5-1983

An Analysis of the Hiskey Nebraska Test of Learning Aptitute for Navajo Children with Suspected Learning Difficulties

Susan Louise Sawyer
Utah State University

AN ANALYSIS OF THE HISKEY NEBRASKA TEST OF LEARNING APTITUDE FOR NAVAJO CHILDREN WITH SUSPECTED LEARNING DIFFICULTIES

by

Susan Louise Sawyer

A thesis submitted in partial fulfillment

of the requirements for the degree

of

MASTER OF SCIENCE

in

Psychology

Approved:

Major Professor	Committee Member
Committee Member	Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

1983

ACKNOWLEDGEMENTS

I would like to thank my mother, Mary Bryan Sawyer, who instilled the idea at an early age that I could do anything I wanted to. This thesis proves her right.

Susan Louise Sawyer

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

ABSTRACT

An Analysis of the Hiskey Nebraska Test of Learning Aptitute for Navajo Children with Suspected Learning Difficulties

by

Susan Louise Sawyer, Master of Science

Utah State University, 1983

Major Professor: Dr. Marvin Fifield
Department: Psychology

The purpose of the study was to evaluate the appropriateness of the Hiskey Nebraska Test of Learning Aptitude as a measure of ability for Navajo students with suspected learning difficulties. Item analytic procedures were employed for the 61 students in the younger group (ages 5-10) and for the 57 students in the older group (ages 11-17).

Using Hoyt's analysis of variance routine, reliability estimates of .87 (younger group) and .81 (older group) were obtained. The majority of the items discriminated in the appropriate direction for both subtest and total test scores.

It was concluded that the Hiskey Nebraska Test of Learning Aptitude would give acceptably reliable and consistent results for Navajo children with suspected learning difficulties. Further research to demonstrate validity of the instrument for this population was discussed.

(80 pages)

CHAPTER I

INTRODUCTION

Individual psychoeducational evaluation is an essential part of the identification and placement process of handicapped students. However, the use of psychoeducational testing information, a significant component of student evaluation, to place minority children in special education has become a controversial issue (Reschly & Phye, 1979). Inadequacies of conventional psychometric techniques for assessing minority children has been well documented (Jensen, 1980; Sattler, 1978).

One setting in which the problems in the assessment and placement of minority children is particularly acute is in schools serving large numbers of Native American children. Havinghurst (1981) estimated that 218,500 Native American students, 80% of all Native American students, are attending Bureau of Indian Affairs (BIA) schools or other schools located in very rural, remote areas. The rural, remote nature of the schools had led to difficulty in recruiting and retaining qualified professionals to serve the handicapped (Fifield, 1981). This difficulty in retaining professionals has meant that most BIA schools contract for assessment services with psychological consulting individuals or firms located in universities or cites surrounding the reservations (Fifield, 1981). A disadvantage with this arrangement is that the contracted psychologists are employed on a short-term basis and subsequently, have only minimal commitment. Some of the problems these consulting psychologists face are lack of appropriate norms for

available tests and scarcity of culture-fair instruments. In addition, testers often have limited familiarity with the Indian culture and the constraints placed on the service system delivery in these schools. Often, there is not adequate space and/or appropriate facilities in the schools for individual evaluations. Also, the psychologists are often confronted with the over- or underscheduling of students for assessment. Thus, the task of performing pscyhological assessment of BIA students by contracted psychologists is a difficult one at best. Nonetheless, BIA agencies are under the jurisdiction of Public Law 94-142. Therefore, in order for their students to be placed in special education classes or to receive supportive services, pscyhological testing must take pace.

In August, 1980, a contract was awarded to Utah State University, Exceptional Child Center to provide 170 psychoeducational assessments of Native American children referred for special education services in the Fort Defiance agency of the Navajo area office of the Bureau of Indian Affairs. In addition to providing the contracted testing, it was the purpose of this contract to study the procedures for individual psychoeducational tests in the Fort Defiance agency and provide recommendations to improve the accuracy and relevance of such testing.

In preparing for this contact, Exceptional Child Center staff reviewed Public Law 94-142, the guidelines established by the BIA, and procedures for the selection of unbiased testing instruments. Since virtually every child in the BIA schools is either bilingual or non-English speaking, psychologists providing individual testing on the

Navajo Reservation generally select the Performance section of the Revised Weschler Intelligence Scale for Children (WISC-R) which does not require verbal response from the child (McShane & Plas, 1982; Teeter, Moore, & Peterson, 1982). However, using the Performance section of the WISC-R only addresses part of the problem, since the instructions must be given in English. To eliminate such bias, other non-language dependent tests were reviewed.

The Hiskey-Nebraska Test of Learning Aptitude (HNTLA) is made up of items very similar to the performance section of the WISC-R. However, since it was designed to be used with deaf students, the instructions are presented in pantomime, either by demonstration or by hand signals. Taken at face value, the HNTLA appears to be an excellent choice for the assessment of intelligence of Navajo students with suspected learning difficulties. The literature also suggests that the HNTLA is an appropriate instrument for the ability assessment of Navajos (Shutt & Hannon, 1974). Furthermore, the HNTLA has been used with bilinguals and other populations that appear to be penalized by verbal intelligence tests (Burors, 1978). Based on the preceeding data, the HNTLA was selected by Exceptional Child Center project staff as one of the ability measurement instruments for the Indian students referred for assessment in the Fort Defiance contracts.

The Problem

Although the literature suggests that the HNTLA is an appropriate instrument for ability assessment of Navajos (Shutt & Hannon, 1974), clinical analysis of the HNTLA protocols collected for the Navajo

children referred for assessment in the Fort Defiance contract for the Exceptional Child Center showed marked inconsistencies.

For children with suspected learning difficulties (the primary reason for referral in this sample) one would expect to find most students with an average to low average intelligence score and a two or more year deficiency in academic skills (Bureau of Indian Affairs, Office of Indian Education Guidelines, adopted January, 1980.) However, in examining the test protocols of the Fort Defiance referees, it was found that 22.9% had achievement scores that were higher or equal to their HNTLA learning age score.

The HNTLA consists of 12 subscales. Subtests 1-8 are administered to children ages 3-10 (younger group). Subtests 6-12 are administered to children ages 11-17 (older group). The data indicated that it was the "younger group" which had the majority (63%) of the inconsistent achievement results noted above. Furthermore, the 118 Navajo pupils who were given the HNTLA, 66.1% had a variation in the subtests of five years or more.

A variety of reasons can be examined as possibly contributing to this inconsistency:

1. Examiner effects, (i.e., bias due to cultural and language differences) may have affected the test results (Brodie, 1970; Palomares & Johnson, 1966). However, examiner bias should have been equally distributed with other groups and other tests administered to the same student. This was not the case. Other measures of ability did not show such inconsistencies.

2. Poor tests administration may have also contributed to these findings. Error in administration could have also occurred in other tests given at the same time, however, this was not found.

3. The achievement measures may have been inadequate for the assessment of achievement in this group. However, achievement test scores were rather consistent with teacher ratings and other measures of ability. In addition, several different measures of achievements were utilized.

In summary, it was concluded that the HNTLA, as utilized in this situation, did not appear to be an appropriate measure of ability.

The question is: Why didn't the HNTLA predict more accurately the achievement level of Navajo students referred for suspected learning difficulites? The HNTLA may be inappropriate for several different reasons: The test items may not discriminate, they may be overly easy, overly difficult, or they may be measuring other things than what the test purports to measure.

The literature indicates that the HNTLA has not been item analyzed for Navajo students. Thus, the discriminative ability of the test items has not been determined and it is quite possible that certain test items or subtests are just as culturally loaded or even more culturally loaded than many verbal tests.

Justification

Information relating to the nature of individual item statistics obtained when the HNTLA has been applied to culturally different individuals would seem to be of value for a number of reasons.

1. Psychologists, in addition to considering the total performance, often make clinical inferences based upon particular item responses (Anastasi, 1976). The discriminative ability of an item or subtest would provide additional data to improve the accuracy of such clinical inferences.

2. Psychologists are interested in a scale which differentiates between normal and subnormal groups, but also reveals individual differences within the special group. To make such inferences, it is necessary to determine the discriminating power of individual test items, since the total test performance is a function of the aggregate performance on the individual items comprising the scale.

3. The order of difficulty of items when the scale is applied to special groups is also important. The subtests are administered such that when a specific number of items have been failed, the subtest is discontinued. The assumption, based on the standardization group's performance, is that all further items would also be failed. If the order of difficulty of items for the special group is different than for the standardization group, it is possible that after the specified number of items have been failed, and the subtest is discontinued, that later items, not given, could be passed. Thus, the total test scores may be affected (lowered) by the order of presentation of items.

Purpose of the Study

The purpose of the present study was to evaluate the appropriateness of the HNTLA as a measure of ability for Navajo students with suspected learning disabilities as determined by an item analytic

study, to measure the item difficulty, discriminative power, and internal consistency of the items of this test for this population.

This study collected data to answer the following specific questions:

1. Using item analytic techniques, what is the item difficulty, discriminative power, and internal consistency of the items of the HNTLA for Navajo students ages 6 to 12 referred for suspected learning problems.

2. Is there a difference in difficulty of the subtests or total test of the HNTLA for different sexes?

3. Is there a difference in item difficulty, discriminative power, or internal consistency for the items in the subtests given to the younger group (ages 5-10) versus the older group (ages 11-17)?

CHAPTER II

REVIEW OF LITERATURE

There is a vast amount of professional literature, studies, and opinion that address the use of test information in the placement of children in special education services. Therefore, the literature review for this study will be broad and selective and divided into two major areas of concern: 1) nonbiased assessment, 2) item analytic studies of intelligence tests.

Nonbiased Assessment

> "Even though many school systems have tried to minimize the importance of standardized tests, these tests still play an important role in who succeeds in our society." (Ford, Jr., 1980)

Testing and test results are an inherent part of education in this country. Public Law 94-142, Education for all Children's Act of 1975 (Federal Register, 1977b) mandated that certain procedures be followed in the process of identifying and educating handicapped children. One requirement of the law is that all children referred for special education placement have, a statement of the present levels of educational performance. Individually administered psychological and educational tests are the most frequently used method for identifying a student's current level of educational performance.

Intelligence tests are deeply entrenched in this process of evaluation and placement as a result of their role in predicting scholastic achievement. The ability of intelligence tests to predict

academic achievement has been the source of much controversy. However, the evidence for IQ tests' predictive validity is substantial and generally consistent (Lunemann, 1974).

Empirical support for this use of intelligence tests has done very little to abate the persistent concern over the insensitivity of intelligence tests in differentiating the learning potential of minority and lower socio-economic children. Questions of bias in assessment and placement of children with special needs has become a major focus of concern by those interested in equality in the schools (Bailey & Harbin, 1980).

It has been evident for many years that minority group children generally do not score as well on achievement and intelligence tests as do Anglo children (Ford, Jr., 1980). The controversy is over the inference that the poor performance of lower socioeconomic status and minority children on intelligence tests is an index of learning ability. Minority groups contend that they (minority groups) are proven inferior by tests that are based on false criteria. The ethnic concern is that the existing criteria in psychological testing does not encompass a multi-cultural society, but only serves to maintain the status quo within a framework of White supremacy (Drew, 1973). Responses to their accusations are frequently defensive in nature and nearly always provide data pointing out that it is not the tests themselves that are bias but the society, or environment and that tests are merely the vehicle that conveys this information. The dissonance fosters emotion to a level of confusion so great that the issues often seem unresolvable.

Although the concerns about bias in tests are not new and did not arise only with school psychology, school psychology and intelligence tests have been among the most targeted areas of expression of concern.

Perhaps the most dramatic expression of concern has occurred in the form of litigation over the past 10 years. The litigation bearing upon bias in tests is concerned with special education placement, principally the over-representation of minority group children in classes for the mildly retarded (Chandler & Spakos, 1969; Mendoza-Friedman, 1973; Mercer, 1971). Dunn (1968) stated that over 50% of those enrolled in classes for the retarded in this country are ethnic minority children. Mercer (1971) found, in Riverside, California, that three times more Chicanos and two-and-one-half times more Blacks than would be expected from their percentage in the population were placed in mentally retarded classrooms. A number of cases have been before the courts on the placement issue (Diana v. State of California, 1970; Larry P. v. Riles, 1972; Guadelupe v. Tempe Elementary District, 1972). The net effect of these cases has been to raise serious questions about traditional assessment practices, to institute a variety of protections for parents and children, and to establish guidelines for the assessment of minority children (Tractenberg & Jacoby, 1977; Turnbull, 1978).

A number of unresolved issues are apparent from the examination of litigation over special education placement. These issues, for example, labeling effects, bias in tests, and effectiveness of special education programs are all sources of intense debate within the scientific and professional communities. None of these issues can be resolved "beyond a shadow of a doubt" with the currently available

empirical data (Reschly & Phye, 1979). For example, although over-representation of minorities has been regarded as inherently suspicious by the courts and represents a major emphasis of the Federal Office for Civil Rights, other facets of the data have been noted. Reschly and Phye, (1979) point out that there is a tendency for the over-representation data to be misinterpreted and/or exaggerated in the attacks on intelligence tests with minority populations. For example, in the Larry P. case, Black students constituted 28.5% of the total district enrollment, but nearly 66% of the enrollment in special education classes for the mildly retarded. This data has sometimes been understood to mean that two-thirds of all Black students were diagnosed as mentally retarded through the use of intelligence tests. Reschly and Phye (1979) note that, in fact, a smaller percentage, 2%, of Black students were actually diagnosed as mildly retarded.

Despite the contradictions and misinterpretations in data available in the area of nonbiased assessment, court decisions were made and new legislation ensued. Two recent federal laws provide nearly identical guidelines for the assessment and placement of handicapped children. Section 504 of the Rehabilitation Act of 1973 (Federal Register, 1977a), and Public Law 94-142, The Education for All Children's Act of 1975 (Federal Register, 1977b), reflect the clear influence of the special education placement litigation. Most pertinent to the area of nonbiased assessment is the following requirement from Public Law 94-142 (1975):

> "Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory (p. 42496)."

This requirement is one of the most important features of the legislation (Reschly & Phye, 1979). However, it has potential weakness in that no clear definition of the meaning of bias in assessment appeared in the rules and regulations. There are differing interpretations and contradictory evidence on bias in tests and assessment. Therefore, the practical effects of the Public Law 94-142 requirement are unpredictable (Reschly & Phye, 1979).

Perhaps in response to the potential confusion, the Bureau of Education for the Handicapped (BEH) awarded a contract to the Coordination Office of Regional Resource Centers (CORRC) to explore the meaning of bias in assessment. However, efforts of CORRC have not resulted in clearly stated definitions of bias or specific guidelines for eliminating bias in assessment. Both the CORRC reports and Federal rules and regulations have apparently concluded that although bias cannot be defined unequivocally, evaluation procedures are likely to be less biased if procedural safeguards are followed and a broad variety of information is gathered and considered (Mowder, 1980).

Bias in tests: Differing conceptions and empirical results. The literature indicates there have been may efforts to define bias in tests and assessment (Jensen, 1980), but consensus on theoretical or research criteria and agreement on practical implications has not been achieved. Analyses of bias in specific tests have ranged from speculative judgments about specific items to sophisticated statistical examinations of test results and prediction systems. The conclusions of these efforts are largely contradictory. Throughout the discussion of test bias, conclusions are confounded by confusion about the meaning of IQ tests (Jensen, 1980).

However, the inadequacies of conventional psychometric techniques for assessing the intellectual functioning of minority group children has been well documented (Anastasi, 1976; Jensen, 1980; Sattler, 1978). Much of the criticism toward the use of standardized tests in the evaluation of minority populations is a question about the validity of the instruments and the techniques available. Standardized tests have been criticized as: (1) being highly loaded with White, middle-class values, and experiences (Williams, 1970); (2) standardized tests penalize children with linguistic styles different from that of the dominant culture (Bailey & Harbin, 1980); and (3) standardized intelligence tests sample cognitive styles that are directly opposed to those found in low income families or culturally diverse groups (Cohen, 1969).

In addition to these noted criticisms, it is the practice of test publishers to report validity coefficients only for the total population of the standardization sample and not for identifiable subgroups of these populations. For example, the WPPSI and the 1973 restandardization of the Stanford-Binet (Thorndike, 1973) included in the norming samples children from various occupational, geographic, and non-White groups, according to their distribution in the population. However, the correlations between the WPPSI and its criteria, or validity coefficients, were reported only for the standardization population as a whole, and not for each previousy identified subgroup. Similarly, the Stanford-Binet results were not presented separately for the various groups included in the population.

There is evidence that reliabilty coefficients, as well as validity coefficients are affected by subcultural differences (Vance & Gaynor, 1976). The test manuals for the widely used tests seldom give separate reliability data for specific minority groups. As with the validity coefficients, test users often assume that the reliability coefficients provided are reliable enough to use with the minority group as well.

In addition to test bias, a frequent criticism is that the atmosphere of the testing situation is unfair to minority children. Examiner effects have been recognized as damaging for some time (Brodie, 1970; Palomares & Johnson, 1966). Taylor and White (1981) found that test scores tend to be more accurate if the examiner is trained in proper testing procedures. Brodie (1970) found that the performance of minority children was affected by such examiner variables as bilingualism, ethnic group membership, and style of administration.

Assessment strategies for minority children. Past measurement trends have been proven to be insufficient for the assessing of intellectual ability of minority children. Subsequently, current efforts are being made to reduce this bias in the assessment of minority children. Over the years, test developers have proposed various solutions to the problem of discriminatory evaluation (Mowder, 1980), such as culture-fair tests, translating assessment instruments, adaptive behavior scales, and pluralistic assessment techniques.

The problem encountered in attempting to implement decisions that mandate fairer testing procedures for minority children is that the guidelines presented are stated in general terms with few specific

suggestions for practitioners (Reschly & Phye, 1979). Confounding the problem is the fact that a firm theoretical and empirical basis for the current evaluation options has yet to be established (Bailey & Harbin, 1980). There is a scarcity of tests that are comparable across languages and cultures.

The most common strategy employed by test developers to eliminate bias has been an attempt to minimize the cutural and verbal components of testing (Bailey & Harbin, 1980). Generally, nonverbal or performance tests have been regarded as less culturally loaded. The primary thrust is a search for instrumentation capable of assessment across cultural subgroups that would not reflect cultural difference (Drew, 1973). Examples of such attempts include the Davis-Eells Games (Davis & Eells, 1953), the Cultural-Fair Test (Cattell, 1950), and Raven's Progressive Matrices (Raven, 1962). However, not all minority children perform better on these nonverbal tests (Reschly & Phye, 1979).

In summary, there is no one concept of nondiscriminatory evaluation nor is there one strategy for assessment of minorities that is agreed upon as being the most effective. Nonbiased assessment is a complicated issue, and even the best efforts have proven to be inadequate (Mowder, 1980). Despite the confusion centering around nonbiased assessment, local and state agencies must continue to assess minority children. They are required to demonstrate compliance with the nonbiased assessment provision of the law.

Item Analysis of Intelligence Test

Allegations of cultural bias in the items used on conventional tests have and continue to be the most popular of the criticisms of standardized tests (Reschly & Phye, 1979). Examples of subjective judgments of item bias are numerous (Dent, 1976; Williams, 1971). However, relatively little evidence or critical examinations of the allegations of item bias has appeared in the literature. Subjective judgments of item bias are not necessarily consistent with empirical data. Subjective judgments are not always accurate, and revisions of current tests, either in the direction of greater or less cultural loading, might have the effects of maintaining group differences and reducing validity (Reschly & Phye, 1979).

Item difficulty has also been assessed by subjective judgments. In a study by Sandoval and Millie, (1980) two methods for determining item difficulty were examined; empirical item analytic evidence and rational subjective judgments. In this study, a) judges were not able to detect items that were more difficult for a minority child than for an Anglo-American child, and b) the ethnic background of the judge made no difference in accuracy of item selection for minority children. The value of item analysis, according to Lehman and Mehrens (1976), is that it helps to judge the quality of a test and is of aide in test revision and adaptation. Shortening a test (randomly, or through subjective judgment) lowers validity and reliability (Anastasi, 1976). However, when a test is shortened or adapted by eliminating the least satisfactory items, the sort test may be more no valid and reliable than the original, longer version (Lehman & Mehrens, 1976). Anastasi,

(1976) notes that the validity and reliability of any test is dependent ultimately on the characteristics of its item.

Since the literature review contains no example of an item analysis on the HNTLA, excluding the standardization data, the Weschler Intelligence Scale for Children will be used to illustrate how item analytic data contributes in the evaluation of appropriateness of the instrument for the measuring of intelligence for different populations.

An Illustration of Item Analysis on the Weschler Intelligence Scale for Children - Item Analysis Illustrated

The Weschler Intelligence Scale for Children-Revised (WISC-R, [Weschler, 1974]) has received increasing acceptance for evaluating the general intelligence of children, as had the WISC prior to its revision. Item analysis of the WISC scores obtained by 194 boys and 172 girls between the ages of 7-6 and 15-11 were reported by Carleton and Stacey (1955). This work is probably the most quoted study dealing with the internal consistency of the WISC as related to subnormal groups. The mean full scale IQ for these subjects was 67.3 and the mean chronological age was 12.3 years. Biserial coefficients between dichotomized item responses and the total score were estimated. Results of this study indicated that the rank order correlation computed for each of the various subtests ranged from .85 to 1.0. This indicates that the obtained order of difficulty tends to approximate rather closely the orders specified by the manual. Carleton and Stacey (1955) contend that relatively few items are misplaced in respect to

order of presentation, and that the misplacements that do occur are not sufficient enough as to affect the total score. Relatively few items approach the 50-50 ideal split which would contribute most to total item variance. The average index of difficulty for all test items except Coding was .37.

Results of Carleton and Stacey's study indicated that the subtests of the WISC would not be particulary sensitive to revealing individual differences chosen from a comparable population (Vance & Gaynor, 1976). From their population, 41 out of 183 items were completely nonfunctional, as no subject in the group was successful in responding to them, and a large number of items contribute relatively little to the total variance. For the Carleton and Stacey population, only 40 items were within the .30 to .70 difficulty range, a standard many test developers to be the most desirable (Ebel, 1979).

Ninety disadvantaged children's WISC protocols from different cultural backgrounds whose full scale (IQ's ranged from 80 to 95 were collected and analyzed by Vance and Gaynor (1976). Total scale analysis evidence from the study indicated that the majority of the WISC items satisfied the major requirements of internal consistency and reliability when used with disadvantaged children. The results indicated that while item difficulties compared rather well with those from the standardization sample, there did appear a need for a rearrangement of many subtest items, when examining children from culturally different groups (Vance & Gaynor, 1976).

Although the HNTLA has been suggested as appropriate for the measuring of intelligence for Navajo pupils, (Shutt & Hannon, 1974) an

item analysis has not been carried out with Navajo students to test the assumption that the items appropriately discriminate, or that items are at an appropriate level of difficulty for Navajo pupils. This study was designed to address this research gap by providing an item analyses on the HNTLA on the performance of 118 Navajo students which were referred by their classroom teacher for psychological assessment.

CHAPTER III

METHODOLOGY

The purpose of the present study was to conduct an item analysis of the Hiskey Nebraska Test of Learning Aptitude (HNTLA) in order to evaluate the appropriateness of the test for a sample of Native American children with suspected learning difficulties. The item analysis was computed to evaluate the item difficulty, discriminative power, and internal consistency of the items of HNTLA for the sample population. The follow chapter outlines the procedures utilized in this study.

Sample

A contract was awarded to Utah State University Affiliated Exceptional Child Center in August, 1980, to provide 170 individual psychoeducational assessments of Native American children referred for special education services in two boarding schools (Chuska, Tohatchi) in the Fort Defiance area of the Navajo Reservation. In compliance with the contract, 118 HNTLA were administered. The sample used in this study consists of those 118 students who were administered the HNTLA as part of their psychoeducational assessments. All were Navajo students, either bilingual or non-English speaking. Descriptive information concerning the sample is presented in Table 1. Seventy-seven male and 41 female students between the ages of 5 and 17 comprise the sample; the mean age was 11-1. Educational level for the students ranged from the first grade to the eighth, with a mean grade level of

4.6, excluding the six students who were identified solely as special education students.

Table 1

Sample Characteristics

AGE	N	GRADE	N
6.0-6.11	1	1	2
7.0-7.11	9	2	11
8.0-8.11	12	3	24
9.0-9.11	15	4	21
10.0-10.11	24	5	18
11.0-11.11	20	6	15
12.0-12.11	15	7	11
13.0-13.11	8	8	10
14.0-14.11	7	Special	
15.0-15.11	6	Educ.	6
16-0-16.11	0		118
17.0-17.11	1		
	118		
SCHOOL	N	SEX	N
Chuska	79	Female	41
Tohatchi	39	Male	77
	118		118

Measure

The HNTLA is a revision of the Nebraska Test of Learning Aptitude for Young Deaf Children originally developed for the deaf. Prior to the publication of the Nebraska Test of Learning Aptitude for Young Deaf Children in 1941, there was no available individual test of ability which had been specifically designed for young deaf children, and standardized upon such children in this country. Research revealed

the inappropriateness of scales designed for and standardized upon a hearing population (Hiskey, 1966). It was to meet this need that the Nebraska test was published.

Norms were established for hearing children in 1954-55 and, thus the test came to be regarded as useful with hearing children. Due to its nonverbal nature, it has been used with subjects which have traditionally been penalized by verbal intelligence tests. This includes such groups as bilingual and speech impaired clients.

The HNTLA scale consists of a series of performance tasks that are organized in a presumed ascending order of difficulty within 12 subscales. The subtests 1-8 are administered to children ages 3-10 (younger group). Subtests 6-12 are administered to children ages 11-17 (older group). The total test Learning Age score is calculated by determining the median level of the examinee's performance on the set of subscales administered (Hiskey, 1966).

Two sets of norms for translating raw subscale scores to Learning Ages are provided. When the pantomimed directions are used, the deaf norms are generally appropriate, while the use of verbal directions dictates the employment of norm conversions based on hearing subjects. Burors, (1978) notes that standardization samples used are sufficiently large and broadly representative of the respective populations to provide reasonable confidence in the arrived scores. However, no breakdowns by race or sex are provided.

Split-half reliability estimates for the young (ages 3-10) and the older ages (11-17) subsamples of deaf and hearing subjects all exceed .90. The author presents four kinds of validity data: subtest

intercorrelations, correlations between age ratings on the subtests and the median Learning Age, correlations with the 1960 Stanford-Binet (S-B) and the WISC for hearing children, and correlations with performance on achievement tests and teacher ratings for the deaf. For the hearing, the intercorrelations range from .32 to .78 (median .55) for ages 3-10 and .25 to .46 (median .34) for ages 11-17. The author reports a correlation of .86 (IQ's) with the S-B for 99 students ranging in age from three through ten and a correlation of .78 for ages 11-17. For the WISC, the correlation between IQ's is .82 on 52 subjects (age range 5-11) for three groups.

In an effort to assess the HNTLA as an appropriate instrument for the placement of bilingual children in appropriate special education, Shutt and Hannon (1974) correlated the HNTLA with the performance of the WISC-R for a Navajo and Mexican-American sample. The correlation between the Learning Age score of the HNTLA and the mental age score of the WISC-R was .73. From this, they concluded that the HNTLA is a valid instrument for the assessment of the intelligence of Navajo and Mexican-American pupils.

Procedures

The procedures utilized in the present study are outlined below. To explore the appropriateness of the various tests used in the assessment battery administered as part of the Fort Defiance assessment project, a table of scores for all tests administered was prepared. The table listed the descriptive information (age, grade, sex, school) for each student tested and the subtest and total test scores for all

tests administered. A portion of this table is presented in Appendix A.

By examining Table I, the inconsistency discussed in Chapter I can be observed. The inconsistent test results found in Table 1 suggested a need for further examination of the tests used. An informal analysis was performed on the table scores to examine, on a clinical basis, an estimate of the predictability of the HNTLA for achievement skills. The findings, as presented in Chapter I, suggested that the HNTLA may not be appropriate for ability assessment for this population. To examine the reasons for these unusual inconsistencies, it was determined that an analysis was needed to evaluate the appropriateness of the test for this population. The following steps were taken:

1. The 118 HNTLA protocols for the sample population were examined, item-by-item to ensure that all test protocols were complete, correctly administered, and scored.

2. The Laboratory of Educational Research Tests Analysis Package (LERTAP) was selected as the most appropriate test analysis program to analyze the data as needed for this study.

3. To prepare the raw data from the test protocols for the LERTAP, the Computer Card Coding Sheets were prepared listing the descriptive information for each student and the individual item of the HNTLA.

4. From the Computer Card Coding Sheets, a summary of the data was prepared, listing student and subject descriptive information, as well as comments pertaining to the accuracy of the test protocol. This summary sheet is presented in Appendix B.

5. The data for each subject was transferred from the coding sheets to IBM cards. A print-out of the totals for the data cards was obtained and examined.

6. The computer control cards for the LERTAP were prepared. The item analysis was computed at Utah State University and the data was examined.

Data Analysis

The LERTAP item analysis program was selected because it appeared to most adequately meet the needs for this study. LERTAP evolved from programs written for the Venezuelan Ministry of Education and the Laboratory of Educational Research of the University of Colorado. LERTAP has been adapted for a variety of machine installations in a number of countries. The program was written to meet a demand for an easy-to-use, flexible and powerful item and test analysis routine. LERTAP provides the following information:

Index of difficulty. The index of item difficulty is determined by the proportion or percentage of test-takers who make the correct response. The higher the difficulty index, the easier the item. The difficulty levels provide information about the value of an item since it is related to item and total score variance. The dispersion of scores on an item is reduced if only few people are getting the item correct and also if only a few get the item correct. Consequently, item difficulty sets limits on the discriminative ability, since no item that is too easy or too difficult can show good discrimination

power. (A small item variance places constraints on the correlation of the scores with the total test score.)

Discriminative indices. LERTAP also provides discriminative indices for each item option, i.e., the correct item and distractors. To determine the discriminative indices, the correlation coefficients are computed between the item option and the subtest as well as the distractors.

Correlation coefficients. LERTAP provides indices to determine the quality of each item option: point-biserial and biserial-correlation coefficients. The point-biserial is computed by coding those who choose the option as "1's" and those who do not as "0's. The point-biserial is, then, the correlation between these codes and scores (either subtest or total). A biserial correlation coefficient differs from the point-biserial in that it assumes that the dichotomy of 0's and 1's is artificial, i.e., it assumes there is a continuum of response probability underlying the dichotomy.

Internal test reliability. Test reliabilty, using Hoyt's analysis of variance routine was employed. Reliability refers to the accuracy of the measures taken by the test. A "reliability coefficient" reflects the accuracy of the measuring process: the higher the value of the coefficient, the greater the accuracy of the process. The lower the value of the coefficient, the more error there is in the measuring process. A perfectly reliable test is said to have a reliability coefficient of unity (1.00). However, most test constructors are reasonably well satisfied if their tests yield reliability coefficients in the vicinity of .90.

Measurement of the standard error. The standard error of measurement for the sample is computed using Hoyt's analysis of variance routine.

Precoded variables. One of the research questions referred to whether there was significant differences in the scores of the test for males and females. LERTAP also included the "precoded" variable of sex, which was correlated with subtest and total test scores. This was used to test the hypotheses that the populations represented: (male/female) do not differ with respect to the subtest or total test score.

CHAPTER IV

RESULTS

It was the purpose of this study to evaluate the appropriateness of the Hiskey Nebraska Test of Learning Aptitude as a measure of ability for Navajo students with suspected learning difficulties. Item analytic procedures were used to provide a measure by which the research questions could be tested. Separate item analyses were run for the "younger group" (ages 5-10) and the "older group" (ages 11-17). This study sought to answer the following specific questions:

1. What is the item difficulty, discriminative power, and internal consistency of the items of the HNTLA for this population?

2. Is there a difference in difficulty of the individual subtests or for different sexes?

3. Is there a difference in item difficulty, discriminative power, or internal consistency for the item in the subtests given to the younger group (ages 5-10) vs. the older group (ages 11-17)?

The findings of the item analytic techniques employed in evaluating this instrument are presented in this chapter.

Results of the two-item analyses are presented in Tables 2-15. For subtests 1, 2, 4, 5, and 8 of the younger group (Tables 2, 3, 5, 6, and 9) and subtests 3, 6, a-d 7 of the older group (Tables 12, 15, and 16) in which the items are dichotomous (correct or incorrect) the results are tabulated in five columns: 1) the difficulty index, or the percentage of students who answered the item correctly, 2) the point-biserial correlation coefficient between each item and the

subtest, 3) point-biserial correlation coefficient between each item and the total test, 4) the average subtest score earned by those individuals who selected the correct response; and 5) the average total test score of those who selected the correct response.

For the subtest 3, 6, and 7 of the younger group (Tables 4, 7, and 8) and subtests 1, 2, 4, and 5 of the older group (Tables 10, 11, 13, and 14) which have response weightings. For example, where three points is better than one point, and only zero points is considered incorrect, the results are tabulated in three columns; 1) the difficulty index; the percentage of students who chose each item option, 2) the product-moment correlation coefficient of the item with the subtest, and 3) the product-moment correlation of the item with the total test. Items are presented in Tables 2-5 in the same sequence as they appear in the HNTLA protocol. Subtest statistics--number of items, highest score, lowest score, number of individuals, mean, standard deviation--are also presented for each subtest.

The following are tables and discussions of the item analyses for each subtest in the HNTLA as to what they tell us about the item difficulty, discriminative power, and internal consistency of the items for the subtest.

Subtest 1 - Bead Patterns

Evidence of item placement for this sample of children according to difficulty does not differ from that of the item placement found in the manual, as is shown in Table 2 which follows. There appears to be a dramatic increase in difficulty from item #4 through the remaining

Table 2

Subtest 1: Bead Patterns

Younger Group

Item Number	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Mean $\overline{ST}$[d]	Mean $\overline{TT}$[e]
1	98.4	.10	-.20	4.88	111.95
2	98.4	.43	.05	4.93	112.33
3	96.7	.53	.14	4.98	112.56
4	91.8	.53	.15	5.05	112.80
5	60.7	.63	.31	5.46	115.27
6	27.9	.70	.36	6.18	119.06
7	8.2	.55	.46	7.00	130.60
8	4.9	.42	.13	7.00	119.00

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]$\overline{ST}$ = Average subtest score of individual who selected the correct response.

[e]$\overline{TT}$ = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals = 61
Mean of subtest = 4.87
Standard deviation = 1.18
Number of items = 8
Highest score = 8
Lowest score = 1

Hoyt Estimate of Reliability = 0.55
Standard Error of Measurement = 0.75

items on the subtest. The first four items (first half) of the subtest were poor discriminators of the subtest and total test since almost all of the subjects responded correctly to them. For this sample, all of the items discriminated in the appropriate direction for both the subtest and the total test.

Subtest 2 - Memory for Color

For this subtest, as is shown in Table 3, there appears to be a relatively graduate shift from items which seem to be quite easy to items which seem to be quite difficult. Items 4, 6, and 7 did not contribute to the discriminatory power of the subtest, since

Table 3

Subtest 2: Memory for Color

Younger Group

Item Number	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	98.4	.02	.00	13.32	112.27
2	98.4	.37	.26	13.42	112.67
3	98.4	.37	.26	13.42	112.67
4	100.0	.00	.00	13.31	112.26
5	98.4	.08	-.02	13.33	112.23
6	100.0	.00	.00	13.31	112.26
7	100.0	.00	.00	13.31	112.26
8	98.4	.37	.10	13.42	112.42
9	88.5	.48	.21	13.69	113.15
10	78.7	.35	.05	13.71	112.54
11	88.5	.36	.14	13.59	112.85
12	65.6	.56	.42	14.20	115.85
13	72.1	.48	.31	13.95	114.57
14	57.4	.50	.41	14.26	116.43
15	29.5	.62	.35	15.39	118.61
16	13.1	.35	.35	15.25	122.88
17	16.4	.45	.28	15.50	119.30
18	14.8	.43	.33	15.56	121.67

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals	= 61	Number of items	= 19
Mean	= 13.13	Highest score	= 18
Standard deviation	= 2.20	Lowest score	= 7

Hoyt Estimate of Reliability = 0.65
Standard Error of Measurement = 1.27

their correlations with the subtest were .00. This is due to the fact that all of the subjects responded correctly to them. Item #5 was the only other item which did not discriminate. For this item, the point-biserial correlation coefficient between the item and the subtest was .08. The correlation between the item and the subtest was .08. The correlation between the item and the total test was -.02. However, since 98.4% of the students answered the item correctly, these statistics are based on only one person's score.

Subtest 3 - Picture Identification

Item placement according to difficulty, did not differ from that of the placement in the HNTLA manual, as is shown in Table 4.

Table 4

Subtest 3: Picture Identification

Younger Group

Item	Option	Difficulty[a] Index	Correlations ST[b]	TT[c]
1	0	0.0	.619	.266
	1	3.3		
	2	1.6		
	3	95.1		
2	0	0.0	.544	.204
	1	1.6		
	2	1.6		
	3	96.7		
3	0	0.0	.630	.213
	1	3.3		
	2	0.0		
	3	96.7		
4	0	1.6	.629	.208
	1	1.6		
	2	3.3		
	3	93.4		
5	0	3.3	.330	.465
	1	9.8		
	2	19.7		
	3	67.2		
6	0	8.2	.365	.399
	1	19.7		
	2	26.2		
	3	45.9		
7	0	16.4	.176	.371
	1	32.8		
	2	24.6		
	3	26.2		

[a]Difficulty Index = Percentage of students who chose each item option.

[b]ST = Product-moment correlation coefficient between the item with the subtest.

[c]TT = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 61	Number of items	= 7
Mean	= 24.90	Highest score	= 28
Standard deviation	= 2.74	Lowest score	= 12

Hoyt Estimate of Reliability = 0.65
Standard Error of Measurement = 1.51

Item #7 appeared to contribute little to the discriminatory power of the subtest, since the correlation coefficient was only .76 between the item and the subtest. Items #1, 2, 3, and 4 were poor discriminators of the total test since almost all of the subjects responded correctly to them. The correlation coefficients were .226, .204, .213, and .208, respectively.

Subtest 4 - Picture Association

There does not appear to be a sudden shift from item to item within the subtest, as is shown in Table 5. Item #2 did not add to

Table 5

Subtest 4: Picture Association

Younger Group

Item	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	86.9	.44	.25	9.58	113.42
2	100.0	.00	.00	9.21	112.26
3	93.4	.24	.08	9.35	112.51
4	78.7	.55	.43	9.83	114.90
5	75.4	.53	.32	9.87	114.46
6	82.0	.50	.36	9.72	114.26
7	70.5	.45	.30	9.84	114.56
8	78.7	.46	.42	9.73	114.35
9	63.9	.41	.28	9.87	114.79
10	45.9	.52	.41	10.43	117.54
11	60.7	.27	.10	9.68	113.49
12	41.0	.27	.30	9.92	116.60
13	21.3	.30	.24	10.46	117.85
14	23.0	.16	.02	9.86	112.79

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals	= 61	Number of items	= 14
Mean	= 9.21	Highest score	= 13
Standard deviation	= 2.18	Lowest score	4

Hoyt Estimate of Reliability = 0.53
Standard Error of Measurement = 1.43

the test's discriminative ability for this sample, since all of the children succeeded in passing. It appears that the items tend to lose discriminative power as the items become more difficult. The last two items, 13 and 14, are particularly poor. A percentage of 21.3 of the students chose the correction option A to item #13. The point-biserial correlation coefficient computed between item #13 and the subtest was .30 and with the total test was .24. While these correlations are not exceptionally low, a point-biserial correlation coefficient of .23 was computed between distraction option C and the total test, indicating that option C, an incorrect response alternative, attracted students that attained high total test scores. The correct option B for item #14 also lacks discriminative power, with a point-biserial correlation coefficient between the item and the subtest of .16 and with the total test of .02. Two the distractor options, C and D, correlated higher with the total test than did the correct option.

Subtest 5 - Paper Folding

Item #8 in Table 6, seems to be particularly difficult for this sample. Discriminative indices of .00 were computed between items #1 and 2 for the subtest. Thus, items #1 and 2 did not contribute to the discriminative power of the subtest. Inspection of the difficulty index reveals that these .00 discriminative indices are due to the fact that all subjects responded correctly to them. Except for these first two items, the items of this

Table 6

Subtest 5: Paper Folding

Younger Group

Item	Difficulty Index[a]	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	100.0	.00	.00	6.93	112.26
2	100.0	.00	.00	6.93	112.26
3	96.7	.38	.02	7.03	112.31
4	95.1	.42	.14	7.07	112.64
5	93.4	.55	.13	7.14	112.67
6	73.8	.71	.19	7.63	113.58
7	57.4	.64	.32	7.71	115.67
8	16.4	.46	.30	8.40	120.30
9	60.7	.69	.23	7.73	114.46

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals	= 61	Number of items	= 9
Mean	= 6.93	Lowest score	= 9
Standard deviation	= 1.44	Highest score	= 2

Hoyt Estimate of Reliability = 0.60
Standard Error of Measurement = 0.96

subtest operate as positive discriminators with both the subtest and the total test.

Subtest 6: Visual Attention Span

Analysis of the findings regarding the placement of the items on this subtest, as shown in Table 7, indicates that for this sample, item #9 appears to be more difficult than item #4. There is a sudden increase in difficulty from item #5 to item #6. All of the items are positive discriminators. However, the majority of the items are poor discriminators with the subtest and the total test. All of the items except item #5 have correlations

Table 7

Subtest 6: Visual Attention Span

Younger Group

Item	Option	Difficulty[a] Index	Correlations ST[b]	Correlations TT[c]
1	0	1.6	.207	.122
	1	98.4		
2	0	1.6	.207	.122
	1	98.4		
3	0	3.3	.121	.131
	1	34.4		
	2	62.3		
4	0	1.6	.234	.282
	1	14.8		
	2	83.6		
5	0	9.8	.316	.431
	1	34.4		
	2	55.7		
6	0	63.9	.034	.144
	1	19.7		
	2	16.4		
7	0	73.8	.296	.343
	1	18.0		
	2	8.2		
8	0	93.4	.270	.228
	1	6.6		
	2	0.0		
9	0	95.1	.204	.154
	1	4.9		
	2	0.0		

[a]Difficulty Index = Percentage of students who chose each item option.

[b]ST = Product-moment correlation coefficient between the item with the subtest.

[c]TT = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 61	Number of items	= 9
Mean	= 16.82	Highest score	= 22
Standard deviation	= 1.81	Lowest score	= 13

Hoyt Estimate of Reliability = 0.41
Standard Error of Measurement = 1.31

with the subtest below .30. The poorest subtest discriminators are items #3 and 6. The correlation coefficient computed for item #3 with the subtest was only .121. The correlation coefficient computed with the subtest for item #6 was only .034.

Items #1, 2, 3, 6, and 9 were poor discriminators with the total test. The correlation coefficients for these (.122, .122, .131, .144, and .154 respectively) items were below .20. The reliability estimate for this subtest was .41, the lowest subtest reliability estimate for the younger group.

Subtest 7 - Block Patterns

Placement of items in this subtest for this sample approximates the order which appears in the HNTLA manual, as is shown in Table 8. All of the items discriminate the appropriate direction. However, items #1 and 2 were poor discriminators for the total test due to the fact that almost all of the students reponded correctly to them. Items #12 and 13 were poor discriminators because few students were able to respond correctly to them. The correlation coefficients for the four items and the total test were .187, .170, .187, and .187 respectively.

Subtest 8 - Completion of Drawings

Item placement was not of significance for this subtest, as is shown in Table 9, since all drawings were given. All items discriminate in the appropriate direction for both the subtest and the total test. Items #2, 4, 5, 6, and 7 contribute little to the discriminative power of the subtest and the total test, since almost all subjects responded correctly to them. Item #12 appears to be a poor discriminator of the total test. The correlation coefficient of .07 was computed between item #12 and the total test.

Table 8

Subtest 7: Block Patterns

Younger Group

Item	Option	Difficulty[a] Index	Correlations ST[b]	Correlations TT[c]
1	0	0.0	.652	.187
	1	98.4		
	2	0.0		
	3	0.0		
2	0	1.6	.595	.170
	1	96.7		
	2	0.0		
	3	0.0		
3	0	9.8	.648	.436
	1	88.5		
	2	0.0		
	3	0.0		
4	0	29.5	.407	.335
	1	68.9		
	2	0.0		
	3	0.0		
5	0	29.5	.417	.323
	1	67.2		
	2	0.0		
	3	1.6		
6	0	39.3	.399	.486
	1	8.2		
	2	11.5		
	3	39.3		
7	0	67.2	.291	.501
	1	11.5		
	2	19.7		
	3	0.0		
8	0	91.8	.376	.313
	1	1.6		
	2	4.9		
	3	0.0		
9	0	93.4	.468	.377
	1	1.6		
	2	1.6		
	3	1.6		
10	0	96.7	.347	.220
	1	0.0		
	2	0.0		
	3	1.6		
11	0	96.7	.527	.322
	1	0.0		
	2	1.6		
	3	0.0		
12	0	98.4	.673	.187
	1	0.0		
	2	0.0		
	3	0.0		
13	0	98.4	.673	.187
	1	0.0		
	2	0.0		
	3	0.0		

[a]Difficulty Index = Percentage of students who chose each item option.

[b]ST = Product-moment correlation coefficient between the item with the subtest.

[c]TT = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 61	Number of items	= 13
Mean	= 19.33	Highest score	= 28
Standard deviation	= 3.64	Lowest score	= 0

Hoyt Estimate of Reliability = 0.73
Standard Error of Measurement = 1.81

Table 9

Subtest 8: Completion of Drawings

Younger Group

Item	Difficulty Index[a]	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	90.2	.36	.41	17.38	113.89
2	96.7	.22	.03	17.05	112.32
3	82.0	.44	.28	17.74	113.80
4	100.0	.00	.00	16.89	112.26
5	93.4	.20	.00	17.11	112.26
6	93.4	.28	.32	17.19	113.23
7	96.7	.11	.14	16.97	112.56
8	91.8	.25	.20	17.20	112.98
9	91.8	.38	.39	17.36	113.64
10	77.0	.35	.35	17.68	114.55
11	52.5	.50	.40	18.84	116.84
12	65.6	.36	.07	17.98	115.33
13	86.9	.46	.36	17.62	113.91
14	49.2	.52	.42	19.07	117.40
15	59.0	.43	.45	18.39	116.75
16	59.0	.23	.27	17.67	114.94
17	65.7	.43	.35	18.47	115.94
18	78.7	.40	.31	17.75	114.17
19	50.8	.45	.29	18.74	116.62
20	42.6	.38	.45	18.69	118.46
21	27.9	.44	.34	19.32	119.32
22	29.5	.56	.44	20.00	120.44
23	16.4	.57	.47	22.20	124.90
24	14.8	.43	.31	21.22	121.00
25	29.5	.45	.28	19.78	117.44
26	26.2	.32	.24	19.13	117.00
27	31.1	.36	.20	19.11	116.74

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals = 61
Mean = 16.89
Standard deviation = 4.19
Number of items = 27
Highest score = 26
Lowest score = 9

Hoyt Estimate of Reliability = 0.78
Standard Error of Measurement = 1.92

TOTAL TEST STATISTICS

Number of individuals = 61
Mean = 112.26
Standard deviation = 12
Number of items = 106
Highest score = 137
Lowest score = 88

Hoyt Estimate of Reliability = .87
Standard Error of Measurement = 4.32

Subtest 1 - Visual Attention Span

As with the younger group, in the older group the majority of the nine items of this subtest are poor discriminators, as is shown in Table 10. Items #1-5 have correlations with the subtest below .30.

Table 10

Subtest 1: Visual Attention Span

Older Group

Item	Option	Difficulty[a] Index	Correlations ST[b]	Correlations TT[c]
1	0	0.0	0.000	0.000
	1	100.0		
2	0	0.0	0.000	0.000
	1	100.0		
3	0	3.5	-0.015	0.092
	1	35.1		
		61.4		
4	0	3.5	0.050	0.217
	1	3.5		
	2	93.0		
5	0	5.3	0.232	0.385
	1	15.8		
	2	78.9		
6	0	42.1	0.174	0.218
	1	12.3		
	2	45.6		
7	0	59.6	0.393	0.213
	1	15.8		
	2	24.6		
8	0	73.7	0.366	0.193
	1	15.8		
	2	10.5		
9	0	84.2	0.412	0.162
	1	8.8		
	2	7.0		

[a] Difficulty Index = Percentage of students who chose each item option.

[b] ST = Product-moment correlation coefficient between the item with the subtest.

[c] TT = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 57	Number of items	= 9
Mean	= 18.49	Highest score	= 25
Standard deviation	= 2.32	Lowest score	= 13

Hoyt Estimate of Reliability = 0.46
Standard Error of Measurement = 1.51

The correlation coefficients for the total test were below .30 for all items except item #5. The reliability estimate for this subtest, .46, was the lowest for the older group, as it was for the younger group.

Subtest 2 - Block Patterns

Placement of the items in this subtest, as is shown in Table 11, approximate the order that appears in the HNTLA manual. Discriminative indices of .00 were computed for items #1, 2, and 3 due to the fact that all subjects responded correctly to them. The remaining items, #4-13, discriminate in the appropriate direction for both the subtest and the total test. Items #10, 11, and 13 were poor discriminators for both the subtest and the total test, since few of the subjects were successful in responding to them. Correlation coefficients computed for these tiems were below .30.

Subtest 3 - Completion of Drawings

Item placement is not of significance for this subtest, as is shown in Table 12, since all drawings were given. Items #1 and 4 were negative discriminators with the subtest. Items #4, 10, and 21 were negative discriminators with the subtest. The correlation coefficients computed between item #1 and the subtest was -.20 and for item #4 was -.09. Items #4, 10, and 21 were negative discriminators with the total test. The correlation coefficients were -.28, -.05, and -.11 respectively. Discriminative indices of .00 for the subtest and the total test comprised for items #2, 7, and

Table 11

Subtest 2: Block Patterns

Older Group

Item	Option	Difficulty[a] Index	Correlations r_{bis}[b]	Correlations r_{tt}[c]
1	0	3.5	.000	.000
	1	100.0		
	2	0.0		
	3	0.0		
2	0	0.0	.000	.000
	1	100.0		
	2	0.0		
	3	0.0		
3	0	0.0	.000	.000
	1	100.0		
	2	0.0		
	3	0.0		
4	0	10.5	.348	.424
	1	80.0		
	2	0.0		
	3	0.0		
5	0	12.3	.283	.253
	1	86.0		
	2	0.0		
	3	0.0		
6	0	26.3	.551	.597
	1	0.0		
	2	24.6		
	3	49.1		
7	0	59.6	.415	.440
	1	0.0		
	2	24.6		
	3	17.5		
8	0	40.4	.606	.446
	1	0.0		
	2	12.3		
	3	0.0		
9	0	86.0	.442	.433
	1	0.0		
	2	0.0		
	3	1.8		
10	0	96.5	.155	.144
	1	0.0		
	2	0.0		
	3	1.8		
11	0	93.0	.291	.278
	1	0.0		
	2	1.8		
	3	3.5		
12	0	96.5	.356	.323
	1	0.0		
	2	3.5		
	3	0.0		
13	0	96.2	.227	.252
	1	0.0		
	2	0.0		
	3	1.8		

[a] Difficulty index = Percentage of students who chose each item option.

[b] r_{bis} = Product-moment correlation coefficient between the item with the subtest.

[c] r_{tt} = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals = 57
Mean = 22.11
Standard deviation = 3.64

Number of items = 13
Highest score = 31
Lowest score = 16

Hoyt Estimate of Reliability = 0.66
Standard Error of Measurement = 2.04

Table 12

Subtest 3: Completion of Drawings

Older Group

Item	Option	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	0	3.5	.20	-.01	23.00	87.00
	1	96.5	-.20	.01	20.22	87.47
2	0	0.0	0.00	0.00	0.00	0.00
	1	100.0	0.00	0.00	20.32	87.46
3	0	10.5	-0.02	-0.01	20.17	87.17
	1	89.5	0.02	0.01	20.33	87.49
4	0	1.8	.09	.28	22.00	109.00
	1	98.2	-.09	-.28	20.29	87.07
5	0	1.8	-.17	-.21	17.00	71.00
	1	98.2	.17	.21	20.33	87.75
6	0	3.5	-.06	-.10	19.50	82.00
	1	96.5	.06	.10	20.35	87.65
7	0	0.0	0.00	0.00	0.00	0.00
	1	100.0	0.00	0.00	20.32	87.46
8	0	1.8	-.27	-.01	15.00	87.00
	1	98.2	.27	.01	20.41	87.46
9	0	0.0	0.00	0.00	0.00	0.00
	1	100.0	0.00	0.00	20.32	87.46
10	0	5.3	-.15	.06	18.67	89.67
	1	94.7	.15	-.06	20.41	87.33
11	0	26.3	-.41	-.15	18.53	84.93
	1	73.7	.41	.15	20.95	88.36
12	0	29.8	-.39	-.19	18.76	84.47
	1	70.2	.39	.19	20.98	88.72
13	0	5.3	-.33	-.22	16.67	77.67
	1	94.7	.33	.22	20.52	88.00
14	1	84.2	.27	.22	20.63	88.46
15	1	91.2	.30	.19	20.56	88.06
16	1	78.9	.29	.23	20.71	88.69
17	1	64.9	.40	.17	21.08	88.78
18	1	96.5	-.13	.08	20.39	87.62
19	1	71.9	-.18	.10	20.61	88.10
20	1	63.2	.55	.13	21.42	88.47
21	1	40.4	-.16	-.11	20.83	86.09
22	1	47.4	.40	.27	21.41	90.37
23	1	29.8	.36	.08	21.76	88.65
24	1	26.3	.36	.24	21.87	91.60
25	1	33.3	.38	.20	21.74	90.42
26	1	42.1	.55	.28	22.00	90.83
27	1	50.9	.40	.27	21.34	90.17

[a] Difficulty Index = Percentage of students who answered item correctly.

[b] PB-ST = Point-biseral correlation between item and subtest score.

[c] PB-TT = Point-biseral correlation between item and total test score.

[d] ST = Average subtest score of individual who selected the correct response.

[e] TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals = 57
Mean = 20.32
Standard deviation = 2.63
Number of items = 27
Highest score = 26
Lowest score = 14

Hoyt Estimate of Reliability = 0.52
Standard Error of Measurement = 1.79

9, since all students responded correctly to them. Items #3, 5, 6, 10, 18, 19, and 21 were poor discriminators with the subtest, since correlation coefficiencies computed for them were below .20.

Subtest 4 - Memory for Digits

For this subtest, there is a relatively gradual shift from items that appear to be quite easy to items that appear to be quite difficulty, as is depicted in Table 13. All of these items discriminate in the appropriate direction for both the subtest and the total test.

Subtest 5 - Puzzle Blocks

Evidence of item placement for this sample of children, according to item difficulty, as is shown in Table 14, does not differ from that of the item placement found in the HNTLA manual. All of the items discriminate in the appropriate direction for both the subtest and the total test.

Subtest 6 - Picture Analogies

Fairly high correlation coefficients were computed between the items and the subtest, as is shown in Table 15. However, items #1, 3, 6, 7, 11, and 12 were poor discriminators for the total test. Their correlation coefficients between the items and the total test were below .20. Item #6 was a negative discriminator for the total test. The point-biserial correlation coefficient computed between item #6 and the total test was -.30.

Table 13

Subtest 4: Memory for Digits

Older Group

Item	Option	Difficulty[a] Index	Correlations	
			ST[b]	TT[c]
1	0	0.0	.608	.306
	1	1.8		
	2	98.0		
2	0	0.0	.564	.261
	1	0.0		
	2	94.7		
3	0	1.8	.593	.301
	1	1.8		
	2	93.0		
4	0	14.0	.485	.219
	1	10.5		
	2	75.4		
5	0	38.6	.617	.521
	1	8.8		
	2	52.6		
6	0	68.4	.443	.467
	1	12.3		
	2	19.3		
7	0	75.4	.472	.485
	1	5.3		
	2	19.3		
8	0	86.0	.327	.326
	1	12.3		
	2	1.8		

[a]Difficulty Index = Percentage of students who chose each item option.

[b]ST = Product-moment correlation coefficient between the item with the subtest.

[c]TT = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 57	Number of items	= 8
Mean	= 17.37	Highest score	= 23
Standard deviation	= 3.74	Lowest score	= 6

Hoyt Estimate of Reliability = 0.80
Standard Error of Measurement = 1.57

Table 14

Subtest 5: Puzzle Blocks

Older Group

Item	Option	Difficulty[a] Index	Correlations	
			S_T[b]	T_T[c]
1	0	1.8	.358	.320
	1	1.8		
	2	19.3		
	3	77.2		
2	0	14.0	.363	.476
	1	14.0		
	2	26.3		
	3	45.6		
3	0	47.4	.402	.478
	1	24.6		
	2	26.3		
	3	1.8		
4	0	42.1	.567	.519
	1	14.0		
	2	38.6		
	3	5.3		
5	0	73.7	.358	.496
	1	8.8		
	2	10.5		
	3	7.0		
6	0	78.9	.435	.474
	1	10.5		
	2	8.8		
	3	1.8		
7	0	77.2	.471	.453
	1	8.8		
	2	12.3		
	3	1.8		

[a]Difficulty index = Percentage of students who chose each item option.

[b]S_T = Product-moment correlation coefficient between the item with the subtest.

[c]T_T = Product-moment correlation coefficient between the item with the total test.

Subtest Statistics

Number of individuals	= 57	Number of items	= 7
Mean	= 14.88	Highest score	= 24
Standard deviation	= 3.67	Lowest score	= 7

Hoyt Estimate of Reliability = 0.70
Standard Error of Measurement = 1.85

Table 15

Subtest 6: Picture Analogies

Older Group

Item	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	68.4	.28	.15	7.59	88.51
2	89.5	.35	.23	7.45	88.25
3	87.7	.33	.16	7.46	88.08
4	91.2	.34	.29	7.42	88.38
5	75.4	.55	.24	7.84	88.86
6	71.9	.30	-.03	7.59	87.24
7	52.6	.40	.15	7.97	88.93
8	45.6	.57	.40	8.46	91.92
9	47.4	.53	.39	8.33	91.74
10	40.4	.49	.23	8.39	90.39
11	12.3	.20	.07	8.29	89.29
12	38.6	.26	.06	7.86	88.23

[a]Difficulty index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals	= 57	Number of items	= 12
Mean	= 7.21	Highest score	= 12
Standard deviation	= 2.02	Lowest score	= 1

Hoyt Estimate of Reliability = 0.48
Standard Error of Measurement = 1.39

Subtest 7 - Spatial Reasoning

Correlation coefficients computed with the subtest, as is shown in Table 16, were above .30, for all of the items on this subtest except #6 and 8. However, the items were poor discriminators with the total test. Correlation coefficients computed with the total test were below .30 for all of the items except #9. Items #4 and 7 are especially poor discriminators for the total test. Item #4's correlation coefficient with the total test was .00. Item #7's total test correlation was -.02.

Table 16

Subtest 7: Spatial Reasoning

Older Group

Item	Difficulty[a] Index	PB-ST[b]	PB-TT[c]	Means ST[d]	Means TT[e]
1	73.7	.42	.26	3.95	89.07
2	75.4	.36	.17	3.88	88.44
3	71.9	.50	.26	4.05	89.12
4	24.6	.35	.00	4.50	87.43
5	15.8	.33	.12	4.89	90.33
6	22.8	.26	.08	4.31	88.92
7	13.3	.37	-.02	4.73	87.09
8	8.8	.29	.13	5.00	91.80
9	31.6	.39	.31	4.44	92.11
10	14.0	.35	.27	4.88	94.38

[a]Difficulty Index = Percentage of students who answered item correctly.

[b]PB-ST = Point-biseral correlation between item and subtest score.

[c]PB-TT = Point-biseral correlation between item and total test score.

[d]ST = Average subtest score of individual who selected the correct response.

[e]TT = Average total test score of those who selected correct response.

Subtest Statistics

Number of individuals = 57
Mean = 3.58
Standard deviation = 1.51
Number of items = 10
Highest score = 7
Lowest score = 0

Hoyt Estimate of Reliability = 0.30
Standard Error of Measurement = 1.20

TOTAL TEST STATISTICS

Number of individuals = 57
Mean = 103.95
Standard deviation = 11.19
Number of items = 86
Highest score = 126
Lowest score = 80

Hoyt Estimate of Reliability = 0.81
Standard Error of Measurement = 4.88

Internal Test Reliability

Employing Hoyt's analysis of variance routines, the coefficients of reliability were found to be .87 for the Navajo 5 to 10 year group and .81 for the Navajo 11 to 17 year group.

For the standardization sample, Hiskey used the split-half method and the Spearman-Brown formula and found coefficients of reliability of .933 for the hearing 3 to 10 year group and

.964 for the hearing 11 to 17 year group. Thus, the HNTLA appears to be slightly less reliable for the Navajo sample than for the standardization sample.

Subtest reliability estimates for the two Navajo groups are presented in Table 17. Separate subtest reliability estimates were

Table 17

Summary of Subtest and Total Test Reliability Estimates for the HNTLA

Subtest #		
NAVAJO YOUNGER GROUP (Ages 5-10)		
1	Bead Patterns	.55
2	Memory for Color	.65
3	Picture Identification	.65
4	Picture Association	.53
5	Paper Folding	.60
6	Visual Attention Span	.41
7	Block Patterns	.73
8	Completion of Drawings	.78
	TOTAL TEST	.87
NAVAJO OLDER GROUP (Pages 11-17)		
1	Visual Attention Span	.46
2	Block Patterns	.66
3	Completion of Drawings	.52
4	Memory for Digits	.80
5	Puzzle Blocks	.70
6	Picture Analogies	.48
7	Spatial Reasoning	.30
	TOTAL TEST	.81

not reported by Hiskey on the standardization group. The reliability estimates computed for the Navajo sample are fairly low, as would be expected due to the few number of items within many of the subtests. Considering the length of the subtests and their relative unreliability, there is little support for the potential of the subscales for differential diagnosis.

Intercorrelations

Hiskey, (1966) holds that the correlations between subtests are important because the scale consists of separately scored parts of subtests. The subtest intercorrelations for the standardization group are listed in Table 18. The subtest intercorrelations

Table 18

Norm Group Correlation Matrix

SUBTESTS	BP	MC	PI	PA	PF	VAS	BLP
Ages 3-10 (Younger Group)							
Bead Pattern							
Memory for Color	.554						
Picture Identification	.434	.667					
Picture Association	.529	.736	.617				
Paper Folding	.378	.717	.777	.668			
Visual Attention Span	.530	.413	.716	.397	.621		
Block Patterns	.728	.599	.523	.551	.480	.608	
Completion of Drawings	.659	.440	.371	.477	.315	.505	.644
Ages 11-17 (Older Group)							
	VAS	BLP	CD	MD	PB	PA	
Visual Attention Span							
Block Patterns	.395						
Completion of Drawings	.371	.407					
Memory for Digits	.402	.458	.265				
Puzzle Blocks	.298	.409	.367	.375			
Picture Analogies	.320	.395	.348	.387	.369		
Spatial Reasoning	.253	.439	.289	.366	.360	.331	

for the Navajo sample are listed in Table 19. Out of the 28 intercorrelations for the younger group, only two intercorrelations

Table 19

Correlation Matrix for Navajo Sample

SUBTESTS	BP	MC	PI	PA	PF	n=16 VAS	BP CD
Ages 3-10 (Younger Group)							
Bead Pattern							
Memory for Color	.455						
Picture Identification	.068	.077					
Picture Association	.278	398	.337				
Paper Folding	.182	.059	232	.340			
Visual Attention Span	.309	.379	.185	.365	.011		
Block Patterns	.139	.254	.157	.231	.215	.206	
Completion of Drawings	.234	.428	.386*	.520*	.219	.270	.372
Precoded Variable - Sex	.087	.018	-.109	-.234	-.059	-.043	-.182
Ages 11-17 (Older Group)							
	VAS	BLP	CD	MD	PB	n=57 PA	SR
Visual Attention Span							
Block Patterns	-.171						
Completion of Drawings	-.207	.188					
Memory for Digits	.372	.194	-.085				
Puzzle Blocks	.237	*.511	.240	.362			
Picture Analogies	.027	.192	.293	.167	.141		
Spatial Reasoning	-.523	.236	.276	.041	.132	.376*	
Sex	.24	-.258	.171	.221	.024	-.056	-.199

*The intercorrelations that are higher for the Navajo sample than for the norm group.

(Completion of Drawings as it relates to Picture Identification and Picture Association) for the Navajo group are higher than the

intercorrelations for the norm group. Out of the 21 intercorrelations for the older group, also only two (Puzzle Blocks with the Block Patterns and Picture Analogies with Spatial Reasoning) are higher than the norm group. This seems to indicate that the subtests are operating more independently for the Navajo group than for the norm group.

Subtest Correlation with Total Test

The correlation coefficients computed between each subtest and the total test for both the Navajo and the standardization sample are presented in Table 20. In general, the subtest correlations were

Table 20

Correlations Between Subtests and the Total Test for Both Navajo and Standardization Groups

Subtests	Navajo		Norm	
	Group I Ages 5-10	Group II Ages 11-17	Group I Ages 3-10	Group II Ages 11-17
Bead Pattern	.440		.755	
Memory for Color	.608		.621	
Picture Identification	.548		.514	
Picture Association	.706*		.619	
Paper Folding	.372		.544	
Visual Attention Span	.514	.415	.667	.574
Block Patterns	.628	.704*	.774	.646
Completion of Drawings	.802*	.296	.723	.577
Memory for Digits		.609		.635
Puzzle Blocks		.765*		.677
Picture Analogies		.470		.650
Spatial Reasoning		.391		.626

*The subtest correlations that were higher for the Navajo sample than for the norm group.

younger Navajo group, two of the eight correlations (Picture Association and Completion of Drawings) were higher than the standardization sample. For the Older Navajo group, two of the seven correlations (Block Patterns and Puzzle Blocks) were higher.

Correlations Between Subtest and Sex of Subjects

The correlation between subtests and sex of Navajo subjects are presented in the correlation matrix in Table 19. The correlations were not significant at the .05 level.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The primary purpose of this study was to evaluate the appropriateness of the Hiskey Nebraska Test of Learning Aptitude as a measure of ability for Navajo students with suspected learning difficulties. To accomplish this, item analytic procedures were employed for both the younger (ages 5-10) and older (ages 11-17) groups. This chapter will discuss and draw conclusions and implications from the data presented in Chapter IV.

Discussion of Results

1. The estimated reliability for the Navajo sample was .87 for the younger group and .81 for the older group. The reliability estimates of the standardization group were .93 for the younger group and .90 for the older group. The HNTLA appears to be slightly less reliable for the Navajo sample than for the standardization group. However, two factors may have affected these reliability estimates. First, the standardization sample contained a broader range of ability than did the Navajo sample. The Navajo students were referred for assessment due to suspected learning difficulties, where the standardization sample was randomly selected. Secondly, the norm group included children down to the age of three. There were no children in the Navajo sample younger than five years old. The reliability coefficient for the younger group may have been lowered by this truncation of range.

2. Subtest intercorrelation matrices for the younger and older subsamples of the standardization group suggest that a substantial portion of the variation in subtest scores are associated with an underlying performance dimension. The subtest intercorrelations for the Navajo sample are generally lower than for the norm group. In some cases, the differences are substantial. This seems to indicate that the subtests operate more independently for the Navajos than they did for the standardization group.

3. In general, the subtests did not correlate as well with the total test for the Navajo sample. This gives evidence that the test may be measuring a different underlying factor for this sample, and calls the validity of this test, for this group, into question.

4. The correlations between the subtests and sex of subject were not significant. It does not appear that there is sex bias associated with this test.

5. The items with low discriminative ability, due to extreme difficulty indices (.00-.10 and .90-.100) appear necessary to provide representative sampling for the age range encompassed by the instrument. These items provide adequate ceiling and floor to the test. In addition, as Lehmann and Mehrens (1976) note some easy items are needed to instill proper motivation in the examinees, and thus, these items are valuable."

6. It is the items with low or negative discriminative indices, in which the difficulty indices are <u>not</u> extreme, that the item is "red-flagged" and is in need of further examination. For, in the case of negative discriminating items, one has a situation in which

the higher ability students get the item incorrect and the lower ability students get the item correct. Low and negative discriminating items with average difficulty indices will be examined for both the younger and older Navajo group.

Younger group. The correlation coefficient computed between Item 7 (Picture Identification) and the subtest was low. The item presents a series of five houses, and the subject is to match a presented house with one of the five in the series before him/her. The low discriminative ability on this item could be due to the fact that many children on the reservation live in "hogans" and are not familiar with the type of house presented.

Items 13 and 14 of the Picture Association subtest were poor discriminating items. In this subtest, the subject is called to match one of four pictures with two associated pictures. With Item 13 (hands holding objects), no explanation is attempted. However, with Item 14, the pictures are of underwater animals. Navajo children seldom travel off the reservation, and thus, they may not be familiar with animals associated with the ocean. Item 12 of the Completion of Drawings was a poor discriminator of the total test for the younger group. The student is presented with a picture of a tricycle with one of the rear wheels missing. It could be that Navajo children are unfamiliar with tricycles.

Older group. Many of the items in the Completion of Drawings subtest were poor discriminators with the subtest or total test. The reasons speculated is that the Navajos are unfamiliar with the type of objects presented.

Item 6 (Picture Analogies) was a negative discriminating item for the total test. The analogy is between articles of clothing and the parts of the body with which they are associated. The Navajo children may be unfamiliar with the type of clothing articles presented. The traditional style of dress for the Navajo culture is much different than the type presented in this analogy.

Items 4 and 7 (Spatial Reasoning) were poor discriminators of the total test. No explanation is attempted.

7. The reliability estimates computed for the Visual Attention Span subtest were the lowest for both younger and older Navajo groups. In this subtest, the subject is called to reproduce a sequence of objects from memory. The lack of consistency with this subtest could be due to a cultural difference in the ability required for this subtest. A more probable explanation of the inconsistent and unreliable results of this subtest is due to the fact that the majority of the items (such as violin, lawnmower, camera) are based on white, middle class experiences.

Conclusions

The findings of this study are: (1) reliability estimates of .87 and .81 indicate a high reliability of the instrument; (2) the majority of the items discriminate in the appropriate direction for both subtest and total test scores; (3) the intercorrelation matrices indicate that the subtests may be operating even more independently for this group; (4) there is a lack of sex bias associated with the test; and (5) subtest correlations with the total test are adequate. Based on these findings, it was concluded that the Hiskey Nebraska Test of

Learning Aptitude will give acceptably reliable and consistent results for Navajo children with suspected learning difficulties.

However, it is validity, not reliability that is the ultimate measure of the quality of a test. As stated by Stanley and Hopkins (1974, p. 101): "Regardless of all merits of a test, if it lacks validty for a particular task, the information provided is useless." Validity may be defined as the accuracy with which the test measures what it is intended to measure. This is in contrast to reliability, which is usually defined as the accuracy with which the test measures whatever it does measure. Thus, the HNTLA may be measuring accurately, but measuring something other than "learning ability." It seems that only half of the job is done in terms of assessing the appropriateness of the test for Navajo students. Further research is needed to demonstrate that the test is valid for use with these children.

Limitations

1. The Navajo students in this sample were not randomly selected, but were referred for assessment for suspected learning difficulties. This restriction in range of ability affects the reliability estimates as well as the generalizability of this study.

2. There were no children from ages 3-5 included in the Navajo sample. The standardization group for the HNTLA included children of these ages. There are limitations on the comparisons that can be drawn between the two groups, due to the truncated range of the Navajo sample.

Recommendations

Based on the results obtained and the conclusions drawn, the following recommendations were made:

1. Further item analytic studies should be employed with a more representative sample of Navajo students and with other culturally different populations.

2. If, from further item analytic studies of the HNTLA with this population, the pattern of negative discriminators continues, adaptation of the HNTLA for the Navajo population would be useful. This could be accomplished by either systematically eliminating the non-discriminatory items or by replacing the present items which are negative discriminators with items of higher discriminative ability which the Navajo children are more familiar.

3. Factor analytic studies would provide information into what the HNTLA is measuring.

4. Research is needed to determine the validity of the HNTLA for the Navajo population; predictive, construct, and concurrent.

REFERENCES

Anastasi, A. Psychological testing (4th Ed). New York: MacMillian, 1976.

Bailey, D. B., & Harbin, G. L. Nondiscriminatory evaluation. Exceptional Children, 1980, 56(8), 590-596.

Brodie, J. G. Language tests and linguistically different learners: The sad state of the art. Elementary English, 1970, 47, 814-828.

Bureau of Indian Affairs. Gudielines for the education of Indian Students. U.S. Office of Interior, January 1980.

Burors, O. K. The eighth mental measurement yearbook, Highland Park, New Jersey: Gryhon Press, 1978.

Carleton, F. G., & Stacey, C.S. An item analysis of the WISC. Journal of Clinical Psychology, 1955, 11, 149-154.

Cattell, R. B. Culture-free intelligence test. Champaign, Il.: Institute for Personality and Ability Testing, 1950.

Chandler, J. T., & Spakos, J. Spanish-speaking pupils classified as educable mentally retarded. Integrated Education, 1969, 7(2), 28-33.

Cohen, R. A. Conceptual styles, culture conflict, and nonverbal tests of intelligence. American Anthropologist, 1969, 71, 828-856.

Davis, A., & Eells, K. Davis-Eells test of general intelligence or problem solving ability. Yonkers, N.Y.: World Book, 1953.

Dent, H. Assessing Black children for mainstream placement. In R. Jones (Ed.), Mainstreaming the minority child. Reston, Va.: Council for Esceptional Children, 1976.

Diana V. State Board of Education C-70 37 REP, District Court for Northern California, February, 1970. In D. J. Reschly & G. D. Phye, School psychology: Perspectives and issues. New York: Academic Press, 1979.

Drew, C. J. Criterion-referenced assessment of minority group children. Journal of School Psychology, 1973, 11(4), 323-329.

Dunn, L. Special Education for the mildly retarded: Is much of it justified? Exceptional Children, 1968, 35, 5-22.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

Ebel, R. L. Essentials of education measurement. Englewood Cliffs, New Jersey: Prentice-Hall, 1979.

Federal Register. Nondiscrimination on basis of handicap regulation implementing Section 504 of the Rehabilitation Action of 1973. May 4, 1977, pp. 22676-22702.(a)

Federal Register. Education of handicapped children. Regulations implementing education for all handicapped children's act of 1975, August, 23, 1977, pp. 42474-42518.(b)

Fifield, M. G. Improving the utilization and educational relevance of individual psychoeducational assessment reports in the placement of and IEP development for handicapped Native American children. Proposal submitted to the U.S. Department of Education, January 9, 1981.

Ford, C., Jr.. School testing programs and minorities. The Clearing House, 1980, 53(6), 300-301.

Guadelupe V. Tempe Elementary School District. District Court for Arizona, 71-435, Janaury, 1972. In Reschly, D. J. & G. D. Phye (Eds.), School psychology: Perspectives and issues. New York: Academic Press, 1979.

Harris, M. L. Some methodological suggestions for construction of an objective measuring instrument. In L. R. Nelson, Guide to LERTAP use and interpretation. Dunedin, New Zealand: Department of Education, University of Otago, 1974.

Havinghurst, R. J. Indian education: Accomplishments of the past decade. Phi Delta Kappan, January 1981, 62(5), 329-331.

Hiskey, M. S. Hiskey-Nebraska test of learning-aptitude (Revised Manual, Lincoln, Neb.: Union College Press, 1966.

Jensen, A. R. Bias in mental testing. New York: The Free Press, 1980.

Larry, P. et al. V. Riles, W. et al. United Stated District Court, Northern District of California, Case No. C-71-2270 RFP, 1972, 1974. In D. J. Reschly & G. D. Phye (Eds.), School psychology: Perspectives and issues. New York: Academic Press, 1979.

Lehman, I., & Mehrens, W. Mesurement and evaluation in education and psychology. Holt, Rinehart and Winston, 1976.

Lunemann, A. The correlational validity of IQ and a function ethnicity and desegregation. Journal of School Psychology, 1974, 4,(12), 28-32.

McShane, D., & Plas, J. Weschler scale performance patterns of American Indian children. Psychology in the Schools, 1982, 19(1), 39-44.

Mendoza-Friedman, M. Spanish bilingual students and intelligence testing. Thrust for Educational Leadership, 1973, 3(2), 20-33.

Mercer, J. R. Sociocultural factors in labeling mental mandates. Peabody Journal of Education, 1971, 48(3), 188-203.

Mowder, B. A. A strategy for the assessment of bilingual handicapped children. Psychology in the Schools, 1980, 17(1), 7-11.

Nelson, L. R. Guide to LERTAP use and interpretation. Dunedin, New Zealand: Department of Education, University of Otago, 1974.

Palomares, U. H., & Johnson, L. C. Evaluatioin of Mexican-American pupils for EMR classes. California Education, 1966, 3(8), 27-29.

Public Law 94-142. Education for all Handicapped Children Act, 1975.

Raven, J. C. Coloured progressive matrices. New York: Psychological Corp., 1962.

Reschly, D. J., & Phye, G. D. School psychology: Perspectives and issues. New York: Academic Press, 1979.

Sandoval, J., & Millie, M. Accuracy and judgments of WISC-R item difficulty for minority group. Journal of Consulting and Clinical Psychology 1980, 48(2), 249-253.

Sattler, J. M. Assessment of children's intelligence. Philadelphia: W. B. Sanders Company, 1978.

Shutt, D. L., & Hannon, T. A. The validity of the HNTLA for the evaluation of abilities of bilingual children. Educational and Psychological Measurement, 1974, 34, 429-432.

Stanley, M., & Hopkins, L. In L. R. Nelson, Guide to LERTAP use and interpretation. Dunedin, New Zealand: Department of Education University of Otago, 1974.

Taylor, C., & White, K. Effects of reinforcement and training on student test taking skills. Journal of Educational Measurement, 1982, 19, 199-208.

Teeter, A., Moore, C., & Peterson, J. WISC-R verbal and performance abilities of Native American students referred for school learning problems. Psychology in Schools, 1982, 19(1), 39-44.

Thorndike, R. L. Stanford-Binet intelligence scale, form L-M, 1972, norm tables, (3rd ed.). Boston: Houghton-Muffin, 1973.

Tractenberg, P. L., & Jacob, E. Pupil testing: A legal view. Phi Delta Kappan, 1977, 60, 249-254.

Turnbull H. The past and future impact of court decisions in special education. Phi Delta Kappan, 1978, 60, 523-572.

Vance, H. B., & Gaynor, P. E. Reliability scores of the Weschler Intelligence Scale for Children Subtests with culturally different children. Journal of General Psychology, 1976, 95, 132-139.

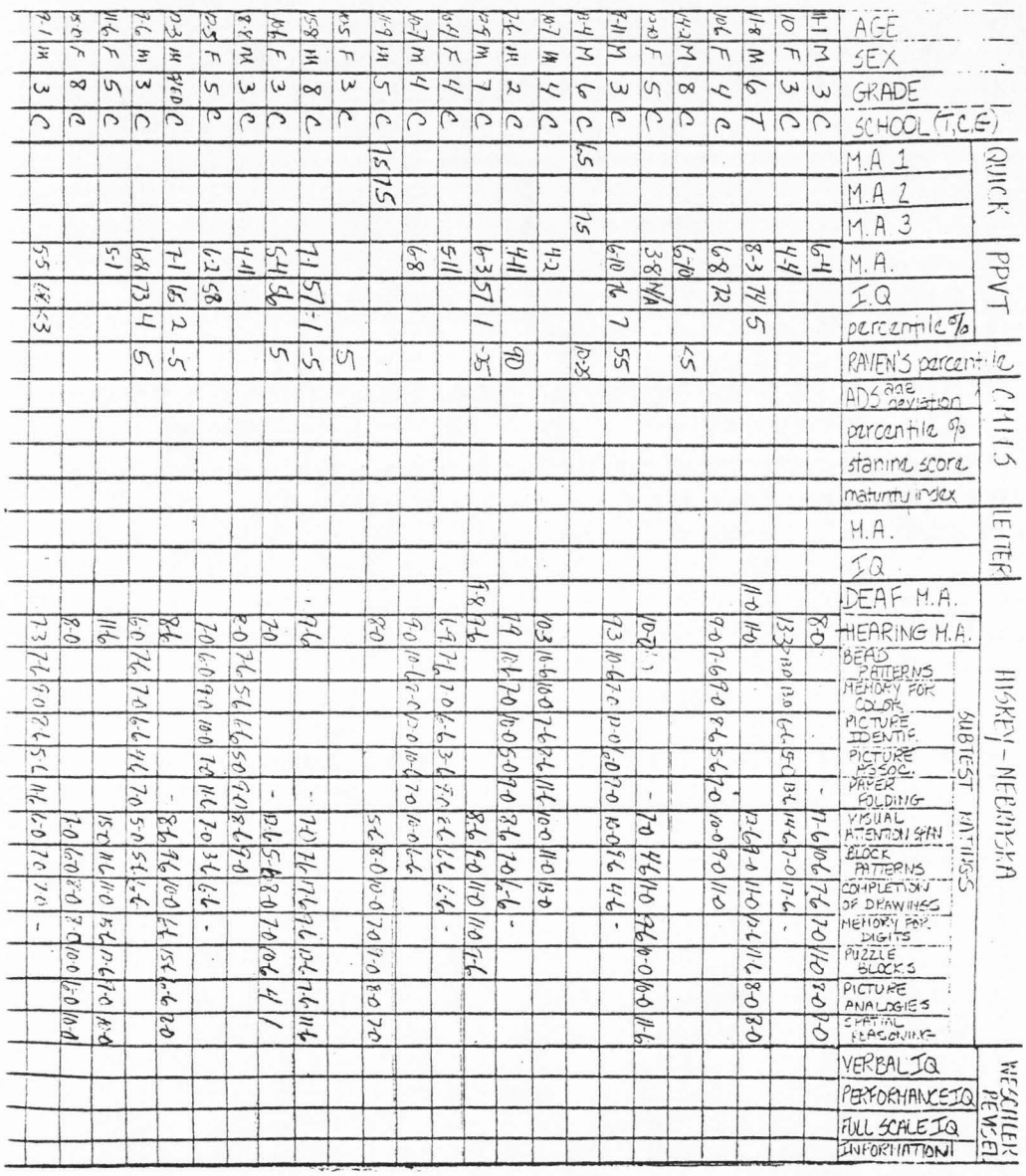
Weschler, D. Weschler Intelligence Scale for Children-Revised. New York: Psychological Corporation, 1974.

Williams, R. Abuses and misuses in testing Black children. The Counseling Psychologist, 1971, 2, 62-73.

APPENDIXES

Appendix A

Table Developed for Tests Administered for the Fort Defiance Assessment Project



WISC-R: Verbal IQ	WISC-R: Performance IQ	WISC-R: Full Scale IQ	Subtests: Information	Subtests: Similarities	Subtests: Arithmetic	Subtests: Vocabulary	Subtests: Comprehension	Subtests: (Digit Span)	Subtests: Picture Completion	Subtests: Picture Arrangement	Subtests: Block Design	Subtests: Object Assembly	Subtests: Coding	Subtests: (Mazes)	Bender: age equivalent	Bender: percentile %	VMI: age equivalent	Visual Auditory Digit Span: X error score	Visual Auditory Digit Span: Y error score	WRAT spelling: grade rating	WRAT spelling: standard score	WRAT spelling: percentile %	WRAT arithmetic: grade rating	WRAT arithmetic: standard score	WRAT arithmetic: percentile %	WRAT reading: grade rating	WRAT reading: standard score	WRAT reading: percentile %	Brigance	Woodcock Word I.D.: grade score	Woodcock Word I.D.: easy reading level	Woodcock Word I.D.: failure reading level	Woodcock Passage Comp.: grade score	Woodcock Passage Comp.: easy reading level	Woodcock Passage Comp.: failure reading level
															8-6 8-11	20		7/30	9/10							3.4									
																	7-2	10/30	2/10																
																	7-4	10/30	1/10																
																	7-11	2/30	9/10																
																	7-11	5/30	4/10																
																	7-4	4/30	0/10																
	96								10	9	11	12	6		X	X		4/30	1/10																
																	9-4	0/30	9/10											1.8	1.7	1.7	1.9	1.6	2.2
															X	X	8-7												X						
																	8-7	7/30	9/10																
																	7-11	3/30	1/10																
																	6-8	1/30	0/11																
																	7-4	1/30	0/10																
	60		1	1	2	1			7	1	2	6	3				8-8	1/30	2/10	2.4			2.5							1.6	1.5	1.7	1.6	1.4	1.9
																	8-7	7/30	1/10																
																	7-10	0/30	2/10																
																	8-7	5/30	3/10																
	80								10	1	7	9	8				6-3	1/30	1/10																
																	10-2	16/30	4/10	2.2			2.9			1.9									
																	8-7	12/30	0/10																
																	7-6	5/30	3/10																
	84								6	7	9	8	8		X	X		3/30	0/10	5.0			3.9							5.2	2.9	3.6	2.7	2.2	3.2
																	6-0	2/30	3/10																

DIAGNOSTIC ARITHMETIC			WJ TEST COGNITIVE ABILITY		WOODCOCK-JOHNSON TEST OF ACHIEVEMENT																	
					READING			MATHEMATICS			WRITTEN LANG											
addition	subtraction	multiplication	grade score	age score	grade score	instructional range	age score	grade score	instructional range	age score	grade score	instructional range	age score	DRAW-A-PERSON	PRIMARY SELF-CONCEPT TEST TOTAL SCORE	SIGNIFICANT CONFLICTS	TEACHER RATING SCALE	HAND TEST				
					2.6	3.3 / 3.1	7-8	2.6	2.1 / 3.2	7-9	1.8	1.5 / 3.2	7-2	10-3								
					1.9	1.6 / 2.2	7-1	1.8	1.1 / 3.2	7-0				9-9								
X	X	X			2.3	2.0 / 2.8	7-6	4.2	3.6 / 5.2	9-8	3.0	2.5 / 3.7	8-3	13+			X					
					1.7	1.5 / 2.0	7-0	3.2	2.7 / 3.9	8-5				7-6								
					2.3	2.0 / 2.8	7-6	2.6	2.1 / 3.2	7-9				8-3	14							
					1.5	1.3 / 1.8	6-10	2.2	1.8 / 2.8	7-5	1.0	1.0 / 1.0	4-11	11-3								
					2.8	2.4 / 3.2	8-0	3.5	2.9 / 4.3	8-4	2.3	1.8 / 3.0	7-9	13+		X						
								2.4	1.8 / 2.9	7-6	1.3	1.0+ / 1.5	6-4	9-9								
														11-3								
					2.3	2.0 / 2.8	7-6	2.4	1.8 / 2.9	7-6	1.6	1.4 / 2.0	7-0	10-8								
					1.3	1.2 / 1.6	6-9	1.5	1.3 / 1.9	6-9	1.5	1.2	1.7	8-3	12			1				
					1.8	1.5 / 2.1	7-0	3.2	2.6 / 3.8	8-4	1.7	1.5 / 2.0	7-11	10-0		X		•				
					1.3	1.2 / 1.5	6-8	2.0	1.5 / 2.4	7-2				10-9								
														9-0								
					3.2	2.6 / 4.1	8-4	3.4	2.7 / 4.0	8-7	3.0	2.3 / 3.7	8-2									
					1.4	1.2 / 1.6	6-9	2.2	1.8 / 2.7	7-5	1.7	1.5 / 2.0	7-1	8-0				1				
					1.7	1.5 / 2.0	7-0	2.0	1.6 / 2.8	7-2				8-0								
					2.3	1.9 / 2.7	7-5	1.8	1.4 / 2.3	7-0				7-0				1				
														11-3								
					1.3	1.2 / 1.5	6-8	1.5	1.2 / 2.0	6-9	1.3	1.0 / 1.5	6-4	7-9				1				
					2.7	2.3 / 3.3	7-11	4.7	4.0 / 5.6	10-0				11-2	16							
					2.4	2.1 / 3.0	7-7	2.4	1.9 / 3.2	7-6	1.4	1.1 / 1.6	1.4	9-9		X		1				

Appendix B

Computer Card Coding Sheet

Column	Description	Comments
1-3	Identification number	
4-7	Age (years and months)	
8	Grade	blank = special education placement
9	School (1 = Chuska 2 = Tohatchi)	
10	Sex (1 = male 2 = female)	
	Subtests	
11-18	Bead Paaterns (ages 3-10) credited with highest level obtained 1 = correct 0 = incorrect	Item 11 is credited as one if subject strings eight beads or more
19-37	Memory for Color (ages 3-10) 1 point scored for each correct response 1 = correct 0 = incorrect	
38-44	Picture Identification (ages 3-10) 1 point socred for each picture matched correctly. Three points possible for each item. (0, 1, 2, 3)	
45-58	Picture Association (ages 3-10) Choice of (a, b, c, d) 1 point scored for each series completed correctly.	Some examiners marked the item only correct or incorrect. Therefore, the incorrect distractor choice was not known, and was marked "f" on the computer card.

Column	Description	Comments
16	Completion of Drawings (all ages) One point scored for each drawing completed correctly. (0, 1)	Only the total subtest score is listed in the test protocol. To run the analysis on the items, rescoring of the 27 drawings was required.
44-51	Memory for Digits (ages 11 and above) Two points scored for the exact reproduction of any part of each series. One point if the best performance on any part of the series is the selection of the proper digits but not in the proper order. (0, 1, 2)	
52-58	Puzzle Blocks (ages 11 and above) One point scored if the subject completes the cube within the time limit. One bonus point scored if cube is completed in correct color. One bonus point scored if blocks 1-3 constructed correctly in 30 seconds and blocks 4-7 in 60 seconds. (0, 1, 2, 3)	
59-70	Picture Analogies (ages 11 and above) Choice of (a, b, c, d, e); 1 point scored for each analogy completed correctly.	Some examiners marked the item only correct or incorrect. Therefore, the incorrect distractor choice was not known, and was marked "f" on the computer card.
71-80	Spatial Reasoning (ages 11 and above) Choice of (a, b, c, d); 1 point scored for each series completed correctly.	

Appendix C

Hiskey Nebraska Test of Learning Aptitude

Test Protocol

HISKEY — NEBRASKA TEST OF LEARNING APTITUDE

Record Booklet

Name_____ Sex_____ Date of Test Yr_____ Mo_____ Day_____

Address_____ Date of Birth Yr_____ Mo_____ Day_____

School_____ Grade_____ Age Yr_____ Mo_____ Day_____

Examiner_____ Deaf L.A._____ Hearing M.A._____

Father_____ Occupation_____

Mother_____ Occupation_____

SUMMARY

Subtest	Rating	Subtest	Rating
Bead Patterns	_____	Block Patterns	_____
Memory for Color	_____	Completion of Drawings	_____
Picture Identification	_____	Memory for Digits	_____
Picture Association	_____	Puzzle Blocks	_____
Paper Folding	_____	Picture Analogies	_____
Visual Attention Span	_____	Spatial Reasoning	_____

SUMMARY OF BEHAVIORAL FACTORS

Check the appropriate blank. During the period of observation and testing the subject exhibited the trait:
(1) Frequently, (2) Occasionally, (3) Seldom or never

I. RELATIONSHIP WITH EXAMINER			
Cooperative	1 2 3	Fearful	1 2 3
Self-Reliant	1 2 3	Nervous	1 2 3
Pleasant	1 2 3	Negative	1 2 3
Sense of Humor	1 2 3	Boisterous	1 2 3
Confident	1 2 3	Withdrawn	1 2 3
II. REACTIONS TOWARD TASK			
Motivated	1 2 3	Needs Encouragement	1 2 3
Attentive	1 2 3	Directions must be repeated	1 2 3
Persistent	1 2 3	Impulsive	1 2 3
Purposeful Activity	1 2 3	Reliant on Trial and Error	1 2 3
Assured	1 2 3	Failure is Threatening	1 2 3
Intrinsic Reinforcement	1 2 3	Perseveration	1 2 3
Absorbed in Task	1 2 3	Hyperactive	1 2 3
III. SPEECH AND MOTOR OBSERVATIONS			
Superior Diction	1 2 3	Omissions in Sound	1 2 3
Sentences of Average Length	1 2 3	Substitutions in Pronunciation	1 2 3
Fluent	1 2 3	Distortions in Wording	1 2 3
Skillful in Gross Movements	1 2 3	Maladroit in Fine Movements	1 2 3

OVERALL RATING OF TESTING CONDITIONS

Excellent Good Average Minimal Impairing

BEAD PATTERNS Ages 3-10

Bead Stringing: Number strung in one minute ------

Bead Patterns: Visual I-----II-----III----- (Time limit - 2 minutes
Memory IV-----V-----VI-----VII----- per pattern)
Score 1 point for each pattern completed correctly TOTAL _____

MEMORY FOR COLOR Ages 3-10 Order: Black, Blue, Orange, Yellow, Green, Brown, White, Red

A. Black ----
Yellow ----
Green ----
Brown ----
Orange ----
White ----
Red ----
Blue ----

B. White, Green ----
Black, Yellow ----
Blue, Orange ----

C. Brown, Yellow, Blue ----
White, Red, Green ----

D. Red, Orange, Blue, White ----
Brown, Yellow, Black, Green ----

E. Red, Blue, Orange, Yellow, Black ----
Brown, Green, Black, Blue, White ----

F. Green, Orange, Brown, Blue, Red, White ----
Black, Yellow, Red, Brown, Green, Orange ----

Score 1 point for each correct response (color only) TOTAL _____

PICTURE IDENTIFICATION Ages 3-10

	Series Score		Series Score
1. a-b-c	----------	5. a-b-c	----------
2. a-b-c	----------	6. a-b-c	----------
3. a-b-c	----------	7. a-b-c	----------
4. a-b-c	----------		

Score 1 point for each picture matched correctly TOTAL _____

PICTURE ASSOCIATION Ages 3-10

1. a-b-c-d 4. a-b-c-d 7. a-b-c-d 10. a-b-c-d 13. a-b-c-d
2. a-b-c-d 5. a-b-c-d 8. a-b-c-d 11. a-b-c-d 14. a-b-c-d
3. a-b-c-d 6. a-b-c-d 9. a-b-c-d 12. a-b-c-d

Score 1 point for each series completed correctly TOTAL _____

PAPER FOLDING Ages 3-10

1. -----
2. -----
3. -----
4. -----
5. -----
6. -----
7. -----
8. -----
9. -----

Score 1 point for each pattern completed TOTAL _____

VISUAL ATTENTION SPAN All ages

1a ----- 2a ----- 3 ----- 5 ----- 7 -----
1b ----- 2b ----- 4 ----- 6 -----

Score 1 point for a correct selection in 1a and 1b. For the remaining series score 2 points for each series in correct order and 1 point if the correct selections are made but not in proper order. Discontinue if 2 successive series are failed completely. TOTAL _____

BLOCK PATTERNS All Ages

	Time	Score		Time	Score
1.		-----	X.	(Demonstration)	
2.		-----	8.	----	-----
3.		-----	9.	----	-----
4.		-----	10.	----	-----
(Give subject remaining blocks)			11.	----	-----
5.		-----	12.	----	-----
6.	----	-----	13.	----	-----
7.	----	-----	14.	----	-----

Score 3 points for the correct reproduction of designs 6-14 in 30 seconds or less; 2 points in 31-60 seconds, and 1 point for all others completed. Discontinue if the subject fails three consecutive patterns. TOTAL _____

COMPLETION OF DRAWINGS All ages

Score 1 point for each drawing completed correctly TOTAL _____

MEMORY FOR DIGITS Ages 11 and above

	Series Score		Series Score
A1---A2---A3---	----------	E1---E2---E3---	----------
B1---B2---B3---	----------	F1---F2---F3---	----------
C1---C2---C3---	----------	G1---G2---G3---	----------
D1---D2---D3---	----------	H1---H2---H3---	----------

Score 2 points for the exact reproduction of <u>any part</u> of each series. Give 1 point if the best performance on any part of the series (as B1, B2, or B3) is the selection of the proper digits but not in the proper order. Discontinue after 2 successive levels (for example D & E) have been failed <u>completely</u> TOTAL _____

PUZZLE BLOCKS Ages 11 and above

	Time	Score	Limit		Time	Score	Limit
1.	----	-----	2 min.	4..	----	-----	4 min.
2.	----	-----	2 min.	5.	----	-----	4 min.
3.	----	-----	2 min.	6.	----	-----	4 min.
				7.	----	-----	4 min.

Score 1 point if the subject completes a cube within the time limit. Give 1 bonus point each for cubes completed with correct color. Give an additional bonus point (each) if blocks 1-3 are constructed correctly in 30 seconds and blocks 4-7 in 60 seconds. Maximum score is 3 points per cube. TOTAL _____

PICTURE ANALOGIES Ages 11 and above

1. a-<u>b</u>-c-d-e	4. a-b-c-<u>d</u>-e	7. a-b-<u>c</u>-d-e	10. <u>a</u>-b-c-d-e
2. a-b-c-<u>d</u>-e	5. a-b-c-<u>d</u>-<u>e</u>	8. a-b-<u>c</u>-d-<u>e</u>	11. <u>a</u>-b-<u>c</u>-d-e
3. a-<u>b</u>-c-d-e	6. a-b-c-<u>d</u>-<u>e</u>	9. a-b-<u>c</u>-d-<u>e</u>	12. <u>a</u>-b-<u>c</u>-d-e

Score 1 point for each analogy completed correctly TOTAL _____

SPATIAL REASONING Ages 11 and above

1. a-b-c-<u>d</u>	3. <u>a</u>-b-c-d	5. a-b-<u>c</u>-d	7. a-b-<u>c</u>-d	9. a-b-c-<u>d</u>
2. a-<u>b</u>-c-<u>d</u>	4. <u>a</u>-b-<u>c</u>-d	6. <u>a</u>-b-c-d	8. a-b-c-<u>d</u>	10. a-<u>b</u>-c-<u>d</u>

Score 1 point for each series completed correctly. TOTAL _____

NORMS FOR DEAF CHILDREN

Learning Age	3-0	3-6	4-0	4-6	5-0	5-6	6-0	6-6	7-0	7-6	8-0	8-6	9-0	9-6	10-0	10-6	11-0	11-6	12-0	12-6	13-0	13-6	14-0	14-6	15-0	15-6	16-0	16-6	17-0	17-6	18-0	18-6
Bead Pattern	4	5	6	7	I	II		III			IV					V				VI												
Memory for Color	4-5	6-7	8	9	10	11		12		13		14			15			16				17										
Picture Identif.	5	6-7	8-9	10-12	13-14	15		16	17		18		19			20				21												
Picture Association	2	3	4	5		6	7	8	9	10		11			12				13												
Paper Folding	1	2		3	4	5		6					7					8				9									
Visual Att. Span	1	2	3	4		5	6	7			8				9				10						11				12		13
Block Patterns	1		2		3		4		5	6	7	8	9		10	11	12	13		14	15		16	17		18		19			20	
Compl. of Drawing			1	2	3-5	6-8	9-10	11-12	13-14	15	16	17	18		19	20	21		22			23				24				25		26
Memory for Digits										4	5		6		7	8			9			10			11			12		13		14
Puzzle Blocks									2		3	4		5		6	7		8		9		10			11			12		13	
Picture Analogies									4		5			6		7			8			9				10			11			12
Spatial Reasoning										2		3		4			5			6				7			8			9		

NORMS FOR HEARING CHILDREN

Mental Age	3-0	3-6	4-0	4-6	5-0	5-6	6-0	6-6	7-0	7-6	8-0	8-6	9-0	9-6	10-0	10-6	11-0	11-6	12-0	12-6	13-0	13-6	14-0	14-6	15-0	15-6	16-0	16-6	17-0	17-6	18-0	18-6
Bead Pattern	4-5	6	7	I	II		III			IV						V					VI											
Memory for Color	5-6	7-8	9	10		11		12	13		14		15		16			17			18											
Picture Identif.	8-9	10-11	12-13	14	15	16		17		18		19			20				21													
Picture Association	4-5	6		7	8	9	10		11			12				13				14												
Paper Folding	1	2	3	4		5			6				7					8				9										
Visual Att. Span	1-2	3	4		5	6	7		8			9			10				11					12					13		14
Block Patterns	1	2		3		4		5	6		7		8	9		10	11	12	13		14	15		16	17		18		19			20
Compl. of Drawing	2	3	4-5	6-8	9-10	11-12	13	14-15	16	17	18		19		20		21		22		23				24					25		26
Memory for Digits									8		9			10			11			12			13			14	15		16			
Puzzle Blocks										2	3		4	5	6	7		8		9			10			11				12		
Picture Analogies										6	7		8		9				10				11					12				
Spatial Reasoning									2		3				4			5			6					7				8		9

COMPLETION OF DRAWINGS