations, and results indicate that subjects with superior intelligence referral reports, read previously to the examiner's scoring the test, obtained significantly more credit for ambiguous responses on the Comprehension, Similarities and Vocabulary subtests of the WAIS than did subjects with average intelligence referral reports.

Another study by Dickstein and Kephart (1972) examined the effect of explicit examiner expectancy of good test performance on subject's performance on the WAIS. It was found that female college students who were administered the WAIS with a high expectancy performance level did significantly better on the Performance subtests than did female college students who were not. Implications of the results are presented by the researchers with the suggestion
that if assessment of an individual's maximum capability on an intelligence test is the goal, the examiner must make a serious effort to arouse maximum motivation. The effect of the manipulation of instructions (the explicit examiner expectancy in this study) was limited, however, to the Performance subtests. Failure to obtain significant differences on the Verbal subtests could reflect the fact that the verbal skills of a college population are stable and insensitive to situational motivational factors.

Examiner Personality Characteristics

Other examiner variables investigated by various researchers include a broad spectrum of less global, more specific aspects of examiner bias than race, gender and expectancy. A discussion of these will be more brief than the discussion of examiner race, gender and expectancy, as the bulk of the research on examiner bias so far, has been done on race, gender and expectancy. However, the rest of the examiner variables that will be discussed have all been identified as relevant, contributing factors to examiner bias.

The first of the variables to be discussed is that of personality characteristics of the examiner. Included in this discussion will be rapport building, affective and
cognitive aspects of the examiner's personality, and general nonspecific examiner effects.

Sacks (1952) found that preschool children tested by examiners who had established a "good social relationship" prior to testing scored higher on the Stanford-Binet than a comparable group tested by an examiner who had established a "poor social relationship". A good social relationship consisted of prior acquaintance with the examiner and the establishment of rapport. Thomas et al. (1971) did an interesting study aimed at examining the effect of rapport on WISC scores. Specifically, they were interested in the influence of rapport on the manner and style of coping with cognitive demands of lower socioeconomic status Puerto Rican children. The results of their study indicate that examiner differences in rapport building do elicit significantly different levels of performance on standardized intelligence tests of Puerto Rican working class children.

Examiner differences were examined in the ways (informal and formal) examiners initiated and conducted the testing situation. The examiners who elicited higher test scores spent more time with the children prior to testing and conducted the testing in a less formal atmosphere than did the other examiners. An examiner operating in a less formal atmosphere was more spontaneous, gave encouragement
after eliciting an "I don't know" response and was instructed to appear warm, interested and friendly, while providing a game-like testing atmosphere.

The IQ scores elicited by neutral examiners (those who operated in a more formal test environment, who didn't give encouragement, who were not spontaneous and who were less affectively warm) correlated better with reading achievement, as administered by standardized group tests administered in school, than did the test scores elicited by the warm, less formal examiner. The researchers suggested this finding might exist because the procedures used by most classroom teachers and standardized group testing situations most closely resemble the interaction and atmosphere established by the neutral examiner.

Kinnie and Sternlof (1971) looked at examiner factors described as nonintellective that might affect test performance. Nonintellective was defined as factors present in testing situations, and which influence test scores, but which aren't related to the subject's skills or knowledge. In their study, they examined pretest examiner familiarization with the subject as one nonintellective factor and results indicated that on the WPPSI, FSIQ and PIQ can be increased significantly for middle class white children, and lower class white and black children, if pretest examiner familiarization is employed, and also,
pretest familiarization with WPPSI-like items. Contrary to the results just discussed, one study examined (Irons, 1981) did find that pretest examiner familiarity with first and fifth graders who took the WISC, had no significant effect on either Verbal, Performance or Full Scale IQ.

Personality characteristics of examiners have been frequently investigated as an area of examiner bias. Holder, Drasgow and Pierce (1970) examined the effect of the "good guy" examiner role, consisting of interpersonal characteristics of empathy, respect, genuineness, and concreteness in differences in objective task-oriented test situations; specifically, the General Aptitude Test Battery. Results suggest that personality characteristics of examiners (level of interpersonal functioning involving empathy, respect, genuineness, and concreteness) do account for some of the variability in objective test results heretofore attributed to individual differences.

Donahue and Sattler (1971) evaluated the effects of four examiner and subject variables, including hostility, dominance, liking and warmth, on undergraduate college students who were administered the WAIS. Their results, with respect to warmth, failed to support those of Rosenthal (1966) who reported a significant, positive correlation between subject's rating of the examiner's warmth and subject's test scores. Donahue and Sattler's (1971) results
suggest however, that examiner bias is present, in that examiners who liked their subjects and who found them warm gave them more credit than they gave to subjects they liked less and found less warm.

Egeland (1967) investigated whether or not children differing in measured anxiety respond differentially on the WISC to examiners previously classified as high- or low-anxious. Specifically, Egeland wanted to examine whether WISC mean IQ scores were significantly different for children tested by a high-anxious examiner, and whether there was an interaction existing between high- and low-anxious examiners and high- and low-anxious subjects. The results showed that the examiner effect was most apparent on the Verbal scales where low-anxious subjects tested by high-anxious examiners received the highest test scores. These results could be interpreted as confirming the previous findings of Sarason (1963) which indicated that low-anxious subjects were generally superior to high-anxious subjects on measures of intelligence, and also as offering support to the validity of the argument that high-anxious examiners do have an influence on test performance. In Egeland's study (1967) however, the fact that the subjects responded differentially to high- or low-anxious examiners was dependent upon the measured anxiety level of the subjects and the nature of the WISC subtest. The
examiner-anxious effect was found on the Verbal subtests of Comprehension, Similarities and Vocabulary, which are less structured than the Performance subtests, where no examiner effect was found. A possible explanation for these results is that the high-anxious examiner tended to increase the drive level or motivation of the subjects he/she tested, especially the low-anxious ones. A motivational or high-anxious examiner would not increase the motivation of high-anxious subjects to any significant degree since these subjects are generally operating under a high level of anxiety anyway.

**Examiner Verbal Behavior**

Another area of examiner effect that is discussed in the literature is that of type of verbal comments and verbal reinforcement used in administering an individual intelligence test.

Terrell, Taylor and Terrell (1978) found that the use of culturally relevant social information such as "nice job young blood," "good work young soul" and "nice job little brother" improved the performance of black children on the WISC. In a more recent study by the same authors (Terrell, Terrell & Taylor, 1980), attempts were made to compare the use of examiner nonreinforcement, a tangible reward (candy), a traditional social reward ("good", "fine") and a culturally relevant social reinforcement ("nice job
young blood," "good work young soul" and "nice job little brother"), on the short form of the WISC-R. The subjects were 120 black males, nine to eleven years old. Results showed that for a black examiner, traditional social reinforcement was the least effective, with the tangible and culturally social reinforcement equal in raising scores. For the white examiner, the tangible reward was the most effective, and the traditional social reinforcement the least effective.

Piersel (1977) wanted to examine the effects of three types of examiner administration on the WISC-R shortened version test scores of 63 black and Mexican American children. The first type of administration consisted of a situation where subjects were told, in the testing situation, whether their responses were correct, mostly correct or incorrect, and also the amount of points they received for their responses. Subjects also kept track of their own points on a sheet provided by the examiner. In the second type of administration, the subjects viewed a videotape of an examiner giving positive verbal feedback. This was followed by a standard WISC-R, shortened version, administration. The third type of administration consisted of a standard WISC-R administration. Results of the comparison of the test scores elicited in each of the three
types of administrations indicate that a pretest vicarious condition of an examiner giving positive verbal reinforcement elicits significantly higher test scores than the two conditions to which it was compared. Terman and Merrill (1937, 1960) also substantiate the contention that examiners giving verbal approval can significantly increase test scores on the WISC and the Stanford-Binet.

Since Witmer, Bernstein and Dunham (1971) had already demonstrated the sensitivity of four WISC subtests to examiner verbal effect (Arithmetic, Block Design, Picture Completion, and Digit Span), Saigh and Payne (1976) decided to use these four subtests to investigate the effect of examiner verbal comments on test scores elicited from 40 educable mentally retarded students. Their results indicate that two of the four subtests, Block Design and Digit Span, were sensitive to verbal approval examiner conditions, and that praise elicited higher scores these subtests. Saigh and Payne do suggest that given the implications of their work, more research is needed on the topic, especially in the area of Full Scale IQs. Saigh followed his own suggestion on this issue, and in 1981, published a study examining the effects of positive examiner verbal comments on the total WISC-R performance of intellectually superior elementary and secondary school students. He found that the scale scores
for five of the eleven subtests, in addition to Verbal, Performance and Full Scale IQ scores, were significantly higher when positive examiner verbal comments were employed.

Other studies refute the significant effect of examiner praise on intelligence test performance, including one by Isenberg and Bass (1974). They examined the effectiveness of nonverbal and verbal reinforcement on the WAIS performance of college undergraduates. They found no significant difference in scores as a result of type of reinforcement used. They did suggest that the trend of their results, although nonsignificant, indicates that subjects could possibly score better on an intelligence test under verbal and nonverbal reinforcement than they would under test conditions of no reinforcement at all. Another study by Galdieri, Barcikowski and Witmer (1972) also did not find differences between the effects of verbal and no verbal reinforcement on intelligence test scores. They looked at the effect of verbal approval on WISC test scores of 168 low and middle class white children. Results indicate that non significant differences exist between the verbal and no verbal modes of test administration. Ziv (1972), using the Raven Progressive Matrices, found no significant differences between the examiner utilizing praise as a reinforcer and the examiner using blame.
Two additional studies investigated the effect of verbal behavior of the examiner, but not in the sense of specific language used or type of verbal reinforcement. Klingelhofer (1971) administered the Progressive Matrices in both Swahili and English to investigate the effects associated with examiner, the school the child attended and the language in which the test was administered, on test scores. The results indicate that none of the variables had any significant effect on test scores. Swanson and DeBlassie (1979) examined the effects of the use of an interpreter in the administration in Spanish on the performance on the WISC of Mexican American children, as compared with performance on the test when administered in English or in English with a Spanish interpreter. The results indicate that the method of administration of the WISC, in terms of language used, is an important factor in determining the level of performance on the total Verbal phase of the WISC with the WISC in English yielding higher scores than the administration in Spanish only or in Spanish with an interpreter, while the Performance subtests yielded optimum performance when given in Spanish only.

**Examiner Nonverbal Behavior**

An area of examiner bias in intelligence testing that is related to the effect of verbal reinforcement is that of nonverbal reinforcement. Some interesting studies
have explored this examiner variable, although the paucity of the research has created a real need for more investigation on this topic. One study (Saigh, 1980) examined the positive nonverbal examiner effects of good eye contact, smiling and nodding after each test response, with the neutral nonverbal examiner effects of inconsistent eye contact and a bland facial expression, on the Digit Span and Similarities subtests of the WISC-R. The results indicate that a warm smile and a nod from an adult authority figure during a testing situation can facilitate rapport, lower levels of subject anxiety and improve test scores on both subtests. Since the subject group used in this study were all high achievers and of Arab parentage only, Saigh suggests replicating the study with intellectually and culturally diverse groups.

An interesting study by Delprato and Jackson (1975) looked at the nonverbal examiner behavior as represented by type of examiner seating arrangement in relationship to subject seating arrangement. They compared two seating arrangements, that of subject and examiner face to face, and subject and examiner corner to corner, and their effects on test scores on the WAIS Digit Span and Digit Symbol subtests. They found that Digit Span scores were not affected by seating arrangement but that Digit Symbol scores of male subjects were significantly higher in the corner to
corner arrangement. It would seem worthwhile to see if a significant difference could be found on any other subtests, or on the Verbal, Performance and Full Scale IQs.

Another interesting examiner nonverbal behavior area investigated (Saigh, 1979) was that of explicit perception of examiner religion. In this study, examiners administering the WISC Digit Span subtest wore either a large cross on a chain around their neck, a large symbol of the Koran on a chain around their neck, or wore no neck chain at all. The subtest was administered to Lebanese Christian and Moslem nine to twelve year old children. Results indicate that the perceived religion of the examiner clearly has a significant effect on the performance of children who have the same religion as the perceived religion (as represented by the neck chain symbol) of the examiner. These children scored significantly higher on the subtest regardless of type of religion. Saigh suggests that religious affiliations conveyed symbolically in a testing administration may have a significant impact on attitudes, interpersonal relations and test performance, although cautions that the results may apply only in the Middle East where the study was done, or in countries where more than one form of religion is practiced.

One topic tangentially related to examiner nonverbal behavior was explored by Saigh and Islam (1979). They wanted
to determine if an experimentally induced delay in subject's responses could significantly affect performance on the WAIS. Using the WAIS subtests of Picture Completion and Similarities, they found that the subjects kept from responding for a certain period of time, ranging from ten to thirty seconds, scored significantly higher on both subtests than the subjects allowed to respond normally. Saigh comments that an imposed latency period helps subjects mobilize their intellectual resources and perform at qualitatively higher cognitive levels. Subjects in his study, he argues, conceptualized better solutions, more carefully formulated their answers, rarely searched for words and seemed to verbalize their answers with more ease than those subjects allowed to respond when they wanted to. Saigh views the imposed latency technique as a "possible validity enhancing vehicle" (p. 172) in standardized, individually administered intelligence tests.

Davis, Peacock, Fitzpatrick, and Mulhern (1969) looked at the effect of the interaction between variations in prior experiences and differences in examiner characteristics, on the WAIS Arithmetic subtest scores of adult males. Examiner variables, in this instance, were age, experience and social status, in addition to the personality of each examiner. The results indicate no main effects as a function of variations in examiner. However, a significant
interaction was found between examiners and the treatment condition of prior failure experience. An analysis of the Arithmetic scores revealed a significant performance decrement as a function of a prior failure experience, with the Arithmetic subtest scores, following a nonsense, insoluble task, significantly lower than Arithmetic scores alone or merely following the Comprehension subtest.

**Examiner Administration**

Another examiner variable that has been examined is that of examiner administration versus machine administration. Mishra (1971) investigated the effect of this comparison variable on test scores of 40 college level subjects on the Verbal subtests of the WAIS. Results indicate that test scores presented under the two modes of administration, machine and human administration, were comparable, except on the Information subtest. Scores on the Information subtests, when administered by the examiner were significantly higher than those on the same subtest administered by machine. Mishra points out however, that the possible influence of other factors, such as subject's unfamiliarity with the machine method of test taking, inconsistency of motivational level during the machine administration, and examiner cues during the examiner administration, needs to be examined before it can be said conclusively that the examiner method of administration on
the Information subtest really was more appropriate and effective. In another study on a similar topic, Mishra (1982) investigated the effects of examiner administration and machine (tape recorder) administration on test scores of the Verbal scale of the WAIS given to high- and low-anxious graduate students. The results suggest that the subject's performance on the Verbal subtests was influenced by test administration procedures on two of the subtests, Information and Arithmetic. Subtests scored higher with the examiner administration than with the machine administration. Mishra (1983) also investigated the effects of examiner administration and machine (tape recorder) administration on test scores of six selected subtests of the McCarthy Scales of Children's Abilities. The results suggest that the type of administration produced no significant differences on test results.

Naturally Occurring Examiner Differences in Scoring and Administration

One final area of examiner bias that needs to be explored is that of naturally occurring examiner differences in scoring and administration. Naturally occurring differences result from no explainable differences in examiners, just merely their being different people with different personal characteristics that present themselves during a test administration, and during interaction between
the examiner and the subject. The general, nonspecific examiner effects have been the focus of the following several studies. Schwartz and Flanigan (1971) examined the test scores of three well-trained examiners who administered the WISC, WPPSI and Stanford-Binet, Form L-M, to 21 elementary school and preschool children. Results revealed significant differences in examiner scoring. Attempts to investigate the etiology of these differences resulted in the discovery of no definitive patterns in scoring. It was apparent that the differences were due to individual examiner scoring styles and nothing else. The researchers stated that "the results of this study should be a subject of concern to anyone who uses test reports in the formulation of placement decisions that affect the lives of children" (p. 265).

Rothman (1974) investigated whether the previous observed sensitivity of the WISC subtests to general, nonspecific examiner effects would appear under ordinary testing conditions. Results indicated the presence of both general and differential vulnerability of subtests, especially Comprehension, Similarities and Vocabulary, to examiner effects, consisting of either differences in administration or scoring or some combination of both, during a normal WISC test administration. Smith and May (1967) examined these nonspecific examiner effects on test
scores of the Stanford-Binet and the Illinois Test of Psycholinguistic Abilities, given to 96 Caucasian six year olds. Results indicate significant differences between scores obtained by the different examiners on both the Stanford-Binet and the ITPA, as well as on seven ITPA subscales. No reasons were given for these differences other than nonspecific examiner differences in test administration and scoring, but suggestions were made for more research in the area of individual examiner testing styles. Franklin, Stillman, Burpeau, and Sabers (1982) investigated examiner error in scoring and testing on the WAIS, with results indicating that significant examiner error existed in scoring and administration, particularly on the Digit Symbol and Digit Span subtests. On the Information, Comprehension and Vocabulary subtests, two classes of examiner error were found including: (1) a failure to credit responses with accurate point assignments, and (2) improper termination of the subtests. The implications of the results of this study underscore the need for thorough and competent training of intelligence test examiners. Errors of the magnitude found in the Franklin study could well result in the misplacement of students in appropriate programs, especially for students presenting atypical and/or problematic response patterns.
Sattler (1973) studied the examiner effect, as represented by the variables of scoring accuracy, ability level and personality scores, that might be associated with differences in scoring styles and could find no significant relationships. It was not possible to ascertain why scoring differences existed among examiners.

**Disability and Related Areas of Examiner and Examinee**

A few studies have also investigated the area of disability and considerations in administering intelligence tests to the physically disabled, although none have specifically looked at the topic of the presence of a disability, either on the part of the examiner or examinee, operating as bias in the testing situation. It is clear that this is an area that has been neglected and is one that deserves attention.

One study investigated whether an examinee's physical condition and/or report of his/her academic achievement influenced examiner's scoring of that examinee's test responses. Grossman (1978) utilized the WISC-R and a portrayal of a nondisabled examinee, and a disabled, cerebral palseyed examinee, on video-tape to determine if significant interactions between disability and test scores existed. Results indicated that significant differences in the Arithmetic test scores did exist between the disabled
and nondisabled examinees, with the disabled examinees achieving lower test scores.

The paucity of studies in the area of disability operating as bias surely indicates the need for further research and discussion before any interpretations can be made with respect to its impact and significance in the area of intelligence testing.

An examination such as the one just presented gives a clear picture of the current status in research on the subject of examiner and examinee bias operating in the testing situation of the individually administered intelligence test, and leaves little room for doubt as to the conclusions that can be drawn relative to the significant and nonsignificant effects of the influence of these bias effects in a test administration.

A review of these conclusions seems pertinent. The bulk of the research indicates that the race of the examiner is not a factor in the intellectual performance of either black or white children, contrary to the pervasive myth that exists in testing circles, that black children's test scores are sensitive to the race of the examiner. The literature has demonstrated, however, that minority groups score lower on intelligence tests than the white groups used for the standardization samples and to the white groups to which they are compared (Loehlin, Lindzey & Spuhler, 1975).
Current interest in the concept of culture free testing has stimulated research into the reasons for these differences in scores; however, due to the paucity of the available literature on this subject, a discussion of this concept will not be presented at this time.

Concerning the variable of the gender of the examiner, of the research discussed, only one study suggested results supporting the hypothesis that the gender of the examiner does not exert an effect on intelligence test scores. The rest of the studies reviewed, suggest strongly that both examiner gender and examinee gender impact significantly on intelligence test scores. To understand the direction of that impact, it is necessary to review the studies individually.

Of the literature discussed, only one study found examiner expectancy bias having no significant influence on intelligence test scores. The results of the rest of the research indicate that when examiners possess positive expectancies prior to testing, test scores are significantly higher than when the expectancies are negative.

The major conclusions of the literature reviewed for the purposes of this discussion dealt with the examiner effects of race, gender and expectancies. Other examiner and examinee variables were discussed, with research on those
reaching more tentative conclusions based on sparser and more contradictory results.

The implications of the findings in the literature are many. Wessler (1970) in his study examining the correlates of examiner estimated and obtained IQs, sums up the general feeling in the research community, regarding examiner bias in intelligence testing when he stresses, emphatically, the need for more extensive investigation of examiner differences in the administration and scoring of tests and measurement instruments. As Mishra (1980) states,

Because an...examination is designed to create conditions that elicit the subject's maximal performance, an examiner not only must learn his instrument and standard directions for administering the test well, he must also be aware of his own personal and professional attributes as factors that affect performance, especially on individually administered ability tests (p. 825-826).

Since the testing environment represents the interaction between examiner, examinee and situational factors, in addition to test content and test administration procedures, all of these represent areas of potential bias and should be examined. It appears the research findings encourage agreement with Egeland (1967) when he states,

even though rigorous attempts at standardization of test items and administration procedures have been made, it appears that certain examiner characteristics may interact with subject characteristics in such a way as to bias and
contaminate the results of an individual measure of intelligence (p. 414).

Another implication presented in the literature is a need to explore other areas of bias that heretofore have received little research attention. A potentially rich topic for exploration is that of examiner disability and operant bias in the testing situation.

Related to this topic of bias in testing, and the significance of disability as it might impact as bias, is the topic of psychosocial aspects of disability and the characteristics of disability that might correlate significantly to bias.

Goffman (1963) approaches this topic of disability by suggesting that the person with a visible or known disability has a spoiled identity. He used the term stigma interchangeably with disability and defines both as an attribute that is deeply discrediting—something that makes one different from others. He discusses three types of stigma that include physical deformities, character weaknesses (dishonesty, for example) and tribal stigmas of race or religion. Those who don't possess a stigma he labels as normals.

Goffman (1963) feels that one's social identity, or the demands given to one's character, if based on stigma (or
disability) can create negative feelings that can impact on interactions, creating tension, and he suggests that in those interactions between a normal and a disabled person, both persons are likely to experience uneasiness. Goffman further suggests that in this type of interaction, a situational impropriety might exist, with the normal wanting to withdraw from contact as a means of adjustment and to help lower uneasiness.

Lewis (1971) agrees with this and indicates that most nondisabled people don't treat disabled people in a normal way--they either underreact or overreact. The idea conveyed from some professional literature concerning this reaction toward the disabled is that these reactions then create feelings that constitute a lack of social acceptance that in turn, has a negative psychological effect on the disabled person (Barker, 1953; Wright, 1960). Roessler and Bolton (1978) suggest that these feelings are really negative attitudes toward stigma (disability) and differentness. They assert that whether these negative attitudes reflect some kind of natural aversiveness to deviations in physique or behavior, or whether they can be explained by a developmental social learning theory, there is no doubt that these attitudes exist. They go on to say
that these negative attitudes constitute a virtually insurmountable barrier to fair treatment.

Sillers (1976) summary of the studies that have examined these negative attitudes presents several general conclusions with respect to the specifics of these attitudes. These conclusions are: (1) attitudes toward disabled persons, while highly varied, are frequently negative. Public, verbalized attitudes are on the average, mildly favorable while evidence suggests that deeper, nonverbalized attitudes are frequently more hostile, (2) negative attitudes toward the disabled result in greatly restricted vocational and social opportunities for the disabled, (3) these attitudes are multidimensional and composed of approximately 10-12 somewhat independent attitudinal components, and (4) attempts to modify these negative attitudes have been generally unsuccessful.

According to English (1971), studies conducted on the attitudes of the general public toward the disabled suggests that nearly one-half of the nondisabled persons hold to be true a series of myths and stereotypes about the disabled--including that they are helpless, unproductive and incompetent. He contends that these myths tend to form invisible constraints operating in most interactions between
the disabled and the nondisabled. Gellman (1959) further describes the pervasiveness of unfavorable attitudes toward disabled. He states:

Prejudice toward handicapped (disabled) persons with their open or hidden rejection by the nonhandicapped occurs at all socio-economic levels and in all regions of our country. It is evident in the social, educational and vocational discriminations which hamper disabled persons. It is obvious in the institutional gates which separate the severely disabled from the community of the nondisabled. It is apparent in the difficulties which the handicapped face in securing employment. It is clearly manifest in the self-depreciation of the disabled (p. 20).

Gellman (1959) also attempts to explain the etiology of most negative attitudes toward the disabled. He ascribes these attitudes to four sources that include: (1) social customs and norms that emphasize youth, wholeness, and bodily perfection, (2) child rearing practices that stress normalcy and health, (3) recrudescence of neurotic childhood fears in anxiety-provoking situations, and (4) prejudice by invitation--discrimination provoking behaviors by the disabled. Roessler and Bolton (1978) believe Gellman's sources of negative attitudes can be organized into a three-step sequence that helps to explain prejudice against disabled persons. This sequence involves the following: (1) nondisabled persons are fearful of loss of self-control, disablement, etc., (2) therefore, contact with disabled
persons causes intense discomfort and arouses anxiety, and (3) disabled persons are therefore avoided, and negative attitudes develop concerning the disabled person.

Davis (1961) takes this one step further and discusses those interactions that he believes involve the types of situations between a disabled person and a nondisabled person, in which negative attitudes, uneasiness, or as he terms it, interactional discomfort, are most acutely felt. These include: (1) face-to-face contact, (2) prolonged enough contact to permit more than a fleeting glimpse but not so prolonged that close familiarity ensues, and (3) contact intimate to the extent that the parties must pay more than perfunctory attention to one another but not so intimate that social graces can be dispensed with. As long as this interactional discomfort exists, the disabled person has substantial reason to believe that she/he is being denied the status of social normalcy that she/he aspires to. Even when the disabled person appears competent and normal in all respects except for the disability, discordance exists, resulting in a strain of the interactions between the nondisabled person and the disabled person.
To examine more specifically how these negative attitudes impact on both the nondisabled and the disabled, a review of some pertinent research is warranted. Worthington (1979) studied the significance of personal space and stigma effect of disability held by nondisabled persons toward persons with an obvious physical disability (wheelchair and nostril breathing tube). The results of her study lend support to Goffman's (1963) notion of spoiled identity. She found a contamination effect to be operative in the interactions between the nondisabled and disabled persons with the nondisabled subjects feeling a need to increase distance between themselves and the disabled person.

Smits (1965) studied the reaction of nondisabled adolescents to a group of 200 male and female disabled adolescents, and attempted to delineate the specifics of negative attitudes that might exist toward the disabled adolescents. His results indicated that nondisabled female adolescents were significantly less accepting of disabled persons than were male adolescents, and that disabled adolescents, as a group, were given significantly lower ratings of acceptance than were their nondisabled counterparts.
Malka and Miron (1983) investigated the attitudes of high school students toward disabled persons by assessing their attitudes from responses to a questionnaire measuring acceptance. Results indicated that females tended to have more accepting attitudes toward disabled person and to be less concerned with their differentness than males. The degree of familiarity with disabled persons in general was not found to be significant concerning attitudes. These results appear to contradict those of Smits (1965) although the samples of the two studies differed in nationality, which may have been related to the difference in results.

Thomson (1970) studied the psychosocial interaction between nondisabled and visibly disabled persons to explore the interaction strain (feelings of anxiety, fear, aversion and constraint) that some nondisabled experience in the presence of visibly disabled persons. He tested the hypothesis that interaction strain effects could disrupt the cognitive and social functioning of nondisabled college subjects on tasks including the WAIS Arithmetic, Digit Span and Digit Symbol subtests. The disabling condition utilized was a subtest examiner with a hand prothesis and with an artificial arm. The examiners were legitimately disabled. Results indicated that no significant interaction effects
existed as measured due to the disability of the examiners. And finally, English (1971) reviewed the available literature on the demographic correlates of negative attitudes toward the disabled and presented the following findings:

1. although the literature is mixed, in general, females display more favorable attitudes toward physically disabled persons than do males.
2. high school and college students demonstrate more positive attitudes at each higher grade level.
3. little or no relationship exists between age and attitudes toward the disabled.
4. extensive research indicates that there is virtually no relationship between specific disability variables of type, extent, severity and age at onset, and negative attitudes toward disability.
5. marital status, urban-rural residence, nationality, and race have found to be unrelated to negative attitudes toward the disabled.
6. religion appears to be significantly related to attitudes with Catholics tending to be more accepting of disabled persons than Protestants and Jews.
Concerning the topic of bias in testing, it could be postulated that what Davis (1961) and the others are discussing could very easily apply to the interactions between examiner and examinee in an individually administered testing situation between the nondisabled and a disabled person. It could further be suggested that the negative attitudes and their implications discussed by these researchers could exist as bias to impact on test results. Certainly the characteristics of the types of situations described by Davis (1961) when he discussed interactional discomfort between a disabled and nondisabled person appear to correlate well with the type of interaction found between examiner and examinee in the testing situation of an individually administered intelligence test.

It appears that given the necessity of equal employment, social and personal opportunities for all disabled and nondisabled individuals, surely the disabled professional, as a test examiner, and the disabled examinee, shouldn't be neglected in the researcher's attempts to investigate bias, and to determine the true scope and impact of such bias. Given the continued interest in the topic of bias as a source of error in individually administered intelligence tests, and given the need to orient effort at
eradicating this error, research on the topic of the interaction of disability and attitudes, beliefs and test scores in a testing situation is definitely timely.
CHAPTER 3

THE STUDY

Research Question
The study attempted to answer the general research question: Does the variable of presence or absence of an obvious physical disability on the part of the examiner and the variable of gender of the examiner and gender of the examinee exert a significant influence on the outcome of test scores of selected subtests of the Wechsler Adult Intelligence Scale-Revised?

Null Hypotheses
The following null hypotheses were tested:

$H_1$ There is no difference ($p=.05$) between test scores of examinees tested by a disabled examiner and those of examinees tested by a nondisabled examiner.

$H_2$ There is no difference ($p=.05$) between the test scores of examinees tested by a female examiner and those tested by a male examiner.

$H_3$ There is no difference ($p=.05$) between the test scores of male examinees and female examinees.
$H_4$ There is no difference (p=.05) in the test scores of examinees caused by the interaction between examiner disability and examiner gender.

$H_5$ There is no difference (p=.05) in test scores caused by the interaction of examiner disability and examinee gender.

$H_6$ There is no difference (p=.05) in the test scores caused by the interaction of examiner gender and examinee gender.

$H_7$ There is no difference (p=.05) in the test scores caused by the interaction of examiner gender, examiner disability, and examinee gender.

Null hypotheses were used to posit a no difference direction for the variables of gender of examiner, gender of examinee and the ensuing interactive effects of each, because the literature reviewed concerning research results of bias in these areas is contradictory and does not lend itself to clear results. Not only do the results of the studies differ, with either female or male examiners eliciting higher scores and sometimes no difference occurring or various interactive effects occurring, but in almost every study, aspects such as demographics of the sample, the intelligence test utilized and the makeup of independent variables, in addition to gender, differed. In
addition, none of the studies used the WAIS-R as the intelligence test being researched. On the basis of the summary just provided, it would be difficult to posit a direction other than no difference. If a direction were to be postulated, it would have been that gender of examiner, especially female, would have a positive effect on test scores, raising them significantly.

Sample and Description of Subjects

One hundred and one nondisabled university students were used in this study. They were recruited from University of Arizona Introductory Psychology 101 classes and were randomly assigned to treatment and control groups. The treatment group was administered the Arithmetic and Picture Completion subtest of the WAIS-R by a disabled, wheelchair bound examiner. The control group was administered the same subtests by a nondisabled examiner. University populations were used in this study due to their accessibility to the researcher, and because their familiarity with test taking could influence their willingness to serve as participants in this study. Introductory Psychology 101 students were utilized because of the representative sample of university students they comprised and because of their probable
limited exposure to academic instruction on the topics of disabilities and test taking.

**Research Instrumentation**

The instruments used in this research were the Picture Completion performance subtest and the Arithmetic verbal subtest of the WAIS-R, an individually administered intelligence test. These two subtests were selected, as representative of each of the Performance and Verbal portions of the WAIS-R, as each can be objectively scored, each can be objectively administered and each requires no extensive training in test protocol concerning administration. In addition, Saigh and Islam (1979) indicated that Picture Completion is one of only three culture free subtests on the WAIS. Additional reasons for utilizing the Arithmetic verbal subtest include the fact that on other Wechsler scales, the WISC and WISC-R, it correlates most highly, .5-.57, with the Full Scale test. Both subtests were administered as contained in the complete version of the WAIS-R.

Each examiner administered the Verbal subtest first and the Performance subtest second, on his/her first test administration, and then continue to alternate the order of administration of the subtest in each subsequent administration. The WAIS-R was chosen as the instrument of
choice because it appears to be the most common individually administered intelligence test utilized in the research to date, that explores bias in adult intelligence tests.

Data Collection (Procedure)
The following procedure was used:

1. The examiners consisted of one male and one female graduate student in the Department of Rehabilitation, at the University of Arizona, Tucson, Arizona. Each was given identical training in the administration of the Arithmetic and Picture Completion subtests, and each was given identical training in his/her portrayal of a disabled test examiner. (See Appendix A for protocol and instructions regarding verbal and nonverbal behaviors of examiners as they pertain to complete test administration, including references to the presence of the wheelchair. See Appendix B for protocol and instructions for the use of the wheelchair concerning its mechanical operation.) Each examiner portrayed a disabled examiner in a wheelchair, and a nondisabled examiner not in a wheelchair.
2. Each examiner was randomly assigned to 1/2 of the experimental group of subjects posing as a disabled examiner in a wheelchair, and 1/2 of the control group of subjects as a nondisabled examiner. Each examiner posed as either a disabled or nondisabled examiner first (this was determined randomly) and administered the subtests (in sequential order, with the Verbal and Performance subtest order determined randomly for the first administration, and then followed in an orderly sequential arrangement after that) as that examiner and then, changed roles, and administered the next series of subtests, in the fashion just described, in that examiner role. When both roles were completed, that examiner's role as participant in the study, as examiner, was completed, and no further participation was required.

3. An n of 120 was considered an ideal number, due to cell size requirements for the ANOVA statistical procedure. However, due to examinees nonattendance in certain conditions, an n of 101 was utilized, with cell sizes meeting homogeneity of variance requirements.
4. Each examiner participated without knowledge of the purpose of this study.

5. Each examinee was given an Informed Consent (See Appendix D) to read and sign, and each examinee was asked to participate voluntarily.

6. Each examinee, and both examiners were read a Debriefing Statement (See Appendix D) after his/her participation in the study was completed.

7. All examinees participated only once, being administered both the Picture Completion subtest and the Arithmetic subtest.

8. Test administrations were scored blind by one graduate student trained in the objective scoring of these two subtests, as per standardized scoring procedures for the complete administration of the WAIS-R. The current test booklet of the WAIS-R was utilized in the scoring.

9. Data were recorded from the scored protocols and analyzed by two three-way ANOVAs.

10. The results of this study were available to all participants upon completion of the study.

Design and Variables

The independent variables in this study were gender of examiner, gender of examinee and examiner disability. The
dependent variables were test scores of the Picture Completion performance subtest and the Arithmetic verbal subtest. A Campbell and Stanley (1963) posttest only control group design was utilized. Since the most adequate assurance of lack of initial biases between groups is randomization (Campbell & Stanley, 1963), random assignment of subjects (examinees) to experimental and control groups was utilized, and no pretest intelligence test scores measure was utilized.

Treatment conditions of examiner disability versus examiner not having a disability were maximized. Disabled examiners participated in the testing situation in a wheelchair. Nondisabled examiners did not. It is not known how the difference between the two treatment conditions could have been made any more obviously different. Efforts were made to control for systematic variance through the random assignment of examinees to treatment groups, and through the participation of each examiner in both the disabled and nondisabled examiner role. Subjects were not tested beforehand to determine individual levels of intellectual functioning. However, it is felt that this absence of a pretest might have enhanced the error variance somewhat. A reliable measure of intelligence was used however, to determine the effects of bias of test scores, and this was seen as an effort to minimize error variance.
**Treatment of the Data**

The data were analyzed using two 3-way ANOVAs, to examine the effects of gender of examiner, gender of examinee and disability of examiner on test scores of the Picture Completion performance subtest and the Arithmetic verbal subtest.

**Assumptions of the Study**

The sample for the study is regarded as homogeneous and as a representative sample of a subsequent university population. The administration of the objective measures of the portion of the WAIS-R utilized standardized methods of test administration for all examinees.

**Limitations of the Study**

The threats to internal validity include instrumentation. How the examiners portrayed their disability (even though an attempt was made to control for this through the utilization of a wheelchair) was not monitored during the actual test administration, so differences might have existed in the disabled examiner role.

The threats to external validity include interaction of testing and treatment, the interaction of selection and treatment, history, and reactive arrangement.
The interaction of testing and treatment variables involves the examinees taking only two of the WAIS-R II subtests, and the limiting effect this might have on the ability to generalize the results of this study. The interaction of selection and treatment variables involves the probable inability of generalizing the results of this study to other than college adults. The results of this study are only validly demonstrated for the population from which the subject groups were selected.

Reactive arrangement effects could not be controlled for as all subjects know that they were participating in an experimental procedure. Although examinees will not be given the actual reason for this study, their knowledge that they were participating in an experimental procedure could not have been controlled for without violating research ethics.

Effects of history were controlled for as much as possible. However, this study took place over a three week time interval, and intervening variables among examinees' and examiners' experiences during that three week period could have exerted an influence on the results.
CHAPTER 4

RESULTS AND DISCUSSION

Introduction

The purpose of this study was to investigate examiner bias in the areas of examiner disability, and examiner and examinee gender on test scores of selected subtests of the WAIS-R. This chapter presents the findings along with a brief discussion of these results. The first section details the statistical analysis relating to each hypothesis.

Results

Hypothesis 1: There is no difference (p<.05) between test scores of examinees tested by a disabled examiner and those of examinees tested by a nondisabled examiner.

Two three-way ANOVAs were calculated to test hypothesis 1. Test results from the Arithmetic subtest and the Picture Completion subtest of the WAIS-R constituted the dependent variables for each analysis. In each, examiner disability, examiner and examinee gender were tested as independent variables. As shown in Tables 1 and 2, nonsignificant effects were found for the disability manipulation, an indication that test scores did not differ significantly between those tested by a nondisabled examiner
Table 1. Analysis of variance of Arithmetic subtest scores.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender</td>
<td>.90</td>
<td>1</td>
<td>.90</td>
<td>.18</td>
<td>.67</td>
</tr>
<tr>
<td>Examiner Disability</td>
<td>.00</td>
<td>1</td>
<td>.09</td>
<td>.02</td>
<td>.89</td>
</tr>
<tr>
<td>Examinee Gender</td>
<td>24.69</td>
<td>1</td>
<td>24.69</td>
<td>5.03</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Two-Way Interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender X Examiner Disability</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.99</td>
</tr>
<tr>
<td>Examiner Gender X Examinee Gender</td>
<td>.25</td>
<td>1</td>
<td>.25</td>
<td>.05</td>
<td>.82</td>
</tr>
<tr>
<td>Examiner Disability X Examinee Gender</td>
<td>.16</td>
<td>1</td>
<td>.16</td>
<td>.03</td>
<td>.86</td>
</tr>
<tr>
<td><strong>Three-Way Interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender X Examiner Disability X Examinee Gender</td>
<td>.00</td>
<td>1</td>
<td>3.42</td>
<td>.70</td>
<td>.41</td>
</tr>
<tr>
<td>Residual</td>
<td>456.62</td>
<td>93</td>
<td>4.90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Analysis of variance of Picture Completion subtest scores.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender</td>
<td>16.87</td>
<td>1</td>
<td>16.87</td>
<td>3.46</td>
<td>.07</td>
</tr>
<tr>
<td>Examiner Disability</td>
<td>10.70</td>
<td>1</td>
<td>10.70</td>
<td>2.20</td>
<td>.14</td>
</tr>
<tr>
<td>Examinee Gender</td>
<td>21.21</td>
<td>1</td>
<td>21.21</td>
<td>4.35</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Two-Way Interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender X Examiner Disability</td>
<td>4.17</td>
<td>1</td>
<td>4.17</td>
<td>.86</td>
<td>.36</td>
</tr>
<tr>
<td>Examiner Gender X Examinee Gender</td>
<td>1.64</td>
<td>1</td>
<td>1.64</td>
<td>.34</td>
<td>.56</td>
</tr>
<tr>
<td>Examiner Disability X Examinee Gender</td>
<td>.86</td>
<td>1</td>
<td>.86</td>
<td>.18</td>
<td>.68</td>
</tr>
<tr>
<td><strong>Three-Way Interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Gender X Examiner Disability X Examinee Gender</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.98</td>
</tr>
<tr>
<td><strong>Residual</strong></td>
<td>453.34</td>
<td>93</td>
<td>4.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and those tested by an examiner posing as disabled. The results presented in Table 1 for the Arithmetic subtest scores and in Table 2 for Picture Completion subtest scores report nonsignificant F ratios, Table 1, (F [1,99] = .02; p>.05) and Table 2, (F [1,99] = 2.20; p>.05), for the main effect of examiner disability. Examination of cell means as shown in Appendix E (Table 3) indicate the mean Arithmetic subtest score for those tested by a disabled examiner to be 10.12, while those tested by a nondisabled examiner to be 10.12. Appendix F (Table 4) reports the mean Picture Completion subtest score for those tested by a disabled examiner to be 10.06, while for those tested by a nondisabled examiner it is 9.32. Based on this analysis, null hypothesis 1 was retained.

**Hypothesis 2:** There is no difference (p<.05) between the test scores of examinees tested by a female examiner and those tested by a male examiner.

Based on the results of the three-way ANOVAs presented in Tables 1 and 2, gender of the examiner does not appear to have any mediating effect on the dependent variables of Arithmetic and Picture Completion subtest scores. As shown in Table 1, (F [1,99] = .18; p>.05) and in Table 2, (F [1,99] = 3.46; p>.05), nonsignificant F ratios were obtained for the main effect of examiner gender, with
test scores on both subtests not differing significantly between examinees tested by either a male or female examiner. As shown in Appendix G (Table 5), the mean Arithmetic subtest score for all those tested by a male examiner ($\bar{X} = 10.00$) does not differ significantly from the mean Arithmetic subtest score for all those tested by a female examiner ($\bar{X} = 10.24$). Appendix H (Table 6) presents the mean Picture Completion subtest scores for all those tested by a male examiner ($\bar{X} = 9.24$) and a female examiner ($\bar{X} = 10.14$). Based on these results, null hypothesis 2 was retained.

**Hypothesis 3:** There is no difference ($p<.05$) between the test scores of male examinees and female examinees.

The primary test for hypothesis 3 is the main effect of examinee gender. As shown in Table 1, the $F$ ratio for gender of examinee for Arithmetic subtest scores was significant ($F [1,99] = 5.03; p<.05$). Examination of cell means as shown in Appendix I (Table 7) indicates that the mean Arithmetic subtest scores of male examinees ($\bar{X} = 10.63$) were found to be significantly higher than those of the female examinees ($\bar{X} = 9.64$). This finding could be postulated as a source of examinee bias operating to influence test scores on the WAIS-R Arithmetic subtest or could represent males being more adept on this subtest. As
is also indicated in Table 1, examiner gender and examiner disability had no interactive effect with examiner gender in terms of higher test scores, \( (F[1,99] = .70; p>.05) \).

The F ratio for examinee gender was significant, as well, for the Picture Completion subtest scores \( (F[1,99] = 4.35; p<.05) \). Once again, group means as shown in Appendix J (Table 8) indicate that Picture Completion subtest scores of the male examinees \( (X = 10.20) \) were significantly higher than the Picture Completion subtest scores of the female examinees \( (X = 9.21) \). This finding could lend support to the assumption that examinee bias, in the form of gender of examinee, is a viable variable operating to impact on test scores, or that perhaps, males simply do better on this subtest. On the basis of this analysis, null hypothesis 3 was rejected.

**Hypothesis 4:** There is no difference \( (p<.05) \) in the test scores of examinees caused by the interaction between examiner disability and examiner gender.

As shown in Table 1, nonsignificant F ratios, \( (F[1,99] = .00; p>.05) \) were obtained for the interactive effects of examiner disability and examiner gender, with nonsignificant differences in subtest scores for the Arithmetic subtest resulting from the interaction of these independent variables. As shown in Table 2, the F ratios \( (F \)
obtained for the interactive effects of examiner disability and examiner gender on the Picture Completion subtest scores were also nonsignificant. As a result, null hypothesis 4 was retained.

**Hypothesis 5:** There is no differences ($p < .05$) in test scores caused by the interaction of examiner disability and examinee gender.

The three-way ANOVA results yielded nonsignificant differences for the Arithmetic subtest and the Picture Completion subtest scores due to the interactive effects of examiner disability and examinee gender. As is shown in Table 1, a nonsignificant F ratio was obtained, ($F_{[1,99]} = .03; p > .05$) for the Arithmetic subtest scores, and as is presented in Table 2, a nonsignificant F ratio ($F_{[1,99]} = .86; p > .05$) was also obtained for Picture Completion subtest scores. Null hypothesis 5 was retained.

**Hypothesis 6:** There is no difference ($p < .05$) in the test scores caused by the interaction of examiner gender and examinee gender.

Nonsignificant differences were found in both the Arithmetic subtest scores ($F_{[1,909]} = .05; p > .05$) and the Picture Completion subtest scores ($F_{[1,99]} = .34; p > .05$) between those male examinees tested by a male or female examiner and those female examinees tested by a male or
female examiner (see Tables 1 and 2). Since the interaction of examiner gender and examinee gender did not significantly alter subtest scores, this null hypothesis was retained.

**Hypothesis 7:** There is no difference (p<.05) in the test scores caused by the interaction of examiner gender, examiner disability and examinee gender.

As is shown in Table 1, a nonsignificant F ratio ($F_{[1,99]} = .70; p>.05$) was obtained for the interactive effects of examiner gender, examiner disability and examinee gender with no significant differences in test scores for the Arithmetic subtest resulting from the interaction of these three independent variables. As is presented in Table 2, the F ratio ($F_{[1,99]} = .00; p>.05$) obtained for the interactive effects of these three independent variables on the Picture Completion subtest scores was also nonsignificant. As a result, null hypothesis 7 was retained.

**Discussion**

The purpose of this study was to investigate examiner bias in the areas of examiner disability, and examiner and examinee gender on test scores of selected subtests of the WAIS-R. The primary questions being investigated were: (1) whether or not the presence or absence of an overt physical disability on the part of the examiner in the testing situation would significantly alter
test scores, (2) whether or not the gender of either the examiner or the examinee would create a significant difference in the test scores, and (3) whether or not any interactive effects of examiner disability, gender of the examiner and gender of the examinee would create a significant difference in the test scores.

The results of the study indicate that for this sample, gender of the examiner and disability of the examiner had no significant effect on Arithmetic and Picture Completion subtest scores. Results also indicate that interactive effects of any of these three variables did not significantly alter the subtest scores.

The no difference findings for gender of examiner is not consistent with the bulk of the research reviewed concerning examiner gender and its impact as bias in intelligence testing. Of all the studies reviewed, only one (Cieutat & Flick, 1967) did not support the contention that examiner gender is a possible source of bias in testing. It is interesting to note, however, that in this study, although nonsignificant F ratios were obtained for the main effect of examiner gender on both the Arithmetic and Picture Completion subtests, $p=.07$ for examiner gender on Picture Completion subtest scores approaches significance and should be noted. Had perhaps a larger sample been utilized, as is
discussed in recommendations for future research, a significant F ratio would have been obtained, and this study could have lent support to the conclusion of most of the literature that examiner gender appears to be a consistent source of bias in intelligence testing.

The fact that examiner disability did not appear to be functioning as a source of bias impacting on the test scores of examinees in this study does not necessarily dispute the conclusion of most of the literature that the existence of negative attitudes of nondisabled toward the disabled and their disabilities is pervasive, and interferes with many types of interactions between the nondisabled and the disabled. It could be postulated that negative attitudes of examinees existed in the testing situation of this study, but did not interfere with the examiner's cognitive functioning. Or, perhaps the examinees in this study were enlightened participants who did not have negative attitudes toward disability, and in particular toward the disabled examiner who tested them.

Significant differences were found, however, between the subtest scores of both the Arithmetic and Picture Completion subtests of the male examinees and those of the female examinees, with the males having significantly higher subtest scores. This resulted in the rejection of null hypothesis 3.
The literature to date has presented mixed findings with respect to whether males or females have better scores on the various intelligence tests, including the WAIS-R and all of its 11 subtests. As was indicated in the review of the literature, of the 12 studies that did investigate differences in test scores due to the independent variable of gender of examinee, only 3 reported significant main effects, and the results of those three are mixed. Only one study (Pederson, Shinedling & Johnson, 1968) investigated the Arithmetic subtest (WISC Arithmetic subtest), and nonsignificant relationships for gender of examinee and test scores were reported. Only one of the 12 studies utilized university Introductory Psychology students, the same type of sample used in the present study, and in that study, the female examinees scored better on the WAIS, than their male cohorts.

With such mixed findings in the literature concerning the effect of examinee gender and test scores, it is difficult to postulate viable reasons for the significant results found in the present study. Perhaps differences found are indigenous to the particular sample utilized, and as has been suggested previously, had a larger sample been used, the study would have assured a more viable attempt at using a representative sample, with less differences
peculiar to the sample. In addition, had all 11 WAIS-R subtests been used, with only Full Scale scores being evaluated, perhaps differences among subtests would have evened out among male and female examinees, resulting in findings that more truly represented the overall mediational impact of the independent variables. Based on the results of the present study, it is difficult to state with any surety that male examinees were demonstrating quantitative or discrimination skills significantly better or more sophisticated than the female examinees. It can be stated that the differences in subtest scores were significant between the male and female examinees, and that further research is needed.

Caution must be taken when considering the findings, then, of this study for the following reasons:

1. Limitations of the design. These are discussed in detail in Chapter 3 of this study. They include instrumentation, interaction of test and the experimental variable, interaction of selection and the experimental variable, reactive arrangement and history.
2. Lack of research to support the finding that male examinees score significantly higher on the WAIS-R Arithmetic and Picture Completion subtests than females.

Conclusion

In summary, no statistically significant relationships were found between gender of examiner, examiner disability and the scores of the WAIS-R Arithmetic and Picture Completion subtests. This could indicate that gender of examiner and disability of examiner are not viable aspects of examiner bias in individually administered intelligence tests. This finding does contribute to the bulk of mixed findings with respect to gender of examiner and examiner bias but does not support most of the studies that find a significant relationship between differences in test scores and examiner gender. The presence or absence of an examiner disability and its relationship to WAIS-R subtest scores has not heretofore been studied, so no definitive results are available to support or challenge the finding of this study that examiner disability does not exist as examiner bias influencing intelligence test scores. One study did investigate the relationship of interaction strain effects and WAIS subtest scores, and found that no
significant interaction effects existed as measured due to the disability of the examiner (Thomson, 1970).

A significant relationship was found, however, between scores of both the Arithmetic and Picture Completion subtests and the gender of examinee, with male examinees having significantly higher subtest scores than the female examinees.

Based on the threats to both internal and external validity, the size of the sample, and lack of research to support the findings of this study, caution must be exercised when considering these results.

This study can be considered only a part of a larger body of research regarding the relationship between bias in testing and test scores. Because this represents a topic in the field of testing thought to be extremely significant in terms of the impact test scores has on the lives of all who take them, additional research needs to be conducted regarding this topic.
CHAPTER 5

SUMMARY AND RECOMMENDATIONS

Introduction
This chapter is presented in two sections. Section one is a general summary of the project. Section two is a list of recommendations for future research.

General Summary
The review of literature presented mixed findings with respect to error or bias in testing. Four sources of error, including examiner bias, examinee bias, scoring bias, and test environment bias have been investigated by various researchers to determine the significance of their impact on intelligence test scores, particularly the individually administered intelligence tests. The contradictory results suggest that although bias exists as a viable influence on test scores, the direction of that influence is varied. A review of pertinent research on attitudes toward disability indicated that, in general, negative attitudes on the part of nondisabled people toward disabled people are pervasive. These negative attitudes also appear to create interaction strain in interactions between the disabled and the nondisabled, and appear to influence the outcome of the interactions. An area not fully addressed prior to this
study concerns examiner bias in the area of examiner disability. The present study identifies the need for more research concerning the question of whether or not the presence or absence of a physical disability on the part of the examiner is related to examiner bias.

The Purpose

The purpose of this study was to investigate bias, in the areas of examiner disability and examiner and examinee gender on test results of selected subtests of the WAIS-R.

The Sample

One-hundred-one male and female University of Arizona students enrolled in Introductory Psychology 101 classes, fall semester 1985, participated as subjects in this study.

The Procedure

Subjects were randomly assigned to treatment and control groups. The treatment group was administered the Arithmetic and Picture Completion subtests of the WAIS-R by a wheelchair bound examiner, (a nondisabled examiner posing as a disabled examiner). The control group was administered the same subtests by a nonwheelchair bound examiner.
Statistical Treatment

The data were analyzed by two three-way ANOVAs. The three-way analysis of variance tested for significant relationships between main effects of the three independent variables - examiner disability, examiner gender and examinee gender, and the scores of the Arithmetic and Picture Completion subtests, and also for two-way and three-way interactive effects of the three variables and the subtest scores.

Results

Overall results indicated that in terms of bias operating to influence the individually administered subtest scores, only examinee gender was found to have a significant effect, with subtest scores for male examinees significantly higher than those for female examinees on both Arithmetic and Picture Completion subtests. This resulted in the rejection of null hypothesis 3. All other null hypotheses, 1, 2, 4, 5, 6, and 7, were retained. Nonsignificant relationships were found between the other two independent variables of examiner disability and examiner gender, and subtest scores. The presence of a visible, physical disability on the part of the examiner, and the gender of the examiner were not seen to be operating as bias and were not seen to be impacting on subtest scores. These findings
need to be examined with caution, however, due to limitations of the study and due to lack of supportive research in the area of examinee gender as bias operating to influence the scores of individually administered intelligence tests.

**Recommendations for Future Research**

Based on the results of this research, and on the acknowledged limitations of the study, future research is suggested. It is recommended that:

1. Similar research be conducted with a larger sample.
   
   Results of this study indicated that male examinees performed better on the WAIS-R Arithmetic and Picture Completion subtests than did their female counterparts. Differences in subtest scores may have been indigenous to this sample. The use of a larger sample more representative of a university population could have controlled for these differences.

2. Similar research be conducted with a sample drawn from a more heterogeneous population.
   
   The subjects used in this study were all enrolled in Introductory Psychology 101 classes at a local university. The rationale used for employing a university population included accessibility to the researcher, familiarity with test taking, assumed limited exposure to academic
instruction on the topics of disability and the representative sample of university students they comprised. In particular, Introductory Psychology students were very accessible as subjects, much more so than any other university population this researcher attempted to access. However, it could be suggested that perhaps the assumptions of this subject selection rationale were not all viable, and that these particular subjects were not as unsophisticated with test taking as assumed, and perhaps, had more exposure to the topic of disability and bias than anticipated. Using a more heterogeneous sample would help alleviate the problems associated with these assumptions.

3. Similar research be conducted using an authentically disabled examiner rather than a nondisabled examiner posing as disabled.

It is acknowledged that although the two examiners received extensive training (see Appendices A & B) concerning their participation in the study and in particular, their individual portrayal of a disabled person, their actual performances as examiners in the testing sessions could not be monitored. Because of this, differences in performances by the two examiners could have existed, and thus impacted as bias and influenced subtest scores. In this way, procedure appears to exist as a viable
threat to internal validity, one that could be controlled for, to some extent, by using an authentically disabled person as one of the examiners. The use of one authentically disabled examiner, however, would limit external validity and thereby, decrease generalizability. In addition, all subjects were debriefed after their participation in the study. It is possible that the debriefed subjects could have discussed their participation in the study with other potential subjects. Discussion could have ensued concerning the examiners and their disabilities with differences noted by subjects between which examiner was disabled and which was not. This could have contributed to significant pretest information for some subjects and thereby, could have existed to bias results. The use of only one disabled examiner, who is authentically disabled, could somewhat control for this.

4. Similar research be conducted utilizing the complete WAIS-R intelligence test rather than one performance subtest and one verbal subtest.

It is acknowledged that the use of only two of the eleven WAIS-R subtests exists as a threat to external validity in this study, thereby limiting the generalizability of results. In addition, it is not known whether treatment effects could have made more of a
difference on the subtest scores had all subtests been administered. Perhaps examinee performances in this study were indigenous to only the two subtests utilized.

5. Similar research be conducted utilizing the administration of another test instrument.

While a review of the literature clearly shows the WAIS-R and WAIS as the most frequently utilized intelligence tests in the study of bias in intelligence testing, several other tests have also been used. These include the WISC-R, Peabody Picture Vocabulary Test, Ravel Progressive Matrices and the Stanford-Binet. A need clearly exists for these tests to be examined more extensively as they pertain to the topics of bias in general, and in particular, disability and bias.

6. Similar research be conducted utilizing a disability other than wheelchair bound as one of the independent variables.

Since the topic of disability and bias in testing currently exists as a relatively unexplored research area, it would seem appropriate to encourage study concerning all types of disabling conditions. Any variable that might exist as bias in testing should be explored, and there are many types of disabling conditions that could be easily researched.
7. Similar research be conducted investigating the effects of a disabling condition on the part of the examinee who is being administered the intelligence test.

There is minimal research investigating the impact of examinee disability on intelligence test scores (Grossman, 1978). The results of Grossman's study indicated that significant differences did exist in the Arithmetic subtest scores, with the disabled examinees receiving lower subtest scores. The paucity of studies in this area surely indicates the need for further research and discussion before any interpretations can be made.
APPENDIX A

PROTOCOL AND INSTRUCTIONS FOR VERBAL AND
NONVERBAL EXAMINER BEHAVIORS DURING
WAIS-R SUBTEST ADMINISTRATION
Please read carefully and feel free to respond to any of the statements listed below. I will be happy to answer any questions you may have.

1. Each of you will be administering the Picture Completion and Airthmetic subtests of the WAIS-R.

2. Each of you will be given the instructions from the WAIS-R manual as provided in the manual, to read, regarding the administration of the two subtests. After each of you have read them once, you will be provided approximately three hours of instruction, by the principle researcher of this study, regarding your participation as an examiner in the administration of the subtests. The instruction will consist of the following:
   a. a rereading of the instructions of the WAIS-R Manual.
   b. a hands-on demonstration of the administration of both subtests by the principle researcher of this study.
   c. three practice trials by each examiner of both subtests, for a total of six trials per each examiner.
d. a rehearsal of allowed and nonallowed verbal and nonverbal behaviors during the administrations, to include the following:

(1) **Allowed and Required Behaviors**

(a) Smiling at examinee.

(b) A greeting at the start of each administration to consist of the following:

1) "Hello, my name is ____________.

   Today (tonight) I will be asking you some questions. If you have any questions regarding your participation in this study, I will not be able to answer them, but will refer you to Susan Voskuil, the principle researcher. If at any time you want to stop what we are doing and discontinue your participation, you may do so."

(c) Introduce each subtest verbatim per the instructions in the WAIS-R Manual.

(d) At the end of the two administrations say, "Thank you for your participation in this study. Should you have any questions, please refer to the Informed
Consent that you signed, and Susan Voskuil. You may leave now."

(e) Stay seated in the room and wait for the next examinee.

(2) **Nonallowed and Prohibited Behaviors**

(a) Engaging in social conversation with the examinee.

(b) Identifying personal information about yourself to the examinee.

(c) Asking the examinee any personal information about himself or herself.

(d) Verbalizing comments about the examinee's performance such as:
   "You did very well."
   "That's really good."
   "You did not do well."

(e) Commenting on the examinee's appearance or personality.

(f) Revealing any information about the examiner's participation in the study.

(g) Commenting on the presence of the wheelchair. If you are asked about the wheelchair or your involvement with the wheelchair, offer as little information
as possible or say, "I can't answer that. I will have to refer you to the principle researcher."

(h) At any time during the administration, supplying any part of the answers to the questions, to the examinees.

(3) Each of you will be administering subtests 1/2 of the time from a wheelchair and 1/2 of the time from a desk chair. Each of you will be given instructions concerning the mechanical operation of the wheelchair (See Appendix B).

(4) Once each of you has completed your role as examiner, no further participation in this study will be required of you. When your participation is completed, feel free to ask any questions regarding the significance of not only the outcome of the study, but also of your participation. You will each be given a Debriefing Statement (Appendix D).
APPENDIX B

PROTOCOL AND INSTRUCTIONS FOR THE USE OF THE WHEELCHAIR CONCERNING ITS MECHANICAL OPERATION
1. One wheelchair will be provided for each examiner.

2. Each examiner will position himself/herself in the wheelchair before the examinee enters the testing room to begin the subtest.

3. Each examiner will remain seated during the complete administration of each subtest for which a wheelchair is provided.

4. Once the subtest begins and the examiner is in the wheelchair, the examiner is not to move any portion of his/her body from the waist down.

5. Once the subtest is completed, the examiner is not to move the wheelchair away from the table or toward the door. The examiner will say goodbye to the examinee but remain seated in the wheelchair and will not attempt to move the wheelchair. The examiner will simply place his/her hands on the table, or in his/her lap.

6. The wheelchair brake is to be kept on at all times to prevent movement.
Dear Participant:

I am requesting your voluntary participation in a question and answer interview session as part of this research study. The purpose and objective of this research is to study interview sessions of simple math and discrimination skills. All information collected during your participation will be treated with anonymity and confidentiality, and upon completion of the research, all data collected will be destroyed.

The study in which you are participating is part of a research project for my Doctoral Degree in Rehabilitation Psychology from The University of Arizona, College of Education. The data gathered from this research will be used to determine the significance of several research hypotheses involving math and discrimination skills and assessment administration, and statistical summaries of the data may eventually reach publication. If you are willing to participate in this research study, it should take about fifteen to twenty minutes.

At no time do I anticipate that your participation in this research will present any physical or social risk to you. If you would want to withdraw from participation in this study at any time, you may do so. Your participation in this study is completely voluntary.

Your agreement to willingly cooperate in this study will be indicated by your signing and returning to me the consent form at the bottom of this letter. Please keep the letter for your reference.

Sincerely,

Susan L. Voskuil

----------------------------------------------------------

I have received a complete explanation of the research project and procedures to be carried out by Ms. Susan L. Voskuil and agree to participate in the research study.

_________________________  _______________________
Signature                  Date
APPENDIX D

EXAMINER AND EXAMINEE DEBRIEFING
The assessment session in which you have just participated has been part of a research study attempting to determine the significance of examiner bias in individually administered intelligence tests. Bias in testing can be simply defined as any variable which impacts on the results of the testing in such a way as to positively or negatively impact on those results. Results are usually defined as scores.

Research to date has indicated that both verbal and nonverbal behaviors of both examiner and examinee (subject) exist as bias and can impact on test scores in individually administered test batteries, whether those batteries are of a subjective or objective nature. This research study is attempting to determine whether the gender of the examiner (male or female), the gender of the examinee (male or female), and the presence or absence of a wheelchair will influence the subject's test scores to such an extent that there results in a significant difference between the test scores of those subjects tested by a disabled, wheelchair bound male examiner, a disabled wheelchair bound female examiner, a nondisabled male examiner and a nondisabled female examiner. The general research question is: Does the variable of presence or absence of an obvious physical disability on the part of the examiner, and the variable of
gender of examiner and gender of examinee exert a significant influence on the outcome of test scores in selected subtests of the WAIS-R. Seven null hypotheses will be tested, with the researcher postulating a no difference direction for all the results. It is expected that there will be no difference between results for all subjects, as all the literature to date has presented mixed results.

A Campbell and Stanley post test only control group design is being utilized. One treatment and one control group is being utilized. The treatment group is the group being tested by a wheelchair bound examiner, and the control group is the group being tested by a nondisabled examiner. If you were tested by a wheelchair bound examiner you were tested by a nondisabled person who was posing as a disabled examiner. Two nondisabled Department of Rehabilitation graduate students acted as examiners in this study, with treatment conditions of examiner disability versus examiner not having a disability being maximized. It is not known how the difference between the two treatment conditions could have been made any more obviously different.

Data in this study will be analyzed utilizing two three-way ANOVAs. The limitations of this study include threats to internal and external validity. The threat to internal validity includes instrumentation, involving how
the examiners portrayed their disability. Attempts were made to control for this through the utilization of a wheelchair. However, the actual test administrations were not monitored, so differences might exist between the various test administrations. The threats to external validity include interaction of testing and X, the interaction of selection X reactive arrangement and history. The interaction of testing and X involves the subject taking only two of the 11 subtests of the WAIS-R, and the limiting effect this might have on the ability to generalize the results of the study. The interaction of selection and X involves the probable inability of generalizing the results of this study to other than college adults. Reactive arrangements could not be controlled for (the fact that you as a subject have awareness that you are participating in an experiment) as you volunteered for this study and signed an Informed Consent prior to your participation. History involves the length of time from onset to end of the study and the presence of variable existing to influence subject's participation in the study during that time, and hence, influence the outcome of the study.
APPENDIX E

TABLE 3
EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER DISABILITY ON ARITHMETIC SUBTEST
Table 3. Examination of cell means for independent variable of examiner disability on Arithmetic subtest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability</td>
<td>10.12</td>
<td>2.07</td>
<td>51</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.96</td>
<td>2.42</td>
<td>24</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.27</td>
<td>2.24</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.62</td>
<td>2.62</td>
<td>13</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.26</td>
<td>1.74</td>
<td>27</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.93</td>
<td>1.90</td>
<td>15</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.42</td>
<td>1.08</td>
<td>12</td>
</tr>
<tr>
<td>No Disability</td>
<td>10.12</td>
<td>2.75</td>
<td>50</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>10.04</td>
<td>2.80</td>
<td>26</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.67</td>
<td>3.62</td>
<td>12</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.50</td>
<td>1.82</td>
<td>14</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.21</td>
<td>1.79</td>
<td>24</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.55</td>
<td>1.03</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.92</td>
<td>2.25</td>
<td>13</td>
</tr>
</tbody>
</table>
APPENDIX F

TABLE 4
EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER DISABILITY ON PICTURE COMPLETION SUBTEST
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability</td>
<td>10.06</td>
<td>2.07</td>
<td>51</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.38</td>
<td>1.95</td>
<td>24</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>9.91</td>
<td>1.04</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>8.92</td>
<td>2.43</td>
<td>13</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.67</td>
<td>2.01</td>
<td>27</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.87</td>
<td>2.41</td>
<td>15</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>10.42</td>
<td>1.44</td>
<td>12</td>
</tr>
<tr>
<td>No Disability</td>
<td>9.32</td>
<td>2.41</td>
<td>50</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.12</td>
<td>2.62</td>
<td>26</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>9.83</td>
<td>2.72</td>
<td>12</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>8.50</td>
<td>2.47</td>
<td>12</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>9.54</td>
<td>2.18</td>
<td>24</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.00</td>
<td>1.41</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.15</td>
<td>2.67</td>
<td>13</td>
</tr>
</tbody>
</table>
APPENDIX G

TABLE 5
EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER GENDER ON ARITHMETIC SUBTEST
Table 5. Examination of cell means for independent variable of examiner gender on Arithmetic subtest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Examiner</td>
<td>10.00</td>
<td>2.60</td>
<td>50</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.48</td>
<td>2.98</td>
<td>23</td>
</tr>
<tr>
<td>Disability No wheelchair</td>
<td>10.67</td>
<td>3.62</td>
<td>12</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>10.27</td>
<td>2.24</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.59</td>
<td>2.21</td>
<td>27</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>9.50</td>
<td>3.37</td>
<td>14</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>9.69</td>
<td>2.63</td>
<td>13</td>
</tr>
<tr>
<td>Female Examiner</td>
<td>10.24</td>
<td>1.75</td>
<td>51</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.77</td>
<td>1.58</td>
<td>26</td>
</tr>
<tr>
<td>Disability No wheelchair</td>
<td>10.56</td>
<td>1.04</td>
<td>11</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>10.93</td>
<td>1.91</td>
<td>15</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.68</td>
<td>1.77</td>
<td>25</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>9.92</td>
<td>2.25</td>
<td>13</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>9.42</td>
<td>1.08</td>
<td>12</td>
</tr>
</tbody>
</table>
APPENDIX H

TABLE 6
EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINER GENDER ON PICTURE COMPLETION SUBTEST
Table 6. Examination of cell means for independent variable of examiner gender on Picture Completion subtest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Examiner</td>
<td>9.24</td>
<td>2.31</td>
<td>50</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>9.87</td>
<td>2.05</td>
<td>23</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>9.83</td>
<td>2.73</td>
<td>12</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>9.91</td>
<td>1.05</td>
<td>11</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>8.70</td>
<td>2.42</td>
<td>27</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>8.50</td>
<td>2.47</td>
<td>14</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>8.92</td>
<td>2.43</td>
<td>13</td>
</tr>
<tr>
<td>Female Examiner</td>
<td>10.14</td>
<td>2.15</td>
<td>51</td>
</tr>
<tr>
<td>Examinee Male</td>
<td>10.50</td>
<td>2.06</td>
<td>26</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>10.00</td>
<td>1.41</td>
<td>11</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>10.87</td>
<td>2.42</td>
<td>15</td>
</tr>
<tr>
<td>Examinee Female</td>
<td>9.76</td>
<td>2.22</td>
<td>25</td>
</tr>
<tr>
<td>Disability No Wheelchair</td>
<td>9.15</td>
<td>2.67</td>
<td>13</td>
</tr>
<tr>
<td>Disability Wheelchair</td>
<td>10.42</td>
<td>1.44</td>
<td>12</td>
</tr>
</tbody>
</table>
APPENDIX I

TABLE 7

EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE
OF EXAMINEE GENDER ON ARITHMETIC SUBTEST
Table 7. Examination of cell means for independent variable of examinee gender on Arithmetic subtest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Examinee</td>
<td>10.63</td>
<td>2.33</td>
<td>49</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheelchair</td>
<td>10.65</td>
<td>2.04</td>
<td>26</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>10.27</td>
<td>2.24</td>
<td>11</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.93</td>
<td>1.91</td>
<td>15</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No wheelchair</td>
<td>10.61</td>
<td>2.66</td>
<td>23</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>10.67</td>
<td>3.63</td>
<td>12</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.56</td>
<td>1.04</td>
<td>11</td>
</tr>
<tr>
<td>Female Examinee</td>
<td>9.64</td>
<td>1.99</td>
<td>52</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheelchair</td>
<td>9.56</td>
<td>2.00</td>
<td>25</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.69</td>
<td>2.63</td>
<td>13</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>9.42</td>
<td>1.08</td>
<td>12</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No wheelchair</td>
<td>9.70</td>
<td>2.02</td>
<td>27</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.50</td>
<td>1.83</td>
<td>14</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>9.92</td>
<td>2.25</td>
<td>13</td>
</tr>
</tbody>
</table>
APPENDIX J

TABLE 8

EXAMINATION OF CELL MEANS FOR INDEPENDENT VARIABLE OF EXAMINEE GENDER ON PICTURE COMPLETION SUBTEST

<table>
<thead>
<tr>
<th>Gender</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>85</td>
</tr>
<tr>
<td>Female</td>
<td>90</td>
</tr>
</tbody>
</table>
Table 8. Examination of cell means for independent variable of examinee gender on Picture Completion subtest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Examinee</td>
<td>10.20</td>
<td>2.06</td>
<td>49</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheelchair</td>
<td>10.46</td>
<td>1.99</td>
<td>26</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.91</td>
<td>1.05</td>
<td>11</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.87</td>
<td>2.42</td>
<td>15</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Wheelchair</td>
<td>9.91</td>
<td>2.15</td>
<td>23</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>9.83</td>
<td>2.73</td>
<td>12</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.00</td>
<td>1.42</td>
<td>11</td>
</tr>
<tr>
<td>Female Examinee</td>
<td>9.21</td>
<td>2.36</td>
<td>52</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheelchair</td>
<td>9.64</td>
<td>2.12</td>
<td>25</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>8.92</td>
<td>2.43</td>
<td>13</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>10.42</td>
<td>1.45</td>
<td>12</td>
</tr>
<tr>
<td>Disability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No wheelchair</td>
<td>8.82</td>
<td>2.54</td>
<td>27</td>
</tr>
<tr>
<td>Examiner Male</td>
<td>8.50</td>
<td>2.47</td>
<td>14</td>
</tr>
<tr>
<td>Examiner Female</td>
<td>9.15</td>
<td>2.67</td>
<td>13</td>
</tr>
</tbody>
</table>
REFERENCES


