A Comparison of the Matrix Analogies Test-Expanded Form and the Raven's Standard Progressive Matrices: European and United States Norms

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of the Ohio State University

by

Jane A. Welch, B.A.

* * * * *

1989

Masters Examination Committee: Approved by

Jack A. Naglieri Adviser
James L. Collins College of Education
To My Parents, Betty and Jim Welch
ACKNOWLEDGMENTS

I want to thank Dr. Jack A. Naglieri for his guidance, patience, and encouragement and Dr. James L. Collins for his assistance in the completion of this project as a member of the review committee.

I also wish to thank the students, parents and staff at the Ohio School for the Deaf not only for allowing me to conduct this research, but for their cooperation in scheduling and participating. A special thanks is extended to J. Michael Weber for his continued interest and support.
VITA

November 18, 1960 ........................................... Born-Springfield, Ohio

1983. ............................................................... B.A. Wilmington College
               Wilmington, Ohio

1984-1985 ............................................................ Youth Counselor, Marion
               County Juvenile
               Detention Center

1985-1987 ............................................................ Caseworker, Marion
               County Children's
               Services

1987-1988 ............................................................ Graduate Research
               Associate, Ohio State
               University. Research in
               reading disabilities,
               Buckeye Youth Center,
               Columbus, OH

1988-1989 ............................................................ Graduate Research
               Associate, Ohio State
               University. Huelsman-
               Peters Clinic

FIELDS OF STUDY

Major Field: School Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Figure 1, Sample Matrices Test Item</td>
<td>4</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>MAT-EF with Normals</td>
<td>7</td>
</tr>
<tr>
<td>MAT-EF and Special Populations</td>
<td>9</td>
</tr>
<tr>
<td>MAT-EF and Cross Cultural Populations</td>
<td>12</td>
</tr>
<tr>
<td>Raven's Progressive Matrices (RPM)</td>
<td>13</td>
</tr>
<tr>
<td>RPM and Normals</td>
<td>13</td>
</tr>
<tr>
<td>RPM and Special Populations</td>
<td>18</td>
</tr>
<tr>
<td>RPM and Cross Cultural Studies</td>
<td>22</td>
</tr>
<tr>
<td>Norms Information for RPM</td>
<td>26</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Difference between Raven's European and US Norms</td>
<td>37</td>
</tr>
<tr>
<td>2. Means and Standard Deviations</td>
<td>40</td>
</tr>
<tr>
<td>3. Pearson product moment correlations</td>
<td>41</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURES

1. Sample Matrices Test Item . . . . . . . . . . 4
CHAPTER I

INTRODUCTION

The assessment of intelligence of hearing impaired children poses special problems (Sattler, 1982, Moores, 1981, Levine, 1974). For example, their difficulty with oral communication makes typical questioning of the subjects problematic. The assessment of hearing impaired children requires consideration of the type and degree of hearing loss, the age of onset, and the etiological components, for example, in order to understand the extent to which these variables may have influenced development as well as the test scores.

Etiological components of hearing impairments are important in determining the likelihood of additional handicaps. Maternal rubella, Rh factor complications, and prematurity all increase the risk of additional handicaps. These additional handicaps not only impede the children's ability to learn and interact with the world, but also compound educational and assessment problems.

The adventitiously deaf, or children deafened after two years of age, have varying degrees of sound awareness and may understand the importance of sound in communication (Vess and Gregory, 1985). In contrast, congenitally deaf children are unaware of the existence of
sounds, do not recognize the contribution of sound to communication, and have not acquired adequate speech or language skills. Brill et al (1986) specify the following hearing threshold classifications: mild (26-54 decibels), moderate (55-69 decibels), severe (70-89 decibels), and profound (90 decibels and above).

According to Karchmer, Milone, and Wolk (1979), profoundly deaf youngsters comprise the largest grouping among hearing impaired children in special education classes and account for an increasing percentage of special class enrollment as they grow older. Moderate to Severely hearing impaired children generally enter special education at an older age than profoundly deaf students and are frequently mainstreamed. Mildly hearing impaired children generally have sufficient hearing to acquire language and speech.

According to Sattler (1988), developing competence in a standard natural language is of critical importance for hearing impaired children. Levine (1974) states that overreliance on oral communication is a major weakness in evaluating the deaf. Many instruments used to measure intelligence contain large verbal components which make it difficult to obtain an accurate picture of the intelligence level of hearing impaired students. Other areas of difficulty surrounding the testing of hearing impaired students include a lack of tests that are appropriate given their oral communication difficulties, the absence of norms for the hearing impaired, problems of interpretation and assessment, and the need for
specifically trained examiners (Levine, 1974).

To meet the special demands of assessing the mental ability of the hearing impaired, psychologists need instruments that are appropriate for the population, in that they can be administered to a wide range of subjects, normal, hearing impaired, physically handicapped, and so on. The Performance Scales of the Wechsler tests (WISC-R, WPPSI, and WAIS-R) and the Leiter International Performance Scale are the most frequently used measures of intelligence for the hearing impaired. The Hiskey-Nebraska Test of Learning Aptitude (Hiskey, 1966) was developed as an individually administered intelligence test designed for and standardized on deaf children. Additionally, nonverbal tests such as the Draw-A-Person, Bender Gestalt, and matrices tests are also used in the assessment of the hearing impaired.

Because the language deficits of hearing impaired children represent a major obstacle to the assessment process, greater emphases is often placed on nonverbal measures. For this reason, progressive matrices tests such as those developed by Raven (1956) and more recently by Naglieri (1985) which do not involve language, and have minimal oral directions, and allow for non-verbal responses (subjects point to the correct options) are appropriate for use with individuals with language and hearing limitations.

As a measure of nonverbal ability, matrices tests depend primarily on
an individual's ability to examine features and relationships among figural diagrams and to discover the abstract rules which govern the differences among the elements. For example, in the figure below:

![Test Example:]

1 2 3 4 5 6

FIGURE 1
Sample Matrices Test Item

Raven's Standard Progressive Matrices test is one measure of nonverbal ability that is frequently used as part of the test battery for hearing impaired students. Studies have been conducted which support that matrices tests are a good measure of "g". (Sattler 1988). According to Jensen (1980), this "g" represents the general factor of mental ability
that is measured in common by all of the intercorrelated mental ability tests as first hypothesized by Spearman (Jensen, 1980).

There are three Raven's Progressive Matrix tests which are frequently included in the assessment battery for hearing impaired children as measures of nonverbal intelligence. These are the Standard Progressive Matrices (Raven, 1938); the Coloured Progressive Matrices (Raven, 1947a); and the Advanced Progressive Matrices (Raven, 1947b). Of the three tests, the Standard Progressive Matrices (SPM) is the most widely used (Matthews, 1988).

Despite widespread use, however, the Raven's SPM has received some criticism. Anastasi (1982) noted that Raven's manual lacks vital information on reliability and validity. Sattler (1988) cites the lack of US norms as a problem. Although the SPM was recently re-standardized using American samples, the relationship between Raven's old and new norms has not yet been determined.

In order to meet the need for a well normed, well constructed nonverbal test of intelligence, the Matrix Analogies Test (MAT) was developed (Naglieri, 1985). Various criterion related validity studies have been conducted which compared the MAT with achievement, with the WISC-R on normal children, and with other nonverbal tests on deaf children (Naglieri, 1985).
No research has been done which compares the MAT with Raven's new US norms and no research has been done to compare the Raven's European and US norms. It is important to understand the similarities and differences between these tests in order to make competent decisions in the assessment of nonverbal ability. The purpose of this study is to meet the need for knowledge about the similarities and differences between Raven's SPM and MAT-EF.
CHAPTER II
LITERATURE REVIEW

The literature pertaining to the Matrix Analogies Test (MAT) and Raven's Progressive Matrices (RPM) will be discussed in this chapter. Research conducted on the MAT will be presented first followed by studies dealing with the RPM.

MAT with Normals

The Matrix Analogies Test-Expanded Form (MAT-EF) and Wechsler Intelligence Scale for Children-Revised (WISC-R) were compared on a sample of 82 normal children in grades 1 and 2, 5 and 6, and 10 and 11 (Naglieri, 1985b). Forty-three percent of the sample were male, and fifty-seven percent were female. The WISC-R and MAT-EF were administered in counterbalanced order. Forty-eight children were administered the WISC-R first and thirty-four were administered the MAT-EF first.

Results indicate that the MAT-EF correlated highest with the WISC-R FSIQ (.52), then PIQ (.41), then VIQ (.37). All correlations were significant ($p<.001$). In addition, the mean standard score on the MAT-EF was lower than the WISC-R VIQ, PIQ, and FSIQ. Such a difference can be
expected when an older test is compared to a more recently normed test (Kaufman, 1979).

The MAT was found to correlate significantly with Raven's Coloured Progressive Matrices (CPM) on a sample of 200 normal children (Naglieri, 1985b). The sample consisted of 100 first graders and 100 second graders with an equal number of males and females who attended a regular elementary school. Subjects were administered the MAT-EF and CPM in counterbalanced order in one sitting.

Two sets of analyses were performed in order to evaluate the relationship between these measures of nonverbal ability, one using raw scores and one using standard scores. The correlation of the total raw scores on the sample of normal children was .71 (p<.001). This result indicates that the MAT has good concurrent validity when compared to the RCPM. Additionally, Naglieri found that the MAT mean standard score (109.8) was lower than the CPM mean standard score (117.8) which suggests that test users risk interpreting results incorrectly when using a test normed in another country.

Results of a study of the relationships among the MAT-EF (Naglieri, 1985b), Draw A Person: A Quantitative Scoring System (Naglieri, 1988), and the Differential Ability Scales (Elliott, 1989) using a sample of 32
students in grades three through nine, revealed significant correlations ($r= .71$) between the MAT-EF and the DAS (Lillis, 1987). These results provide concurrent validity evidence for the MAT-EF in relation to measures of verbal and nonverbal ability as measured by the DAS.

Anderson (1987) compared the MAT-EF to the WISC-R (Wechsler, 1974) and PPVT-R (Dunn & Dunn, 1981) on a sample of 39 gifted students. The sample was comprised of 22 males and 17 females who attended third grade at an upper middle class suburban elementary school. Correlations among the variables indicated that the MAT-EF correlated significantly at the .05 level with the FSIQ ($r=.33$).

MAT and Special Populations

The MAT was compared to a number of nonverbal ability measures using a sample of 70 hearing impaired children who attended a school for the deaf (Naglieri, 1985b). This sample was comprised of 43 males and 27 females who ranged in age from 6 through 17 years. The MAT was administered by two advanced doctoral students in school psychology to a large percentage of the school population over a two-month period. Information regarding achievement and ability test scores was obtained from the students' files.

The nonverbal ability tests administered prior to the MAT testing consisted primarily of WISC-R Performance IQs (54 cases). Other tests
included the WAIS-R (5 cases); K-ABC (1 case), Leiter International Performance Scale (8 cases), and the Nonverbal Test of Cognitive Skills (2 cases). The MAT was found to correlate highly with these measures ($r=.72$). This study provides support for the MAT as a measure of nonverbal ability.

The relationship between the MAT-EF, WISC-R PIQ, and Stanford Achievement Test was conducted using a sample of 40 hearing impaired students (Bardos and Weber, 1988). Results indicate that the MAT-EF correlates significantly ($p<.01$) with the WISC-R PIQ ($r=.77$) and with achievement ($r=.57$).

A study conducted by Naglieri and Wisniewski (1988) examined divergent validity among the WISC-R (Wechsler, 1974), MAT-EF (Naglieri, 1985a), and Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981). Subjects were fifty-one students, 34 boys and 17 girls, in grades one through twelve who were referred due to academic difficulties. The tests were administered in counterbalanced order to each student.

The significance of the divergent validity was tested to determine to what extent practitioners should expect such results. In addition, the differences required for significance between these tests were also calculated. The differences between standard scores required for
significance for the WISC-R compared to the MAT-EF and PPVT-R were calculated following Anastasi's formula (1982).

The differences required for significance at the .05 and .01 levels, respectively, when comparing the following tests were as follows: WISC-R PIQ with MAT-EF (12.2 and 16.0); and MAT-EF with PPVT-R (16.0 and 21.0). Divergent validity of the MAT-EF with the WISC-R PIQ and VIQ was not significant as the MAT-EF correlated significantly with both the Verbal and Performance Scales of the WISC-R.

A significant relationship (p<.01) was found between MAT-Short Form (MAT-SF) and the Kaufman Test of Educational Achievement (K-TEA) (Prewett, Bardos, and Naglieri, 1988). This study used a sample of 77 fourth and fifth graders and found MAT/K-TEA correlations of .5 and .43 for normal and reading disabled students respectively.

The results illustrate that the MAT-SF and K-TEA standard scores are significantly related to achievement in reading in normal and reading disabled samples. This supports the validity and use of the MAT-SF as a screening test.

The MAT-SF also correlated significantly (p<.01) with the K-TEA on a sample of 85 normal and developmentally handicapped (DH) fourth and fifth grade students (Prewett, Bardos, and Naglieri, 1988). The sample was comprised of 46 normals (r=.50), and 39 DH students (r=.56). These results support the use of the MAT as a screening test for the
developmentally handicapped.

**MAT and Cross Cultural Studies**

Results of a cross cultural study of the MAT-SF using a sample of 200 Greek elementary students in grades one through six indicate similar Greek and U.S. mean standard scores with significant \( r \) scores at only the third and fourth grade levels (Bardos and Naglieri, 1986). The lower means for the Greek sample were small and likely due to the lack of educational programming for exceptional children in Greece.

A significant correlation was found between MAT and WISC-R PIQ \( (r=.43) \), and RCPM \( (r=.64) \) for a sample of 114 Native American children (Naglieri, 1985b). The sample consisted of 57 males and 57 females in grades three through eight. It was found that the MAT mean standard score (68.1) was lower than both the mean WISC-R PIQ (97.7) and RCPM mean standard score (96.3). Also, there was no significant difference between the correlation between MAT/WISC-R PIQ \( (r=.43) \) and RCPM/WISC-R PIQ \( (r=.50) \) which indicates that MAT and RCPM do not differ significantly from each other in their relationship to the WISC-R.

In a study that compared the performance of matched samples of 407 Canadian and U.S. children aged five \( (n=95) \), six \( (n=22) \), nine \( (n=71) \), thirteen \( (n=110) \), and seventeen \( (n=109) \) years on the MAT-SF it was
found that the mean standard scores for the Canadian sample were similar to the American norms at each age (Naglieri and Bardos, 1988). The mean raw score differences between these samples ranged from less than one to almost three raw scores.

Raven's Progressive Matrices (RPM)

RPM and Normal Populations

Rock and Nolen (1982) compared the traditional and computerized versions of the Raven's Coloured Progressive Matrices Test (CPM) using a sample of fifteen children (4 girls and 11 boys) referred for educational assessments to the University of Washington Psychoeducational Center. Subjects ranged in age from 7 to 14 years (mean = 10.33, SD = 2.44). Their WISC-R Full Scale IQ scores ranged from 65 to 126 (mean = 105.21, SD = 15.27). All subjects were individually tested.

Results indicate that the mean scores on the computer-administered CPM (21.3, SD = 7.3) compared to mean scores for the normative sample (24.9, SD = 5.8) were not statistically significant (t = 1.62, df = 38, p > .05). The correlation between computerized CPM score totals and WISC-R FSIQ scores was .59 (p < .05) which was smaller in magnitude than the coefficient obtained between the standard Raven and the Terman-Merrill scale, Form L (r = .66) reported by Raven, et al. (1977) for a 9 year old
British sample.

These results indicate that the traditional CPM and the computerized version are not substantially different in terms of total scores and normal total-score composition. In addition, the correlations between scores obtained from the experimental version and the WISC-R were similar in sign and magnitude to the correlations reported by Raven (1977). These results should be interpreted with caution due to the small sample size.

Burke and Bingham (1969) examined the validity of the Raven's Progressive Matrices (RPM) for a sample of 91 males at the East Orange Veterans Administration Hospital. Subjects ranged in age from 19 to 59 years with a mean age of 35.1 (SD 8.7). The RPM, the Wechsler Adult Intelligence Scale (WAIS), and parts of the Army General Classification Test (AGCT) were administered. In addition to scoring the WAIS by the standard method, the test was also scored for Cohen's factors which include Verbal Comprehension, Memory, and Perceptual Organization. The AGCT subtests used in this study were Vocabulary, Arithmetic, and Block Counting.

The sets of test scores were intercorrelated and subjected to a principal components analysis and a varimax rotation. Results indicate that the RPM correlated slightly higher with the WAIS Performance Scale (r = .76) than it did with either the WAIS Verbal Scale (r = .70) or the
AGCT ($r = .71$). This provides evidence for concurrent validity of the RPM. In order to examine the construct validity of the RPM, intercorrelations among the RPM, Cohen's WAIS factors, the AGCT part scores, and age were submitted to a principal components analysis. All variables except age loaded heavily on a single factor of general intellectual functioning.

Factor analytic studies such as those conducted by Burke and Bingham (1969) offer evidence that the Raven's Progressive Matrices (RPM) is a good measure of the general factor (g) of intellectual functioning. As subsequent studies indicate, the RPM may not be as pure a test of g as generally thought in that different subsections may tap different mental capabilities (Carlson, 1973).

Carlson (1973) investigated the relationship between performance on the Raven's Coloured Progressive Matrices (CPM) and performance on a subtest of the Inhelder-Piaget Standard Matrix Test. The sample was comprised of 33 third grade children who ranged in age from 8 years, 1 month to 9 years, 10 months (mean age = 8.3 years, SD = .45 years).

Each subject was first administered the Inhelder-Piaget Matrix Test. The subject was considered to have solved the problems operationally if they gave a correct solution, justified it, and refused to change on countersuggestion. The subject was considered to have solved the problem by perceptual similarity or graphic collection if they gave
correct initial solutions but either could not justify them or changed their minds in the face of countersuggestion. All subjects were administered the RPM, Forms A, Ab, and B, immediately after the administration of the Inhelder-Piaget Test.

On the Inhelder-Piaget Test, there were 11 subjects in the operational group and 22 in the graphic collection group. Performance on the RPM and the graphic and operational groups were compared for total and subsection scores.

Statistically significant differences were found for set B ($t = 3.04, p < .01$) and for the total score ($t = 2.64, p < .05$). No significant differences were found for sets A or Ab. These results indicate that the total score difference is primarily attributable to difference in performance on set B. The results provide support for the notion that different parts of the CPM tap different processes.

A study conducted in 1974 by Carlson and Goldman builds on the previous study in an attempt to clarify the operational content, as defined by Piaget, in the Raven's Coloured Progressive Matrices (CPM). The sample consisted of 103 fourth graders whose mean age was 8 years, 6 months. The children were selected from a school district in a working-class community in Riverside, California. The subjects' mean
IQ, based on the Lorge-Thorndike Group Intelligence Test, was 93.0 (SD = 17.5).

Subjects were individually administered the CPM followed one week later by the Inhelder-Piaget Standard Matrix Test (SMT). Operational responses only were considered correct on the SMT. Total scores were calculated for the SMT. The CPM total score was obtained and the CPM was then separated into two categories: conceptual, which included items which required abstract reasoning; and perceptual, which included items that could be solved by pattern completion.

Pearson product-moment correlations were calculated between the SMT and CPM Total, $r = .58$ ($p < .01$), the SMT and CPM Conceptual, $r = .54$ ($p < .01$), and the SMT and CPM Perceptual, $r = .61$ ($p < .01$). Multiple regression equations were calculated to compare the predictive importance of the two sets of CPM items. Although differences were small, the conceptual items were found to be more useful predictors of SMT performance. These results provide evidence that there is a significant association between performance on the SMT and performance on the CPM.

A study by Weidl and Carlson (1976) further examines the factorial structure of the Raven Coloured Progressive Matrices Test (CPM) using a sample of 180 first, second, and third graders in Trier, West Germany.
The average age was 7 years, 5 months. Sixty children from each grade level were chosen and there were approximately equal numbers of males and females.

The CPM was individually administered to each subject and the data were factor analyzed from phi-coefficients by the principal component method. Three significant factors emerged: 1) concrete and abstract reasoning, 2) continuous and discrete pattern completion, and 3) pattern completion through closure. These results support the heterogeneity of the CPM although a replication of the study using larger samples is recommended in order to clarify the nature of the constructs portrayed and lead to a better understanding of the factorial validity of the test.

RPM and Special Populations

Corman and Budoff (1974) examined the factor structure of Raven's Coloured Progressive Matrices (RCPM) through two studies of normal and educably mentally retarded children. The sample for the first study was comprised of 243 normal children from low income areas of an urban community, ranging in age from 7-12 years. Approximately sixty percent of the sample were male and 40 percent were black. The retarded sample for the first study included 399 EMR students who attended special
classes in New England public schools. These students ranged in age from 7-15 years and the mean Stanford Binet IQ that was available for 242 individuals in the sample was 71.0 ($SD = 7.8$). Sixty percent of the sample were male, twenty five percent were black, and approximately 75 percent were from working class families.

The normal subjects selected for the second study were 379 students ranging in age from 6 to 11 years who attended schools in upstate New York. Fifty-five percent of the sample was male and fifty-three percent were black. Twenty-eight percent of the subjects' fathers were professionals while the majority were blue collar workers. The EMR sample for the second study consisted of 174 EMR students who attended special classes in schools in upstate New York. They ranged in age from 5 to 14 years. Forty-three percent were black, and 95% were of low socio-economic status. Their mean IQ on the WISC or Stanford-Binet was 68.4 ($SD = 9.0$).

Raven's Progressive Matrices were group administered to all subjects in both studies and the test items were factor analyzed. Similar factor structures were obtained for two independent samples of normal subjects. Four factors emerged which accounted for 42 percent of the variance. In addition, four factors were obtained for the retarded subjects. These results provide evidence of the stability of the factorial composition of the test for retarded as well as nonretarded subjects.
This finding supports the factorial invariance of RPM with children of different intellectual levels.

James (1984) examined the correlation coefficients between the RPM and the WISC-R Performance Scale for a sample of 64 hearing impaired children. Approximately forty percent of the children lived at the Atlantic Provinces Resource Center for the Hearing Impaired in Amherst, Nova Scotia, while the remaining sixty percent were nonresidents who attended the school.

The sample was comprised of children who were referred for psychological assessment and who had been administered the Raven's Matrices and WISC-R as part of a test battery. The subjects ranged in age from 6 to 16 1/2 years. Hearing losses for the group ranged from 40 to 110 dB. Thirty-four subjects between the age of 6 and 11 years were given the Coloured Progressive Matrices (CPM) Book Form. Fifty subjects over the age of 11 were administered the Book Form the the Standard Progressive Matrices (SPM).

Pearson product moment correlations were calculated for each group. Correlation coefficients between the CPM and WISC-R for the younger group was .87. For the older group, the correlation coefficients between the SPM and WISC-R was .78. Results of this study indicate a high correlation between RPM and WISC-R which suggests that the RPM can
provide a reasonable indicator of a subject's level of intellectual functioning. Caution should be exercised in applying these findings to the hearing handicapped in general, due to the relatively small sample.

Carlson and Dillon (1978) examined the effects of testing the limits procedures in measuring the intellectual capabilities of a group of hearing impaired children. The sample was comprised of 70 children enrolled in the California School for the Deaf who ranged in age from 6 years, 1 month to 10 years, 11 months. The mean age was 9.1 years, standard deviation, 1.54 years. Mean IQ scores of 98.3 (SD = 11.92) were obtained from the Leiter International Performance Scale, which has an expected mean value of 95 and SD of 20.

The subjects were randomly assigned to one of five testing conditions in which either the booklet or puzzle form of the Raven's Coloured Progressive Matrices was administered. Scoring was the same in each condition, the first response was scored as either correct or incorrect. The testing conditions were as follows: 1) Standard Condition, 2) Simple Feedback, 3) Elaborated Feedback, 4) Elaborated Feedback Plus Reason Given by the Child, and 5) Performance Dialogue.

The results of the 2 (test version) x 5 (testing condition) ANOVA indicates that significant main effect differences due to testing condition were found. A post hoc analysis of mean differences was then conducted to localize the differences in means. This analysis, which
utilized the Scheffe method indicated that conditions 1 and 2 did not differ from each other, but differed significantly from conditions 3, 4, and 5 ($p < .01$). No difference between conditions 3, 4, and 5 was indicated. As the results indicate, neither the standard procedure nor the simple feedback procedure lead to higher levels of performance on the CPM. Increased levels of performance resulted, however, when elaborated feedback (condition 3), elaborated feedback plus reasons given by child (condition 4), or a combination of these (condition 5) was used.

These approaches support the efficacy of certain testing the limits procedures for the assessment of intellectual abilities of young hearing impaired children. The effect of these approaches in improving performance is suggested to imply greater accuracy in measurement of estimation of intellectual competence (Carlson and Dillon, 1978).

The RPM and Cross Cultural Studies

Corman and Budoff (1974) examined the factor structure of the Coloured Progressive Matrices (CPM) by comparing the responses of a sample of Spanish-speaking students to a sample of English-speaking students. The Spanish-speaking sample was comprised of 228 students who ranged in age from 6 to 17 years (mean age =11 years, 2 months).
The subjects were approximately evenly divided by sex and were selected from three low-income urban areas in Massachusetts.

The English-speaking sample consisted of 243 subjects ranging in age from 7 to 12 years. Approximately sixty percent were male and twenty-five percent were black. These subjects were also selected from low-income areas of urban Massachusetts.

All subjects were administered Raven's CPM in class size groups. The item scores were submitted to a principal components analysis with 1.00 in the diagonal, followed by a varimax rotation of the four factors obtained. This procedure was carried out separately with item scores of Spanish-speaking and English-speaking subjects.

Factor structures of Spanish-and English-speaking subjects on the CPM were found to be generally comparable. The four factors accounted for 46.6 percent of the variance for Spanish-speaking subjects and 42.5 percent for English speaking subjects. An examination of items with high loadings revealed that item loadings on Factors I (Continuity and Reconstruction of Simple and Complex Structures), II (Discrete Pattern Completions), and III (Reasoning by Analogy) were comparable for both groups of subjects. The results provide evidence of the factorial invariance of the CPM with children from different cultural backgrounds.

Powers and Barkan (1986) examined the adequacy of Raven's
Progressive Matrices as a concurrent measure of academic achievement for a sample of Hispanic and Caucasian students. The sample was comprised of 99 seventh grade Hispanic students (53 boys and 46 girls) and 93 seventh grade Caucasian students (49 boys and 44 girls) who attended a large urban school district in the Southwest.

Each student was administered Raven's Standard Progressive Matrices (SPM) and the Reading, Language, and Mathematics subtests of the California Achievement Test (CAT). The mean SPM standard score differences for the Hispanic students (mean = 39.38, SD = 7.16) was not statistically significant (t = .52) from the Caucasian students (mean = 40.00, SD = 9.16). Correlations between the CAT Reading, Language, and mathematics scores and the SPM, examined separately for the Hispanics and Caucasians, were all significant at the p<.001 level. Validity coefficients were also compared for reading, language, and mathematics. No significant differences appeared between the reading validity coefficient (ρ = .95), the language validity coefficient (ρ = .53), and the mathematics validity coefficient (ρ = .56) of the two groups.

The results indicate that the SPM has concurrent validity as a measure of academic achievement for Hispanic and Caucasian seventh grade students. Additionally, the study provides evidence of the comparability of the validity coefficients. The similarity of the coefficients of the Hispanic and Caucasian students suggests that the
value of the correlations does not depend on ethnic classification. These findings support the continued use of the SPM in public schools for students from different ethnic and cultural backgrounds.

Sidles and MacAvoy (1987) conducted a correlational study between Raven's Standard Progressive Matrices (SPM), a Primary Language Questionnaire, and the Comprehensive Test of Basic Skills (CTBS) on a sample of Navajo adolescents. The sample was comprised of one hundred and twenty-four seventh and eighth grade bilingual Navajo students ranging in age from 14 to 16 years. Males and females were equally represented and all attended regular classes. Bilingualism was determined by the results of the Primary Language Questionnaire which was completed prior to testing.

All SPM testing was conducted in small group format one month prior to the administration of the CTBS. Raw scores obtained for the SPM and CTBS were transformed into derived scores. Correlations between SPM and CTBS range from the upper .30's to the upper .40's and are all significant at the $p < .01$ level which supports the use of the SPM in assessing the academic potential of Navajo adolescents. Correlations computed between SPM and CTBS scores for males and females were not significant which suggests relatively equivalent ability levels between the sexes. In addition, there was no significant difference between the
correlations computed for the SPM and Primary English Speaking adolescents and the SPM and Primary Navajo speaking adolescents.

This finding supports the applicability of the SPM with Navajo Indian adolescents. Results suggest that the SPM is a viable means of estimating the intellectual potential of adolescent Navajo students.

**Norms Information for RPM**

Jensen, Saccuzzo, and Larson (1968) rescaled the raw scores on the Standard (SPM) and Advanced (APM) forms of the Raven Progressive Matrices (RPM) in order to create a scale that accommodates a wider range of talent than do the raw scores of either form. This study describes the process of equating the two forms of the RPM.

The sample was comprised of 261 undergraduates who were enrolled in a required course in psychology in a large state university (San Diego State University). The sample was selected to represent the full range of scholastic aptitude comprised by both academically highly selective universities and colleges for which a high school diploma is the only entrance requirement. All subjects were administered both the APM and SPM and the Otis-Lennon Mental Ability Test, Advanced Level, Form J. The three tests were administered on separate occasions in the following order: Otis-Lennon, APM, SPM.

The two forms of the RPM were equated by means of equi-percentile
equating. Raw scores on both forms of the RPM were converted to percentile ranks. Percentile ranks were converted to normalized $z$ scores (mean = 0, SD = 1). Hence raw scores from either form that had the same percentile rank had the same $z$ score. Finally, $z$ scores were converted to an IQ standard score (mean = 100, SD = 16), as in the Otis-Lennon Standardization.

Karnes, Lee, and May (1982) conducted a study to determine the correlation between the scores on the 1966, 1973, and 1979 norms of the Raven's Standard Progressive Matrices (RPM) from a sample of economically disadvantaged elementary students. The sample was composed of 140 economically disadvantaged students in Grades 3 through 5 who attended an elementary school in the Southeastern region of the United States. The subjects included 77 males and 63 females.

The RPM was administered to class size groups by grade. Raw scores were calculated and converted to percentiles for the 1966, 1973, and 1979 norms. A nonparametric procedure, subprogram NONPAR CORR of SPSS (Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975) was used to determine the Spearman Rho correlation for percentile data on each set of norms. The following means, standard deviations, and percentiles were calculated for each set of norms: 1966 Norms: mean = 36.59, SD = 26.60, percentile range = 3 to 94; 1973 Norms: mean = 42.60, SD =
28.29, percentile range = 2 to 95; 1979 Norms: mean = 22.90, SD = 20.41, percentile range = 3 to 82.

There was a significant positive correlation (.88 p < .01) between the 1966 and 1973 norms and a significant positive correlation between the 1966 and 1979 norms (.85 p < .001) and the 1973 and 1979 norms (.85 p < .001). These data suggest that the 1966, 1973 or 1979 norms can be used with economically disadvantaged students.

Vincent and Cox (1974) conducted a study to assess the appropriateness of using Raven's Progressive Matrices (RPM) 1948 British norms on Vocational Rehabilitation and general adult populations in the United States. Using a sample of 380, all of whom had been administered the untimed RPM, records were drawn from the psychological files of the Houston Diagnostic Unit of the Texas Rehabilitation Commission. This sample compared favorably with the general population of the Houston-Harris County area and the United States on variables of sex, education, age, and race, with the exception that blacks were overrepresented.

RPM percentile ranks were converted to deviation IQ scores (mean = 100, SD = 15). Because blacks were overrepresented in the sample and there was a significant 14 point difference between black (mean = 91.98) and white (mean = 106.01) performance, the 1948 British norms should be used with caution when making inferences to the general U.S.
population, as they may tend to be an overestimate of IQ. Results indicate that the RPM 1948 British norms are appropriate for the Vocational Rehabilitation sample (sample mean = 100.63, SD = 16.79).

A study by Paul (1965-66) compares normative data for the Raven's Advanced Progressive Matrices (APM) based on a sample of American university students to the normative group of students included in the APM standardization. Paul's sample was comprised of 300 students who attended the University of California, Berkeley. The average age was 21 years with a standard deviation of 32 months. There were 190 females, and 110 males in the sample. The APM was individually administered to each subject. One hundred fifty of the subjects were also individually administered the Terman Concept Mastery Test (CMT), and a different set of 62 subjects were also individually administered the Wechsler Adult Intelligence Scale (WAIS).

Results indicate that the Berkeley sample scored much higher overall than the normative sample of Raven's 1962 edition of the APM. The mean total scores for the sample of 300 students was 27.0 with a standard deviation of 5.14. The mean score of the normative group of 170 university students presented by Raven (1965) was 21 (SD = 4). In addition, the 95th percentile corresponds to a total score between 34 and 35 for the Berkeley sample. The 95th percentile value based on Raven's normative group with similar ages is between 23 and 24. In
addition, there was a moderate correlation ($r = .44$) between the APM and CMT and a correlation of .84 between the APM and WAIS. These results indicate that the APM, CMT, and WAIS similarly rank.
CHAPTER III

METHODS

Sample

The sample for this study comprised of thirty-four predominantly white hearing impaired children enrolled in a state residential school for the deaf. The thirty-four subjects were selected from forty-four returned consent forms. Ten students were not tested due to scheduling difficulties. Twenty-six males (76%) and eight females (24%) participated in the study. Mean age of the students was 164.8 months or 13 years 7 months with a standard deviation of 26.1 months. There were 31 white subjects (91%) and 3 nonwhite subjects (9%) in the sample. Testing was conducted between the hours of 5:30 and 8:30 PM.

Procedures

An advanced graduate student in school psychology individually administered the Raven's Standard Progressive Matrices (RSPM) and Matrices Analogies Test-Expanded Form (MAT-EF) to each student. All administrations were conducted in one session of approximately twenty to thirty minutes. The instruments were administered in counterbalanced order to control for practice effects or fatigue, that
may influence the scores. Seventeen students were administered the RSPM followed by MAT-EF and seventeen were administered MAT-EF followed by RSPM. Percentile rank scores for the RSPM were interpolated because the SPM norms tables used in the analysis report percentile ranks for only a limited number of points over the distribution. Percentiles were then converted to standard scores using a standard conversion table.

Instruments

**Matrix Analogies Test-Expanded Form**

The Matrix Analogies Test-Expanded Form (Naglieri, 1985) is a nonverbal test of intellectual ability which involves minimal verbal comprehension and minimal motor movement. It is intended to provide an individually administered measure of nonverbal ability and has no verbal content, minimal verbal instructions and little or no verbal response by the student. The test is, therefore especially useful for an individual with limited knowledge of English, a handicap such as deafness, and motor problems such as cerebral palsy, a communication disorder, or someone who is unwilling to interact verbally.

The MAT-EF utilizes abstract designs of the standard progressive matrices type, printed one per page. The test is composed of 64 items
presented in a multiple choice format. It is divided into four item
groups: pattern completion, reasoning by analogy, serial reasoning, and
spatial visualization. The MAT-EF yields a Total Test standard score
with a mean of 100 and standard deviation of 15.

The test was standardized on a total sample of 5,718 American
students between the ages of 5 and 17 years and was stratified
according to age, sex, race, geographic region, community size, and
socioeconomic status. Naglieri (1985) reports that standardization was
conducted in two phases.

Subjects in both phases were selected to be representative of the U.S.
population according to the 1980 U.S. Census data on the basis of
geographic region, sex, age, race, geographic region, community size and
socioeconomic status. In Phase 1, a sample of 4,468 students were
administered 34 of the 64 items in class sized groups by their regular
teachers. Phase 2 consisted of the individual administration of all 64
items to a sample of 1250 students. The large sample obtained by
combining Phase 1 and Phase 2 provided more stable and precise
estimates of distributions of scores at each age interval.

Naglieri (1985) reports internal consistency reliability coefficients
that range from .88 (age 5) to .95 (age 6) for the Total Test, with the
median coefficient being .93 (Cronbach alpha). This reliability
coefficient, which utilizes a single administration of a single form and
is based on the consistency of all responses to all items of the test (Anastasi 1986), is high (Bracken, 1987). Similarly, test-retest reliability, determined by repeating the identical test on a second occasion, was also high (Bracken, 1987). An MAT-EF test-retest median Total Test reliability coefficient of .77, which was obtained by administering the test two times to a sample of 65 fifth grade students with a four week interval between administrations, is also reported (Naglieri, 1985).

**Raven's Standard Progressive Matrices Test**

The Raven’s Standard Progressive Matrices was designed to measure nonverbal mental ability through the assessment of an individual’s abstract reasoning. The test has been used as a measure of nonverbal ability and appears to be appropriate for individuals with handicapping conditions such as hearing impairments, expressive/receptive language disabilities, or cerebral palsy. The matrices are constructed so that a logical relationship exists between the horizontal and vertical dimensions of the item. The individual’s task is to uncover these relationships to determine which option best fits the missing location in the matrix. The RSPM (Raven, 1947b) contains five sets of twelve matrices for a total of 60 items, presented one item per page.
Raven's European Norms

The RSPM has been standardized for representative samples of British people, aged 6-65 years (Raven, 1948) and for Irish children aged 6-12 years (Gill and Burt, 1973). Information on the Irish standardization is presented in the manual and is based on a total sample of 3,464 Irish students between the ages of 6 and 12 years. Data on father's occupation, number and sex of siblings, and location of home were also obtained. To account for the limited age range of the Irish norms, the manual contains extrapolated norms covering ages 3 1/2 to 100 years.

Reliability of the RSPM appears to be high. Burke (1973) reports internal consistency reliability coefficients ranging from .89 to .97 depending on age of over 500 adults in the United States. Similarly, Raven (1948), reports test-retest reliability coefficients ranging from .83 to .93.

Raven's US Norms

Raven's test was recently administered in a group of local norms development projects using American samples. The result of this process is presented in a research supplement called A Compendium of
North American Normative and Validity Studies (Raven, 1986). This compendium is a collection of approximately 10 local norms tables and a table described as providing norms for the United States (Raven, 1986, p.15). The manual provides little information usually given about a normative sample. This makes determination of the degree to which these norms are consistent with the population characteristics of the US difficult. In addition, the local norms were collected using sampling procedures which "vary widely as a result of local constraints and the support available." (Raven, 1986 p.8), making interpretation tentative.

**Comparison of Raven's European and U.S. Norms**

In order to understand how scores on Raven's tests may be similar or different, the differences between raw scores that correspond to the same percentile points were examined. An analysis of scores from Raven's European Norms, Table SPM XIV (Raven, 1949, p. 31) and Raven's U.S. Norms, Table RS3SPM6 (Raven, 1986, p.15) is presented in the following table:
### TABLE 1
Differences between Raven's European and US Norms

<table>
<thead>
<tr>
<th>PERCENTILE RANK:</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>-2</td>
<td>-4</td>
<td>-3</td>
<td>0</td>
</tr>
<tr>
<td>8.5</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>-4</td>
<td>-5</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>0</td>
<td>-5</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>9.5</td>
<td>1</td>
<td>-1</td>
<td>-3</td>
<td>-5</td>
<td>-4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>-1</td>
<td>-4</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>10.5</td>
<td>0</td>
<td>-1</td>
<td>-3</td>
<td>-5</td>
<td>-1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11.5</td>
<td>1</td>
<td>-2</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>-2</td>
<td>-3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12.5</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: *Difference = European - US Raven raw scores*
Data Analysis

Since the Raven's SPM norms table used in the analysis reports percentile ranks for only a limited number of points over the distribution, interpolation was required. Percentiles were then converted to standard scores using a standard conversion table. The standard scores were analyzed by computer using the Macintosh Statview 512+ software package. Standard score, means, and standard deviations were calculated in addition to a Pearson Product Moment intercorrelation matrix, and a repeated measures ANOVA.

Research Questions

The following questions were asked in this study:

1. Do MAT-EF standard scores correlate significantly with Raven's SPM standard scores obtained using the European norms? This question was answered using a Pearson Product-Moment Correlation (p < .05).

2. Do MAT-EF standard scores correlate significantly with Raven's SPM standard scores obtained using U.S. norms? This question was answered using a Pearson Product-Moment correlation (p < .05).
3. Do the standard scores obtained using the Raven's European and U.S. norms correlate significantly? This question was answered using a Pearson Product-Moment correlation.

4. Is there a significant difference between the mean standard scores on the MAT-EF, the Raven's SPM European Norms and the Raven's U.S. Norms? This question was answered using a repeated measures ANOVA.
CHAPTER IV
RESULTS

Mean scores for each test as presented in Table 2 indicate that the subjects generally earned scores that fell within the Low Average range of intellectual functioning. On the MAT-EF, mean scores ranged from 55 to 116. Mean scores for the Raven’s European Norms ranged from 71 to 121 and for Raven’s U.S. Norms mean scores ranged from 61 to 121.

TABLE 2
Means and Standard Deviations for MAT-EF, Raven’s European Norms, and Raven’s U.S. Norms

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAT-EF</td>
<td>84.5</td>
<td>16.4</td>
</tr>
<tr>
<td>Ravens European Norms</td>
<td>89.7</td>
<td>14.0</td>
</tr>
<tr>
<td>Ravens US Norms</td>
<td>84.7</td>
<td>15.8</td>
</tr>
</tbody>
</table>

40
In order to examine the relationships between the three tests, Pearson Product-Moment correlations were calculated and are presented in Table 3.

### TABLE 3

Pearson product-moment correlations for MAT-EF, Raven's European Norms, and Raven's U.S. Norms

<table>
<thead>
<tr>
<th></th>
<th>2)</th>
<th>3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) MAT-EF</td>
<td>.79**</td>
<td>.79**</td>
</tr>
<tr>
<td>2) Ravens European Norms</td>
<td></td>
<td>.98**</td>
</tr>
<tr>
<td>3) Ravens U.S. Norms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * = p < .05, ** = p < .01.

These results indicate a significant positive correlation between the MAT-EF and Raven's European Norms (r = .79). There was also a significant positive relationship between the MAT-EF and Raven's U.S. Norms (r = .79). Similarly, there was a significant positive correlation between the Raven's European and Raven's U.S. Norms (r = .98). All
correlations are significant at the .01 level and indicate high positive relationships between the MAT-EF, Raven's European Norms, and Raven's U.S. Norms.

In order to determine if there is a significant difference between the mean standard scores on the MAT-EF, Raven's European norms and Raven's U.S. Norms, a one-way repeated measures analysis of variance was applied. A Scheffe F-test yielded an F value of 6.03 which indicates a significant difference (p < .001) between the MAT-EF and Raven's European Norms. There was also a significant difference between Raven's European Norms and Raven's U.S. Norms (F=5.62, p < .001). There was no significant difference between the mean scores of the MAT-EF and Raven's European Norms.

While these results suggest that the MAT-EF and the Raven's U.S. Norms may be used interchangeably with a hearing impaired population, it is important to consider other factors before choosing either test. These factors will be discussed in detail in the next chapter.
CHAPTER V
DISCUSSION

The findings of this study will be discussed in this section. The relationships between the MAT-EF, Raven's SPM European Norms, and SPM U.S. Norms will be discussed with an emphasis on the limitations and suggestions for future research.

**Intercorrelations and Mean Scores**

The significant correlation between the MAT-EF and the SPM European Norms ($r = .74$) indicates a high degree of relationship between these two instruments with this sample. A similar correlation ($r = .79$) with the MAT-EF and the SPM U.S. Norms also indicates a high degree of relationship between these instruments for this sample. A significant correlation ($r = .92$) was also found between the SPM European Norms and the SPM U.S. Norms. The high degree of relationship between these tests is to be expected since they represent essentially the same test.

When mean scores for each instrument were compared using a repeated measures ANOVA, the mean MAT-EF standard score was found to be significantly lower than the mean standard score of the Raven's SPM European Norms. Similarly, the Raven's SPM U.S. Norms mean standard
score was found to be significantly lower than the mean standard score of the Raven's SPM European norms. This is an expected result when a newer test is compared to an older one (Kaufman, 1979).

Although the results of this study might be interpreted to suggest that the MAT-EF and the SPM U.S. Norms may be used interchangeably with a hearing impaired population, further analysis of Raven's normative data suggest they should not. An examination of the Raven's norms is warranted in order to point out the limitations this study as well as to provide suggestions for future research.

As evidenced in Table 1 (see page 35), there is a considerable amount of inconsistency between Raven's European and US norms. The pattern of scores changes as a function of age and IQ score. For example, at age 13, the scores are lower for the Raven's US norms than for Raven's European norms, a pattern that would be expected when a newer test is compared to an older one (Kaufman, 1979). At age 8, however, the pattern is almost completely opposite, as scores for Raven's European Norms are consistently lower than the more recent US norms.

It becomes apparent that the age of subjects (mean = 13 years, 7 months SD = 26.1 mo.) in the present sample had an influence on the results obtained in this study. The age of the sample in this study corresponded to the age at which scores on Raven's US Norms were lower
than scores obtained using Raven's European Norms. As an analysis of Raven's normative data has shown, such a correspondence is not always the case. It is therefore recommended that this study be repeated on samples of different age levels and with different IQ levels.

It is important to consider various factors when selecting a test. Anastasi (1962) suggests that the appropriateness of a test should be based on such information as the title of the test, author, date of publication, time required to administer, and cost. In addition, there should be a description of the purpose and nature of the test which includes the population for which the test was designed, the nature of content and the identification of subtests, item types, and types of scores yielded by the test. There should also be a practical evaluation based on qualitative features such as clarity of directions and ease of administration and a technical evaluation based on information pertaining to norms, reliability and validity.

When evaluating the MAT-EF and the SPM U.S. Norms based on Anastasi's criteria, it becomes apparent that the MAT-EF conforms much more closely to Anastasi's standards. The MAT-EF manual is clearly written and organized in a manner that provides easy access to both practical (i.e. directions for administration) and technical (reliability, validity, and standardization) information. In addition, all the tables and information needed for scoring are conveniently located and identified in
Limitations

The small sample size was a limitation in the present study since it did not allow for a by-age comparison of the MAT-EF and Raven's US Norms. A by-age comparison would provide greater insight into the inconsistencies between these two sets of norms.

In addition, the lack of IQ level comparisons also represents a limitation of the current study. Such a comparison would enable practitioners to view a distribution of scores and determine the consistency between IQ level and the normative data of the MAT-EF and Raven's.

Future Research

The results and limitations of the present study provide directions for future research involving the comparisons of the MAT-EF and Raven's European and US Norms. One suggestion is to extend the present study to look at comparisons between these tests and other IQ tests (i.e. WISC-R) and achievement tests. It would also be interesting to examine the comparison between these tests with other groups (i.e. Hispanics, blacks, bilingual).
REFERENCES


