A Study of a Prekindergarten Screening Test: The Project Intercept Inventory

Francis J. Rutowski
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/2243
A STUDY OF A PREKINDERGARTEN SCREENING TEST:  
THE PROJECT INTERCEPT INVENTORY

by

Francis J. Rutowski

A dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Educational Leadership

Western Michigan University
Kalamazoo, Michigan
June 1987

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A STUDY OF A PREKINDERGARTEN SCREENING TEST:
THE PROJECT INTERCEPT INVENTORY

Francis J. Rutowski, Ed.D.
Western Michigan University, 1987

The emphasis on the importance of early childhood education has brought attention to techniques used in determining the readiness of young children to successfully undertake kindergarten programs.

The Project Intercept Inventory is a criterion-referenced prekindergarten screening test that has been used in the Grand Rapids, Michigan public schools for 12 years. The test consists of 21 items distributed among seven developmental areas. The test was locally developed and no previous analyses had been done to determine the reliability and validity of the test.

The information gained from the use of this test has been used in making educational programming recommendations for thousands of entering kindergarten children.

The purpose of this study was to examine the Project Intercept Inventory for item difficulty, item discrimination, reliability, construct validity, content validity, and criterion validity. An examination was also done to determine the presence or absence of bias in the test items relating to gender and to race.

The data were gathered from the screening results of over 500 children taken in the spring of 1980. This group of children was intended to enter kindergarten in the fall of 1980. Nine hypotheses were developed to aid in examining the screening test.
The study results indicate that: (1) the majority of the test items are appropriately difficult to screen for kindergarten readiness, (b) the majority of test items would aid in discriminating among children's abilities, (c) the test can be regarded as reliable, (d) the test has adequate content validity, (e) the test has adequate construct validity, (f) the test has adequate criterion validity, (g) the test can adequately predict future school success as measured by later educational achievement tests, (h) the majority of test items did not reflect bias due to gender, and (i) the majority of test items did not reflect bias toward majority or minority test takers.

The major conclusion reached was that the use of the Project Intercept Inventory can provide beneficial information for educators for use in preparing educational recommendations that affect kindergarten children as they are about to enter school.
INFORMATION TO USERS

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. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., 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 as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A study of a prekindergarten screening test: The project intercept inventory

Rutowski, Francis John, Ed.D.
Western Michigan University, 1987
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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
DEDICATION

To my wife, Madeline, whose loving support and encouragement provided the motivation to bring this project to fruition.

To my children—be all that you can be.
ACKNOWLEDGEMENTS

The writer wishes to express his sincere appreciation to the various people who contributed to the development of this study.

I am particularly grateful to Dr. Lawrence Schlack, Chairperson of my doctoral committee. His understanding and guidance were instrumental in the completion of this work. In addition, I wish to thank Dr. David Cowden and Dr. Dorothy Bladt, committee members, for their scholarly advice and assistance. Their encouragement and suggestions proved to be most valuable.

The writer wishes to express thanks to all respondents who participated in this study.

A special note of gratitude is made here for the input and encouragement provided by Mr. Robert Korte. He and other staff members of the Curriculum Planning and Evaluation department of the Grand Rapids Public Schools provided a great deal of support and technical assistance for this project.

Lastly, I am greatly indebted to my parents who insisted that their children constantly strive to attain ever-higher goals.

Francis J. Rutowski
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ............................................................... ii

**LIST OF TABLES** .......................................................................................... v

**CHAPTER**

I. THE STUDY PROBLEM

Need for the Study ............................................................. 1

Problem Statement ............................................................ 2

Consideration ............................................................. 4

Definition of Terms .......................................................... 5

Research Hypotheses ....................................................... 6

II. REVIEW OF SELECTED LITERATURE

Introduction .................................................................... 11

Screening of Preschool Children .................................... 13

Factors Important in School Readiness ......................... 18

Some of the Effects of Prekindergarten Screening .............. 19

Negative Aspects of Preschool Screening ....................... 21

Provision of Programs .................................................... 26

International Research ..................................................... 28

Requirements of Prekindergarten Screening Instruments .... 31

Need for Valid and Reliable Instruments ......................... 35

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

The Criterion-Referenced Approach ......................... 40
The Project Intercept Inventory ......................... 42
Summary ................................................. 43

III. REVIEW OF SELECTED LITERATURE

Introduction ................................................. 46
Description of the City and School Populations ......... 47
Description of the Sample and Data Collection ........ 49
Project Intercept ........................................ 51
The Project Intercept Inventory ....................... 52
The Analyses Method ..................................... 55

IV. RESEARCH FINDINGS

Introduction ................................................. 70
Statement of Hypotheses and Findings ..................... 71

V. SUMMARY AND CONCLUSIONS

Introduction ................................................. 127
Chapter Summarizations .................................. 128
Limitations of the Study .................................. 139
Implications of the Study ................................ 141
Final Comments .......................................... 144

APPENDIX .................................................. 147

BIBLIOGRAPHY ............................................. 153
# LIST OF TABLES

1. Item Difficulty Analysis ........................................... 73
2. Item Discrimination Analysis ........................................ 77
3. Factor Analysis Results .............................................. 86
4. Gender Item Characteristic Curve Item 1-
   Identify Human Body Parts ........................................ 92
5. Gender Item Characteristic Curve Item 2-
   Balance Beam Walking .............................................. 93
6. Gender Item Characteristic Curve Item 3-
   Jumping ..................................................................... 94
7. Gender Item Characteristic Curve Item 4-
   Hopping .................................................................... 94
8. Gender Item Characteristic Curve Item 5-
   Folding a Paper Triangle .......................................... 95
9. Gender Item Characteristic Curve Item 6-
   Paper Cutting ............................................................ 95
10. Gender Item Characteristic Curve Item 7-
    Circle Drawing .......................................................... 96
11. Gender Item Characteristic Curve Item 8-
    Square Drawing .......................................................... 97
12. Gender Item Characteristic Curve Item 9-
    Triangle Drawing .......................................................... 98
13. Gender Item Characteristic Curve Item 10-
    Open-Ended Square and Circle Drawing ....................... 98
14. Gender Item Characteristic Curve Item 11-
    Identifying Colors ..................................................... 99
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Gender Item Characteristic Curve Item 12-Identifying Geometric Shapes</td>
<td>100</td>
</tr>
<tr>
<td>16</td>
<td>Gender Item Characteristic Curve Item 13-Positions in Space</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>Gender Item Characteristic Curve Item 14-Visual Recognition of Big and Little</td>
<td>101</td>
</tr>
<tr>
<td>18</td>
<td>Gender Item Characteristic Curve Item 15-Sequencing Objects</td>
<td>101</td>
</tr>
<tr>
<td>19</td>
<td>Gender Item Characteristic Curve Item 16-Naming Objects From Memory</td>
<td>102</td>
</tr>
<tr>
<td>20</td>
<td>Gender Item Characteristic Curve Item 17-Repeating Digits From Memory</td>
<td>102</td>
</tr>
<tr>
<td>21</td>
<td>Gender Item Characteristic Curve Item 18-Repeating Directions</td>
<td>103</td>
</tr>
<tr>
<td>22</td>
<td>Gender Item Characteristic Curve Item 19-Following a Series of Directions</td>
<td>104</td>
</tr>
<tr>
<td>23</td>
<td>Gender Item Characteristic Curve Item 20-Verbal Associations</td>
<td>105</td>
</tr>
<tr>
<td>24</td>
<td>Gender Item Characteristic Curve Item 21-Drawing a Human Figure</td>
<td>106</td>
</tr>
<tr>
<td>25</td>
<td>Corrected Point-Biserial Results-Gender</td>
<td>107</td>
</tr>
<tr>
<td>26</td>
<td>Race Item Characteristic Curve Item 1-Identify Human Body Parts</td>
<td>110</td>
</tr>
<tr>
<td>27</td>
<td>Race Item Characteristic Curve Item 2-Balance Beam Walking</td>
<td>111</td>
</tr>
<tr>
<td>28</td>
<td>Race Item Characteristic Curve Item 3-Jumping</td>
<td>112</td>
</tr>
<tr>
<td>29</td>
<td>Race Item Characteristic Curve Item 4-Hopping</td>
<td>113</td>
</tr>
</tbody>
</table>
List of Tables--Continued

30. Race Item Characteristic Curve Item 5-
Folding a Paper Triangle ............................................................... 114

31. Race Item Characteristic Curve Item 6-
Paper Cutting .................................................................................. 115

32. Race Item Characteristic Curve Item 7-
Circle Drawing ................................................................................ 115

33. Race Item Characteristic Curve Item 8-
Square Drawing ................................................................................ 116

34. Race Item Characteristic Curve Item 9-
Triangle Drawing ............................................................................... 116

35. Race Item Characteristic Curve Item 10-
Open-Ended Square and Circle Drawing .......................................... 117

36. Race Item Characteristic Curve Item 11-
Identifying Colors ............................................................................ 117

37. Race Item Characteristic Curve Item 12-
Identifying Geometric Shapes .......................................................... 118

38. Race Item Characteristic Curve Item 13-
Positions in Space .............................................................................. 118

39. Race Item Characteristic Curve Item 14-
Visual Recognition of Big and Little ................................................... 119

40. Race Item Characteristic Curve Item 15-
Sequencing Objects ........................................................................... 120

41. Race Item Characteristic Curve Item 16-
Naming Objects From Memory .......................................................... 120

42. Race Item Characteristic Curve Item 17-
Repeating Digits From Memory ......................................................... 121

43. Race Item Characteristic Curve Item 18-
Repeating Directions .......................................................................... 121
List of Tables—Continued

44. Race Item Characteristic Curve Item 19-
   Following a Series of Directions ........................................... 122

45. Race Item Characteristic Curve 20-
   Verbal Associations ............................................................. 122

46. Race Item Characteristic Curve 21-
   Drawing a Human Figure ..................................................... 124

47. Corrected Point-Biserial Results-Race .................................. 126
CHAPTER I

THE STUDY PROBLEM

Need for the Study

An educational concern which has been noted for a very long period of time in American schools has been the early failure of a substantial number of children as they begin the formal schooling process. The term failure as used here denotes a child's inability or unwillingness to perform in a satisfactory fashion in mastering the basics of early preacademic skills.

This concern is noted on national, state, and local levels. As a result of this interest serious attempts have been made to prevent such early failure for children entering kindergarten. Important among these attempts are the processes by which educational decisions are frequently made early in the young child's school career. These decisions, usually made by professional educators in conjunction with the children's parents, affects the child's entrance into formal schooling. The decisions also influence the immediate and subsequent progress that the individual child makes within the educational system (Hall, 1963; Ilg & Ames, 1972; Salvia & Ysseldyke, 1978).

The emotional well-being of children and parents is at risk in this regard. There are also legal and financial responsibilities which school districts must consider. There is great interest and concern relating to early educational decision making for children, for many reasons.

These decisions are important because school districts must provide appropriate school programs for the children involved. These responsibilities
go further to include the need to meet the demands of federal, State, and local rules and regulations. There are laws which dictate provision of services to children. An example would be the federal Education of the Handicapped Act (Public Law 94-142) passed in 1975, and the State of Michigan's Public Act 198 (the Mandatory Special Education Act) which was passed in 1971, and revised in 1977, 1980, and 1983. A further requirement which must be met is the local rule in most States that children must be enrolled in school by the time that they reach age six years.

Problem Statement

The major purpose of this study was to examine the reliability, validity, and possible presence of bias of a prekindergarten screening instrument. The instrument under study is the Project Intercept Inventory (PII), a copy of which is included in the appendix. This Inventory has been used in the Grand Rapids (MI) public schools for the past 13 years. There has never been research done to show the predictive capability of this instrument. The Grand Rapids Public Schools (GRPS) plan to continue the use of this instrument, and must have data to demonstrate that the instrument does predict what is claimed for it. It must also be consistent in what is expected of it.

The problem considered in this dissertation is the examination of the capabilities of the PII. This was done by comparing prekindergarten screening results gained from the PII with pupil performance results in the fourth grade.

The instrument under study is used in the Grand Rapids public school system, but the concerns generated by the application of results gained from the use of such screening instruments are much more widespread. Since
the emotional well-being of children and parents is at risk, as well as the financial and legal responsibilities for providing appropriate educational programs for children, the interest and concern regarding early educational decision making for children is great. It is shared by parents, educators, and local boards of education.

Such early decisions are also important because they directly relate to the financial responsibilities of the local school district which must provide appropriate school programs for the children. The responsibility of the school district also entails legal obligations, since federal, State, and local laws and regulations require that adequate and appropriate schooling be provided. Among such laws are the federal Education of the Handicapped Act (Public Law 94-142) passed in 1975 and the State of Michigan's Public Act 198 which was passed in 1971 and revised in 1977, 1980, and 1983. Another factor which must be considered here is that, while it is true that not every state requires that children attend a kindergarten program, the great majority of states do stipulate that children must be enrolled in a formal school program by the time the children reach a specified age, frequently 6 years of age.

To aid in this decision-making process many school districts use basic standardized and published or locally developed measurement devices. This is done to assist in determining if a child is ready in a global sense to successfully undertake a kindergarten program (Abbot & Crane, 1977; Gesell & Ilg, 1946; Ilg & Ames, 1972; Rogers, 1982). If the child is considered ready for kindergarten, then there usually is little apparent reason for concern. If the child is felt to be unready, then decisions are often made regarding entrance into a special or alternative program. The child may also be recommended to remain at home for another year.
If the choices made at this point are accurate in terms of meeting the needs of the child, much good can be accomplished for the child, parent, and school district. If poor or inaccurate decisions are made, then it is possible that these same parties would suffer from the consequences of these inaccurate judgments.

With these consequences in mind, it was suggested that the basis for many of these decisions was the screening or evaluation instruments used by the educators to aid in making their conclusions. These instruments can become key components in the early development of the whole procedure put in place to make these educational determinations. The results of the administration of these tests are used to support and substantiate the recommendations made by the educators during the screening of the children for school entrance. From this point, the recommendations regarding the children's educational program placements may be made immediately. The evaluation process can also be extended to gain more information. However, the process by which the decisions will be made is well underway. It can be seen that the screening instruments become an important factor in the total process. They can be considered pivotal.

Considerations

The entrance into formal schooling is a positive and successful experience for many children. Note should be taken of the use of the phrase many children. Not all of the children who enter kindergarten will be ready to successfully begin this process. Some of them will not be ready in a developmental sense. Socially, motorically, cognitively, experientially, and in a number of other ways, some children will not be ready to undertake the
skill acquisition and learning processes that will be expected of them. Some of these children will fail to meet the expectations set for them. They will begin their formal educational process with failure as the first result of a school experience.

A substantial number of school districts in this country have established procedures intended to eliminate the incidence of these children who are programmed for early failure within the school system. A frequent strategy used for this purpose is the prekindergarten screening process. Such a process most often employs screening instruments of a standardized or nonstandardized nature.

With so much at stake, it is imperative that school districts be fully aware of the effectiveness and accuracy of the screening tests they use in this process. It is necessary that the instruments used in the screening procedure are reliable and valid. Kerlinger (1973, p. 442) indicates that the reliability of a diagnostic instrument is its accuracy of precision, while its validity is determined by the degree to which the instrument measures what it purports to measure (Kerlinger, 1973, p. 457). It is also necessary that such a screening device be able to accurately predict the global readiness of the individual child to successfully undertake a kindergarten program. It should also be able to provide the early warning signals that school readiness has not yet been reached by the child.

Definition of Terms

Terms and phrases frequently used in this study are defined below.

Device: the screening and/or evaluation instrument used in the evaluation of children; each is standardized or nonstandardized.
Educational program placements: school settings in general education, special education, or an alternative setting, e.g., prekindergarten.

Formal school process: the educational cycle which begins with the kindergarten year; many children enter the kindergarten program when they have reached their 5th year of chronological age.

Global: taken as a whole without attempt to separate parts or functions, as in the global readiness of a child to undertake formal schooling.

Kindergarten failure: the circumstance in which children who attend kindergarten but who progress at a less than satisfactory rate, for whatever reasons, and who are recommended by school personnel for retention in kindergarten. This term will also include those children who are advanced to the first grade but who have not acquired the necessary skills to succeed at that grade level.

Screening instruments: the tests used in the process of surveying children for kindergarten readiness; for the purposes of this study, this term referred most often to the Project Intercept Inventory (PII).

Research Hypotheses

This study undertook an examination of the reliability, validity, and the possible presence of bias of a prekindergarten screening instrument. The instrument involved was the Project Intercept Inventory (PII).

To examine the PII the following hypotheses were developed.

**Item Difficulty**

Hypothesis 1. The p-value of individual items of the Project Intercept Inventory will be less than 0.2, greater than 0.8. (Null hypothesis)
Alternative. The p-value of individual items of the Project Intercept Inventory will be greater than 0.2 and equal to or less than 0.8.

**Item Discrimination**

**Hypothesis 2.** The individual test items of the Project Intercept Inventory will have correlation coefficients less than 0.15 when compared with total test scores, as measured by the corrected point-biserial technique for dichotomous items, and the corrected Pearson product-moment correlation coefficient for nondichotomous items. (Null hypothesis)

**Alternative.** The corrected point-biserial correlation coefficients between individual items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with dichotomous items.

**Alternative.** The corrected Pearson product-moment correlation coefficients between individual test items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with nondichotomous items.

**Reliability**

**Hypothesis 3.** The alpha coefficient for the total test of the Project Intercept Inventory will be less than 0.6. (Null hypothesis)

**Alternative.** The alpha coefficient for the total test of the Project Intercept Inventory will be greater than or equal to 0.6.
Validity

**Hypothesis 4.** There will be no significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique. (Null hypothesis)

**Alternative.** There will be significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique.

**Hypothesis 5.** No more than one factor of the Project Intercept Inventory will be found with an eigenvalue equal to or greater than one, based upon the application of factor analysis. (Null hypothesis)

**Alternative.** There will be two or more factors of the Project Intercept Inventory found with an eigenvalue equal to or greater than one, based upon the application of factor analysis.

**Hypothesis 6.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test (1978) taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

**Alternative.** There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test (1978) taken 4 years later by the
children in the study, based on the application of the Pearson product-moment correlation coefficient.

Hypothesis 7. There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test (1978) taken 4 years later by the children in the sample, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

Alternative. There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient.

Bias

Hypothesis 8. The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for boys and girls. (Null hypothesis)

Alternative. The general shapes of the item characteristic curves of the Project Intercept Inventory for boys and girls will not be substantially different.

Hypothesis 9. The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for racial majority and minority children. (Null hypothesis)
Alternative. The general shapes of the item characteristic curves for the Project Intercept Inventory for racial majority and minority children will not be substantially different.

In Chapter II, Review of Selected Related Literature, the examination of research pertinent to this project will be undertaken.
CHAPTER II

REVIEW OF SELECTED LITERATURE

Introduction

It is thought by educators that success in formal schooling is often highly dependent upon a good start. Beginning on a positive note allows the individual child to interact with his or her peer group, relate to a new situation in which a strange adult (teacher) becomes a major factor in the child's life, and allows the development of a positive feeling of self-worth. This initial involvement will also prepare the child for the process of gaining the basic academic skills which will be needed in the elementary grades.

The need for this successful start in schooling is important because the need for education in our general society is known to be imperative. Regardless of the socioeconomic status of the family unit, all elements of the American public will acknowledge the basic premise that education is a major need to be successful in earning a living and providing for one's family and one's self. The vast majority of American children are exposed to a great deal of statistical evidence which indicates to them, and to their parents, that the thriving individual will be the one who has received a good education. This impression is emphasized in modern living with a highly technical age upon us. America as a nation is well advanced into the age of technology, often referred to as the postindustrial age.

This type of information is not new. Americans are being made aware at an ever-increasing rate that education is more important now than it has
ever been in our history. Every segment of our nation, regardless of culture, has an urgent need to gain and develop an educational fund. At the same time, the information presented by many public school districts in this country indicates that the numbers of school dropouts have continued to grow and that they remain a national disgrace. With the urgency of an insatiable need for better educated young people, it has become glaringly obvious that strategies of a positive and direct nature must be implemented to remedy this situation.

Some strategies being used in school districts throughout the United States speak to this problem on a long-range basis. The thinking in this regard goes back to the beginning in the sense that a closer look is being taken in many school districts at the children who will be entering kindergarten classrooms. These appraisals are intended to take a careful look at the readiness of the children to successfully undertake a school program. Lichtenstein and Ireton (1984) point out that school readiness is the consideration of a wide range of skills and behavior needed by children to meet the demands of a school situation. Bloom (1976) indicates that children, in terms of an attitudinal impression of entering school for the first time, tend to be similar in their outlooks. Differences are influenced primarily by parents, siblings, and peers. Other writers (Ames, 1967; Moore & Moore, 1975; Rudolph & Cohen, 1964) stipulate the need for viewing the developmental progress of the child prior to entrance into formal schooling at the kindergarten level. Kagan and Klein (1973) point out that most children are able to make adjustments to schooling, and will learn what is expected when they are placed in situations which are appropriate for them. When the demands placed upon children are commensurate with their abilities, they
will perform at their best. Furth and Wachs (1974) argue that children are born with greatly differing dispositions, and they they will enhance their strengths and modify their weaknesses when they are allowed to do so in the fullness of time and their natural development.

With such an approach the excessive number of school dropouts could be inhibited. Children placed in school programs for which they are ready have an improved opportunity to perform at a successful rate. They will go further in their formal schooling.

Screening of Preschool Children

Individual assessment can be done with individual children prior to school entrance. This should be done to determine the capabilities of the individual child for a successful kindergarten experience. Cohen (1972) refers to Piaget's postulate (Piaget, 1952) that certain kinds of concepts cannot be understood by children before some degree of maturing has taken place, no matter how much teaching of these concepts is done. The exact point at which the stages or sequences in this developmental pattern take place varies with individual children.

Assessment to aid in making these determinations is not new in the field of education. Assessments of many kinds are made every school day. The origins of many of these evaluative techniques go back for decades. The assessment of individual strengths and weaknesses, for example, is considered a primary endeavor in the fields of education and psychology. Tyler (1970) indicates that human beings differ from one another in a variety of ways, and that individual assessment can aid the individual in the effort to function effectively in society. The type of individual evaluation that
relates to consideration for entrance into a successful kindergarten experience, however, is a newcomer to the field. Halpern (1978) points out that screening of these young children is increasingly becoming a major activity throughout the nation. She suggests that high standards for screening must be established and maintained if the process is to be successful. Perlman (1981) strongly argues the importance of conducting longitudinal studies in the public schools pertaining to the subsequent progress of the children who are screened. This is based on their studies in the Chicago public schools.

There is considerable awareness of what children should be able to accomplish, and the skills and developmental levels which should be demonstrated at particular stages of life. Gesell and Ilg (1946), for example, identify behavior characteristics of the young child at various ages of life. Another aspect of this same concept is depicted by Harris (1963) as he describes the drawings of children and the distinct developmental features that are demonstrated in a largely uniform manner at various ages. While the most rapid period of overall development in children is noted in the years from 1 to 5, Bloom (1964) points out that individuals do not change in a uniform fashion as they encounter the same environment, due primarily to their innate differences and uniqueness. Klausmeier and Goodwin (1966) agree, stipulating that readiness for school among children of the same age for particular learning tasks varies greatly because of differences among the children in the cognitive, psychomotor, and affective areas of their development. These offerings are supported by Moore and Moore (1975).

Children are different, and it is these differences that educators must discover in the individual child if they are to provide the best possible
opportunity for education for each child. The individual child presents a complex system of heterogeneities to the educator who is charged with the task of teaching. Precisely because these dissimilarities are present, the educator needs to be aware of them in the specific way presented by each child. Screening for school readiness has become a prevalent tactic for helping to discover these individual differences in children. This pertains especially to global development of the normal child. It has been inferred (Zeitlin, 1976) that the screening process in many parts of the country had quickly involved itself with younger and younger children, as educators became more convinced of the need for awareness of individual differences in children as they entered kindergarten.

Souweine (1980) found that screening detected problems which demanded remedial programming as children advanced in school. Such findings are well received by many educators since there are a number of reasons for determining the early detection of real or potential learning difficulties in children. Ilg and Ames (1972) found that teachers who receive these children should have time to obtain the screening information and to utilize it in preparing for their teaching responsibilities. Other than the immediate learning welfare of the children, such practical reasons as the legality of the matter have to be considered. Governmental agencies have indicated that the States and local boards of education must make a concerted effort to detect handicapping conditions in children at the earliest possible time. Lombard (1980) noted that this type of mandate is in place, and also points out the need for greater awareness on the part of people who operate prekindergarten screening programs of the physical readiness needs of children.
Magliocca, Rinaldi, Crew, and Kunzelmann (1978) support such techniques as a frequency sampling to gauge school readiness. They feel this could produce adequate predictive validity by means of a screening process. It is pointed out by Reynolds (1979) that early childhood assessment is necessary to aid in demonstrating measurable gains of specific intervention programming. Other writers suggest that there are other benefits to be gained from early screening of children. A study done by DeHirsch, Jensky, and Langford (1966) indicated that the detection of distinct and identifiable patterns of perceptomotor and oral language deficits are predictive of difficulties in reading, spelling, and writing in later schooling. Oberklaid (1978) reported on the inclusion of a medical/physical component to a prekindergarten screening project. This produced effective results in identifying children at risk from a medical standpoint. The results of the study tended to correlate at an acceptable level with the psychological assessment devices used in the same project.

Further support for prekindergarten screening is noted by Berger and Perlman (1973). Their research over a 3-year period with a prekindergarten screening project indicated a good success factor in terms of an intervention strategy. Casey (1976) reported on the development of a prekindergarten screening device developed by school personnel which could be used to detect the educational needs of children entering kindergarten. The device demonstrated significant positive correlations with two published instruments. Similar results were announced by Moore and Sunal (1978) as they studied the findings of a teacher-made device. Their outcomes were not considered as conclusive due to the small number of children involved in the study. But it was observed that children who performed low in certain
skill areas also exhibited deficiencies in the same skill areas in the classroom.

The study and work done in this area is necessitated in part by governmental mandates at several levels (e.g., federal Public Law 94-142). These require that Child Find operations be put into place by States and local school districts. The purpose of these operations is to identify as early as possible those children who are experiencing handicapping conditions, regardless of variety or origin. The next step is to provide the appropriate educational support for these individuals (Reynolds, 1979). The implementation of prekindergarten screening projects can be seen as a consistent extension of the governmental injunctions for Child Find programs (Sahin, 1978).

Further implications can be seen in the area of general education. Teachers and building administrators who will receive these children as kindergarten students need as much information regarding developmental levels and the capabilities of the individual children as possible. Parents often seek the reassurance that their children are ready to enter school, and that they can anticipate that the children will experience a positive entrance into school.

For those children who are determined via the screening process to need more time to develop and mature before entering school, the general education personnel often have alternative programs to offer. The number of children in need of such a program must also be made known to general education. Appropriate provision must be made programmatically for these children.
Factors Important in School Readiness

Researchers have developed a substantial body of information regarding the factors that are considered relevant for considering a child as ready to begin kindergarten. The importance of these factors is summarized by Gardner (1964), indicating that the developmental level of the typical 5-year-old child is a special stage in the process. The children at this point are ready for a less formal experience than one that requires instruction in the basics of reading and writing. Elkind (1976) discusses the Piagetian emphasis upon the concept of optimal time when considering the growth of certain abilities. To attempt to hurry skill development in young children can often produce disastrous results for the children.

Researchers have spent considerable periods of time in attempting to isolate the factors that are considered necessary to normal early child development. The application of these considerations to readiness for kindergarten is also considered. A case in point is the reporting by Ilg, Ames, Haines, and Gillespie (1978). They pointed out that the Gesell preschool tests have a long history, and they contend that four facets of human behavior are most important for measurement purposes at the school entrance level. These areas are motor, adaptive, language, and personal-social development.

Various sources of information regarding the factors attempted to be measured by standardized and published preschool screening devices indicate that many elements are considered by the authors of such instruments. Buros (1978) provides a source for extensive review of such material. Davidson, Lichtenstein, Canter, and Cronin (1977) as well as Strully (1977) elaborate upon the factors that are to be measured by many of these devices.
The number of factors vary greatly, and no single instrument can be considered to measure more than a few. Commonalities can be found, however. The factors which were most often listed as being included in a single screening tool were: language, memory skills, math concepts, fine motor skills, gross motor skills, perceptual functioning, cognition, and social consciousness.

Other skills which were enumerated occasionally were drawing a human figure (boy or girl), self-help capabilities, and various copying tasks with pencil and paper.

Some of the Effects of Prekindergarten Screening

The chief effect of the prekindergarten screening process is that decisions are made that will directly impact upon the educational lives of the children involved. The families of the children, the school district, and the community are also affected. Most of the decisions that will be made will pertain to whether or not children who have reached the necessary chronological age dictated by local custom or regulation are ready to actually begin school. There will also be situations in which parents will seek early entrance into school for their child. This situation relates to children who have not chronologically come of age for school entrance; however, the parents believe the children are ready to undertake the kindergarten program. Since school districts in most of the United States will not receive State aid for early entrants, the school districts are reluctant to allow such admission.

There are other reasons why such children should not be entering school at this point. They would be involved in the classroom environment with peers who, in some cases, would be as much as 1-year older and
more mature. While Green and Simmons (1962) indicate that even 5-year-old children can often profit from formal instruction, they also point out the need to be aware of the individual differences in these young children. DeHirsch (1957) states that children between 5 1/2 and 6 1/2 years make dramatic advances in overall maturation. To push the child into formal schooling prematurely is often a mistake for which the child suffers. Kephart (1971) contends that children go through developmental phases in both a qualitative and quantitative manner and that both aspects of the process are important and necessary. They cannot be hurried.

Carll and Richard (1977) relate numerous findings which tabulate the damage that can occur to the child who is not ready for school, but who is enrolled (the overplaced child). King (1955) would agree with this stance. Her study suggested that having attained a few additional months of chronological age before attempting formal school entrance is an important factor in a child's ability to meet the restrictions and tensions that formal schooling will impose. Hampleman (1959) contends that children will have a better chance for success in reading by starting to school a few months later, rather than a few months earlier. Flavell (1963) states that the child becomes noticeably more testable (in formal experiment) from age 4 or 5 years and up. The child, in Piaget's late preoperational years, also becomes capable of reasoning about progressively more complex and extended problems or displays in the testing circumstance. Almy (1966) points out that the major difference between the kindergarten-aged child and the child attending a middle grade is the matter of the way they organize and systematize, or, in the case of the younger children, fail to organize systematically the things they experience. Carter (1956) and Austin and Postlethwaite (1974) point out
that their investigations indicate that the underage school entrant tends to achieve at a lower level than the child who enters kindergarten at a point of more mature development. These same applications can be made to the situations of children who are chronologically of kindergarten entrance age, but who are not developmentally ready to undertake the schooling process.

Such decisions play an extremely important part in the child's early schooling. When the correct decision is made a great deal of good can ensue and the child stands to benefit. When improper or poorly conceived judgments are made in this process, the situations of these children can become very problematic. This would be true if the wrong decision in screening was instrumental in having the child placed in an incorrect school program, for example. This can play a major part in the application of a label to the child, an issue which is very sensitive for many parents and educators.

Lichtenstein (1982) states that the inevitability of identification errors is not typically considered in designing an early identification program. Yet errors can be made in this process. This can result in the erroneous labeling of children. Coupled with this is the concern of false positives and false negatives. Hobbs (1975) points out that the false positive circumstance effects the child who is declared to have a disorder, when in fact the child does not have a disorder. A false negative relates to the reverse, and a child is deemed to be free of any disorder when in fact a disorder or handicapping condition is present.

Negative Aspects of Preschool Screening

It is also seen that not all educators are convinced that screening for the purpose of detecting developmental deficiencies at an early age is a sound
procedure. Divoky (1977) is critical of screening children. She offered that this identification process consumes time and resources that might better be used in the performance of teaching. Kermoian (1962) advocates teacher appraisals. This procedure would eliminate readiness tests and rely primarily upon the experience and observational abilities of the classroom teacher. The argument would be that such an approach would save time, money, and effort for the school district. Lichtenstein (1982) would further support this opinion. He suggested that the extensive assessment of young children at any single point in time is wasteful of the school district's resources.

Another approach to obtaining beneficial information regarding the readiness of a particular child for entrance into the kindergarten program would be the use of parent questionnaires. This alternative to a screening battery was offered by Colligan (1976). A later study offered by the same writer (Colligan, 1982) again supported the efficacy of parent questionnaires and indicated that this technique was useful and economical.

Wendt (1979) put forward the opinion that the screening process can provide much anxiety for children and parents due to the suggestion that the educators are questioning the abilities of the children. This could develop a negative reaction on the part of the parent and child as regards entrance into formal schooling. Wendt further intimated that the process is too costly in effort and money to the school districts, and that very few good quality prekindergarten assessment programs are in existence at this time. Wendt (1978) further presented that testing for readiness is a questionable practice. He felt that it could provide negative public relations for school districts.

There is no denying that parents do have the legal right to enter their children in school at the kindergarten level when the children attain the
legally stated chronological age. State laws are formulated to require that children are entered into school by a certain age. This conforms to the legislation relevant to mandatory public education.

Parents may accept or reject any recommendations made to them by a school district's educational personnel following their child's involvement in a prekindergarten screening process. A frequent rationale offered by parents who are unhappy with such recommendations is that the screening process was too brief, the child was having a bad day, or the like, and the screening results were not accurate. This reaction is often followed by the indication that the parents wish to give the child a chance to attend kindergarten and then to see what develops.

A frequent concern voiced is that the instruments used in the screening process are not precise, appropriate, or conclusive. Salvia and Ysseldyke (1978) point out that readiness tests must meet the highest technical standards. Wendt (1978) cautions that many such devices are being newly developed, but that many of them will be adopted for use without consideration for philosophical issues which underly assessment. Studies in the area of screening instrument development are cited in a study done by LaCrosse et al. (1970). The writers conclude that more tests of proven validity and reliability are needed in all domains of early child development. Shepard (1986) contends that current tests used in the kindergarten screening process cannot make highly accurate assessments of school readiness. A popular test used in prekindergarten screening in many school districts is the Metropolitan Readiness Tests (MRT) (Nurss & McGauvern, 1976). A study by Ravitch (1985) found it to be among the technically best measures available for screening purposes. However, the MRT was designed for aiding teachers in organizing instruction, not for detecting kindergarten readiness.
Another factor that must be considered in this regard are cultural differences of children. Legal implications that may bear on the local board of education, its administrative staff, and the personnel in the field who are actually involved in the screening process with children must also be considered.

There are still some concerns of a financial nature to be considered. The school district does fund all or most of the expense which accrues to operating the prekindergarten screening process. The largest expense is that of employing the necessary professional personnel, without whom the operation could not take place. This additional expense can be cited as a negative factor when the entire screening process is considered.

Parents may also be involved in additional financial expense in this situation. Where the opportunity is available, individual parents have opted to place their child in a nonpublic (private or parochial) school and pay the additional expense of tuition, books, and fees. This is considered preferable to accepting the recommendation of the public school personnel.

While some educators advocate additional growth time for children who are deemed immature or unready for kindergarten, some studies (May & Welch, 1984; Shepard & Smith, 1985) indicate little difference is subsequently made for children after a few years of schooling. They contend that achievement levels become nearly equal, and cite the negative factor of emotionally harming the child by not allowing the child to enter school with his or her peers.

There are contentions made (Gredler, 1980; Holmes & Matthews, 1984; Miller & Norris, 1967) that the preschool screening process may prove more harmful than helpful to some children. The studies cited indicate that
teacher biases regarding the entrance age of the child may prove to be more influential in the child's early school progress. If the teacher sees the child as too young or immature to successfully meet the curriculum objectives, the child is more likely to be recommended for retention.

Another concern in this same area is that of the referral of children for special education programs, due to their young age or seeming unreadiness. Pugach (1985) states that often children are placed in special education based on teacher referral and that the teachers were actually seeking one-to-one instruction for the child without being aware of the scientific and legal conception of handicap. Thus labels can be placed on children which have dubious validity, and may be totally inaccurate in some cases.

Trotter (1975) points out some negative aspects of the labeling concern. He indicates that labels placed on children in the assessment process are boundary markers for the professions and that the label a child receives, especially in the Special Education procedure, often depends on the professional identity of the person attaching the label. The contention has been made that teachers often have lowered expectations for children who have been labeled during the screening process. The labels are frequently viewed in a negative fashion. It is further pointed out that acceptance by children of certain of these labels as true of themselves often results in lowered school attitudes (Jones, 1972).

Goldstein, Arkell, Ashcroft, Hurley, and Lilly (1975) indicate that labeling becomes something of a necessary evil due to the requirements of federal, State, and local systems. A child must receive a diagnostic label of some kind in order to be eligible for Special Education programs, as well as other types of alternative programming. This situation exists because all the states...
use categorical labels in their statues and prescribe certain minimum standards and guidelines for diagnosis, placement, facilities, and finance. The perspective of the child who receives a label as a result of an assessment process is considered by Guskin, Bartel, and MacMillan (1975). They submit that, while there is no simple predictable consequence of labeling for the individual, many of the possible consequences tend to be negative for the child. These consequences range from ignorance of the implications of a label which indicates that the child is different, to the inability of the person to divest himself or herself of the label. The authors acknowledge that, with accurate labels in place, children are able to gain programmatic support and other types of intervention of a beneficial nature.

Mislabeling may be seen as a serious consequence of making a wrong decision as a result of a prekindergarten screening process. Another effect could be interference with the child's school progress. When a child is recommended for an additional year of preparation, often the feeling is that the child has failed the kindergarten. A more serious side to this issue is the traumatic implications for child and parent which could develop if a faulty decision is made from the screening process. The emotional damage that can occur in such instances produces serious implications for the child and the various family members who may be involved.

Provision of Programs

It has been frequently pointed out in the literature (Book, 1974; Meier, 1975; Wilson and Reichmuth, 1984) that preschool screening processes are of value only if they lead to effective intervention. Such intervention equates to the provision of the proper educational program, and/or supportive
services. Provision is usually the responsibility of the local educational agency (LEA) in the form of the board of education and the administrative personnel of that school district. A very obvious illustration of this responsibility is the provision of Special Education programs and services. Schaer and Crump (1976) suggest that teacher observation and what they term daily evaluation could readily replace the screening process. They feel that this approach would be more accurate and feasible in determining appropriate programming for children with learning disabilities.

A program designed to work with parents and children from the birth of the children to their entrance into kindergarten is reported by Tivnan and Pierson (1982). The program emphasizes the ongoing screening of the children involved. It also works with the parents in a health care and developmental format. The study is of a longitudinal nature. Its results indicate that program participants, who were early identified as at risk, experienced fewer difficulties once they reached school than the comparison group of children. A similar situation is related by Anderson (1981) in another part of the country. It suggests that this type of preschool health and screening program is beginning to become better established in this country.

Support for this contention is offered by Caldwell (1974) who indicates that early intervention with prekindergarten screening has become recognized and accepted as important. A warning is offered as well. A more conservative approach is needed so that the programs and screening processes will better be able to adequately provide services to children and their families.

Grafft (1980) reported on a frequently used type of program called the developmental kindergarten which has demonstrated positive results.
Other studies have related to successes in varying degrees for these programs (Bower, 1974; Carll & Richard, 1977). Petrone (1976) supported using the kindergarten year as a period for diagnosis-instruction-evaluation, a stance similar to that put forth by Kenney (1969). The utilization of the kindergarten program in this fashion would be a reflection of community need and demand. This would be another method of indicating that parents saw this application of the kindergarten program to be something which they felt would be beneficial to their children.

International Research

Included in this review of the literature regarding the prekindergarten screening process, decision making as it relates to this process, and the construction and use of screening instruments was a survey of work done in other nations. The survey relates to the above noted areas of education. Austin (1976) discussed the basis for preschool programming in England, Wales, Canada, Sweden, France, Italy, Belgium, Germany, and the Netherlands. His results indicate that preschool programming in most of these countries was based in response to the health and welfare needs of poor children. He related that middle-class interest in children's cognitive development and achievement have been only recent concerns in terms of preschool programming. This author indicates that a successful kindergarten experience is an important element in the school progress of children. Deasey (1978) was critical of the provision of preschool programming in the British Isles. Comparisons are made in his study with other Western European countries in regard to the programmatic accommodations provided by the Netherlands, Belgium, and Italy. He states that Great
Britain lags far behind these nations in program provision for preschool-aged children.

A study done in Sweden and reported by Gillberg Rasmussen, Carlstrom, Svenson & Waldenstrom (1982) describes an attempt to screen children attending preschool programs in the city of Gothenburg. The screening was done to locate potential learning difficulties of a possible serious nature in young children. Tyler and Foy (1979) reported on research done with a locally developed assessment guide to be used in preschools and nurseries within the city of Keele in England. This approach used observation as well as the incorporation of other aspects of assessment process into the daily routine of the children.

An Australian study by Rowe (1981) was aimed at providing early screening and assessment of preschool children in order to better identify children at risk to aid teachers and parents to improve the children's educational chances for success. This approach downplayed predicting failure for those children who were felt to be at risk in the educational process. This project required that the educators develop a screening device. Several approaches had been considered. It was reported that individually administered Piagetian tasks, while found to be valid indicators of developmental levels in some areas, required highly experienced teachers as well as expensive equipment. This was considered too time consuming for practical classroom use. It was decided that the locally developed instrument would center on specific skills or other modifiable behaviors which could be translated into specific learning requirements.

In the Netherlands, Hermanns (1979) attempted to identify potential school problems in preschool children in order to gain relevant information for systematic screening and early guidance. The process used a pre and post
testing format. It was concluded that changes in cognitive development were significantly related to socioeconomic status and school. Socioeconomic status was strongly related to whether preschoolers at risk for school problems showed improvement in the preschool period. A similar project was begun when the Canadian North York province of Ontario Board of Education in 1978 required that procedures be developed for early identification of children's learning needs and abilities. Crawford (1982) reported on a longitudinal study done as a result of compliance with this Board directive and indicated that the procedures initiated emphasized observation by the teachers. The results of the study indicated that principals were found to be most knowledgeable regarding the procedures to be used. Primary teachers were more satisfied with the program's results than were kindergarten teachers.

Readiness for school, and the determination of this level of development in children, is frequently considered in the literature (Kleisinger, 1973; Telegdy, 1975). Consideration of the instruments to be most effectively used in this process are discussed. Lynch, Mitchell, Vincent, Trueman, and MacDonald (1982) found that a comparison of results gained from their study indicated that intelligence estimates from the McCarthy Scale of Children's Abilities (MSCA) (McCarthy, 1972) were higher in English than in American children when used with preschool children in an assessment process. These results were obtained with American norms. A similar type of study was reported by Hanson (1980) when a standardized screening device was used with a sample of American and German children. The administration was done in the appropriate language for each child. These results revealed similar patterns of strengths and weaknesses between the two groups.
The impression gained from this review is that of a real awareness and interest in Canada and western European countries in the readiness for formal schooling of preschool aged children, as well as the method used in the assessment process. Frequent mention is made of the need for a successful experience for children in kindergarten in order for them to gain a satisfying and rewarding school situation. Also noted (Rowe, 1981) is the initiative taken by American educators in the early 1960s to point out the need for understanding that children must have gained the appropriate developmental levels in order to be successful in kindergarten.

Requirements of Prekindergarten Screening Instruments

The use of a battery of screening instruments is frequently supported as most effective in prekindergarten screening. However, it is often desirable to consider the use of one, or possibly two, screening tests. A survey of work done in the area of preschool screening instrument development points out that several criteria are considered to be key in this process. That is, the application of the instrument or instruments must meet requirements in the spheres of cost, time needed for administration, scoring, and reporting, and personnel restrictions (Barnes, 1982; Goodwin & Driscoll, 1980; Johnson, 1976; Southworth, Burr, & Cox 1981). While some screening devices are able to satisfy certain of these requirements, by far the majority cannot adequately satisfy all of them. A major factor in this regard is the specific and unique circumstances that are relevant to each situation in which prekindergarten screening is to be done. This could relate to individual school districts, as well as to individual school buildings.
While the pragmatic aspects of screening instruments are necessary and impact directly on the screening procedure, other elements of these documents have much greater importance. They are the reliability, validity, and predictive features of the devices. A great deal of work has been done in considering these elements as they concern specific instruments. Wilson and Reichmuth (1984) report that their review of the literature indicated that there were a considerable number of problems noted in describing screening results. Hayes, Mason, and Covert (1975) point out that many of the tests in current use are too long and involved to be effectively administered to young children. Their study emphasized the need for screening instruments which could broadly assess readiness skills, while at the same time giving indication of good reliability and validity. Pascale (1973) cautioned against the use of assessment devices that were not designed for use with preschool children, but had been modified in an attempt to be applied to the preschool child.

Some instruments used in the area of working with children in screening have received considerable attention and a considerable amount of use, in some cases over long periods of time. An example of this type of device would be the Denver Developmental Screening Test (DDST) (Frankenburg, Dodds, & Fandell, 1968). Frankenburg, Camp, and VanNatta (1971) reported that their work indicated that the DDST correlated well with several criterion devices, while its use placed children in three gross categories. A device used even more frequently than the DDST is the Metropolitan Readiness Tests (MRT) (Nurss & McGauvern, 1976). Rubin (1974) found that results gained with the MRT prior to the children's entrance into kindergarten correlated at a 0.65 rate with results obtained when
the instrument was administered to the same children one year later. Nagle (1979) reported substantial validity of the MRT for predicting first grade achievement as measured by the Stanford Achievement Test (Madden, Gardner, Rudman, & Merwin, 1973).

There are a few instruments that have been utilized in the field of prekindergarten screening for considerable periods of time. But there has been a proliferation in recent years of tests for this purpose. Buttrom, Covert, and Hayes (1976) reported on the Hayes Early Identification Listening Response Test (HEILRT) (Buttrom, Covert, & Hayes, in press) and indicated that it was used with children in small groups, producing good correlations with the MRT. Mogoon and Cox (1969) studied the Screening Test of Academic Readiness (STAR) (Ahr, 1966) and reported that it did not fulfill many of the criteria for a school readiness measure. The device was highly related to IQ, and did little to report on developmental levels.

A study reported by Klein (1980) indicated that there was a moderate test-retest reliability found with the application of the Northwestern Syntax Screening Test (NSST) (Lee, 1969), when used as a portion of a screening battery. Meisels (1984a) reported that the use of the Early Screening Inventory (ESI) (Meisels, 1984b) resulted in moderate to excellent accuracy in predicting academic achievement through the second grade.

Some of the dangers encountered when attempting to utilize current tests which were not designed for preschool aged children for the screening process for kindergarten are discussed by Rogers (1982). He used the Slossten Intelligence Test (SIT) (Slossen, 1981) with preschoolers and found that the device tended to overestimate the performance of preschool children. Incorrect educational recommendations could result.
An area of particular concern involved with the prekindergarten screening process is that of the affective (emotional) domain. Behar and Stringfield (1974a) offer that there are very few devices currently in use which are able to detect emotional difficulties in children in this particular age group. These same authors indicate that their study with the Preschool Behavior Questionnaire (PBQ) (Behar & Stringfield, 1974b) suggest that this instrument was found to have criterion validity and high interrater and test-retest reliabilities. Barker (1975) found that the use of the Preschool Rating Scale (PRS) (Barker, Sandler, Borneman, & Knight, 1975) could classify children as typical or nontypical with a high degree of accuracy.

The Classroom Behavior Description (CBD) (1979) reviewed by Aaronson (1979) points out that another way to approach this aspect of early childhood development is to use the observational capabilities of the classroom teacher. The authors feel that this strategy produced impressive correlations between CBD ratings, cognitive measures, and other behavior ratings. Harper and Richman (1979) found that the Behavior Problem Checklist (BPC) (Quay & Peterson, 1976) indicated that the frequency of behavior problems becomes apparent at an early age and remains consistent through-out the preschool years. This device was reported as being of value in aiding educational professionals to select intervention strategies for the preschool child.

Walker (1973) summed up this aspect of prekindergarten assessment by stating that until the major theoretical questions and issues are answered within a comprehensive theory of socioemotional development, socioemotional measures for young children cannot be meaningfully developed.

Miller and Linder (1982a) state that their work with the Miller Assessment for Preschoolers (MAP) (Miller & Linder, 1982b) indicates that it
is one of the few nationally standardized instruments available that identifies the full spectrum of severe-to-mild delays in preschoolers. This test allows for both screening conclusions and supplemental observations, which the authors contend are important for providing appropriate intervention strategies with children.

In sum, there are very few devices to be found at this time that do what Lichtenstein (1981) has indicated that screening devices have as their explicit purpose. That purpose is the assigning of individuals to an initial classification of being at risk or not at risk.

Need for Valid and Reliable Instruments

A great deal of effort and expense has been expended in recent years to develop preschool screening instruments. This effort is due to the need to fill the void that the demand for screening of prekindergarten children in the public schools has created. It is also due to the influence that these tests bring to the entire process of making educational decisions for children. A great deal of research and development has taken place to develop screening tests which can make the screening process effective. Klein (1977) indicates that there is a need for an effective and efficient method for accurately predicting the academic achievement of preschool children. She further states that there are a number of standardized instruments available at the kindergarten level but that what seems to be missing is sufficient valid information to enable users to make rational choices among the instruments available. An example of a screening device would be the Developmental Indicators for the Assessment of Learning (DIAL) (Mardell & Goldenberg, 1975a). Mardell and Goldenberg (1975b) reported that this instrument was
the result of felt need by the educators involved for a single test for screening prekindergarten children. The test needed to contain the essential elements that would disclose developmental patterns in children. Later, Mardell and Goldenberg (1976) reported that the DIAL, based on the results of a longitudinal study, did identify children who were in need for further evaluation prior to actual enrollment in a kindergarten program. Metrick (1980) reported that a study done in Louisiana indicated the McCarthy Screening Test (MST) (McCarthy, 1970) proved to be more effective than the DIAL and several other instruments for screening purposes. This was based on cost, time, ease of administration, number of children tested, and usefulness of the data obtained. Bondy, Constantino, Norcross, and Sheslow (1982) found that the McCarthy device produced more informative and accurate results than the Slossen (Slossen, 1981) test. However, work done by Mardell-Czudnowski (1980) indicated that the DIAL (Mardell & Goldenberg, 1975a) seemed to be maintaining relatively high levels of concurrent as well as predictive validity.

Dukes and Buttery (1982) declared that a battery of tests must be considered more effective than single readiness tests. They found in their investigation that score correlation between the Meeting Street School Screening Test (MSSST) (Hainsworth & Sigueland, 1969) and the Gesell Developmental Test (GDT) (Ilg, Ames, & Haines, 1964) were at a moderate level, indicating that either test could be used in a screening battery. Ireton, Shing-Lun, and Kampen (1981) supported the descriptive, predictive, and intervention elements of the Minnesota Preschool Inventory (MPI) (Ireton & Thwing, 1975). Their study results found that this multifaceted approach to the screening of preschool children resulted in a convenient means for
obtaining developmental information. It also identified those children who are at risk for kindergarten failure. It was stated that a battery of screening instruments is the better approach to attempting to identify potential learning problems in preschool children. Rubin (1976) found that the Metropolitan Readiness Tests (Nurss & McGauvern, 1976) results when compared with results from the Stanford Achievement Test (SAT) (Madden, et al., 1973) indicated that a greater reliance could be placed on the use of high readiness scores as predictors of good academic performance than on use of low readiness scores as predictors of poor performance. Powers (1974) found that the Vane Kindergarten Test (VKT) (Vane, 1968) lacked sufficient validity for use in assessment and program planning for individual children. Aaronson (1979) indicated that the Classroom Behavior Description (1979) provided results that helped to identify children who could profit from early prescriptive intervention to improve cognitive development and scholastic achievement.

Use of a preschool screening battery takes into account the need to take a global view of the individual child's developmental levels. Few single instruments in use today can provide the broad survey of the child that is needed to gain accurate predictions of kindergarten readiness. A major item in this awareness is the differences of children (Gesell & Ilg, 1946; Hobbs, 1975; Ilg & Ames, 1972, Piaget, 1973) which must be taken into account in order to consider whether or not a child is ready to enter kindergarten.

Gender would be listed as an obvious difference in children. The reasons why young boys seem in general to be less ready to do well in early schooling are many (Anastasi, 1958; Ilg & Ames, 1972; Walker, 1973). A strong contention is that young boys simply are not as developmentally
ready at the customary school entrance age as are young girls (Bentzen, 1963). Some studies, however, suggest little difference in school readiness components between boys and girls (Dappen & Reynolds, 1981; Dziuban & Mealor, 1982; Obrzut, Bolocofsky, Heath, & Jones, 1981) based on results of readiness screening devices which had been employed. Reynolds (1980) found in his research that when common measurement instruments were employed the apparent organization of the abilities underlying performance was constant across both race and sex. His study undertook factor analytic research of a preschool battery used with white and black males and females.

Another variable in terms of the differences in children would be that of race. Adler (1973) describes the need to re-examine the results of the accumulated data which suggests the inferiority of the poor and culturally different child when compared with the results of the majority child. Valencia and Rankin (1983) report that their research indicated that the Kaufman short form version is a suitable screening test for preschool and kindergarten Mexican-American children. Their study indicated that the short form would also be recommended with monolingual Spanish-speaking Mexican-American children. The contention is that there are instruments available which may be used effectively with bilingual (Hispanic) children for purposes of screening for kindergarten readiness.

Jensen (1980) states that ethnic differences do not vary across test items, while status differences do vary across test items. Lord (1977) offers a method of comparison of the results of whites and blacks who are at the same level in terms of verbal skills in working with the issue of cultural bias in assessment instruments. His direction is the use of item characteristic curves for this purpose. Another method is offered by Linn...
and Harnisch (1981). Their approach would be most applicable when only modest sample sizes are available for study. This strategy employs the three characteristic curves for this purpose. Another method is offered by Linn and Harnisch (1981). Their approach would be most applicable when only modest sample sizes are available for study. This strategy employs the three parameter logistic model to identify items which could be biased for members of a particular group. A technique for diagnosing cultural differences and comparing different types of groups was developed by Angoff (1972). This approach involved the construction of a scatter plot for the two groups represented.

Another important factor to be considered is that of socioeconomic status (SES). There are strong contentions (Oakland, 1978) that SES plays a major part in assessment in general, and to screening for school readiness in particular. In looking at several specific screening devices in current use, Oakland pointed out that his findings indicate that one cannot assume that the validity coefficients for the total population adequately represent the validity coefficients for persons from various racial or SES groups. Anastasi (1958) points out that minority group membership tends to be associated with low socioeconomic status and inferior education.

The variable of age is also one that must be considered. The earlier review of material pertaining to this item indicates that the developmentally-ready child will fare better on the results of current screening devices (Carll & Richard, 1977; Flavell, 1963). Children who meet chronological age guidelines for school entry but have not as yet gained the necessary developmental levels, will suffer by comparison (Ilg, Ames, Haines, & Gillespie, 1978).
The Criterion-Referenced Approach

The key to the entire process of screening for kindergarten readiness are the instruments used in the screening procedure. From the interpretation by educators of the results gained from these instruments, decisions are made which directly affect the educational lives of children. The instruments which are frequently used in this practice use an approach which compares the individual child’s screening results with a standard that has been set for the purpose of determining the gaining of certain skills and developmental levels. Such an approach is referred to as criterion-referenced testing.

The development of criterion-referenced (CRT) assessment is a new appraisal technique in education and psychology. Its application to the process of screening children's readiness for school entrance is newer still. Hambleton et al. (1978) indicate that Glass and then Popham and Husek were the first to introduce and to popularize the field of criterion-referenced testing. Their motive was to provide the type of test score information needed to make a variety of individual and programmatic decisions arising in objective-based instructional programs. Norm-referenced tests were seen as being less than desirable for producing the type of test score information that was needed. Popham (1975) indicates that a criterion-referenced measure provides express information as to what the individual can and cannot do. The intent is to determine an individual's status relative to a well-defined behavior domain. This approach to human assessment is very appropriate for use with young children to determine their readiness for entrance into a kindergarten program. Criterion-referenced measurement classifies examinees into mastery states or categories on the objectives that are being
assessed (Swaminathan, Hambleton, & Algina, 1974).

Popham and Husek (1969) indicate that both norm-referenced and criterion-referenced techniques are important. In many cases one or the other may be used and both can provide useful information, yet each also has some psychometric properties which render them most appropriate for the purposes for which they were designed.

All writers and practitioners in the field are not enthralled by the use of either criterion-referenced or norm-referenced approaches to assessment. A time-referenced approach (Drew, 1973, for example, is suggested which uses continuous measurement and is concerned with the rate of progress of children. This strategy was offered in a situation in which learning disabled students were being studied. Further concerns with criterion-referenced assessment as it relates to learning disabled children, are offered by Proger and Mann (1973). Glass (1978) submits criticisms of the competency tests utilized in the State of Florida to determine minimum competencies in the State's high school seniors.

Although the criterion-referenced approach does have its detractors, the preponderence of material indicates a positive inclination for the use of this method of assessment. This depends upon the use to which the results are to be put. Bloom, Hastings, and Madaus (1971) indicate that the use of criterion-referenced scores is automatic in that the scores themselves detail what the student can do and what skills are still to be mastered. Hambleton and Eignor (1978) point out that much has yet to be done in the entire area of criterion-referenced test development. They indicate that a great deal needs to be done to resolve issues of disagreement found among members of the test and measurement field. Hambleton et al. (1978) offer that there are few criterion-referenced tests available that can meet today's standards
for test development, validation, and usage. They also suggest that today's technology has developed to the point that it can greatly improve the situation. An instance of the benefits of the application of the criterion-referenced approach is reported by Hunt and Kirk (1974). Their work extended the concept of criterion-referenced tests from the goals of teaching-learning situations to the domain of school readiness. This is where information regarding the mastery of concepts and skills of individual children can be transmitted and applied to meaningful teaching strategies and motivational systems.

The Project Intercept Inventory

The Project Intercept Inventory is a criterion-referenced test used in prekindergarten screening. This instrument is used by the Grand Rapids, (MI) Public Schools in their screening process for determination of school readiness on the part of incoming kindergarten children. There has been little done to determine the reliability, validity, and predictive capability of this instrument. The need is present to make such an effort. A better understanding of this document's capacity and efficiency was required. The approach taken with this test was a criterion-referenced strategy, since the administration of the test yielded results which indicate the child's mastery of specific skills and concepts. These domains were determined by the assessment of the expectations which Grand Rapids Public Schools teachers had for children entering kindergarten. These expectations were relative to the concepts, skills, and various aspects of development which the teachers determined the children should possess as they entered school for the first time. The attainment of these minimal expectations was considered necessary for the child to enter kindergarten.
Since so much of importance is dependent on the interpretation of results gained from this test, it was proposed that a study be done to determine the effectiveness and accuracy of the test. While the impression was present that the instrument was reliable and was a good predictor of readiness for kindergarten, essentially only informal material of a subjective nature had been obtained to substantiate this impression. Further, an earlier study (Cain, 1976) considered the construct and content validity of the PII, and indicated the need for further evaluation and study of the reliability and predictive validity of the instrument.

Summary

Schooling must be successful for children. Educators agree that it is very important to experience success early in the formal schooling process. Future success in the academic environment will build on initial success. The importance of a positive experience in kindergarten can hardly be overrated. Under these circumstances, readiness for children to successfully undertake the kindergarten experience has received a great deal of notice in educational circles.

Many school districts across this country have established screening programs to aid in determining the readiness level of the individual child to enter kindergarten and to master the kindergarten curriculum. Another motivation for this type of procedure is the legislation in place at the federal and State levels. There are requirements that a Child Find approach to locating learning and/or physically handicapped individuals at the earliest possible time be undertaken. Appropriate educational programming must also be provided for identified handicapped individuals.
Not everyone involved agrees with the concept of screening children prior to their entrance into kindergarten. Some educators take issue with this tactic, as do some parents. Parents especially point out their legal right to enter their children in school when the children have reached the State's chronological age requirement for that purpose. Such factors as the fallibility of screening devices, cultural differences which may affect the process, and the threat to parental self-esteem if their child is pronounced unready are pointed out.

These are important considerations. Important decisions which affect the lives of parents and children are made as a result of the interpretation of results gained from the prekindergarten screening process. Should accurate decisions be made, all involved stand to gain short and long-term benefits. Should poor decisions and recommendations be made with erroneous information, a great deal of damage could be done.

Among other ramifications of this process are the negative considerations of the labeling issue if wrong decisions are made; the distinct possibility that a child's progress through the schooling process may be impeded; and the emotional and possibly financial anxieties that could be experienced by parents. An element which also becomes an issue is that of the early entrance of children into school. This situation has two facets: entry of a child who is not developmentally ready for kindergarten, even though meeting the chronological age requirements; and attempted entry of a child who, ready or not, does not meet the minimal age requirements put in place by the State for school entrance.

Following the identification process, it becomes necessary to provide appropriate educational programs. This activity has implications at the
federal, State, and local levels. The needs of children, parents, and the community are involved in this area.

The entire process of screening and follow-up activities is dependent upon the assessment procedures that take place as the individual child is screened for kindergarten readiness. The screening is dependent upon the instruments used to make the determination if the child is ready for kindergarten. The entire operation hinges upon the accuracy of the information which the screening instruments can provide.

A substantial number of such tests are in existence. However, little work has been done to determine the reliability and validity of such tests. This is especially true if the document is a locally developed tool. An approach which has become prevalent in recent years is the application of the criterion-referenced model to this type of test. In such an approach the child's results are compared to a standard, based on developmental expectations, to determine kindergarten readiness. This is the approach used in the development of the Project Intercept Inventory, a screening device used for some years by the Grand Rapids (MI) Public Schools. The PII is a major part of the prekindergarten screening process. Since at least 2,000 children each year are enrolled in Grand Rapids schools as kindergartners, the predictive accuracy of this test takes on special importance.

The study which was undertaken here asked questions regarding the reliability, validity, and predictive capabilities of the results gained from the screening use of the Project Intercept Inventory.

In Chapter III, Design and Methodology, the statistical analyses which were used to examine the hypotheses proposed in this research will be listed and discussed.
CHAPTER III

DESIGN AND METHODOLOGY

Introduction

It has been stated that in judging any test, it is the use or interpretation of the scores that determines the appropriate indicators of test reliability and validity. Method follows function (Millman, 1979, p. 75).

The purpose of this study was to examine the reliability, validity, and possible presence of bias of a prekindergarten screening instrument. The instrument in question is named the Project Intercept Inventory (PII). This screening test is used in a large urban school district in western Michigan as a part of the screening process of incoming kindergarten children. It is used to aid in determining readiness for a successful kindergarten experience. The PII is a locally developed screening instrument. It was originally developed by a group of consultants assigned to Project Intercept. The basis for its formulation was an attempt to determine developmental levels in prekindergarten children. The instrument is described in detail later in this chapter.

The PII is a criterion-referenced test (CRT). Such a testing approach is intended to show what a person knows or can do, as opposed to the norm-referenced approach (NRT). The NRT shows where a person ranks in a group of test takers, according to Walker et al. (1979).

Criterion-referenced scores are considered to be meaningful in terms of the degree of learning which the individual test taker possesses. Scores of
other test takers do not enter into the criterion-referenced meaning of test results. Popham (1975) indicates that a student’s score on a CRT measure produces explicit information as to what the individual can and cannot do. Such a test is used to determine an individual’s status with respect to a well defined behavior domain.

A major use of the CRT approach is to provide information for specific decision strategies, according to Berk (1976) and Cronbach (1970). The power to accurately classify students at the point where a decision is made is taken as an indication of the test’s quality. Since the intent of the PII is to aid in determining the readiness of the individual child for kindergarten entrance, the CRT approach to establishing the presence of readiness-level skills is considered appropriate.

There are also the important questions raised regarding the PII in the areas of reliability, validity and possible presence of bias. To aid in addressing these questions, the Standards for Educational and Psychological Testing (Novick et al., 1985) was used. This document was produced by joint committees of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). It is considered the most current and knowledgeable source of its kind available. The criteria listed in the Standards for Educational and Psychological Testing (SEPT) were applied to the PII.

Description of the City and School Populations

This study undertook an examination of an operative screening test which is used in an ongoing prekindergarten screening program. This program is operated in the Grand Rapids, Michigan public school system.
The set of procedures which make up this program is formally entitled the Preschool and Kindergarten Screening Project, but is more popularly known as Project Intercept (PI). This project has been in place for a number of years. It has undergone some evolutionary changes, particularly logistically. It is continued in use to perform a major function. It is a vehicle used to provide information for early educational decision making for children, regarding the children's placement in initial school programs.

For background purposes, the need for a preschool screening program was determined in Grand Rapids schools early in the decade of the 1970s. The basis for this need was made known by a number of kindergarten teachers who were alarmed by the number of kindergarten children who were unable to accomplish satisfactory mastery of the kindergarten performance objectives. These children were most often recommended for retention at the kindergarten level. There was also concern noted for those children who had not progressed satisfactorily at the kindergarten level, but whose parents chose to reject the recommendation of school personnel for retention. This situation had been present for some time and was not considered to be a new or unique set of circumstances.

The city of Grand Rapids was founded in 1826 on the banks of the Grand River in western Michigan. Grand Rapids has a varied and strong economic basis. It is a typical mid-American city of approximately 181,843 people (1980 census figures) who reside within its metropolitan area. The city has a diverse population make-up. It is predominantly Caucasian with a strong religious bent fostered by its Dutch and Polish ethnic groups. There are several racial minority groups in the city. The largest of these is the Afro-American community. The Hispanic population would be next in
numerical size, followed by the much smaller Native American and Oriental groups. The census figures of 1980 for the city of Grand Rapids, MI were: White—147,220 (80.9%); Black—28,811 (15.8%); Hispanic—5,782 (3.2%); Native American—1,346 (0.07%); Asians—1,415 (0.08%).

The demographic make-up of the Grand Rapids Public Schools for the school year 1979-80 was as follows: White—16,691 (60.3%); Black—8,999 (32.5%); Hispanic—1,341 (4.8%); Native American—329 (1.2%); Asians—334 (1.2%). The total school population for the Grand Rapids Public Schools for the 1979-80 school year was 27,694 enrollees.

Another important consideration with the Grand Rapids educational situation is that of the nonpublic schools (parochial and private). The total number of students in these schools during the 1979-80 year was 10,667 enrollees. Of this number, 6,518 (61.1%) students were enrolled in Catholic schools; 3,234 (30.3%) were enrolled in Christian Reformed schools; and 1,191 (11.2%) were enrolled in schools affiliated with other religious denominations, or in private schools.

All of the nonpublic schools receive the same type of special education and alternative education services that United States law requires and allows. Among these services would be included Project Intercept. The project serves prekindergarten children in all of the elementary schools in the city. This is true of all public elementary schools, and also of all the nonpublic schools that choose to use the service.

Description of the Sample and Data Collection

The population for this study consisted of approximately 1,600 public school prekindergarten children. The ages of the children ranged from 4 1/2
to 6 years of age. They were representative of the population make-up of the city of Grand Rapids, Michigan. A qualification to this statement should be made. The nonpublic schools population is heavily white. Thus, the children of most minority groups attend the public schools, and the public school population then appears to have a higher percentage of minority students than the racial percentage make-up of the city would indicate. The children were screened for kindergarten readiness in the Spring of 1980 using the Project Intercept procedures.

The sample used in this study consisted of 530 prekindergarten children. Their screening results were gained from the administration of the Project Intercept Inventory.

The children were expected to attend kindergarten in the public schools in the city of Grand Rapids, Michigan. The sample reflects a cross section of the city's school population relative to the race and sex of the children. The children were all of kindergarten age, being 5 years old on or before the 1st day of December of the coming school year. There were 530 children who were initially involved in the study. Of this number 303 children were Caucasians, and 227 were members of various minority groups. There were 233 males and 297 females included in this group.

A cluster sampling technique was used to obtain the study sample. The sample selected for this study was representative of the total population of incoming kindergarten children. Various schools within the district were selected as study sites. All of the children screened at these sites who were recommended for kindergarten entrance were included in the study sample.

The data for this study were taken from the results gained through the administration of the PII. Based largely on these results, these children were
recommended to be enrolled in kindergarten. This study then undertook to review the results of the administration of the California Achievement Test (CAT) to these same children several years after the prekindergarten screening. The CAT is regularly administered to children in all grades of the Grand Rapids Public Schools each year. The reading and arithmetic subtests were administered by classroom teachers who were supervised by personnel from the Curriculum Planning and Evaluation office of the Grand Rapids Public Schools. This was done as part of the standard achievement testing program in the public schools.

The intent of this strategy was to determine if the children recommended for kindergarten enrollment due to their performance during the administration of the Project Intercept Inventory were meeting the objective requirements of the CAT, at or above their current grade level.

The attainment of grade level results from the CAT reading and arithmetic subtests could serve as documentation that the PII had functioned as a reliable and valid tool for predicting the success of children in the kindergarten and early elementary grades. This would be reflected by the CAT results.

**Project Intercept**

The services provided by the use of Project Intercept for the prekindergarten students of Grand Rapids consists of screening and diagnostic services. These are performed essentially in a two stage process. The two stages are referred to as Phase I and Phase II. The Phase I procedure consists of a screening process. The Phase I process has for one of its major objectives the screening of every incoming kindergarten child within the
city of Grand Rapids. The process is used in public, parochial, and private schools. Through the screening process children who indicate that they may have serious difficulty in successfully mastering the kindergarten program objectives are recommended for Phase II of Project Intercept, the diagnostic portion. The number of children so referred historically has been one in ten of those screened.

Recommendations for program placement for the children are made to parents. A very high percentage of children screened in Phase I are recommended for kindergarten entrance during the following September. For the much smaller group of children who would subsequently attend Phase II a variety of recommendations can be made.

The Phase II procedures of PI are important and do involve a considerable amount of professional time and effort, as well as financial support from the school district. At the same time it is a complex and multifaceted operation that could well entail thorough study in its own right. This is due to its complexity and the additional issues with which Phase II deals.

The point to be considered is that the children who are recommended to take part in the diagnostic phase of PI are recommended because of their performance results in Phase I. For these reasons, this study did not undertake an examination of the entire Phase II process, but will be concerned specifically with the value of the Project Intercept Inventory as a predictive instrument.

The Project Intercept Inventory

The Project Intercept Inventory is an example of the criterion-referenced approach in a prekindergarten screening test. The PII is currently
being used by the Grand Rapids (MI) Public Schools in the screening process for determination of school readiness on the part of incoming kindergarten children. Since there has been little done to determine the reliability, validity, and possible presence of bias for this test, there was a need to institute such an effort. A better understanding and awareness of this test's capacity and efficiency was needed.

The approach taken with the use of this test was a criterion-referenced strategy. The screening test used was developed locally by consultants assigned to the Project Intercept program. The administration of this test yields results which indicate the child's mastery of specific skills and concepts. The developmental areas which make up the test were determined by the assessment of the expectations which Grand Rapids Public Schools teachers had for children entering kindergarten. These expectations were relative to the concepts, skills, and various aspects of development which the teachers felt the children should possess as they entered school for the first time. These were skills felt to be necessary for the children to successfully master the objectives of the kindergarten curriculum.

There are 21 different tasks which compose the PII. These tasks are distributed through seven developmental areas.

The developmental areas of the PII and specific tasks attached to each area are as follows:

1. Gross motor (large muscle coordination): naming of human body parts; balance beam walking; jumping forward with both feet together; hopping for approximately ten feet.

2. Visual-motor (using eyes and hands together): paper folding; paper cutting; drawing a circle, square, triangle, and open-ended square and circle figures.
3. Perceptual Awareness: identifying colors; recognition of a circle, square, triangle, rectangle, and diamond; awareness of positions in space (in front of, behind, next to, etc.); recognition of size (big–little).

4. Visual memory (remembering what is seen): sequencing objects from memory; identifying missing objects.

5. Auditory skills (listening and understanding): repeating digits; repeating a series of directions; following a series of directions.

6. Verbal associations: responding to verbal contextual stimuli (e.g., A fire is hot; ice is ...).

7. Draw-a-Person: drawing a human figure; scoring based on inclusion of body parts.

The PII is administered to each child on a one-to-one basis. The persons administering the test are Teacher Consultants or Resource Room teachers employed by the Grand Rapids Public Schools. These persons were selected for this work because of their background, training, and experience in test administration and interpretation.

The administration of the test takes approximately 15 to 20 minutes for each child. The children are brought to the screening sites, local elementary school buildings, by their parents. Following the administration of the screening tests, the results are explained to the parents, also on a one-to-one basis. At this time recommendations are made to the parents to either plan to enroll the children in kindergarten in the Fall, or to allow the children to attend Phase II of Project Intercept. This is a procedure that results in a closer diagnostic look at each child who is involved in Phase II.

Results gained from the use of the PII are used in making important educational decisions for children. This study was proposed to aid in
determining the effectiveness and accuracy of the PII. Only informal material of a subjective nature had been obtained to substantiate the accuracy of the PII. Further, an earlier study (Cain, 1976) considered the construct and content validity of the PII, and indicated the need for further evaluation and study of the reliability and predictive validity of the instrument.

The Analyses Method

The Project Intercept Inventory has been used since 1973 as a major part of the kindergarten readiness screening program in the Grand Rapids Public Schools. Decisions regarding the educational futures of children are often based on information gained from the administration of this instrument. Since the instrument has not been studied regarding its reliability, validity, and possible presence of bias, this study undertook several examinations. Item difficulty and item discrimination were included.

All hypotheses were tested at the 0.05 level of confidence, where appropriate. For the purposes of this study, some modifications were employed with the PII. These modifications consisted of the elimination of three of the original test items. Item 12 (recognizing likenesses and differences in shapes) was eliminated since the results consistently were scored so high that it was not considered to be discriminatory regarding kindergarten readiness. Item 18 (verbal fluency) was eliminated because it proved to be too subjective, and was being judged by test administrators who were not trained in this area of development. Item 20 (basic counting skill) was eliminated since it was felt that this skill was developed at a later time in children, through school experience.
Item Difficulty

Item Difficulty Analysis

A portion of the process of item analysis is the assessment of item difficulty (Thorndike, 1971). The most simple and commonly used measure of item difficulty is the p-value—the proportion of correct answers on an item (p. 139). Mehrens and Lehmann (1975) refer to p-value as the item difficulty index and indicate that it is an important aspect of item analysis. This is the case because items that have a p-value of approximately 0.5 provide the most information as regards the difficulty and appropriateness of the test item. In this regard, a range of 0.2 to 0.8 is considered to be appropriate.

Hypothesis 1. The p-value of individual items of the Project Intercept Inventory will be less than 0.2, greater than 0.8. (Null hypothesis)

Alternative. The p-value of individual items of the Project Intercept Inventory will be greater than 0.2 and equal to or less than 0.8.

Item Discrimination

Item Discrimination Analysis

To further assess the test items, an examination of item discrimination was undertaken. This was regarded as item to total test correlation, to gain an impression of how well each item discriminated in terms of the total application of the test. This was done by correlating each item with the total test, after the individual item was removed from the test taken as a whole. The intent was to make a determination as to how well each item
discriminated between students of high and low ability.

In order to analyze the discrimination of each item as it pertained to the readiness of children to undertake a successful kindergarten experience, the corrected point-biserial technique was used, when the item under consideration was dichotomous.

This technique is the special case of the Pearson product-moment coefficient and is used when one variable is measured on an interval or ratio scale and the other variable is a discrete dichotomy. The majority of the test items on the PII are considered to be dichotomous. For the few that were not, the corrected Pearson product-moment correlation coefficient was used.

The use of the corrected point-biserial and corrected Pearson product-moment correlation coefficient was felt to be necessary to aid in eliminating overlap. This could occur if the individual item was included in the correlation, making the correlations spuriously high. In order to perform this analysis, the data was entered on tape and then placed on a disc to be analyzed with the Statistical Package for the Social Sciences-X (SPSS-X) program.

**Hypothesis 2.** The individual test items of the Project Intercept Inventory will have correlation coefficients less than 0.15 when compared with total test scores, as measured by the corrected point-biserial technique for dichotomous items, and the corrected Pearson product-moment correlation coefficient for nondichotomous items. (Null hypothesis)

**Alternative.** The corrected point-biserial correlation coefficients between individual items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with dichotomous items.
Alternative. The corrected Pearson product-moment correlation coefficients between individual items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with nondichotomous items.

Reliability

Reliability, while not the most important facet of measurement, is still extremely important. The Standards for Educational and Psychological Testing (Novick, et al., 1985) indicates that reliability refers to the degree to which test scores are free from errors of measurement. Fundamental to the proper evaluation of a test are the identification of major sources of measurement error, the size of the errors resulting from these sources, the indication of the degree of the reliability to be expected between pairs of scores under particular circumstances, and the generalizability of results across items, forms, raters, administrations, and other measurement facets. Unless the researcher can depend upon the results of the measurement of certain variables, he or she cannot, with any confidence, determine the relationship between the variables (Kerlinger, 1973).

Mehrens and Lehmann (1975) point out that estimates of internal consistency are really indices of the homogeneity of the items in the test, or the degree to which the item responses correlate with the total test score. A major advantage to this method of obtaining reliability estimates is that the work can be done with a single set of data, thus eliminating the need for two or more sets of data. This is a circumstance which very often is not possible or feasible in educational and psychological research (Algina & Noe, 1978; Lord, 1962; Millman, 1979; Shavelson, Block, & Ravitch, 1972; Subkoviak, 1976).
Reliability Analysis

The reliability estimate method used in this study was the alpha coefficient. The alpha coefficient was developed by Cronbach (1951). The author indicates that if a test has substantial internal consistency, it is psychologically interpretable. And, for a test to be interpretable, it is not essential that all items be factorially similar. What is required is that a large proportion of the test variance be attributable to the principal factor running through the test (Vernon, 1950). Alpha coefficient estimates the proportion of the test variance due to all common factors among the items. It reports how much the test score depends upon general and group factors, rather than on specific item factors.

Cronbach (1951) further indicates that alpha coefficient is the general formula of which the Kuder-Richardson-20 (K-R 20) is a special case. It is found to have several important meanings. For example, alpha coefficient is the mean of all possible split-half coefficients. It is the value expected when two random samples of items from a pool like those in the given test are correlated. Alpha coefficient is the lower bound for the coefficient of precision (the instantaneous accuracy of this test with these particular items). It is also a lower bound for coefficients of equivalence obtained by simultaneous administrations of two tests having matched items. Alpha coefficient also estimates the proportion of test variances attributable to common factors among the items. It is an index of common-factor concentration. This index serves purposes claimed for indices of homogeneity.

Alpha coefficient is an upper bound to the concentration in the test of the first factor among the items. Interpretability of a test score is improved if
the score has a high first-factor concentration. A high alpha coefficient is therefore desirable, but items with quite low intercorrelations can yield an interpretable score.

To implement the application of the alpha coefficient to test the reliability facets of the Project Intercept Inventory the data was entered on electronic tape. The computer program applied to the data was the Statistical Package for the Social Sciences-X (SPSS-X).

**Hypothesis 3.** The alpha coefficient for the total test of the Project Intercept Inventory will be less than 0.60. (Null hypothesis)

**Alternative.** The alpha coefficient for the total test of the Project Intercept Inventory will be greater than or equal to 0.6.

**Validity**

The *Standards for Educational and Psychological Testing* (Novick et al., 1985, p. 9) indicates that validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from the test scores. Test validation is the process of accumulating evidence to support such inferences. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. Resources should be invested in obtaining the combination of evidence that optimally reflects the value of a test for an intended purpose.

On this basis, this study undertook to examine aspects of content, construct, and criterion (predictive) validity of the Project Intercept Inventory.
Content Validity Analysis

Hambleton et al. (1978) indicate that the most effective approach in determining content validity involves the judgments of content specialists. This is an approach which Kerlinger (1973) defines as the representativeness or sampling adequacy of the content of a measuring instrument (p. 458). This approach requires that a number of persons examine the test and conclude that it measures the relevant traits in question. If there is disagreement regarding this impression, then content validity is in question.

A strategy put forth by Hambleton et al. (1978) to determine content validity is to develop a matching task for the persons who would examine the test.

This study utilized the matching task approach. Ten persons active in early elementary or preschool education and/or assessment were presented with two lists, one with test items and the other with test objectives. The judges were asked to indicate which objective he or she thought each test item measured (if any). A contingency table was then constructed by calculating the number of specialists matching each item to each objective in the sets of items and objectives which were being studied.

A frequently used method of evaluating this type of data is the chi square test for independence to analyze data that are presented in a contingency table format. That is the method which was utilized in this study.

The data were entered on tape. The material was then placed in a disc format, and was run on the computer with the Statistical Package for the Social Sciences-X as the program for the analyses.
Hypothesis 4. There will be no significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique. (Null hypothesis)

Alternative. There will be significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique.

Construct Validity Analysis

The Standards for Educational and Psychological Testing (Novick et al., 1985) indicates that the evidence classed in the construct-related category focuses primarily on the test score as a measure of the psychological characteristic of interest. The construct of interest for a particular test should be implanted in a conceptual framework. The conceptual framework specifies the meaning of the construct, distinguishes it from other constructs, and indicates how measures of the construct should relate to other variables.

Mehrens and Lehmann (1975) stipulate that construct validity is the degree to which the test scores can be accounted for by certain explanatory constructs in a psychological theory. If an instrument has construct validity, scores will vary as the theory underlying the construct would predict. Messick (1975) offers that construct validation is the process of marshaling evidence in the form of theoretically relevant empirical relations to support the inference that an observed response consistency has a particular meaning (p. 955).
Hambleton et al. (1978) indicate that construct validity should begin with a definite statement of the proposed use of the test scores, arguing that a clearly stated use will provide direction for the kind of evidence that is worth gathering. Establishing construct validity is an ongoing process that involves the verification of predictions made about the test scores.

A specific statement of purpose regarding the Project Intercept Inventory is stated as follows: The intent of the Project Intercept Inventory is to use the instrument in the screening of children to aid in determining their readiness for kindergarten. This is done by presenting the children with a series of developmental tasks. These tasks are considered to be the types of activities that the children should be able to perform in a satisfactory manner in order to consider that the children have developed physically, perceptually, and intellectually to a level where success for them in mastering the objectives listed in the kindergarten curriculum can be predicted.

The Project Intercept Inventory consists of 21 different tasks incorporated into seven developmental categories. The specific constructs which are examined are: gross motor; visual motor (eye-hand coordination); concepts; visual memory; auditory skills; context clues (verbal associations); and Draw-a-Person.

To make a determination of the construct validity of the PII, factor analysis (principal components) was employed. Hinkle, Wiersma and Jurs (1979) indicate that factor analysis is a procedure for determining the number and nature of the constructs or traits (factors) that underlie a set of variables. The variables commonly consist of test scores, or scores on psychological and sociological inventories, although they could be scores on measures of practically any kind. Kerlinger (1973) relates that factor analysis
is a method of determining the number and nature of the underlying variables among larger numbers of measures. Factor analysis is considered an extremely powerful and useful approach to behavioral data. It is a method for determining the underlying variables from various sets of measures. One application of factor analysis, the principal factors method, involves the solution of simultaneous linear equations. The roots obtained from the solution are referred to as eigenvalues. Eigenvalues are related to the percentage of variance which is accounted for through the factoring process. Eigenvectors can also be obtained. After suitable transformation, they become the factor loadings (Kerlinger, p. 669).

In order to implement the factor analysis technique and obtain eigenvalues, the data was entered on tape. The material was then run on the Statistical Package for the Social Sciences-X.

**Hypothesis 5.** No more than one factor of the Project Intercept Inventory will be found with an eigenvalue equal to or greater than one, based upon the application of factor analysis. (Null hypothesis)

**Alternative.** There will be two or more factors of the Project Intercept Inventory found with an eigenvalue equal to or greater than one, based upon the application of factor analysis.

**Criterion Validity Analysis**

The Standards for Educational and Psychological Testing (Novick et al., 1985) indicates that criterion validity demonstrates that test scores are systematically related to one or more outcome criteria. The relationships between test scores and criterion measures may be expressed in various
ways, but the fundamental question is always how accurately can criterion performance be predicted from scores on the test (p. 11).

A predictive study obtains information about the accuracy with which early test data can be used to estimate criterion scores that will be obtained in the future. Berk (1980a) indicates that predictive validity involves using test scores to predict future behavior. In this process, a predictive-validity coefficient is obtained, since criterion-related validity is typically expressed as a correlation coefficient. When a test is used to predict future behavior, predictive validity should be established.

For the purposes of this study, the test data were entered on tape. The material was then analyzed by using the computer and accessing the Statistical Package for the Social Sciences-X program, using the Pearson product-moment correlation coefficient to obtain the correlation coefficients of validity. Mehrens and Lehmann (1975) relate that if two sets of test scores from the same group of people are compared to determine relationships, the Pearson product-moment correlation coefficient is the statistic most often used for that purpose.

The two sets of test scores that were used in this study to aid in determining predictive validity were the screening scores gained from the administration of the Project Intercept Inventory in the Spring of 1980, and the reading and arithmetic scores gained by these same children from the administration of the California Achievement Test (CAT) in their fourth grade school year.

**Hypothesis 6.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the
reading scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

**Alternative.** There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient.

**Hypothesis 7.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

**Alternative.** There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient.

**Bias**

The authors of *Standards for Educational and Psychological Testing* (Novick et al., 1985) designate that when previous research indicates the need for studies of item or test performance difference for a particular kind of test for members of age, ethnic, cultural, and gender groups in a
population of test takers, such studies should be conducted as soon as feasible. Such research should be designed to detect and eliminate aspects of test design, content, or format that might bias test scores for particular groups (p. 27).

There are basically two types of bias noted in test construction. These are bias in the presence of an external criterion, called bias in selection, and bias in the absence of an external criterion, called item bias (Ironson & Subkoviak, 1979; Scheuneman, 1979). This study undertook to examine the possibility that internal (item) bias was present in the Project Intercept Inventory.

A test item is considered biased if individuals with equal ability, but from different groups, do not have the same probability of answering the items correctly. An item may be considered biased if it functions differently for a specific subgroup of students (Plake & Hoover, 1979). An item is unbiased if, for all individuals having the same score on a homogeneous subtest containing the item, the proportion of individuals getting the item correct is the same for each population group being considered (Scheuneman, 1979).

Bias Analysis

Angoff (1982) indicates that any index for evaluating bias should incorporate in it some control on the level of ability of the groups under study. This study used the item characteristic curve (ICC), used in latent trait theory. ICCs depict the relationship between ability and the probability of answering the item correctly. In general, as the former increases, so does the probability of the latter. Three parameters may be used to describe these curves, according to Ironson (1982). The parameters are: the difficulty
parameter, given by the ability level corresponding to the inflection point of the curve; the discrimination value proportional to the slope of the curve at the inflection point; and the probability of examinees of infinitely low ability getting the item right (sometimes called the guessing parameter).

An item that is unbiased should have the same ICC in both of the former groups. Biased items will have different ICCs.

Scheuneman (1979) reported acceptable results when the ICC method was used in comparisons with other item bias evaluation methods, as did Rudner and Convey (1978) and Merz and Grossen (1979). In a study done by Ironson and Subkoviak (1979) the ICC was found to compare well with several other methods which were examined, and found that there was clearly support for this method. The authors also point out that the ICC approach is considered by some to be the most theoretically sound of the methods being used to detect item bias in tests. Rudner, Getson and Knight (1980) found in their study that the ICC with three parameters produced fairly accurate results under all the investigated conditions.

With the ICC analysis to be employed the data was placed in a disc format. The data was entered on tape, and the analysis was accomplished with the Statistical Package for the Social Sciences-X procedure for computer analysis purposes. The program was used to analyze the possibility of both sex and race bias in the items which make up the Project Intercept Inventory.

**Hypothesis 8.** The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for boys and girls. (Null hypothesis)
**Alternative.** The general shapes of the item characteristic curves for the Project Intercept Inventory will not be substantially different for boys and girls.

**Hypothesis 9.** The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for racial majority and minority children. (Null hypothesis)

**Alternative.** The general shapes of the item characteristic curves for the Project Intercept Inventory for racial majority and minority children will not be substantially different.

In Chapter IV, Research Findings, the statistical results from these analyses will be stated and discussed.
CHAPTER IV

RESEARCH FINDINGS

Introduction

The Project Intercept Inventory is a locally developed screening device. Its intent is to serve as a screening test to aid in determining the readiness of children for kindergarten. It is used in the Grand Rapids, MI Public Schools currently, and has been a primary source of school readiness information in that school system for 12 years. Thousands of children have been screened with this test over that period of time, and important initial decisions have been made which impact on the educational futures of children, based in part on the results gained from its use.

No research studies have been done to determine the validity, reliability, and predictive capacity of this test. Because of its direct use in the screening process, it was proposed that a study be done to aid in making some determinations regarding the efficiency and effectiveness of the PII. This writer undertook such a study. The aspects of the PII which were scrutinized were: item difficulty, item discrimination, reliability, content validity, construct validity, criterion validity, gender bias, and racial bias.

This chapter will relate the results gained from the statistical analyses used to examine these various aspects of the PII.

All hypotheses were tested at the 0.05 level of confidence, where appropriate.
Statement of Hypotheses and Findings

**Item Difficulty**

The intention of the item difficulty examination of the PII was to gain more information regarding children. It was felt that examining the individual test items for the level of difficulty would aid in gaining such information. If the test items were appropriately difficult for the age-level children who would be involved, the information gained would be helpful in determining readiness for kindergarten. The information would also be helpful in planning for curriculum development at the kindergarten level, as well as in considering reasonable expectations for children at that level.

The p-value of the items was examined. This strategy considers the proportion of correct answers obtained on individual test items. Mehrens and Lehman (1975) refer to the p-value as the item difficulty index.

The level of difficulty considered most appropriate would be a p-value of 0.5, the level at which item variance is considered to be greatest. The item variance is sought in order to gain information regarding the individual items of the test. A p-value of 0.9, for example, would have little value since it indicates very little variance. A range of 0.2 to 0.8 was considered in this examination to be sufficient to determine if the various test items had an appropriate level of difficulty.

**Hypothesis 1.** The p-value of individual items of the Project Intercept Inventory will be less than 0.2, greater than 0.8. (Null hypothesis)

**Alternative.** The p-value of individual items of the Project Intercept Inventory will be greater than 0.2 and equal to or less than 0.8.
Discussion. The p-values for dichotomous items of the PII indicate that items 5, 8 and 9 (folding a paper triangle, drawing a square, drawing a triangle, respectively) were within the stipulated range of 0.2 to 0.8. The following items had a p-value over 0.8: identifying body parts; walking a balance beam; hopping, jumping; paper cutting; drawing a circle; visually indicating big and small; sequencing objects; identifying missing objects; repeating numbers; repeating directions; following directions (items 1, 2, 3, 4, 6, 7, 14, 15, 16, 17, 18 and 19).

For the nondichotomous items used in the PII, means of the number of responses considered correct were obtained. These results indicate that items 11 (colors recognition), 12, (recognizing geometric shapes), 13 (positions in space), and 20 (verbal associations) were appropriately difficult for the children. Item 21, drawing a human figure, appeared to be moderately too easy, since the mean was calculated at 7.220. However, with this item, the scoring variations employed by the test administrators may have been more susceptible to leniency than with other test items. This is due to the latitude of interpretation provided for this more subjective task. Some test administrators would be more prone to leniency in scoring the appropriateness of body parts, while others would be more stringent.

The statistical results used in examining Hypothesis 1 are listed in Table 1.

The statistical indication that a majority of the PII test items were either too difficult or too easy for the children screened may be due to the unevenness of the distribution of the testing results. The p-value technique can be influenced by such an effect on the distribution.
Table 1

Item Difficulty Analysis
(p-values)

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Valid no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.920</td>
<td>.272</td>
<td>599</td>
</tr>
<tr>
<td>2</td>
<td>.876</td>
<td>.329</td>
<td>599</td>
</tr>
<tr>
<td>3</td>
<td>.967</td>
<td>.180</td>
<td>599</td>
</tr>
<tr>
<td>4</td>
<td>.840</td>
<td>.367</td>
<td>599</td>
</tr>
<tr>
<td>5</td>
<td>.626</td>
<td>.484</td>
<td>599</td>
</tr>
<tr>
<td>6</td>
<td>.912</td>
<td>.284</td>
<td>599</td>
</tr>
<tr>
<td>7</td>
<td>.977</td>
<td>.151</td>
<td>599</td>
</tr>
<tr>
<td>8</td>
<td>.711</td>
<td>.454</td>
<td>599</td>
</tr>
<tr>
<td>9</td>
<td>.419</td>
<td>.494</td>
<td>599</td>
</tr>
<tr>
<td>10</td>
<td>.166</td>
<td>.372</td>
<td>596</td>
</tr>
<tr>
<td>11</td>
<td>6.573</td>
<td>2.300</td>
<td>599</td>
</tr>
<tr>
<td>12</td>
<td>3.087</td>
<td>1.649</td>
<td>596</td>
</tr>
<tr>
<td>13</td>
<td>4.584</td>
<td>1.036</td>
<td>598</td>
</tr>
<tr>
<td>14</td>
<td>.961</td>
<td>.193</td>
<td>592</td>
</tr>
<tr>
<td>15</td>
<td>1.901</td>
<td>.922</td>
<td>598</td>
</tr>
<tr>
<td>16</td>
<td>2.634</td>
<td>.701</td>
<td>595</td>
</tr>
<tr>
<td>17</td>
<td>2.986</td>
<td>.979</td>
<td>566</td>
</tr>
<tr>
<td>18</td>
<td>2.094</td>
<td>.956</td>
<td>597</td>
</tr>
<tr>
<td>19</td>
<td>2.189</td>
<td>.797</td>
<td>594</td>
</tr>
<tr>
<td>20</td>
<td>4.265</td>
<td>1.669</td>
<td>596</td>
</tr>
<tr>
<td>21</td>
<td>7.220</td>
<td>2.824</td>
<td>596</td>
</tr>
<tr>
<td>TOTAL</td>
<td>45.772</td>
<td>9.658</td>
<td>531</td>
</tr>
</tbody>
</table>
Mehrens and Lehmann (1975) point out that in criterion-referenced testing the concept of difficulty is antithetical to the philosophy underlying the mastery concept—that all or nearly all students should pass the item. They suggest that while the application of conventional item difficulty techniques may be appropriate in criterion-referenced testing, the interpretation of data may have to be done somewhat differently.

Based on these results, the null hypothesis cannot be rejected. The majority of the test items of the Project Intercept Inventory did not present as having an appropriate degree of difficulty to adequately assess kindergarten readiness for children between the ages of 4 1/2 and 5 1/2 years of age.

**Item Discrimination**

The intention of this analysis was to determine if the individual items of the PII would discriminate between students of high and low ability, as defined by the total test score.

This determination was made by correlating each item with the total test scores. The individual item result was removed from the test taken as a whole when the comparisons were made. This was done to eliminate overlap, which could make the correlations spuriously high.

**Hypothesis 2.** The individual test items of the Project Intercept Inventory will have correlation coefficients less than 0.15 when compared with total test scores, as measured by the corrected point-biserial technique for dichotomous items, and the corrected Pearson product-moment correlation coefficient for nondichotomous items. (Null hypothesis)
Alternative. The corrected point-biserial correlation coefficients between individual items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with dichotomous items.

Alternative. The corrected Pearson product-moment correlation coefficients between individual items and total test scores for the Project Intercept Inventory will be equal to or greater than 0.15 when used with nondichotomous items.

Discussion. The analysis of the corrected point-biserial correlation coefficient results reveals that all dichotomous items, when individually removed and correlated with the total test, indicate results above the 0.15 level, with the exception of item 7 (drawing a circle). This correlation was at the 0.1048 level. The highest correlation coefficient found was at the 0.5528 level. This was for item 21 (drawing a person). The lowest remaining correlation was at 0.2542 for item 3 (jumping).

These results would strongly suggest that no particular item (with the one exception) would bring down the ability to discriminate among test takers when the test is used.

The analysis of the corrected Pearson product-moment correlation coefficient results reveals that all nondichotomous items, when individually removed and correlated with the total test, indicate results above the 0.15 level. The highest correlation coefficient in these results was 0.5595 for item 18 (verbally repeating a series of directions). The lowest correlation obtained in these results was 0.3847 for item 17 (verbally repeating a series of digits). The nondichotomous correlations could have been higher than the
dichotomous correlations due to the higher variances. The nondichotomous correlations were gained with two continuous variables, and not one dichotomous and one continuous variable, as is the case with calculating the correlations for the dichotomous items. These results would strongly suggest that no particular item would impair the ability to discriminate among test takers when the test is used.

The contrast between item difficulty and item discrimination is noted. There is certainly a relationship between item difficulty and item discrimination, but it is not a perfect relationship. An item can be discriminating in that a majority of the children will successfully complete it within acceptable limits, which makes the item seem possibly not difficult enough. However, the intent of the PII was to screen for children who would experience difficulty in successfully completing the tasks. Such children would be viewed as potential enrollees in prekindergarten readiness programs. If the deficiencies noted from the screening test were serious enough, the alternative programming of Special Education could be considered, after the appropriate diagnostics were accomplished.

The intent of the PII when it was developed was to screen for weaknesses. Those children who successfully complete an item are not categorized as to the degree of success they achieved, but simply that they did accomplish the task.

The results of the correlation coefficients are listed in Table 2.

Based on the analysis of these statistical results, the null hypothesis is rejected. The individual items were found to be discriminating between test takers of high and low ability when compared to the test taken as a whole.
Table 2
Item Discrimination Analysis

<table>
<thead>
<tr>
<th>Item</th>
<th>Corrected Item-total correlation</th>
<th>Alpha if item deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.4145</td>
<td>.7728</td>
</tr>
<tr>
<td>2</td>
<td>.2584</td>
<td>.7746</td>
</tr>
<tr>
<td>3</td>
<td>.2542</td>
<td>.7762</td>
</tr>
<tr>
<td>4</td>
<td>.3838</td>
<td>.7717</td>
</tr>
<tr>
<td>5</td>
<td>.4027</td>
<td>.7693</td>
</tr>
<tr>
<td>6</td>
<td>.2907</td>
<td>.7745</td>
</tr>
<tr>
<td>7</td>
<td>.1048</td>
<td>.7779</td>
</tr>
<tr>
<td>8</td>
<td>.5435</td>
<td>.7662</td>
</tr>
<tr>
<td>9</td>
<td>.4468</td>
<td>.7679</td>
</tr>
<tr>
<td>10</td>
<td>.3038</td>
<td>.7732</td>
</tr>
<tr>
<td>11</td>
<td>.4435</td>
<td>.7681</td>
</tr>
<tr>
<td>12</td>
<td>.4560</td>
<td>.7589</td>
</tr>
<tr>
<td>13</td>
<td>.4350</td>
<td>.7617</td>
</tr>
<tr>
<td>14</td>
<td>.3788</td>
<td>.7745</td>
</tr>
<tr>
<td>15</td>
<td>.4209</td>
<td>.7634</td>
</tr>
<tr>
<td>16</td>
<td>.4700</td>
<td>.7635</td>
</tr>
<tr>
<td>17</td>
<td>.3847</td>
<td>.7650</td>
</tr>
<tr>
<td>18</td>
<td>.5595</td>
<td>.7547</td>
</tr>
<tr>
<td>19</td>
<td>.4840</td>
<td>.7614</td>
</tr>
<tr>
<td>20</td>
<td>.4744</td>
<td>.7572</td>
</tr>
<tr>
<td>21</td>
<td>.5528</td>
<td>.7698</td>
</tr>
</tbody>
</table>
Reliability

The alpha coefficient was selected for use in measuring the reliability of the PII because it is the mean of all possible split-half coefficients (Cronbach, 1951). It is the value expected when two random samples of items from a pool like those in the given test are correlated.

The alpha coefficient is the lower bound for coefficients of equivalence obtained by simultaneous administrations of two tests having matched items. Alpha coefficient also estimates the proposition of test variances attributable to common factors among the items.

Alpha coefficient is an upper bound to concentration in the test of the first factor among the items. Interpretability of a test score is improved if the score has a high first-factor concentration. A high alpha coefficient is, therefore, desirable. The level of 0.6 for the alpha coefficient was used here since research indicates that good, usable tests achieve reliabilities between 0.6 and 0.8. If reliabilities fall below the 0.6 level the test should be regarded as suspect in terms of its reliability.

Hypothesis 3. The alpha coefficient for the total test of the Project Intercept Inventory will be less than 0.6. (Null hypothesis)

Alternative. The alpha coefficient for the total test of the Project Intercept Inventory will be greater than or equal to 0.6

Discussion. The computerized statistical analysis of these data indicated that the alpha coefficient for the total test of the Project Intercept Inventory was found to be 0.7766. This being the case, the results would confirm an acceptable degree of test reliability for the PII, above the
hypothesized level of 0.6. The null hypothesis is rejected.

Validity

Content Validity

In the examination of content validity this study used the matching test approach to determine if the various judges would respond in a consistent fashion in matching PII test items and test objectives. Ten persons active in early elementary or preschool education and/or assessment were presented with lists of the PII items and objectives.

Hypothesis 4. There will be no significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique. (Null hypothesis)

Alternative. There will be significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique.

Discussion. The judges were asked to match each test item with the developmental area he or she thought each test item measured, if any. A contingency table was then constructed by calculating the number of judges matching each item to each objective in the sets of items and objectives which were under study. It was expected that 80% of the judges would appropriately match items and objectives.

A frequently used method of evaluating this type of data is the chi square test for independence. The chi square technique is a nonparametric
test of significance appropriate when the data are in the form of frequency counts. It compares proportions actually observed in a study with the proportions expected, to see if they are significantly different. It is commonly used to analyze data presented in a contingency table format.

The first developmental area considered was Gross Motor. It was anticipated that the first four items of the PII would be placed by the judges in this developmental area. The results gained from the judges indicates that item one (identifying body parts) was not seen by seven of the judges as being appropriately placed in the gross motor area. Items two (balance beam walking), three (jumping) and four (hopping) were placed by all of the judges in this developmental area.

A contingency table was constructed. The chi square value that was calculated was 25.45. The degrees of freedom were set at three, yielding a critical value at the 0.05 level of confidence of 7.815. Since the chi square value exceeded the critical value the null hypothesis was rejected.

The second developmental area considered was Visual Motor. It was anticipated that test items five through ten would be placed in this area by the judges. Item five (folding a paper triangle) was placed by five of the judges in the hypothesized developmental area. Five of the judges placed the item in another developmental area. Items six (cutting a piece of paper), seven (drawing a circle), eight (drawing a square), nine (drawing a triangle) and ten (drawing an open square and circle figure) were placed in the hypothesized developmental area by all ten judges.

A contingency table was constructed. The chi square value that was calculated was 132.99. The degrees of freedom were set at five, yielding a critical value at the 0.05 level of confidence of 11.070. Since the chi square
calculated value exceeded that of the critical value the null hypothesis was rejected.

The third developmental area considered was Perceptual Awareness. It was anticipated that items 11 through 14 would be placed by the judges in this area. Item 11 (identifying colors) was placed by all the judges in other developmental areas. No judge placed this item in the perceptual awareness area.

Item 12 (identifying five geometric shapes) was placed in a different area by two judges, as was item 13 (physical positions in space). Item 14 (identifying big and little objects from a work sheet) was placed in the perceptual awareness area by seven of the ten judges.

A contingency table was constructed. The chi square value that was calculated was 21.04. The degrees of freedom were set at three. The critical value at the .05 level of confidence was found to be 7.815. Since the chi square value exceeded the critical value the null hypothesis was rejected.

The fourth developmental area that was considered was Visual Memory. It was anticipated that two test items would be placed by the judges in this area. Item 15 (sequencing objects from memory) and item 16 (identifying missing objects) were the items so anticipated.

All ten judges placed both of these items in the visual memory developmental area. It was felt that no further analysis would need to be done. The null hypothesis was rejected.

The fifth developmental area that was considered was Auditory Skills. Test items 17, 18 and 19 were anticipated as being appropriately placed in this developmental area. Item 17 (repeating numbers) was placed in the hypothesized area by four of the judges; six judges placed that item in other
areas. Item 18 (repeating directions) was placed in another area by two of the judges, as was item 19 (following directions).

A contingency table was constructed. The chi square value was calculated at 4.81. The degrees of freedom were set at two. The critical value at the 0.05 level of confidence was found to be 5.991. Since the chi square value did not exceed the calculated value the null hypothesis was retained.

Item 20 of the PII (verbal associations) and item 21 (drawing a human figure) were single items each with its own developmental area defined. All ten judges placed each item in the hypothesized developmental area. Therefore, no further statistical analysis was felt to be necessary.

The chi square analyses conducted to measure the content validity of the PII indicate that the majority of the test items were clearly perceived by the judges as being assigned to the hypothesized developmental areas.

There were several items that were not so perceived, indicating a need to reconsider these items. The questions raised pertain to the appropriateness of the item for the developmental area to which they have been assigned, as well as to the appropriateness of the item for inclusion in a pre-kindergarten screening process.

There may be a need for a closer, more refined, definition for some of the developmental areas. The possibility of modifying or deleting these items may also have to be considered.

Item one (identifying body parts) may need to be assigned to another developmental area, even though it also appears to be a valid item for the screening process. Item five (folding a paper triangle) was not seen as a visual-motor task by one half of the judges. Several judges classified it as a visual-memory or auditory skills item. This task may more appropriately be
included in another developmental area. It appears to be an item that shows a good degree of content validity.

Item 11 (identifying colors) was placed in other developmental areas by all of the judges. The majority indicated that this item did not have a match in the categories listed.

Item 17 (repeating numbers) was placed in other developmental areas by six of the ten judges. These judges felt that the item did not constitute an auditory task. The item may not have been presented for the judges' opinions in a proper format, since the item seems to be one that should be categorized as an auditory skill.

The chi square analyses of the content validity inspection of the PII indicate that the null hypothesis is rejected.

Construct Validity

The concept of construct validity focuses primarily on the test score as a measure of the psychological characteristic of interest. The construct of interest for a particular test should be implanted in a conceptual framework. The conceptual framework specifies the meaning of the construct, distinguishes it from other constructs, and indicates how measures of the construct should relate to other variables. If an instrument has construct validity, examinees' scores will vary as the theory underlying the construct would predict.

The Project Intercept Inventory consists of 21 different tasks incorporated into seven developmental categories. To make a determination of the construct validity of the PII, factor analysis (principal components) was used. Factor analysis is a procedure for determining the number and nature of the
constructs or traits (factors) that underlie a set of variables. The variables commonly consist of test scores, or scores on psychological and sociological inventories, although they could be scores on measures of almost any kind.

Factor analysis is considered an extremely powerful and useful approach to behavioral data. It is a method for determining the underlying variables from various sets of measures. One application of factor analysis, the principal factors method, involves the solution of simultaneous linear equations. The roots obtained from the solution are referred to as eigenvalues. Eigenvalues are related to the percentage of variance which is accounted for through the factoring process. Klecka (1970) indicates that the eigenvalue is a special measure computed in the process of deriving the discriminant function. It is a measure of the relative importance of the function, and allows for the possibility of ignoring functions that do not appear to discriminate. The sum of the eigenvalues is a measure of the total variance existing in the discriminating variables. When a single eigenvalue is expressed as a percentage of the total sum of eigenvalues, an easy reference to the relative importance of the associated function is obtained. Since discriminant functions are derived in the order of their importance, this process can be stopped whenever the relative percentage is judged to be too small. There is no fixed rule for determining what is too small.

The use of the eigenvalue technique was used in the analysis of the construct validity of the Project Intercept Inventory.

**Hypothesis 5.** No more than one factor of the Project Intercept Inventory will be found with an eigenvalue equal to or greater than one, based upon the application of factor analysis. (Null hypothesis)
Alternative. There will be two or more factors of the Project Intercept Inventory found with an eigenvalue equal to or greater than one, based upon the application of factor analysis.

Discussion. An analysis of the data indicates that the first factor accounted for 27.1% of variance, with an eigenvalue of 5.69641. The remaining four factors delineated account for 27.2% of the variance, when combined. None of these factors is less than 1.

Factor one, which accounted for 27.1% of the variance was determined to be that of visual-motor skills.

Factor two had an eigenvalue of 1.92710 and accounted for 9.2% of the variance. This factor was determined to be visual and auditory memory skills.

Factor three had an eigenvalue of 1.52761 and accounted for 7.3% of the variance. This factor was determined to be that of perceptual awareness. Some of the tasks involved were recognizing shapes, recognizing colors, and awareness of physical positions in space.

Factor four had an eigenvalue of 1.12526. It accounted for 5.4% of the variance. This factor was determined to be gross motor skills.

Factor five had an eigenvalue of 1.10520, and accounted for 5.3% of the variance. This factor was determined to be a combination of gross and fine motor skills. Tasks involved in this factor were walking a balance beam, cutting a piece of paper with a scissors, and drawing a circle.

The application of the factor analysis technique to the data indicated that five factors were delineated. The test construction indicates that there are seven areas of development that are investigated. The factors of gross
motor skills, visual-motor skills, and perceptual awareness are found in the factor analysis calculations and in the developmental areas listed for the test.

Factor analysis indicates that the second factor delineated was a combination of auditory and visual skills. It would appear that some overlap exists, even though the test construction indicates that these two are separate developmental areas.

Factor analysis does not indicate any evidence of the developmental areas of verbal associations or the person-drawing task. This would seem to indicate that these two areas did not play a substantial role in the overall test results. Table 3 lists these results.

Based on these results, the null hypothesis is rejected.

Table 3

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.69641</td>
<td>27.1</td>
<td>27.1</td>
</tr>
<tr>
<td>2</td>
<td>1.92710</td>
<td>9.2</td>
<td>36.3</td>
</tr>
<tr>
<td>3</td>
<td>1.52761</td>
<td>7.3</td>
<td>43.6</td>
</tr>
<tr>
<td>4</td>
<td>1.12526</td>
<td>5.4</td>
<td>48.9</td>
</tr>
<tr>
<td>5</td>
<td>1.10520</td>
<td>5.3</td>
<td>54.2</td>
</tr>
</tbody>
</table>

Criterion Validity

Criterion validity demonstrates that test scores are systematically related to one or more outcome criteria. The fundamental question is always

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
how accurately can criterion performance be predicted from scores on the test.

Predictive validity involves using test scores to predict future behavior. In this process a predictive-validity coefficient is obtained, since criterion-related validity is typically expressed as a correlation coefficient. When a test is used to predict future behavior, predictive validity should be established.

The Pearson product-moment correlation coefficient was used in this study to obtain the correlation coefficients of validity. If two groups of test scores from the same group of people are compared to determine relationships, the Pearson product-moment correlation coefficient is the statistic most often used for this purpose (Mehrens & Lehmann, 1975).

**Hypothesis 6.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

**Alternative.** There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test taken 4 years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient.

**Discussion.** The data used in this analysis were gained from the results taken from the Project Intercept Inventory administered to the sample population. This screening took place in the Spring prior to the children's entrance into kindergarten the next school year. The data used for
comparison were gained from the total reading results from the administration of the California Achievement Test 4 years later.

The use of the Pearson product-moment correlation coefficient with these data indicated a moderate correlation at 0.4726 between the total results from the Project Intercept Inventory and the total reading results from the California Achievement Test (Hinkle, Wiersma, & Jurs, 1979). This accounts for 22.3% of the variance between these two results. These results proved to be significant at a level of confidence less than 0.05. The results would be regarded as indicating a substantial positive correlation between the two sets of data. This is based on the types of variables being correlated, a predictive screening instrument and an academic achievement test.

The results are considerably better than random chance when used for predictive purposes. This would indicate that the use of the Project Intercept Inventory as a screening instrument can have predictive value when used in the process of screening incoming kindergarten children.

Based on these data, the null hypothesis is rejected, since a significant correlation coefficient at a level of confidence less than 0.05 was found.

**Hypothesis 7.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken 4 years later by the children in the sample, based on the application of the Pearson product-moment correlation coefficient. (Null hypothesis)

**Alternative.** There will be a significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken 4 years later by the
children in the study, based on the application of the Pearson product-moment correlation coefficient.

Discussion. The data used in this analysis were gained from the results taken from the Project Intercept Inventory administered to the sample population. This screening took place in the spring prior to the children's entrance into kindergarten the following school year. The data used for comparison purposes were gained from the total arithmetic results from the administration of the California Achievement Test 4 years later. The use of the Pearson product-moment correlation coefficient technique with the data indicated a moderate correlation at 0.4440 between the total results from the Project Intercept Inventory and the total results from the California Achievement Test arithmetic scores. This accounts for 19.7% of the variance between these two variables. These results proved to be significant at a level of confidence less than 0.05. The outcome would be regarded as indicative of a substantial positive correlation between the two sets of data. Such results would point to a considerably higher predictive results than would be expected from random chance, given that the variables were results from a predictive screening instrument and the results from an academic achievement test.

These results would indicate that the use of the Project Intercept Inventory as a screening instrument can have predictive value when used in the process of screening children for kindergarten entrance.

Based on these data, the null hypothesis is rejected, since a significant correlation coefficient was found at a level of confidence less than 0.05.
Bias

Any index for evaluating bias should incorporate in it some control on the level of ability of the groups under study (Angoff, 1982). Examinees of the same ability level should have the same chance to perform equally well on a test regardless of what particular group they are a member, unless item bias is present in the test. The use of the item characteristic curve approach is considered by some to be the most theoretically sound of the methods used in detecting item bias in tests (Ironson & Subkoviak, 1979; Rudner & Convey, 1978; Scheuneman, 1979).

The item characteristic curve was used in this study to depict the relationship between ability and the percentage of children who answered an item correctly. In general, as the former increases, so does the latter. Ironson (1982) indicates that three parameters may be used to describe these curves. They are: the difficulty parameter, given by the ability level corresponding to the inflection point (that point at which the line bends or angles) of the curve; the discrimination value proportional to the slope of the curve at the inflection point; and the probability of examinees of infinitely low ability getting the item right.

In undertaking the consideration of the presence of bias, real or potential, in the PII, it was considered that test takers of the same ability should do equally well on the test. This would refer to the test as a whole, as well as each individual item. If the PII was unbiased the group to which the individual child belonged should not make a difference in terms of individual results.
Hypothesis 8. The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for boys and girls. (Null hypothesis)

Alternative. The general shapes of the item characteristic curves of the Project Intercept Inventory for boys and girls will not be substantially different.

Discussion. The term ability as used in this series of charts is intended to indicate levels based on raw scores. The raw scores were derived from the number of items to which each child was deemed to have successfully responded. The scores were ranked on a basis of low to high. The scores then were divided into nine levels for the purpose of developing the item characteristic curves. The use of nine levels allows for sufficient length for the item characteristic curves to be developed and demonstrated. The levels were derived by dividing the total number of possible scores by nine.

The information from 252 females' test results was used in this portion of the study. There were 274 males included. The discrepancies noted in the lower portions of many of the ICC graphs appear to be due to the small number of cases that fell in these categories.

Item characteristic curves were developed for this study for each of the 21 items of the Project Intercept Inventory. The ICCs making comparisons between the test results of females and males are discussed in this section. A general review of the graphs indicates little difference in the general shapes of the ICCs between boys and girls. Those portions from the fifth to the ninth levels are particularly similar in shape. This would represent 93.5%
of the female cases and 93.8% of the male cases which were included in the study.

Since such a high percentage of the results are found in the fifth to ninth columns, the analysis of the individual item ICCs will concentrate on those portions of the graphs.

Item 1 (identify human body parts): the ICC indicates very little difference between the results of females and males on this item, although the female results were mildly higher in the levels four through nine. The task was accomplished readily by the large majority of the children.

Item 2 (balance beam walking): the ICC statistics are very similar for this task. The results are virtually identical, with the exception that the males at the lowest portion of the graph had moderately lower results.

Table 4

<table>
<thead>
<tr>
<th>Gender Item Characteristic Curve Item 1- Identify Human Body Parts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
</tr>
<tr>
<td>Value</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>
Item 3 (jumping): again, the large majority of children of both genders recorded very similar ICC results. The males at the very lowest portion of the table had slightly higher results than did the females. Most of the children were able to accomplish this task at a good rate.

Item 4 (hopping): the ICC statistics indicate nearly even results for the majority of children regardless of gender. However, the results for females at the lowest end of the scale (levels one through three) indicated more difficulty in accomplishing the task.

Table 5

<table>
<thead>
<tr>
<th>Gender</th>
<th>Item</th>
<th>Characteristic</th>
<th>Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Item 2</td>
<td>Balance Beam Walking</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>Mean</td>
<td>Valid No.</td>
<td>Value</td>
</tr>
<tr>
<td>3</td>
<td>.8000</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>.7500</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>.7742</td>
<td>31</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>.8750</td>
<td>48</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>.8333</td>
<td>72</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>.9157</td>
<td>83</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
<td>9</td>
</tr>
</tbody>
</table>

Item 5 (folding a paper triangle): the ICC results indicate that the girls did mildly better on this task in a consistent fashion. This was especially the case for levels six through nine.
### Table 6
Gender Item Characteristic Curve
Item 3-Jumping

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Female Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.8000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.7500</td>
<td>12</td>
<td>4</td>
<td>.9000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>.9355</td>
<td>31</td>
<td>5</td>
<td>.8333</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>.9792</td>
<td>48</td>
<td>6</td>
<td>.9655</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>.9583</td>
<td>72</td>
<td>7</td>
<td>.9877</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>83</td>
<td>8</td>
<td>1.0000</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
<td>9</td>
<td>1.0000</td>
<td>27</td>
</tr>
</tbody>
</table>

### Table 7
Gender Item Characteristic Curve
Item 4-Hopping

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Female Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.4167</td>
<td>12</td>
<td>4</td>
<td>.8000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>.5161</td>
<td>31</td>
<td>5</td>
<td>.5833</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>.8125</td>
<td>48</td>
<td>6</td>
<td>.7241</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>.9167</td>
<td>72</td>
<td>7</td>
<td>.8765</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>.9036</td>
<td>83</td>
<td>8</td>
<td>.9737</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
<td>9</td>
<td>1.0000</td>
<td>27</td>
</tr>
</tbody>
</table>
Table 8
Gender Item Characteristic Curve
Item 5-Folding a Paper Triangle

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th>Value</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>4</td>
<td>.2500</td>
<td>4</td>
<td>.4000</td>
</tr>
<tr>
<td>5</td>
<td>.2581</td>
<td>5</td>
<td>.2500</td>
</tr>
<tr>
<td>6</td>
<td>.4792</td>
<td>6</td>
<td>.5862</td>
</tr>
<tr>
<td>7</td>
<td>.4583</td>
<td>7</td>
<td>.6790</td>
</tr>
<tr>
<td>8</td>
<td>.7470</td>
<td>8</td>
<td>.8816</td>
</tr>
<tr>
<td>9</td>
<td>.9565</td>
<td>9</td>
<td>.9630</td>
</tr>
</tbody>
</table>

Table 9
Gender Item Characteristic Curve
Item 6-Paper Cutting

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th>Value</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>3</td>
<td>.6000</td>
<td>2</td>
<td>.6000</td>
</tr>
<tr>
<td>4</td>
<td>.6667</td>
<td>4</td>
<td>.6000</td>
</tr>
<tr>
<td>5</td>
<td>.8387</td>
<td>5</td>
<td>.7917</td>
</tr>
<tr>
<td>6</td>
<td>.8542</td>
<td>6</td>
<td>.8621</td>
</tr>
<tr>
<td>7</td>
<td>.9028</td>
<td>7</td>
<td>.9259</td>
</tr>
<tr>
<td>8</td>
<td>.9880</td>
<td>8</td>
<td>.9868</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>9</td>
<td>.9630</td>
</tr>
</tbody>
</table>
Item 6 (paper cutting): again, the ICC statistics indicate no difference of note, except at the very lowest portion of the table, where the males did mildly better on the task. These results indicated that the large majority of children, regardless of gender, were able to accomplish the task successfully. There was more of a discrimination noted at the lower end of the table for all children in that area.

Item 7 (circle drawing): the ICCs reflect that there was virtually no difference in results for girls and boys, and that the vast majority of the children undertaking this task accomplished it accurately.

Item 8 (square drawing): virtually no difference is noted in the ICC general shape for this task, with the exception of levels two through six where the males scored at a mildly better rate.

Table 10
Gender Item Characteristic Curve
Item 7-Circle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>5</td>
<td>2</td>
<td>1.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.9167</td>
<td>12</td>
<td>4</td>
<td>1.0000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>.9355</td>
<td>31</td>
<td>5</td>
<td>.9533</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>.9375</td>
<td>48</td>
<td>6</td>
<td>1.0000</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>.9722</td>
<td>72</td>
<td>7</td>
<td>.9753</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>83</td>
<td>8</td>
<td>1.0000</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
<td>9</td>
<td>1.0000</td>
<td>27</td>
</tr>
</tbody>
</table>
Item 9 (triangle drawing): a review of the ICC statistics from these results indicates that there is no difference of any note from levels four through nine when gender is compared. The boys' results at the lowest levels were definitely lower.

Item 10 (open-ended square and circle drawing): this task proved to be a difficult one for most of the children, but little difference was noted between the statistics of the ICCs. Male results were mildly higher at levels seven through nine.

Item 11 (identifying colors): the ICC statistics indicate very little difference in the results for boys or girls from levels five through nine. The results for males were slightly better at the lower levels.

Table 11

<table>
<thead>
<tr>
<th>Item Characteristic Curve</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Mean</td>
<td>Valid No.</td>
</tr>
<tr>
<td>3</td>
<td>2.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0.2500</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>0.3548</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>0.6042</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>0.7083</td>
<td>72</td>
</tr>
<tr>
<td>8</td>
<td>0.9277</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 12
Gender Item Characteristic Curve
Item 9-Triangle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th></th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>.2000</td>
</tr>
<tr>
<td>4</td>
<td>.0000</td>
<td>12</td>
<td>.0000</td>
</tr>
<tr>
<td>5</td>
<td>.1290</td>
<td>31</td>
<td>.0833</td>
</tr>
<tr>
<td>6</td>
<td>.2500</td>
<td>48</td>
<td>.1379</td>
</tr>
<tr>
<td>7</td>
<td>.3194</td>
<td>72</td>
<td>.3827</td>
</tr>
<tr>
<td>8</td>
<td>.6145</td>
<td>83</td>
<td>.5789</td>
</tr>
<tr>
<td>9</td>
<td>.9565</td>
<td>23</td>
<td>.8519</td>
</tr>
</tbody>
</table>

Table 13
Gender Item Characteristic Curve Item 10-
Open-Ended Square and Circle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th></th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>.0000</td>
</tr>
<tr>
<td>4</td>
<td>.0000</td>
<td>12</td>
<td>.0000</td>
</tr>
<tr>
<td>5</td>
<td>.0645</td>
<td>31</td>
<td>.0000</td>
</tr>
<tr>
<td>6</td>
<td>.0625</td>
<td>48</td>
<td>.0690</td>
</tr>
<tr>
<td>7</td>
<td>.1250</td>
<td>72</td>
<td>.1111</td>
</tr>
<tr>
<td>8</td>
<td>.2530</td>
<td>83</td>
<td>.1711</td>
</tr>
<tr>
<td>9</td>
<td>.6957</td>
<td>23</td>
<td>.5185</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Item 12 (identifying geometric shapes): the ICCs indicate a consistently slightly higher result for males on this task, with the patterns being virtually identical.

Item 13 (positions in space): very little difference in results is seen on the ICC table between females and males from levels four through nine.

Item 14 (visual recognition of big and little): once again, little difference in results between boys and girls are seen on the ICC table. The results for males at levels two and three were slightly better than female results.

Item 15 (sequencing objects): very little difference is seen in the results of the ICC for this item, with the exception of the very lowest levels where female results are very slightly improved.

Table 14

<table>
<thead>
<tr>
<th>Gender</th>
<th>Item Characteristic Curve</th>
<th>Item 11-Identifying Colors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>Mean</td>
</tr>
<tr>
<td>3</td>
<td>2.8000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4.4167</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>4.5484</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>5.0417</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>6.8611</td>
<td>72</td>
</tr>
<tr>
<td>8</td>
<td>7.7470</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>7.8696</td>
<td>23</td>
</tr>
</tbody>
</table>
Table 15
Gender Item Characteristic Curve Item 12-Identifying Geometric Shapes

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th>Valid No.</th>
<th>Female</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.7500</td>
<td>12</td>
<td>4</td>
<td>1.5000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2.1290</td>
<td>31</td>
<td>5</td>
<td>1.5000</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>2.4167</td>
<td>48</td>
<td>6</td>
<td>2.4483</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2.8889</td>
<td>72</td>
<td>7</td>
<td>2.7531</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>4.1446</td>
<td>83</td>
<td>8</td>
<td>3.6447</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>4.9130</td>
<td>23</td>
<td>9</td>
<td>4.5556</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 16
Gender Item Characteristic Curve Item 13-Positions in Space

<table>
<thead>
<tr>
<th>Value</th>
<th>Male</th>
<th>Valid No.</th>
<th>Female</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.6000</td>
<td>5</td>
<td>2</td>
<td>1.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3.3333</td>
<td>12</td>
<td>4</td>
<td>3.2000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>3.9677</td>
<td>31</td>
<td>5</td>
<td>4.0833</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>4.4167</td>
<td>48</td>
<td>6</td>
<td>4.4483</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>4.6528</td>
<td>72</td>
<td>7</td>
<td>4.7901</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>5.0000</td>
<td>83</td>
<td>8</td>
<td>4.8421</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>5.0000</td>
<td>23</td>
<td>9</td>
<td>4.9630</td>
<td>27</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 17
Gender Item Characteristic Curve Item 14-
Visual Recognition of Big and Little

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.6000</td>
<td>5</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.8333</td>
<td>12</td>
<td>4</td>
<td>.8000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>.8065</td>
<td>31</td>
<td>5</td>
<td>.9167</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>.9792</td>
<td>48</td>
<td>6</td>
<td>.9655</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td>72</td>
<td>7</td>
<td>.9877</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>83</td>
<td>8</td>
<td>1.0000</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>23</td>
<td>9</td>
<td>1.0000</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 18
Gender Item Characteristic Curve Item 15-Sequencing Objects

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.2000</td>
<td>5</td>
<td>2</td>
<td>.8000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.9167</td>
<td>12</td>
<td>4</td>
<td>.8000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>1.0968</td>
<td>31</td>
<td>5</td>
<td>1.4583</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>1.6667</td>
<td>48</td>
<td>6</td>
<td>1.6897</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2.0000</td>
<td>72</td>
<td>7</td>
<td>1.9259</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>2.3133</td>
<td>83</td>
<td>8</td>
<td>2.1842</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>2.5217</td>
<td>23</td>
<td>9</td>
<td>2.4444</td>
<td>27</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## Table 19
Gender Item Characteristic Curve Item 16-
Naming Objects from Memory

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.6000</td>
<td>5</td>
<td>2</td>
<td>1.4000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.9167</td>
<td>12</td>
<td>4</td>
<td>1.5000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2.0968</td>
<td>31</td>
<td>5</td>
<td>2.0833</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>2.6250</td>
<td>48</td>
<td>6</td>
<td>2.2414</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2.7083</td>
<td>72</td>
<td>7</td>
<td>2.7407</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>2.9277</td>
<td>83</td>
<td>8</td>
<td>2.8947</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>2.9130</td>
<td>23</td>
<td>9</td>
<td>3.0000</td>
<td>27</td>
</tr>
</tbody>
</table>

## Table 20
Gender Item Characteristic Curve Item 17-
Repeating Digits from Memory

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Female Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.4000</td>
<td>5</td>
<td>2</td>
<td>1.4000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.6667</td>
<td>12</td>
<td>4</td>
<td>2.2000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2.4194</td>
<td>31</td>
<td>5</td>
<td>2.6250</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>2.7083</td>
<td>48</td>
<td>6</td>
<td>2.5862</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2.7917</td>
<td>72</td>
<td>7</td>
<td>2.9753</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>3.2771</td>
<td>83</td>
<td>8</td>
<td>3.1053</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>3.6522</td>
<td>23</td>
<td>9</td>
<td>3.7407</td>
<td>27</td>
</tr>
</tbody>
</table>
Item 16 (naming objects from memory): the ICC statistics suggest very little difference between results for boys and girls, although the male results are mildly improved at levels one through five.

Item 17 (repeating digits from memory): there are moderate differences noted at the lower portions of the table. The results are mixed in this area. However, the majority of the results indicate virtually no difference in results between the genders.

Item 18 (repeating directions): the ICC statistics show little difference at all in results between girls and boys on this item.

Item 19 (following a series of directions): the ICC indicates that there is very little difference in results for this item between girls and boys, although once again the results for males is very mildly higher than that of females at the levels one through four.

Table 21
Gender Item Characteristic Curve
Item 18-Repeating Directions

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Mean</td>
<td>Valid No.</td>
</tr>
<tr>
<td>3</td>
<td>.4000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.7500</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>1.0968</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>1.7083</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>2.1944</td>
<td>72</td>
</tr>
<tr>
<td>8</td>
<td>2.5542</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>2.7826</td>
<td>23</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 22
Gender Item Characteristic Curve Item 19-
Following a Series of Directions

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Female Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.4000</td>
<td>5</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.5833</td>
<td>12</td>
<td>4</td>
<td>1.1000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>1.5161</td>
<td>31</td>
<td>5</td>
<td>1.6250</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>2.0208</td>
<td>48</td>
<td>6</td>
<td>1.8621</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2.2500</td>
<td>72</td>
<td>7</td>
<td>2.0864</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>2.4699</td>
<td>83</td>
<td>8</td>
<td>2.5789</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>2.5652</td>
<td>23</td>
<td>9</td>
<td>2.5185</td>
<td>27</td>
</tr>
</tbody>
</table>

Item 20 (verbal associations): while proving to be mildly difficult for most of the children undertaking this task, the item showed little difference in results between the genders.

Item 21 (drawing a human figure): the ICC results reflect virtually no difference in results between male and female children on this task, with the exception of levels one through two where the male results were lower.

To obtain another perspective on the possibility of male-female bias being present in the PII test items, the corrected point-biserial technique was also used to make comparisons. This was done to gain an impression of how well each individual item of the PII would predict ability. The term ability used here refers to the total test score. The term predict as used here refers to the correlation between each item and the total test score. The
corrected point-biserial technique was used to correct for the overlap which would occur between the item and the total test score.

Table 23

Gender Item Characteristic Curve
Item 20-Verbal Associations

<table>
<thead>
<tr>
<th>Value</th>
<th>Male Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Female Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.0000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.2500</td>
<td>12</td>
<td>4</td>
<td>2.7000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2.8387</td>
<td>31</td>
<td>5</td>
<td>3.0000</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>3.5417</td>
<td>48</td>
<td>6</td>
<td>2.9655</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>4.5556</td>
<td>72</td>
<td>7</td>
<td>4.2099</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>5.0964</td>
<td>83</td>
<td>8</td>
<td>5.0263</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>5.8261</td>
<td>23</td>
<td>9</td>
<td>5.4444</td>
<td>27</td>
</tr>
</tbody>
</table>

The reason for the use of this technique was to gain additional information regarding the potential for gender bias being present in the various items of the PII. Lower corrected point-biserial numbers would suggest the potential of bias being present in the items. Higher corrected point-biserial numbers suggest better ability being present, and serve as a predictor of later academic success. Thorndike (1971) points out that the point-biserial method will tend to favor items of average difficulty, with proportion correct around 0.5. These items will then tend to make more discriminations between good and poor students than do items with high or low p-values.
This means that the point-biserial correlation coefficient can be said to be a combined measure of item-criterion relationship and of difficulty level.

The procedure used was to calculate two point-biserial correlations, one for each of the groups. Comparisons were then made between the two correlations, to see what differences might occur. An arbitrary cutoff point was set at 0.25 for purposes of making comparisons between the results gained from the use of the corrected point-biserial technique. It was felt that results falling below this cutoff level for any test item could indicate that the item discriminated unfairly for particular groups of students.

A review of the corrected point-biserial results for the male-female categories indicates that 3 of the 21 items suggest the potential for bias being present, based upon the arbitrary cutoff level of 0.25. These items

| Table 24 |
| Gender Item Characteristic Curve Item 21- Drawing a Human Figure |

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Mean</td>
</tr>
<tr>
<td>3</td>
<td>.8000</td>
</tr>
<tr>
<td>4</td>
<td>2.2500</td>
</tr>
<tr>
<td>5</td>
<td>4.3871</td>
</tr>
<tr>
<td>6</td>
<td>5.8750</td>
</tr>
<tr>
<td>7</td>
<td>6.9583</td>
</tr>
<tr>
<td>8</td>
<td>8.5301</td>
</tr>
<tr>
<td>9</td>
<td>10.3913</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Item</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.3844</td>
<td>.4593</td>
</tr>
<tr>
<td>2</td>
<td>.1488</td>
<td>.3772</td>
</tr>
<tr>
<td>3</td>
<td>.1984</td>
<td>.3159</td>
</tr>
<tr>
<td>4</td>
<td>.4281</td>
<td>.3316</td>
</tr>
<tr>
<td>5</td>
<td>.3859</td>
<td>.4160</td>
</tr>
<tr>
<td>6</td>
<td>.2908</td>
<td>.2905</td>
</tr>
<tr>
<td>7</td>
<td>.1318</td>
<td>.0062</td>
</tr>
<tr>
<td>8</td>
<td>.5316</td>
<td>.0583</td>
</tr>
<tr>
<td>9</td>
<td>.4977</td>
<td>.5912</td>
</tr>
<tr>
<td>10</td>
<td>.4687</td>
<td>.4320</td>
</tr>
<tr>
<td>11</td>
<td>.3399</td>
<td>.2668</td>
</tr>
<tr>
<td>12</td>
<td>.3960</td>
<td>.4992</td>
</tr>
<tr>
<td>13</td>
<td>.4371</td>
<td>.4908</td>
</tr>
<tr>
<td>14</td>
<td>.4210</td>
<td>.4526</td>
</tr>
<tr>
<td>15</td>
<td>.4802</td>
<td>.3575</td>
</tr>
<tr>
<td>16</td>
<td>.3899</td>
<td>.5597</td>
</tr>
<tr>
<td>17</td>
<td>.3992</td>
<td>.3675</td>
</tr>
<tr>
<td>18</td>
<td>.5494</td>
<td>.5674</td>
</tr>
<tr>
<td>19</td>
<td>.4558</td>
<td>.5169</td>
</tr>
<tr>
<td>20</td>
<td>.4863</td>
<td>.4737</td>
</tr>
<tr>
<td>21</td>
<td>.5247</td>
<td>.5909</td>
</tr>
</tbody>
</table>
discriminated differently between females and males. The items were balance beam walking, jumping and sequencing of objects.

The statistics for this analysis are listed in Table 25. The preponderance of the PII test items, 18, would seem to be free of the potential for bias for females or males, based on the arbitrary cutoff point of 0.25.

Based on these data, the null hypothesis cannot be accepted. The items of the PII do not appear to be substantially biased as regards gender.

Bias

The use of the item characteristic curve approach is considered by some to be the most theoretically sound of the methods used in detecting item bias in tests (Ironson & Subkoviak, 1979; Rudner & Convey, 1978; Scheuneman, 1979).

The item characteristic curve was used in this study to depict the relationship between ability and the percentage of children who answered correctly a given item. In general, as the former increases, so does the latter.

In undertaking the consideration of the presence of bias in the PII, it was considered that test takers of the same ability should do equally well on the test, regardless of the racial group to which they belong. This would refer to the test as a whole, as well as on individual items.

Hypothesis 9. The general shapes of the item characteristic curves for the Project Intercept Inventory will be substantially different for racial majority and minority children. (Null hypothesis)

Alternative. The general shapes of the item characteristic curves for the Project Intercept Inventory for racial majority and minority children
Discussion. The term ability as used in this series of charts is intended to indicate levels based on raw scores. The raw scores were derived from the number of items to which each child was deemed to have successfully responded. The scores were ranked on a basis of low to high. The scores then were divided into nine levels for the purpose of developing the item characteristic curves. The use of the nine levels allowed for sufficient length for the item characteristic curves to be developed and demonstrated. The levels were derived by dividing the total possible scores by nine.

The information from 309 majority children's test results were used in this study. There were 210 minority children's results included. For the purposes of this study, minority children were considered to be those who had been designated by their parents as being other than Caucasian. This information was gained from the records of the Grand Rapids Public Schools.

The discrepancies noted in the lower portions of many of the ICC graphs are due to the smaller number of cases that fell in these lower levels.

As can be seen in viewing the ICC graphs, there is little difference to be noted in the results from the various test items, especially in the portions which include the ability categories noted as five through nine. These results would represent 97.1% of the majority children studies, and 89% of the minority children who were included in this study. The specific analysis of items will concentrate on the columns from five to nine where by far the largest number of children's results were gathered. Item characteristic curves were developed in this study for each of the 21 items of the Project Intercept Inventory. The ICCs making comparisons between the test results of majority and minority children are discussed in this section.
Item 1 (identifying body parts): the ICC statistics suggest that there is little difference in these results, with the exception of the very lowest portion of the table where majority children results were moderately lower than the results for minority group children. Both groups of children did well on this task.

Table 26
Race Item Characteristic Curve Item 1-
Identify Human Body Parts

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>2</td>
<td>2</td>
<td>.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.5714</td>
<td>7</td>
<td>3</td>
<td>.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.8800</td>
<td>25</td>
<td>4</td>
<td>.6667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.9143</td>
<td>35</td>
<td>5</td>
<td>.6923</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.9885</td>
<td>87</td>
<td>6</td>
<td>.8372</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.9750</td>
<td>120</td>
<td>7</td>
<td>.9677</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
<td>1.0000</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Item 2 (balance beam walking): the ICC statistics for this item indicate very little difference between minority and majority children's results, with the exception of the levels two and three where majority children seemed to do mildly better. Most of the children were able to accomplish this task at a very good rate.
Table 27
Race Item Characteristic Curve
Item 2-Balance Beam Walking

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority</th>
<th>Value</th>
<th>Minority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Valid No.</td>
<td>Mean</td>
</tr>
<tr>
<td>3</td>
<td>1.0000</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>.7143</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.7600</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>.8571</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>.8736</td>
<td>87</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>.9250</td>
<td>120</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
</tr>
</tbody>
</table>

Item 3 (jumping): again, the majority children's results were mildly better at levels one through three. Beyond that point there is very little difference in results. The statistics of the ICCs indicate that most of the children in both groups were able to accomplish this task at a good level.

Item 4 (hopping): the results gained from reviewing the ICC table strongly indicate that the minority children were able to perform this task at a higher rate than the majority children. The minority children's results were consistently higher when compared with the results of the majority group.

Item 5 (paper folding): the ICC results indicate that the largest number of children of both groups performed equally well on this item from levels...
five through nine. However, the minority group children performed better in levels one through four.

Table 28
Race Item Characteristic Curve
Item 3-Jumping

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>2</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.7143</td>
<td>7</td>
<td>3</td>
<td>.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.8800</td>
<td>25</td>
<td>4</td>
<td>.8667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.9429</td>
<td>35</td>
<td>5</td>
<td>.9231</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.9655</td>
<td>87</td>
<td>6</td>
<td>1.0000</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>120</td>
<td>7</td>
<td>.9839</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
<td>1.0000</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Item 6 (paper cutting): the results of the ICC figures indicate that most of the children were able to perform this mildly difficult task at a good rate. The results for the majority group children were somewhat stronger on the lower portion of the graph, from levels two through four.

Item 7 (drawing a circle): there is virtually no difference between the results of the ICC table regarding the results for this task, except for a slight improvement in results for minority children at levels three through six.

Item 8 (drawing a square): there is very little difference to be seen in the results of the ICC graph at any level for this item. The minority
children's group did show mildly improved results in levels two through five.

Item 9 (drawing a triangle): the ICCs reflect that the results of both curves are identical, with a slight improvement in results for minority children in levels four through nine.

Table 29
Race Item Characteristic Curve
Item 4-Hopping

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>.0000</td>
<td>2</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.2857</td>
<td>7</td>
<td></td>
<td>3</td>
<td>.0000</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.4000</td>
<td>25</td>
<td></td>
<td>4</td>
<td>.7333</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.6571</td>
<td>35</td>
<td></td>
<td>5</td>
<td>.7308</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.8391</td>
<td>87</td>
<td></td>
<td>6</td>
<td>.8605</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.9250</td>
<td>120</td>
<td></td>
<td>7</td>
<td>.9516</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td></td>
<td>8</td>
<td>.9756</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Item 10 (drawing an open-ended square and circle figure): the ICC results indicate that the item showed little difference between the two groups. The results suggest that this task proved difficult for both groups of children.

Item 11 (identifying colors): there is close continuity between the ICCs for this item. The results suggest little difference between the groups. The
majority group's results were slightly better from levels five through nine.

Item 12 (identifying a circle, square, triangle, rectangle and diamond): again, the ICCs reflect virtually no difference in the results. The item seems to be equally difficult for both groups, although the results indicate that minority children performed at a slightly better rate in levels one through three.

Table 30
Race Item Characteristic Curve Item 5-
Folding a Paper Triangle

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority</th>
<th>Minority</th>
<th>Value</th>
<th>Majority</th>
<th>Minority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Valid No.</td>
<td></td>
<td>Mean</td>
<td>Valid No.</td>
</tr>
<tr>
<td>3</td>
<td>.0000</td>
<td>2</td>
<td>2</td>
<td>.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.0000</td>
<td>7</td>
<td>3</td>
<td>.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.2000</td>
<td>25</td>
<td>4</td>
<td>.4667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.5429</td>
<td>35</td>
<td>5</td>
<td>.3077</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.5862</td>
<td>87</td>
<td>6</td>
<td>.4884</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.8167</td>
<td>120</td>
<td>7</td>
<td>.5806</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>.9394</td>
<td>33</td>
<td>8</td>
<td>.8049</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Item 13, (positions in space): the table results indicate that the ICCs are nearly identical, except at the lowest portion of the table where minority children's results were slightly better in levels one through three.
Table 31
Race Item Characteristic Curve
Item 6-Paper Cutting

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>2</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.5714</td>
<td>7</td>
<td>3</td>
<td>.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.8000</td>
<td>25</td>
<td>4</td>
<td>.6667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.8571</td>
<td>35</td>
<td>5</td>
<td>.8462</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.9195</td>
<td>87</td>
<td>6</td>
<td>.8605</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.9833</td>
<td>120</td>
<td>7</td>
<td>.9355</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>.9697</td>
<td>33</td>
<td>8</td>
<td>1.0000</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 32
Race Item Characteristic Curve
Item 7-Circle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>2</td>
<td>2</td>
<td>1.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.8571</td>
<td>7</td>
<td>3</td>
<td>1.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.8800</td>
<td>25</td>
<td>4</td>
<td>1.0000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.9714</td>
<td>35</td>
<td>5</td>
<td>1.0000</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.9655</td>
<td>87</td>
<td>6</td>
<td>.9535</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>120</td>
<td>7</td>
<td>.9839</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
<td>1.0000</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 33
Race Item Characteristic Curve
Item 8-Square Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>2</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.1429</td>
<td>7</td>
<td>3</td>
<td>.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.2400</td>
<td>25</td>
<td>4</td>
<td>.2000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.6000</td>
<td>35</td>
<td>5</td>
<td>.3462</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.7011</td>
<td>87</td>
<td>6</td>
<td>.5581</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.9167</td>
<td>120</td>
<td>7</td>
<td>.7581</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
<td>.9756</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 34
Race Item Characteristic Curve
Item 9-Triangle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>2</td>
<td>2</td>
<td>.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.0000</td>
<td>7</td>
<td>3</td>
<td>.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.0000</td>
<td>25</td>
<td>4</td>
<td>.0000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.1714</td>
<td>35</td>
<td>5</td>
<td>.1538</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.3218</td>
<td>87</td>
<td>6</td>
<td>.2558</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.5583</td>
<td>120</td>
<td>7</td>
<td>.4032</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>.8788</td>
<td>33</td>
<td>8</td>
<td>.7073</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>.9333</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 35
Race Item Characteristic Curve Item 10-
Open-Ended Square and Circle Drawing

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.0000</td>
<td>7</td>
<td>3</td>
<td>.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.0000</td>
<td>25</td>
<td>4</td>
<td>.0000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.0857</td>
<td>35</td>
<td>5</td>
<td>.0385</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.1724</td>
<td>87</td>
<td>6</td>
<td>.0465</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>.2167</td>
<td>120</td>
<td>7</td>
<td>.0484</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>.5455</td>
<td>33</td>
<td>8</td>
<td>.2195</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>.6667</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 36
Race Item Characteristic Curve
Item 11-Identifying Colors

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.5000</td>
<td>5</td>
<td>2</td>
<td>.8000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3.4286</td>
<td>7</td>
<td>3</td>
<td>3.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4.8000</td>
<td>25</td>
<td>4</td>
<td>4.4667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>5.1943</td>
<td>35</td>
<td>5</td>
<td>3.9231</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>7.1839</td>
<td>87</td>
<td>6</td>
<td>5.1395</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>7.8417</td>
<td>120</td>
<td>7</td>
<td>6.6774</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>7.8788</td>
<td>33</td>
<td>8</td>
<td>7.7073</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>7.9333</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 37
Race Item Characteristic Curve Item 12-
Identifying Geometric Shapes

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.0000</td>
<td>7</td>
<td>3</td>
<td>1.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1.8000</td>
<td>25</td>
<td>4</td>
<td>1.4667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2.3714</td>
<td>35</td>
<td>5</td>
<td>1.9615</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>2.8966</td>
<td>87</td>
<td>6</td>
<td>2.4884</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>3.9833</td>
<td>120</td>
<td>7</td>
<td>2.7742</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>4.7273</td>
<td>33</td>
<td>8</td>
<td>3.6829</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>4.8000</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 38
Race Item Characteristic Curve
Item 23-Positions in Space

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0000</td>
<td>5</td>
<td>2</td>
<td>1.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3.7143</td>
<td>7</td>
<td>3</td>
<td>2.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3.9600</td>
<td>25</td>
<td>4</td>
<td>3.0667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>4.5429</td>
<td>35</td>
<td>5</td>
<td>4.1583</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>4.7816</td>
<td>87</td>
<td>6</td>
<td>4.3488</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>4.9333</td>
<td>120</td>
<td>7</td>
<td>4.6290</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>4.9697</td>
<td>33</td>
<td>8</td>
<td>4.9024</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>5.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 39

**Race Item Characteristic Curve Item 14—Visual Recognition of Big and Little**

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.5000</td>
<td>2</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.7413</td>
<td>7</td>
<td>3</td>
<td>.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.8800</td>
<td>25</td>
<td>4</td>
<td>.8667</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>.9714</td>
<td>35</td>
<td>5</td>
<td>.8077</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>.9885</td>
<td>87</td>
<td>6</td>
<td>.9535</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>120</td>
<td>7</td>
<td>1.0000</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>1.0000</td>
<td>33</td>
<td>8</td>
<td>1.0000</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Item 14 (identifying big and little): very little difference is noted between the ICC results for this item. It seems to provide only a mild degree of difficulty for children in this age group.

Item 15 (sequencing objects): the ICC was seen to be very close together for this item's results. The indication was that the item was essentially equal in difficulty for both groups of children. The level of difficulty appears to be moderately high for this age group.

Item 16 (recalling objects): the ICC table reflected little change in pattern for the results of the two groups. The majority children's group results were mildly higher in levels two through four, with the minority group's results being slightly higher for levels five through nine.
### Table 40
Race Item Characteristic Curve  
Item 15- Sequencing Objects

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.5000</td>
<td>5</td>
<td>2</td>
<td>.8000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.1429</td>
<td>7</td>
<td>3</td>
<td>.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1.0400</td>
<td>25</td>
<td>4</td>
<td>.7333</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>1.6571</td>
<td>35</td>
<td>5</td>
<td>1.5385</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>1.8851</td>
<td>87</td>
<td>6</td>
<td>1.6744</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>2.2417</td>
<td>120</td>
<td>7</td>
<td>2.0645</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>2.4545</td>
<td>33</td>
<td>8</td>
<td>2.2683</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>2.6000</td>
<td>15</td>
</tr>
</tbody>
</table>

### Table 41
Race Item Characteristic Curve Item 16-  
Naming Objects from Memory

<table>
<thead>
<tr>
<th>Value</th>
<th>Majority Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Minority Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.0000</td>
<td>5</td>
<td>2</td>
<td>4.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.1429</td>
<td>7</td>
<td>3</td>
<td>1.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1.9600</td>
<td>25</td>
<td>4</td>
<td>1.5333</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2.3429</td>
<td>35</td>
<td>5</td>
<td>2.1923</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>2.6322</td>
<td>87</td>
<td>6</td>
<td>1.6047</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>2.9000</td>
<td>120</td>
<td>7</td>
<td>2.8548</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>2.9394</td>
<td>33</td>
<td>8</td>
<td>2.9512</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>3.0000</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 42
Race Item Characteristic Curve Item 17- Repeating Digits from Memory

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.0000</td>
<td>5</td>
<td>2</td>
<td>1.4000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.1429</td>
<td>7</td>
<td>3</td>
<td>2.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2.4400</td>
<td>25</td>
<td>4</td>
<td>1.8000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2.5714</td>
<td>35</td>
<td>5</td>
<td>2.6154</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>2.7816</td>
<td>87</td>
<td>6</td>
<td>2.7442</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>3.1333</td>
<td>120</td>
<td>7</td>
<td>3.0645</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>3.7273</td>
<td>33</td>
<td>8</td>
<td>3.4146</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>3.6000</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 43
Race Item Characteristic Curve Item 18- Repeating Directions

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>2</td>
<td>.2000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>.4286</td>
<td>7</td>
<td>3</td>
<td>.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>.9600</td>
<td>25</td>
<td>4</td>
<td>1.1333</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>1.6000</td>
<td>35</td>
<td>5</td>
<td>1.3846</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>2.0460</td>
<td>87</td>
<td>6</td>
<td>1.6512</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>2.6333</td>
<td>120</td>
<td>7</td>
<td>2.2258</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>2.6970</td>
<td>33</td>
<td>8</td>
<td>2.5122</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>9</td>
<td>2.8667</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 44

**Race Item Characteristic Curve Item 19—Follow a Series of Directions**

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.0000</td>
<td>5</td>
<td>2</td>
<td>.0000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.4286</td>
<td>7</td>
<td>3</td>
<td>.6667</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1.8000</td>
<td>25</td>
<td>4</td>
<td>1.3333</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2.0286</td>
<td>35</td>
<td>5</td>
<td>1.3846</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>2.1494</td>
<td>87</td>
<td>6</td>
<td>1.8837</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>2.5083</td>
<td>120</td>
<td>7</td>
<td>2.1774</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>2.6061</td>
<td>33</td>
<td>8</td>
<td>2.5610</td>
<td>41</td>
</tr>
</tbody>
</table>

### Table 45

**Race Item Characteristic Curve Item 20—Verbal Associations**

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3.0000</td>
<td>5</td>
<td>2</td>
<td>.6000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.7143</td>
<td>7</td>
<td>3</td>
<td>1.3333</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3.3200</td>
<td>25</td>
<td>4</td>
<td>2.3333</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>3.2286</td>
<td>35</td>
<td>5</td>
<td>2.7692</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>4.4368</td>
<td>87</td>
<td>6</td>
<td>3.3721</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>5.0083</td>
<td>120</td>
<td>7</td>
<td>4.2581</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>5.6970</td>
<td>33</td>
<td>8</td>
<td>5.1707</td>
<td>41</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Item 17 (repeating digits): the ICCs for this item indicate a closely identical pattern, suggesting little difference between the two groups in terms of their results. There is a very small increase in the results for minority children on a consistent basis in this pattern.

Item 18 (repeating verbal directions): the ICC pattern suggested that the two groups were nearly equal in performance of this task, with the exception of the lower one third portion of the graph. At this point the minority children performed at a mildly higher rate. The level of difficulty noted for this item was considered moderately high.

Item 19 (following directions): the level of difficulty for this item appeared to be moderately high. Both groups of children performed at nearly identical levels, according to the ICC pattern which was formed.

Item 20 (verbal associations): the ICC reflected a pattern of results which were very close for the two groups of children. The results for the majority children's group was mildly higher in levels two through six. The level of difficulty appeared to be moderately high for all of the children.

Item 21 (drawing a human figure): the results of the ICCs indicate a very close parallel for the two groups of children with this item. The levels from four through nine are virtually identical.

To obtain another perspective on the possibility of bias being present in the PII test items, the corrected point-biserial technique was also used to make comparisons. This was done to gain an impression of how well each individual item of the PII would predict ability. The term ability as used here refers to the total test score. The term predict as used here refers to the correlation between each item and the total test score. The corrected point-biserial technique was used to correct for the overlap which would occur.
between the item and the total test score.

This technique was used to gain additional information regarding the potential for majority-minority bias being present in the various items of the PII. If low corrected point-biserial results were obtained, they could indicate the potential of bias being present in the items. Higher results from the use of the corrected point-biserial technique would be suggested of higher ability among the children being screened, and could also serve as a predictor of later school success.

Table 46
Race Item Characteristic Curve Item 21-
Drawing a Human Figure

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
<th>Value</th>
<th>Mean</th>
<th>Valid No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.0000</td>
<td>2</td>
<td>2</td>
<td>2.4000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.4286</td>
<td>7</td>
<td>3</td>
<td>.0000</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4.4000</td>
<td>25</td>
<td>4</td>
<td>2.4000</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>5.7714</td>
<td>35</td>
<td>5</td>
<td>3.9615</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>7.2644</td>
<td>87</td>
<td>6</td>
<td>6.4419</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>8.6000</td>
<td>120</td>
<td>7</td>
<td>7.1290</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>10.6667</td>
<td>33</td>
<td>8</td>
<td>8.5854</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>10.7333</td>
<td>15</td>
</tr>
</tbody>
</table>

The procedure used was to calculate two corrected point-biserial correlations, one for each of the groups, per test item. Comparisons were then

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
made between the two correlations, to see what differences might occur. An
arbitrary cutoff point was set at 0.25 for purposes of making comparisons
between the results of the groups gained from the use of the corrected point-
biserial technique. Thorndike (1971) points out that the point-biserial
method will tend to favor items of average difficulty, with proportion
correct around 0.5. These items will then tend to make more
discriminations between good and poor students than do items with high
or low p-values.

The results gained from the use of the corrected point-biserial tech­
nique indicated that only one item of the PII was found lower than the 0.25
cutoff level for the minority group children. This was for item seven
(drawing a circle).

The majority group children had corrected point-biserial results below
the 0.25 cutoff level for three items. These were items two (balance beam
walking), three (jumping), and six (paper cutting).

The statistics for this analysis are listed in Table 47.

These results would strongly suggest that the various items of the PII,
with a very few exceptions, do not demonstrate bias when results are com­
pared between majority and minority groups of children.

Based on these results, the null hypothesis must be rejected. The
analyses of data do not suggest the substantial presence of bias in the items
of the Project Intercept Inventory as related to majority and minority
children.

In Chapter V, Summary and Conclusions, a review of the study find­
ings will be done, along with the study implications, limitations of the study
and suggestions for further research.
<table>
<thead>
<tr>
<th>Item</th>
<th>Majority</th>
<th>Minority</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.3359</td>
<td>.4514</td>
</tr>
<tr>
<td>2</td>
<td>.1811</td>
<td>.3023</td>
</tr>
<tr>
<td>3</td>
<td>.2301</td>
<td>.2876</td>
</tr>
<tr>
<td>4</td>
<td>.4592</td>
<td>.3653</td>
</tr>
<tr>
<td>5</td>
<td>.4558</td>
<td>.3339</td>
</tr>
<tr>
<td>6</td>
<td>.2274</td>
<td>.3589</td>
</tr>
<tr>
<td>7</td>
<td>.1986</td>
<td>.0062</td>
</tr>
<tr>
<td>8</td>
<td>.5316</td>
<td>.5608</td>
</tr>
<tr>
<td>9</td>
<td>.4635</td>
<td>.4555</td>
</tr>
<tr>
<td>10</td>
<td>.2588</td>
<td>.3408</td>
</tr>
<tr>
<td>11</td>
<td>.4060</td>
<td>.4291</td>
</tr>
<tr>
<td>12</td>
<td>.4567</td>
<td>.4145</td>
</tr>
<tr>
<td>13</td>
<td>.3970</td>
<td>.4312</td>
</tr>
<tr>
<td>14</td>
<td>.2751</td>
<td>.4439</td>
</tr>
<tr>
<td>15</td>
<td>.3822</td>
<td>.4725</td>
</tr>
<tr>
<td>16</td>
<td>.3804</td>
<td>.5826</td>
</tr>
<tr>
<td>17</td>
<td>.3600</td>
<td>.4233</td>
</tr>
<tr>
<td>18</td>
<td>.5728</td>
<td>.5308</td>
</tr>
<tr>
<td>19</td>
<td>.3745</td>
<td>.5436</td>
</tr>
<tr>
<td>20</td>
<td>.3365</td>
<td>.5700</td>
</tr>
<tr>
<td>21</td>
<td>.5192</td>
<td>.5653</td>
</tr>
</tbody>
</table>
CHAPTER V

SUMMARY AND CONCLUSIONS

Introduction

The major purpose of this study was to examine a prekindergarten screening test to determine if certain factors were present in the test. The test involved in this study was the Project Intercept Inventory.

This test has been used as part of a battery to screen children for kindergarten entrance. Decisions are made, based in part on results gained from the use of this test, which directly affect the education of children at an early point in their schooling. The information gained from the use of this test is used in making judgments as to whether or not an individual child appears ready to successfully undertake the kindergarten experience. Decisions are often made as well that individual children have demonstrated concerns as they undertook the developmental tasks and questions presented to them by the administration of this instrument. When this is the case, parents are asked to allow further evaluation of the child to take place. This evaluation, in turn, frequently leads to the recommendation for placement of the child in an alternative program to kindergarten. The recommended programs are often general education alternatives, but they may also be Special Education programs, if the child qualifies for such programming following the evaluation.

No examination had ever been done to determine if this locally developed test was reliable and valid in its results. This study undertook to
examine these factors in the test, as well as an examination of its predictive results and a further examination of the possible presence of bias in the makeup of the test.

Chapter Summarizations

Chapter I of this study undertook a review of the circumstances under which the testing instrument under discussion, the Project Intercept Inventory, was developed. The problem that was examined in this study was the analysis of the capabilities of the Project Intercept Inventory to accurately screen children for entrance into kindergarten. A list of definitions of terms that would be used frequently in the study was provided.

The hypotheses to be examined in the study were listed. These pertained to: item difficulty, item discrimination, reliability, content validity, construct validity, criterion validity, predictability, and the possible presence of bias pertaining to gender and race.

Chapter II consisted of a review of the literature pertaining to preschool screening and testing instruments that have been developed for prekindergarten screening purposes. The literature has a considerable amount of information in these areas.

Chapter III described the city of Grand Rapids, Michigan and the school population which was included in the study. A description of the sample used in the study as well as the method of data collection was included.

The preschool screening operation which is named Project Intercept was described to aid in making more clear the scope of involvement and importance that the PII effects.

The last portion of this chapter discussed the statistical analyses which would be used to examine the various hypotheses that had been put forth.
Chapter IV described the findings of the statistical analyses used to examine the hypotheses of this study. These findings are summarized as follows:

**Null Hypothesis 1.** Null Hypothesis 1 states that the p-value of individual items of the Project Intercept Inventory will be less than 0.2, greater than 0.8. The hypothesis was concerned with the item difficulty of the PII. If the test items were appropriately difficult for the agelevel children who would be screened, the information gained should be useful in determining readiness for kindergarten.

The statistical analysis used in testing this hypothesis indicated that of the 21 test items examined seven fell within the limits stipulated to determine appropriate difficulty. Twelve of the test items fell above the stipulated maximum range and would be interpreted as too easy for children of this age level, and therefore would not discriminate well in terms of kindergarten readiness. The remaining items were felt to be moderately too easy, even though the results fell within the stipulated range.

These findings indicated that the null hypothesis could not be rejected. A majority of the test items were statistically outside of the parameters set.

**Null Hypothesis 2.** The individual test items of the Project Intercept Inventory will have correlation coefficients less than 0.15 when compared with total test scores, as measured by the corrected point-biserial technique for dichotomous items, and the corrected Pearson product-moment correlation coefficient for nondichotomous items.

The statistical analyses used to examine this hypothesis indicate that of the dichotomous items examined only one (drawing a circle) did not meet
the criterion established. All of the nondichotomous items met the established criterion at the 0.05 level of confidence.

These results strongly indicate that no particular item would impair the ability to discriminate among test takers of various ability levels when the PII is used.

Since the individual items of the PII were found to be discriminating between test takers of high and low ability when compared with results of the test taken as a whole, the null hypothesis was rejected.

**Null Hypothesis 3.** The alpha coefficient for the total test of the Project Intercept Inventory will be less than 0.6.

The alpha coefficient was selected for use in measuring the reliability of the PII because it is the mean of all possible split-half coefficients. It is the value expected when two random samples of items from a pool like those in the given test are correlated.

The computerized statistical analysis of the data indicated that the alpha coefficient for the total test of the PII was found to be 0.7766. Since this was the case, the null hypothesis was rejected.

**Null Hypothesis 4.** There will be no significant agreements among the judges when assigning the test items of the PII to the listed developmental areas, at the 0.05 level of confidence based on the use of the chi square technique.

The hypothesis was concerned with examining the content validity of the PII. The procedure followed was to ask ten judges to match the various items of the PII with the developmental area in which the judges felt the item should be placed. There were seven developmental areas listed.
The data were tested with the chi square of independence technique at the 0.05 level of confidence. The chi square technique compares proportions actually observed in a study with the proportions expected, to see if they were significantly different.

The first developmental area measured was that of Gross Motor. The chi square value was calculated at 25.45. With three degrees of freedom at the 0.05 level of confidence a critical value was found at 7.815. Since the chi square value exceeded the critical value the null hypothesis was rejected.

The second developmental area considered was Visual Motor. The chi square value was calculated at 132.99. The degrees of freedom were set at five, which yielded a critical value at the .05 level of confidence of 11.070. Since the chi square calculated value exceeded that of the critical value the null hypothesis was rejected.

The third developmental area considered was Perceptual Awareness. The chi square value calculated was 21.04. The degrees of freedom were set at three. At the 0.05 level of confidence the critical value was found to be 7.815. Since the chi square value exceeded the critical value the null hypothesis was rejected.

The fourth developmental area considered was that of Visual Memory. Two test items were assigned to this area. All ten judges placed both items in the appropriate developmental area without exception. No further analysis was felt to be necessary.

The fifth developmental area considered was that of Auditory Skills. The chi square value was calculated at 4.81. The degrees of freedom were set at two. At the 0.05 level of confidence the critical value was found to be 5.991. Since the chi square value did not exceed the critical value the null
hypothesis was retained.

The sixth developmental area considered was that of Verbal Associations. This area had one item assigned to it. All judges placed the item in the correct developmental area without exception. No statistical analysis was felt to be necessary.

The seventh developmental area considered was Draw-A-Person. This area had one item assigned to it. All judges placed the item in the correct developmental area without exception. No statistical analysis was felt to be necessary.

The chi square analyses of the content validity inspection of the PII indicate that the null hypothesis was rejected.

**Null Hypothesis 5.** No more than one factor of the Project Intercept Inventory will be found with an eigenvalue equal to or greater than one, based upon the application of factor analysis.

The hypothesis examined the construct validity of the PII. The concept of construct validity focuses primarily on the test score as a measure of the psychological characteristic of interest. The construct of interest for a particular test should be implanted in a conceptual framework. The conceptual framework specifies the meaning of the construct, distinguishes it from other constructs, and indicates how measures of the construct should relate to other variables. If an instrument has construct validity, examinees' scores will vary as the theory underlying the construct would predict.

The technique of factor analysis was used in examining the construct validity of the PII. Factor analysis is considered to be an extremely powerful and useful approach to behavioral data. It is a method of determining the underlying variables from various sets of measures. One application of
factor analysis, the principal factors method, involves the solution of simultaneous linear equations. The roots obtained from the solution are referred to as eigenvalues. Eigenvalues are related to the percent of variance which is accounted for through the factoring process. The sum of the eigenvalues is a measure of the total variance existing in the discriminating variables of a test.

The use of the eigenvalue technique was applied in the analysis of the construct validity of the Project Intercept Inventory.

Factor one (visual-motor skills) was found to have an eigenvalue of 5.25062. This accounted for 25.0% of the variance.

Factor two (visual and auditory memory skills) was found to have an eigenvalue of 1.55767 and accounted for 7.4% of the variance.

Factor three (perceptual awareness) had an eigenvalue of 1.37229 and accounted for 6.5% of the variance.

Factor four (gross motor skills) had an eigenvalue of 1.10812 and accounted for 5.3% of the variance.

Factor five (combination of gross and fine motor skills) had an eigenvalue of 1.05546 and accounted for 5.0% of the variance.

Since five factors of the PII were found with an eigenvalue greater than 1.0 the null hypothesis was rejected.

**Null Hypothesis 6.** There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the reading scores from the California Achievement Test taken four years later by the children in the study, based on the application of the Pearson product-moment correlation coefficient.
The hypothesis examined the criterion validity of the Project Intercept Inventory.

Criterion validity demonstrates that test scores are systematically related to one or more outcome criteria. The fundamental question is always how accurately can criterion performance be predicted from scores on the test.

Predictive validity involves using test scores to predict future behavior. In this process a predictive-validity coefficient is obtained, since criterion-related validity is typically expressed as a correlation coefficient. When a test is used to predict future behavior, predictive validity should be established. The Pearson product-moment correlation coefficient was used in this study to obtain the correlation coefficients of validity.

The use of the Pearson product-moment correlation coefficient with the data indicated a moderate correlation at 0.4726 between the total results from the Project Intercept Inventory and the total reading results from the California Achievement Test. Those results account for 22.3% of the variance between these two results. The results proved to be significant at a level of confidence of 0.05. The null hypothesis was rejected.

Null Hypothesis 7. There will be no significant correlation at the 0.05 level of confidence between the Project Intercept Inventory results and the arithmetic scores from the California Achievement Test taken four years later by the children in the sample, based on the application of the Pearson product-moment correlation coefficient.

The data used in the examination of this hypothesis were gained from the total arithmetic results from the California Achievement Test taken four years after the administration of the Project Intercept Inventory.
The use of the Pearson product-moment correlation coefficient with the data indicated a moderate correlation at 0.4440 between the results. This accounted for 19.7% of the variance between these two variables. These results proved to be significant at the 0.05 level of confidence.

The null hypothesis was rejected.

**Null Hypothesis 8.** The general shape of the item characteristic curves for the Project Intercept Inventory will be substantially different for boys and girls.

The hypothesis was formed to aid in examining the Project Intercept Inventory for the presence of bias in the test relating to gender.

Any index for evaluating bias should incorporate in it some control on the level of the ability of the groups under study. Examinees of the same ability level should have the same chance to perform equally well on a test regardless of what particular group they are a member, unless item bias is present in the test.

The use of the item characteristic curve approach is considered by some to be the most theoretically sound of the methods used in detecting item bias in tests. The item characteristic curve technique was used in this study to show the relationship between ability and the percentage of children who answered an item correctly. In general, as the former increases, so does the latter.

The patterns of the item characteristic curves indicate, in general, that there is very little difference between females and males in results gained from the administration of the items of the PII. The differences noted are mild at best. Where differences are seen, they tend to favor the females. Item five (folding a paper triangle) and item 12 (identifying geometric
shapes) were the two items in which the performance of the females was consistently mildly better than the males.

Based on these data, the null hypothesis cannot be accepted. The items of the PII do not appear to be substantially biased as regards gender.

To obtain another perspective on the possibility of male-female bias being present in the PII test items, the corrected point-biserial technique was also used to make comparisons between the genders.

This was done to gain an impression of how well each individual item of the PII would predict ability. The term ability as used here referred to the total test score. The term predict as used here referred to the correlation between each test item and the total test score. The corrected point-biserial technique was used to correct for the overlap which would occur between the item and the total test score.

The reason for the use of this technique was to gain additional information regarding the potential for gender bias being present in the various items of the PII. Lower corrected point-biserial numbers would suggest the potential of bias being present in the items. Higher corrected point-biserial numbers suggest better ability being present, and could serve as a predictor of later academic success.

The procedure used was to calculate two point-biserial correlations, one for each of the groups, per test item. Comparisons were then made between the two correlations, to see what differences might occur. An arbitrary cutoff point was set at 0.25 for purposes of making comparisons between the results gained from the use of the corrected point-biserial technique.

A review of the corrected point-biserial results for the male-female categories indicated that 3 of the 21 items suggest the potential for bias
being present, based upon the cutoff level of 0.25. These items discriminated
differently between females and males. The items were balance beam
walking, jumping and sequencing of objects.

The preponderance of the PII test items, 18, would seem to be free of the
potential for bias for females or males, based on the cutoff level of 0.25.

The null hypothesis was rejected.

**Null Hypothesis 9.** The general shapes of the item characteristic curves
for the Project Intercept Inventory will be substantially different for racial
majority and minority children.

The use of the item characteristic curves approach is considered by
some to be the most theoretically sound of the methods used in detecting
item bias in tests.

The item characteristic curve approach was used in this study to show
the relationship between ability and the percent of children who correctly
answered a given item. In general, as the former increases, so does the lat­
ter.

In undertaking the consideration of the presence of bias in the PII, it
was considered that test takers of the same ability should do equally well on
the test, regardless of the racial group to which they belonged. This would
refer to the test as a whole, as well as to individual items.

The term ability as used in this examination was intended to indicate
levels based on raw scores. The raw scores were derived from the number
of items to which each child was deemed to have correctly responded.

A review of the item characteristic curve graphs indicated that there is
little difference to be seen in the results between the two groups of children
on the various test items. There were mild improvements in results on a
consistent basis for minority children in item four (hopping) and in item 16 (recalling objects). There was mild improvement in a consistent fashion in the results of majority children on item 20 (verbal associations).

Based on these results, the null hypothesis must be rejected. The analyses of data do not suggest the substantial presence of bias in the items of the Project Intercept Inventory as related to majority and minority children.

To obtain another perspective on the possibility of bias being present in the PII test items, the corrected point biserial technique was used to make comparisons. This was done to gain an impression of how well each item of the PII would predict ability. The term ability used here referred to total test score. The term predict as used here referred to the correlation between each item and the total test score. The corrected point-biserial technique was used to correct for the overlap which would occur between the item and total test score. If low corrected point-biserial results were obtained, they could indicate the potential of bias being present in the items. Higher results from the use of the corrected point-biserial technique could be suggestive of higher ability among the children being screened, and could also serve as a predictor of later school success.

The procedure used was to calculate two corrected point-biserial correlations, one for each group of children, per test item. Comparisons were then made between the two groups, to see what differences might be seen. A cutoff level of 0.25 was set.

The results gained from the use of the corrected point-biserial technique indicated that only one item of the PII was found lower than the 0.25 cutoff level for the minority group children. The item involved was
number seven (drawing a circle). The majority group children had corrected point-biserial results below 0.25 for three items. These were items two (balance beam walking), three (jumping) and six (paper cutting).

These results strongly suggested that the various items of the PIL, with a very few exceptions, do not demonstrate bias when results are compared between majority and minority groups of children.

The null hypothesis was rejected.

Limitations of the Study

The ability to generalize may be considered a limiting factor in this study. The sample was taken entirely from prekindergarten children in the city of Grand Rapids, Michigan. While the sample taken was a substantial one, it was limited to the one city. At the same time, Grand Rapids is considered to be a typical midwestern city, so the possibility of generalizability is improved. The possibility of generalization to other cities of similar size and make-up would seem to be viable.

Some restrictions to the range of the sample are noted. Some of the children from the original group of entering kindergarteners may have been retained at some point between the kindergarten year and the fourth grade. Others may have entered a Special Education program or an alternative general education program before they reached the fourth grade.

It would also seem probable that a certain number of the original entering kindergarten group would have left the city school system and moved to another city or state. It is also possible that some children from this group were kept at home or in some prekindergarten program for a year, thus placing them a year behind their original classmates. Another
real possibility for losing some members of the original class from the public school system in Grand Rapids would be the entrance of some of the children in the nonpublic schools in the city. These multisystems enroll approximately two thirds as many children as the public school system in the city.

The number of test items which comprise the Project Intercept Inventory may also be considered in some quarters as a limitation to this study. The items making up this test number 21. This would not be considered a large number of test items, and that factor could have an effect on the reliability coefficients obtained, the tendency being to make the coefficients lower.

The PII is nonhomogeneous in its make-up, and this factor also would have a tendency to make reliability coefficients lower. The items attempt to screen a number of various readiness factors, and do not concentrate on any one factor.

Another potential limiting feature of this study was the large number of professional educators who are involved in the administration of the screening test. Due to the large number of children who must be screened in a limited time frame, 25 to 30 test administrators may be involved in the screening process at any given time during the period of time provided for this screening procedure. The very number of test administrators involved allows for potential inconsistency in the administration and scoring of the PII, even though each test administrator is carefully trained and periodically inserviced in the use of the screening instrument.

The fact that the vast majority of test administrators were Caucasian also may be seen as a limitation to this study. There are sizeable numbers of
Black, Hispanic, and Oriental children who are involved in the prekindergarten screening process. The make-up of test administrators does not reflect this population distribution.

Implications of the Study

This study undertook to examine the reliability, validity, predictability capabilities, and the presence of bias in a prekindergarten screening test.

While the immediate effect of the study is to examine a testing instrument whose use directly affects only kindergarten-entering children, considerably further-reaching implications can be seen. The relatively recent highly publicized report entitled A Nation At Risk (Gardner, 1983) points out the many concerns associated with the public schooling of America's children. An obvious implication is the need to be concerned about the readiness of children to undertake the educational process before they even enter that system. The trend in recent years in American education has been to demand higher and more stringent requirements of academic excellence for students and teachers. While the greatest pressure is currently applied at the secondary schools level, it has also been introduced in elementary schools.

This more demanding trend in education is now seen in many American school systems at the early elementary and kindergarten levels. School programs are becoming more insistent that their students possess essential skills and that the students can demonstrate that they are able to meet stated academic objectives. Kindergarten children are not excluded from this demand to be able to perform. While the types of skills entering kindergarten children are expected to possess may vary somewhat from...
school system to school system, many of the readiness expectations are very similar.

In such an environment it is imperative that children entering school for the first time are able and ready to meet the demands that will be made upon them to perform. Failure to do so can, and does, have many negative consequences. This set of circumstances speaks eloquently to the need for educators to be able to aid in making determinations regarding the readiness of children to successfully undertake the kindergarten program.

To be able to provide the best possible start in a school program is more important now than ever. It is probable that, for some children, the general education kindergarten program is not the most appropriate educational program. Because children have attained the chronological age of five years does not necessarily mean they can undertake and benefit from today's kindergarten curriculum. Some children will be better served by being enrolled in some alternative-to-kindergarten program, allowing an additional year of growth and experience. Others will be educationally handicapped to a point where they should be served in programs in Special Education. The method most popularly used in American school systems to aid in making determinations regarding readiness for kindergarten is some type of screening instrument, or test. Some school districts may use a battery, two or more, of such tests. Such a procedure, when done properly by trained personnel, can provide considerable information which can be used to make decisions regarding the most beneficial school program for children about to enter kindergarten. In short, the emphasis today in many American school districts is on early identification of children at risk as regards their readiness to enter school and to undergo a successful experience.
However, provision of an effective early identification program is a demanding procedure. The prekindergarten screening program employed by many school districts is highly dependent for effectiveness upon the screening tests put into use. It is incumbent upon educators that the best possible screening devices be used, dependent upon the goals which are to be achieved.

A major challenge in this regard is the difficulty in making accurate appraisals with children at this early chronological age. Besides the expertise and experience of the educators performing the screening, the make-up of the screening test is vitally important. The test must be able to be demonstrated to be reliable and valid. The test should be free of bias, especially as relating to gender and the ethnic backgrounds of the children.

While these factors are of utmost importance in the use of screening tests for kindergarten children, other items must also be considered. The financial cost to school districts is among these items. Many school districts face serious financial shortfalls, especially in periods of an uncertain economy. Despite this consideration, which is serious, the undoubted benefits which will accrue to the school district, as well as to children and parents, will far outweigh the initial expenses of the screening process. The eventual payoff will be in children appropriately placed in school programs and not needing additional, costly services and programs later in their school programs. Another obvious benefit will be in the lowering of the school dropout rate, with more students remaining in school to finish their high school programs and earn their diplomas.

Another item which must be considered when determining a screening test for the prekindergarten level is the amount of time which is
required per child to administer the test. Along with financial considerations, the amount of professional time which must be consumed in this process is an important facet. The number of children who must be screened is a major determinate. However, regardless of the size of the school district and the number of children to be screened, the time of professional educators involved in the screening process is no small consideration. This is especially true if the educators must be called away from their other professional responsibilities to be involved in the screening procedures.

The same rationale would be put forth as was offered for the financial concern. The time spent to do a thorough and effective screening procedure will be repaid many times over in the future benefits gained educationally for children and the school district.

The expertise of the screening personnel is another item which should be given serious consideration in this process. The educators involved in the administration and interpretation of the screening results must have the training and, whenever possible, the experience to do an effective job. Along with the technical skills, the ability to relate with children and to empathize with parents are major requisite points.

Final Comments

The study has provided a considerable amount of information which can be applied with good effect. The immediate effects could well be felt in the closer examination by school authorities in Grand Rapids, Michigan of the PII. Some changes in the test may occur. It is hoped that the information gained from this study will also be useful in other school districts.
where prekindergarten screening procedures are used. There has been considerable attention paid in recent years to kindergarten and early elementary schooling in this country. The emphasis in this important area of school adjustment and development of children is greatly needed. The foundation of the structure must be sound if the structure is to be able to firmly stand.

There continues to be a great need in American education to research the entire process of early childhood development. Particularly, there is great need to examine the methods that are used in making determinations of readiness for children to enter kindergarten. Since this trend seems to be widespread in today's educational environment, it especially points up the obligation to examine carefully what is done in looking at children educationally. It is also vitally important that we consider how these examinations are undertaken.

More research is needed to look carefully at the testing instruments that are used in examining young children. The period covering ages four through six years are very difficult in terms of obtaining accurate assessments of children's capabilities and lags. The area of screening of children at these age levels is even more difficult. What is undertaken in making such evaluations must be most carefully considered.

Another area which is in need of further research and development is that of curricula. It is essential that educators carefully examine what is being done in kindergarten classrooms. The demands made upon young children to achieve are often beyond the capabilities of many of the children who enter kindergarten. Thorough review of what is known regarding early childhood development is needed, as well as continued research.
into this entire area.

A facet of children's overall development as it pertains to progressing in the basic educational environment that warrants considerable research is that of the income of the child's family. A corollary to that aspect is that of the educational background of the parents. These areas of the child's background may be key factors in determining how well the child potentially can do in obtaining an education in today's school atmosphere.
### Gross Motor - large muscle coordination

**Adequate 3 - 4**

1. **Body Parts**
   - Criteria: 2 errors or less = 1 point

2. **Balance**
   - Criteria: may step off beam once = 1 point

3. **Jumping**
   - Criteria: both feet together for one jump = 1 pt.

4. **Hopping**
   - Criteria: hops easily, may put foot down once = 1 pt.

**Needs Help 0 - 2**

- Total Points

### Visual Motor - using eyes and hands together

**Average 1 - 6**

5. **Folding**
   - Criteria: folded within 1/2" = 1 point

6. **Cutting**
   - Criteria: paper cut into two pieces = 1 point

7. **Drawing Shapes**
   - a. **Circle**
   - b. **Square**
   - c. **Triangle**
   - d. **Open Square and Circle**

**Fair 3**

- Criteria: The child has made a predominantly circular line - it need not connect = 1 pt.
- Criteria: The child has made four clearly defined sides. The corners need not be angular.
- Criteria: The child has made three clearly defined sides, with one corner higher than the others.
- Criteria: The drawing shows no more than a slight separation of forms; major distortion of circle and square; both shapes of fairly equal size; bisection of circle passing through corner of square must project into square (see diagrams).

**Poor 0 - 2**

- Correct: 9
- Incorrect: 0

**Total Points**

---

*White - Kindergarten Teacher*
*Canary - Teacher Consultant or Resource Room Teacher*
*Pink - P. I. Director*
Concepts

Average 14 - 19 | Fair 6 - 13 | Poor 0 - 5

8. Colors
- Red
- Green
- Blue
- Yellow
- Black
- Purple
- Brown
- Orange

Criteria: each color correctly named = 1 point

9. Shapes
- Circle
- Triangle
- Diamond
- Square
- Rectangle

Criteria: each shape recognized = 1 point

10. Positions
- in front of
- behind
- next to
- under
- on

Criteria: each position understood = 1 point

11. Size (big-little)

Criteria: all correct = 1 point

Comments

Visual Discrimination - recognizing likenesses and differences:

- Adequate 1
- Needs Help 0

12. Matching Shapes

Criteria: all shapes correct = 1 point

Comments

Visual Memory - remembering what is seen

Average 1 - 6 | Fair 3 | Poor 0 - 2

13. Sequencing
- a) 1 object from memory
- b) 2 objects from memory
- c) 3 objects from memory

Criteria: each part correctly remembered = 1 point

14. Identifying Missing Objects
- a) 1 object
- b) 2 objects
- c) 3 objects

Criteria: each part correctly remembered = 1 point

Comments

Auditory - listening, understanding, and following directions

Average 3 - 12 | Fair 6 - 7 | Poor 0 - 5

15. Repeating Numbers
- 2 numbers
- 3 numbers
- 4 numbers
- 5 numbers
- 6 numbers

Criteria: each sequence correctly repeated = 1 point

Comments
16. a) Repeating Directions
criteria: each set of directions correctly repeated
2 directions _______ 4 directions _______ = 1 point
3 directions _______

b) Following Directions
criteria: each set of directions correctly followed
2 directions _______ 3 directions _______ = 1 point

Comments:

Total Points _______

17. Context Clues - association
criteria: each correct answer = 1 point
Average 4 - 5 Fair 3 Poor 0 - 2
Big - Little _______
Bow-wow - Meow _______
Hot - Cold _______
Yellow - Red _______
Fast - Slow _______
Comments:

Total Points _______

18. Verbal Fluency
criteria: each part present in drawing = 1 point
Adequate 4 - 6 Fair 2 - 3 Poor 1 - 2
Write down exactly what the child says:
a.
Speaks in relevant sentences = 3 points
Speaks in relevant phrases = 2 points
Relevant single word responses = 1 point
No response or irrelevant response = 0
Score a & b separately and combine scores for total points.
b.
Comments:

Total Points _______

19. Draw-a-Person
criteria: each part present in drawing = 1 point
Average - 7 or more Fair 5 - 6 Poor 0 - 4
head  body  nose  feet  eyebrows  clothing
legs  neck  mouth  fingers  arms  eyes
hair  ears  other
Comments:

Total Points _______

20. One-to-One Correspondence
Yes _______ No _______

🌟🌟🌟🌟
BIBLIOGRAPHY


Alston, H. L. & Doughtie, E. B. (1975). Correspondence of constructs measured by the kindergarten screening inventory by sex and ethnic group. Psychology In the Schools, 12, 428-429.


Buttrom, J., Covert, R., & Hayes, M. *Hayes Early Identification Listening Response Test*. Manuscript submitted for publication.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Hanson, I. (1980, March). First grade minimum skills testing of German and American children. Paper presented at the annual meeting of the State of Maryland International Reading Association Council, Baltimore, MD.


