USING A MULTI-DIMENSIONAL ASSESSMENT BATTERY TO SCREEN
FOR LEARNING PROBLEMS: AN EVALUATION STUDY IN A SAMPLE OF
CANADIAN NATIVE INDIAN STUDENTS

by

BRIAN REID

B.A., York University, 1975

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Department of Educational Psychology and Special Education)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1989

© BRIAN REID, 1989
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psychology

The University of British Columbia
Vancouver, Canada

Date April 20, 1989
ABSTRACT

This thesis is an analysis of test scores from a multi-dimensional assessment of Canadian Native Indian students attending an elementary school on a reserve in British Columbia. The intention of the assessment was to determine the incidence of learning problems among the students, and the special educational assistance required.

The testing instruments used included the Metropolitan Reading Readiness Test; the Developmental Test of Visual-Motor Integration; the Peabody Picture Vocabulary Test; the Canadian Test of Basic Skills; and the Canadian Cognitive Abilities Test. Two tests of perceptual acuity were also administered. The single administration of the tests was designed to locate the level of achievement attained by the students and compare this attainment with age and grade placement at time of testing. The intention of the thesis was to determine the appropriateness of the battery of tests for this sample of Native Indian students.

Disparities were found between placement and achievement with evidence of increasing spread in the upper grades. The average difference was approximately one year in grade 3, rising to two or more years in grades 5 and 6.

The conclusion reached was that the instruments were useful in identifying the areas and extent of difference between the sample and the population. Specifically, vocabulary knowledge was low. Incidence of vision and hearing impairment was high; 40% of the students were found to have vision problems, and 21% were diagnosed as having hearing difficulty, 4 times the national average.
# TABLE OF CONTENTS

LIST OF TABLES ......................................................................................................................... vi

LIST OF FIGURES ....................................................................................................................... vii

ACKNOWLEDGEMENTS ............................................................................................................... viii

Chapter I. INTRODUCTION ........................................................................................................ 1
   A. Overview ......................................................................................................................... 1
   B. Purpose of the Thesis ................................................................................................. 5
   C. Research Questions ..................................................................................................... 7
   D. Definitions of Terms Used in the Thesis ................................................................... 7
   E. Delimitation of the Study .......................................................................................... 8
   F. Justification of the Thesis .......................................................................................... 8
   G. Organization of the Thesis ........................................................................................ 10

Chapter II. REVIEW OF LITERATURE ...................................................................................... 11
   A. Considerations in Testing Native Indian Students ...................................................... 11
      1. Performance Decline of Canadian Indian Pupils .................................................. 12
      2. Intelligent Adaptation to the School Environment .............................................. 13
      3. Adaptation to Hostile Environments ...................................................................... 15
      4. Verbal and Non-Verbal Differences in Learning .................................................. 17
      5. Indian Education in Saskatchewan (Age-Grade Retardation) ......................... 19
      6. Factors Affecting District Achievement ................................................................ 20
      7. Indian Education: Lillooet ...................................................................................... 21
      8. Indian Education: Okanagan-Nicola ...................................................................... 23
      9. Quality of Indian Education in Canada ................................................................... 23
     10. Teaching Cognitive Skills ....................................................................................... 24
     11. Tests of Intelligence ............................................................................................... 27
     12. Test Frustration ....................................................................................................... 28
     13. Review of Indian Intelligence Research .................................................................. 30
     14. Cultural Bias in Tests ............................................................................................. 32
     15. Testing Native Indian Students ............................................................................ 33
   B. Summary of Chapter ..................................................................................................... 34

Chapter III. METHODOLOGY ..................................................................................................... 37
   A. Methodology of the Survey of Learning Problems .................................................... 37
      1. Sample ...................................................................................................................... 37
      2. Instruments ............................................................................................................... 37
      3. Testing Procedures ................................................................................................... 43
      4. Assessment Team Data Analysis ............................................................................ 43
   B. Thesis Methodology ................................................................................................. 44
      1. Judgment Model ....................................................................................................... 44
      2. Thesis Data Analysis .............................................................................................. 44
   C. Chapter Summary ...................................................................................................... 50

Chapter IV. RESULTS ................................................................................................................ 51
   A. Results of the Survey of Learning Problems ............................................................. 51
1. Sample Scores on the Battery of Tests

2. Vision and Hearing Problems

B. Results of the Thesis

1. Difference Between Sample Scores and the Norming Population

2. Score Decrements in Higher Grades

3. Age-Grade Retardation

4. Ability Level Differences in Achievement

5. Achievement of Higher Ability Students

6. Language Scores Relative to Other CTBS Components

7. Effects of Perceptual Problems on Test Scores

8. PPVT-R Item Difficulty Rankings for the Sample

9. Underrepresentation of Canadian Native Indians in the Norm Reference Groups

10. Potential Examiner and/or Language Bias

C. Chapter Summary

Chapter V. CONCLUSIONS

A. Summary of the Thesis

1. Purpose

2. Methodology

B. Judgments

1. Does the Battery Achieve Its Goal of Identifying Children Who are in Need of Educational Assistance?

2. Does the Battery Determine the Type and Magnitude of Assistance Required?

3. Does Use of a Multi-Dimensional Battery Lead To Biased Assessment when used in such a sample?

C. Evaluation of Appropriateness

D. Generalizability

E. Limitations of the Study

F. Implications for Future Studies

REFERENCES


APPENDIX 2: SPLIT-HALF RELIABILITY COEFFICIENTS FOR FORM L OF THE PPVT, BY AGE

APPENDIX 3: ALTERNATE FORMS RELIABILITY COEFFICIENTS FOR PPVT BASED ON RAW SCORES OF IMMEDIATE RETEST SAMPLE, BY GRADE

APPENDIX 4: CTBS Internal-Consistency Reliability Coefficients (KR-20)
LIST OF TABLES

Table 1: Distribution of Sample by Gender and Grade ............................................. 38
Table 2: Age of NIRS Students by Year and Month .................................................. 38
Table 3: Differences in Mean Scores Between Grade Placement and CTBS
  Grade Equivalent Scores ......................................................................................... 53
Table 4: Student Frequencies in each CCAT Ability Category .................................. 54
Table 5: Number and Percentage of Students With VMI Scores Below
  Chronological Age .................................................................................................. 56
Table 6: Number and Percentage of Students With PPVT-R Scores One
  Year Below Chronological Age ............................................................................... 57
Table 7: CTBS Achievement of Students with Average or High Average
  CCAT Ability Ratings ............................................................................................. 82
Table 8: Number and Percentages of Students with Perceptual Problems
  Attaining Below Expected on the VMI PPVT-R AND CTBS ................................. 84
Table 9: PPVT-R Form L Item Difficulty Indices for Items 41 to 100 ................. 88
Table 10: PPVT-R Item Difficulty Indices for Items 41 to 70 by Grade ............... 90
LIST OF FIGURES

Figure 1: Judgment Model for Evaluating the Appropriateness of the Standardized Tests .......................................................... 45

Figure 2: Confidence Intervals for the Norming Population and the NIRS Sample on the MRT .......................................................... 60

Figure 3: Confidence Intervals for the Norming Population and the NIRS Sample, Grades K-7, on the PPVT-R ........................................... 61

Figure 4: Confidence Intervals for the Norming Population and the NIRS Sample on the CTBS .......................................................... 64

Figure 5: Confidence Intervals for the Norming Population and the NIRS Sample on the CCAT, Grades 1 & 2 ........................................... 67

Figure 6: Confidence Intervals for the Norming Population and the NIRS Sample on the CCAT, Grades 3-7 ........................................... 68

Figure 7: Percentages of Students Attaining Achievement Scores Commensurate with Chronological Age on the PPVT-R ....................... 73

Figure 8: Percentages of Students Attaining Achievement Scores Commensurate with Chronological Age on the VMI ............................. 74

Figure 9: Percentages of Students Attaining Achievement Scores Commensurate with Grade Placement on the CTBS ............................. 75

Figure 10: Percentages of Students 1 Year Below Age or Grade Level on the VMI, PPVT-R, and CTBS for each CCAT Ability Category .......... 80
ACKNOWLEDGEMENTS

I would like to thank the members of the thesis committee for their assistance in developing and completing this project. The committee consisted of Drs. Robert Conry, Julianne Conry, and Art More. Dr. Julianne Conry provided the expertise in the technical and practical aspects of understanding and interpreting test scores in the special case where the sample consisted of minority students. Dr. Art More's empathy for the problems of Native peoples helped to make this thesis a more humane analysis of individual differences in a specific educational setting. The committee was chaired by Dr. Robert Conry who served as advisor, organizer, advocate and friend, whose personal support I deeply appreciate. Thanks also are deserving to Sharon Reid, my wife, both for her support, and for her cogent comments.
CHAPTER I. INTRODUCTION

A. OVERVIEW

The failure of schooling among Canadian Native Indian children is spectacular (McLeod, 1984). It is estimated that 85% of these students leave school before completing grade 12 (More, 1984a). This dropout rate is a result of low achievement and is evidence of the failure to meet adequately the educational needs of these children. Thomas et al (1979) found average achievement levels to be two or more years behind grade placement. In the Vancouver School District an achievement level two years below grade placement is the criterion used in identifying special needs children. Once identified these children are placed in special education classes and given enhanced individual instruction in the specific content areas where achievement is below expected levels. This process of remediating achievement deficits begins with the identification of students experiencing difficulty. Therefore, accurate assessment is the foundation of determining individual needs from which plans can be generated for more responsive and successful educational programs in schools.

Typically, school educational programs are devised for the entire student population including individuals from all differing ethnic minority groups. This is also true of devices for assessment of learning potential and achievement. However, the appropriateness of using standardized tests devised for the general population may be questioned when assessment must be made of a specific and distinctive group of students that constitutes a small sample of the population (Cole & Bruner, 1971).
These broad-based assessment devices are implemented with varying degrees of success, depending on the familiarity that the testing subjects have with the cultural components of the test. A special case is the administration of tests to samples of Canadian Native Indian groups. For the most part, these children have English as a first language; but their dialects are often not standard, and their cultural heritage is certainly different from that of the dominant North American culture. Nevertheless the educational needs of these children must be understood in order to interpret results fairly. To do so, care should be taken to evaluate the cultural relevance of test materials to the background of Canadian Native Indians.

Assessment devices can be related to the content of the instructional program followed, but it is much more difficult to find tests that have been shown by researchers to be culturally-fair for all racial groups. One assessment technique is the administration of several different tests, which has two benefits: first, to provide more complete and accurate results and second, to diminish the effect of the bias that may be inherent in an individual test. Multiple assessment is defined as the measure of individual standing in each of a number of traits (Anastasi, 1982). The logic behind such multiple assessment is that sampling a number of aspects of achievement and ability will provide the most accurate picture of student intellect. There is often some degree of overlap among dimensions assessed by the various tests, but the tests’ results will then provide mutual corroboration. More importantly, multiple assessment is less likely to lead to mislabelling of children -- an important consideration with any group, but especially when dealing with children from an ethnic minority whose culture varies considerably from the norming sample.
Despite possible cultural differences, assessment devices designed for the general population may have the advantage of giving clues to any ways that achievement of ethnic minority students differs from the rest of the population. Such differences would be of interest to educators to aid in curriculum design and in the planning of Individualized Education Programs (IEP's). Variation in test scores has been shown to be related to ethnic background and socioeconomic status. For example, it has been demonstrated that Caucasian students show higher levels of achievement in statewide studies of minimum standards of basic skills in Massachusetts (Massachusetts Department of Education, 1983). An additional finding was that the percentage of students meeting minimum standards was greatest in the residential suburbs, and lowest in the poorest communities. For students who are neither white nor residents of suburbs, lower levels of achievement might be expected in comparison to nation-wide assessments of minimum standards of basic skills. In the Massachusetts study, achievement levels of the ethnic minorities were found to be greatest at the lowest elementary grades.

Johnstone (1981) identified vocabulary as a key component of achievement related to ethnic group differences. These findings were based on a structural analysis of a cross-cultural administration of the Iowa Test of Basic Skills in grades three through eight. Weaknesses in any content area were found to contribute to lowering scores in other areas, and a conclusion was drawn that student achievement is a set of interdependent components. A main constituent of achievement scores is the vocabulary component; where there is weakness in vocabulary, other scores may be affected. Special attention should be paid to tests of receptive vocabulary such as the Peabody Picture Vocabulary Test-Revised
(PPVT-R); (Dunn & Dunn, 1981) for evidence that vocabulary has a disproportionate role in test scores.

Another factor that may influence test performance is physiology. Physical limitations may have serious implications in the testing situation and in the learning environment. Students must demonstrate the intellectual skills they have acquired by applying sequences of concepts to different classes of situations, as represented by test questions (Gagne, 1984). Not only will students with perceptual problems fail to demonstrate acquired intellectual skills, but also their ability to learn such skills may be seriously impaired. If students are unable to decode the auditory and visual stimuli impinging upon them, they are unlikely to have much success dealing with the concepts embodied in a curriculum.

The problem of perceptual difficulties is acute in samples of Native Indian children. McShane and Plas (1982) found an unusually high incidence of middle ear disease. The percentage of children with loss of hearing due to middle ear disease ranged from 20 to 76%. This compares with an incidence of 20 to 25% of the children from lower socioecomic groups and 5% of the general population. This lead to the conclusion that the American Indian is much more likely to have otitis media than other children and that there is a possibility of mis-labelling a child as learning disabled, when in fact, he has a deficit resulting from middle ear disease. Physiological problems should be considered in testing Native children to determine the possibilty of perceptual impairment affecting performance.

In May and June of 1986, an evaluation with two foci was conducted at a Native Indian reserve school in the province of British Columbia. The first focus was to describe the school and its operations in order to provide
information about changes that could be made to better serve the educational needs of the children. The second was to assess individual achievement and ability of the students in the school, which included kindergarten to grade 7. The assessment results were to be used for educational planning: specifically, for the creation of Individualized Educational Programs (IEP's). A series of standardized tests was administered to provide an accurate assessment of individual differences in ability and achievement. Another component of the multiple assessment was screening for vision and hearing difficulties. The comprehensive screening was designed to identify those children in the sample in need of further observation and assessment. The concern was that an accurate assessment should take place. However, when testing children who are members of an ethnic minority group it is sometimes difficult to gauge the appropriateness of the assessment instruments used. This thesis is based on the data collected from the evaluation project. The information from the test publishers and the data available from the testing of NIRS students will be examined in this thesis. The point will be to first make comparisons between local results and national norms, and second to draw an inference as to whether these tests were indeed appropriate for the sample. Without a determination of appropriateness, the validity of any interpretation of results may be called into question.

B. PURPOSE OF THE THESIS

Whenever assessment of a person is to be undertaken, consideration of the nature of the individual, and of the limitations of test(s) to be used is required. Many commonly used tests have been normed with standardization samples stratified according to the composition of the ethnic components of the population.
For instance, if Hispanics constitute ten per cent of the population, then 10 per cent of the standardization sample would be Hispanic. However, the fact that some ethnic minority groups tend to score lower than others has been taken as evidence that the tests are inappropriate for the lower-scoring samples (Mercer, 1972). Given this perspective, psychologists, when administering tests to ethnic minority groups, must concern themselves with appropriateness of the test(s) for the individuals currently being assessed.

It is difficult to make decisions about which test to use, when it is appropriate and whether the amount of research is sufficient to justify such decisions. Most research on testing Indians in Canada has been hampered by methodological and theoretical problems such as the Pan-Indianism assumption, that members of all nations will demonstrate similar characteristics. In many cases the sample sizes have been small and improper tests have been used. The tests have not been shown to measure the same ability or aptitude with the same measurement units across ethnic groups. Lastly, there is a conspicuous absence of an overarching theory of Indian intelligence upon which appropriate testing may be founded (Chrisjohn & Lanigan, 1984). Appropriateness has been demonstrated for some ethnic minorities, but information regarding Canadian Native Indians is limited.

The goal of this thesis is to evaluate the appropriateness of a multiple assessment battery administered to a particular ethnic minority group of Canadian Native Indian students. The major research question of the evaluation study is:

Are the general purpose instruments used in a multi-dimensional battery, devised for screening learning problems, appropriate to a uniracial ethnic minority group of Native Indians?
To elaborate upon this question, two secondary questions and several subsidiary questions are necessary. The secondary questions are listed below and the subsidiary questions are outlined in the Judgment Model in Chapter 3.

C. RESEARCH QUESTIONS

Appropriateness will be examined by means of the three research questions, each of which deals with an important aspect of the multi-dimensional assessment. An appropriate test is one capable of identifying students in need, revealing the type and magnitude of assistance required, in an unbiased assessment. Therefore, the three research questions are:

1. Does the battery achieve its goal of identifying children who are in need of special educational assistance?
2. Do results of the battery reveal the type and magnitude of assistance required?
3. Does use of a multi-dimensional battery lead to biased assessment when used in such a sample?

D. DEFINITIONS OF TERMS USED IN THE THESIS

The school attended by the children studied in this thesis was located on a reserve for Canadian Native Indians in the province of British Columbia. For simplicity, this Native Indian School will be identified by an acronym, NIRS.

The students themselves were members of four local Indian bands representing an ethnic minority group (EMG) that comprises a portion of the population in British Columbia.

The fact that these students have a relatively common heritage, and that they reside within a small geographical area on nearby reserves, means that
they live and learn in an environment essentially separate from the larger multicultural population in Canada. They live less than 30 kilometres from a major urban centre, so they are not isolated from urban Canada. The language of communication is English although the linguistic heritage of these people is that they are part of the Salishan language family (Foster, 1982). The fact that they currently reside in reserves means that most children are status Indians living in a homogeneous racial community. Because of the common race and the fact that these children are members of an ethnic minority, in this thesis they will be referred to as a uniracial ethnic minority sample (EMG).

E. DELIMITATION OF THE STUDY

The orientation of this thesis is not identical to the purpose of the educational evaluation, which was conducted under contract to the NIRS board of trustees to provide information relevant to their decisions about staffing, facilities expansion, and program goals. Analysis in this thesis focusses on the assessment of individual differences among the students, and not on the program evaluation aspects of the contracted evaluation. The issue here, then, is a particular measurement consideration: the adequacy of the instruments for assessing the ability and achievement of the EMG students.

F. JUSTIFICATION OF THE THESIS

Few standardized tests have been tailored for use with specific minority ethnic populations. It is unlikely that a test designed for use in one EMG would perform well for others. This caveat extends to instances other than differences between visible minorities, such as Blacks and Hispanics. A major criticism of
Indian studies is the assumption that findings from the study of one Indian
nation may be validly generalized to another nation, or to all Indian nations.
There is no evidence of a single Native Indian cognitive structure, rather the
information that is available leads to the inference of substantial differences
across nations (Chrisjohn & Lanigan, 1984). Therefore, valid conclusions drawn
from the results of one Indian group may not transfer to another. Pan-Indianism
is not a fact. Psychologists are commonly faced with choosing the best possible
adaptation of standardized tests for a particular group. This thesis will attempt
to ascertain the appropriateness of a battery of tests for use with the pupils at
NIRS. The appropriateness of tests has ramifications for the development of
individualized instructional programs. The performance of the tests will be
assessed and recommendations made. Since the tests are designed to show levels
of achievement and intellectual ability, any that function well for this sample
may also be beneficial to others involved in planning curriculum and designing
IEP’s. Because of the limited generalizability between national and linguistic
groups, it will be necessary for examiners to compare the characteristics of the
sample used here with those of others for whom testing is planned. Test
administrators will need to decide whether they should use the same group of
tests or change the assessment by deleting or replacing one or more. The
process of assessing individual tests will help in other testing situations by
providing a framework which will permit the evaluation of the testing materials
to be used.

A possible additional benefit may be the use of an existing set of
standardized tests to provide reliable and valid assessments of the achievement
and intellectual ability of Canadian Native Indian students. This information would
be useful to the decision-makers within the administrations of both Native and mixed school districts. Trustees are responsible for adopting policies that reflect the nature and needs of the students within their districts. From policies, programs can be devised and implemented by the allocation of facilities, finances, and personnel. However, in order to make good policy decisions, administrators need accurate and comprehensive information of the special needs of the students in their charge. If this thesis can provide evidence that a set of tests will provide reliable student information, then decision-makers will have been provided with a valuable assessment tool to aid program development and achievement of the band's goals. By accurately assessing the initial state of the learner, the educator can know of any disparity that exists between actual achievement levels and those intended by district policy. Then, it will be possible to program education in order to successfully achieve the goal of meeting the educational needs of these children.

"Policy implementation ... involves the development of programs and procedures designed to guide the behavior of staff and students so that the intent of the policy is realized. It is through the process of implementation therefore, that the school experience can be tailored to provide pertinent program offerings and optimal learning conditions for Native students." (Wilcox, 1984, p.1).

G. ORGANIZATION OF THE THESIS

The balance of this thesis is presented as four chapters. Chapter 2 reviews relevant literature, and in Chapter 3 a "matrix of criteria" is developed and presented. Descriptions of the sample and instruments are also presented in Chapter 3, along with the testing and data analysis procedures used in the evaluation. The findings of this thesis are outlined in Chapter 4 and, finally, conclusions and recommendations are presented in Chapter 5.
CHAPTER II. REVIEW OF LITERATURE

Sternberg (1985) writes of the importance of the "contextual framework for understanding intelligence" and information processing. Schooling defines the general context of learning as the school classroom, and standardized tests of ability and achievement are basic assessment devices for that learning. Contexts vary, partly because of a multitude of ethnographic factors. Cultural factors may play a disproportionate role in the test scores of Indian children in school, and effect our understanding of the intelligence of these students.

A number of studies have discussed the problems of educating native children and evaluating the difficulties inherent in using testing materials as the basis for prescribing educational programs. In this review, 43 articles are cited of which 14 are discussed in depth. Ten articles are concerned specifically with the problems of testing Indian children. Three of the studies report research conducted with Indians in British Columbia.

A. CONSIDERATIONS IN TESTING NATIVE INDIAN STUDENTS

Considerations in testing Native Indian children are discussed under 15 sub-headings in this chapter. The first topic is the decline in performance shown by Canadian Native Indians in school, which is followed by a general discussion of understanding human intelligence in the school environment. Indian adaptation to schooling is dealt with in the third section. A component of intelligence is verbal styles of learning, the importance of which is presented in the fourth section. The specific problems of Indians as identified in the literature are the topics of the next five sections. One suggested solution to some educational problems, the teaching of cognitive skills, is described in the ninth section.
Sections following present topics from the literature such as the results of intelligence testing for Indian children, and the difficulties of using standardized tests with ethnic minorities. A problem of particular interest is the possibility of bias due to culture-related content. The final section will integrate the previous topics and relate them to the topic of testing an Native Indian children.

1. Performance Decline of Canadian Indian Pupils

In the Mackenzie District Norming Study, MacArthur (1968) summarized evidence from a number of construct validity studies of "culture-reduced" intellectual ability measures administered to 792 Indian-Metis and 510 white pupils in the Canadian West and North. Five measures of scholastic aptitude and school achievement were administered including: Progressive Matrices (Colored and Standard); Safran Culture-Reduced Intelligence Test Scale 1; IPAT Cattell; Lorge-Thorndike Non-Verbal Intelligence Test; and the California Short-Form Test of Mental Maturity. A conclusion noted then was that "large proportions of Canadian native pupils of early school age have the general intellectual ability to participate fully in the larger Canadian community" (MacArthur, 1968). On the Progressive Matrices many native students (seven- and nine-years old) performed better than the mean for that of their white cohorts. MacArthur based this conclusion on a comparison of the distribution of scores for the two groups. The 90th percentile score for seven-year-old Indian-Metis students was higher than the 50th percentile score for whites (25 vs 20.5). The comparable figures for nine-year-olds were 31 and 26. Therefore, at both ages there were native students who scored better than half of the white students at the same age. This is clear evidence of considerable intellectual potential. However, results from
the same studies show a decline in the test performance of Canadian Native Indian children at ages eleven and thirteen, where MacArthur reported that few scored as well as average white children. In the intervening two decades tests have changed greatly, and have been adapted for use with different racial groups. The question is whether a decline in performance is still to be found for Indian students. The challenge for educators today is to find appropriate tests to measure native ability and to devise appropriate instruction to avoid the performance decline noted by MacArthur in 1968.

2. Intelligent Adaptation to the School Environment

Sternberg (1985) posits a three-factor model for intelligent mental activity, including purposive adaptation to the real world, selection of the aspects of the environment relevant to one's own life, and shaping these aspects to suit one's abilities. Elementary school students are rarely in the position to select or shape their school environments to personal advantage; rather, their main role is adaptation. Native Indian children must adapt to a school environment that is rarely congruent with the cultural environment to which they have been accustomed. There is a distance between the culture of the students and that which is reflected by the school environment. This distance constitutes a cultural lacuna with which the students must then cope in addition to the problems faced by all learners. The mental activity required to adapt intelligently to the school environment involves translating, rehearsing, and assembling new information, while stimulating the relevant aspects of cognition and monitoring the acquisition
process (Winne, 1985).

Cultural discontinuity aggravates the difficulties inherent in adapting to the environment. The nature of student adaptations become what Simon (1973) has termed "satisficing" which is the selection of a choice of action that is barely satisfactory. The situation that satisfices is not necessarily the one that maximizes the benefits of the learning outcomes, but is the one that is simpler, and less costly in terms of time, training, and cognition (Simon, 1973).

Such cultural discontinuity is a constraint on learning for Indian children not faced by white children in the schools. In North America the dominant culture is caucasian, and this is expressed by the school ecology, in aspects ranging from the curriculum to the race of the teachers. Often teachers are white even in otherwise uniracial reserve schools. Native students must also cope with a shift from essentially non-verbal developmental backgrounds to a verbal environment. Their lack of verbal ability may have an effect on performance in the classroom and also on ability measures that rely on verbal components.

Nonverbal childrearing styles, coupled with nonoral visual modality preferences, may reduce the Indian child's accumulation of standard English language content and efficient non-Indian linguistic capabilities. An emphasis upon social and emotional stimuli and relationships rather than upon transfer of abstract information and factual knowledge may be reinforced through traditional cultural values, low socioeconomic status, and lack of school success. In addition, an extremely high incidence of otitis media and consequent hearing loss in Indian groups may affect psycholinguistic development, reducing auditory verbal reception, association, sequencing, and expressive abilities (McShane & Plas, 1984).

The cultural discontinuity affects whole categories of intellectual components and may tend to depress performance on academic tasks, especially those sampled by standardized tests as being representative of 'primary' mental abilities (Sternberg, 1985).
3. Adaptation to Hostile Environments

"Indian students, on average do not achieve the same degree of success as non-Indians, whether attending all Indian or integrated schools" (Chrisjohn, Towson, & Peters, 1987, p. 2). Unsuccessful performance does not constitute evidence of a deficiency on the part of Native Indians, but results from a reaction of the students to the schooling environment. Chrisjohn et al suggest three contributing factors: historical effects, social interpersonal effects, and mechanical and systematic effects.

Schooling reflects a continuation of the historical conflict between whites and Native Indians in society. Although schools are now more likely to be band-controlled, the fact that teachers are white and operate in a school system developed in the dominant culture means that Indian students are confronted still with a situation not of their making. This historical conflict may be unwittingly exacerbated by the white teachers who see their role as one of aiding the students in a process of assimilation into white society. A negative response by students to schooling may actually reflect a desire to conform to parental attitudes. Many Indian adults have acquired negative attitudes based on their schooling, and their subsequent experience may have lead to the belief that school has not provided them with any benefits. Parental pressure may encourage hostile reactions by students to schooling, a process sometimes seen as making children less Indian.

Social and interpersonal relations in the school environment tend to reinforce parental attitudes. Similarly interaction between Indian students and white teachers has a negative effect since education is thus seen as a white process, with very few Indian teachers, and many of these in support roles. The
structure of schooling may contribute to what Chrisjohn et al have termed "peer labelling". When a teacher praises a student, the intention is to provide reward and motivation. However, individual success is inconsistent with Indian cultural beliefs that eschew humiliation of persons. The student praised is influenced by his peers who remind him that individual exaltation is not acceptable behavior for an Indian. Since most teachers are white, they are not likely to be aware of the Native Indian attitude towards praise.

The schooling system has mechanical factors that are at worst anti-Indian. These factors which are structural to the system and thus rarely challenged by teachers include curriculum, classroom organization, school hours, public evaluation, and competitive lessons. None of these structures were established by Native Indians. They are part of a system built by and for whites, which has been imposed on Indian students wholesale. That the results have been negative should not be surprising. Standardized tests are justifiably used for designing individualized educational programs for white students. The results of these tests are also often used to design programs for Indians "in the complete absence of information on whether the tests used are in anyway applicable to Indians" (Chrisjohn, Towson, & Peters, 1987, p. 5).

The historical, social and mechanistic effects of schooling outlined pinpoint the wide range of problems faced by Indian children when they attempt to adapt to schooling. These conclusions were drawn based on the study of the Kainai band in Southern Alberta, but the problems in Indian education are similar elsewhere. In other places where the Indians have less than complete control of education, negative attitudes may occur, and school may seem to be a hostile environment.
4. Verbal and Non-Verbal Differences in Learning

Greenbaum (1985) investigated the differences in verbal styles between American Indian (Choctaw) and "Anglo" elementary classrooms to determine whether systematic differences in communication behaviors constituted an unfamiliar conversational etiquette for Indians. His inference, which was that interference theory could explain some of the problems Indians face in school, rested on two assumptions: that differences in non-verbal regulation of conversation do exist between Indians and non-Indians, and that these differences cause problems in face-to-face communication and classroom learning. The typical classroom taught by a non-Indian is directed by a "switchboard participation structure" (Philips, 1972). This system, in which the teacher directs who talks and when, was found to be wholly at odds with the communication etiquette of Warm Springs Indians which permits anyone to take part. Thus, the simple fact that conversation is controlled was found to be a rule that is both unfamiliar and contradictory to Indian students.

Greenbaum compared Indian and non-Indian classrooms, both run according to switchboard participation structure. Behaviors were observed in both classes and significant differences were found. These differences showed the reservation students as less likely to speak in class than their non-Indian counterparts. Student utterances by reservation students were shorter but only for individual responses. Choral response, which requires the entire class to respond to a question, was found to be much more frequent in the Indian school structure than in the white school. Greenbaum concluded that reduced duration and frequency of speech by Indians is indicative of an attempt to avoid individual participation, and its attendant possibilities for embarassment. Teachers come to
expect, and to be satisfied with, minimal recitations from the children. The reduced number and length of locutions constitute a verbal difference from white classrooms.

Reserve school teachers displayed utterances different from their white school counterparts: they used shorter utterances; they paused longer following individual responses; and they asked more questions. Choctaw Indians displayed their confusion with the switchboard participation structure by interrupting the teacher more often than white students. Greenbaum concluded that this increased interruption frequency was more a reflection of confusion over communication etiquette than an attempt to dominate. It is difficult to generalize behavior of Choctaw children to that of pupils from other Indian nations, but Greenbaum's observations highlight some of the cultural differences in communication which could lead to misunderstandings. Confusion on the part of teachers and students as to the other's intents can contribute to the problems of learning and instruction.

Greenbaum found a difference in verbal skills between white and Indian students. Teaching these skills could ameliorate this problem and presumably reduce what is essentially a cultural difference in environmental stimuli. The purpose of testing procedures becomes one of identifying and enumerating the cultural deficits so that adequate interventions may be devised for each individual student.
5. Indian Education in Saskatchewan (Age-Grade Retardation)

In a major report, Indian education in Saskatchewan was deemed a failure. The standard for educational success used was employability. The statement of failure was based on the determination that additional schooling does not contribute appreciably to an Indian adult's income-earning potential (Kelly, 1973). Apparently, the education system has not been geared to provide employment training, nor has it been in other ways adapted to the needs of Indians. The Saskatchewan study found that student ages within Indian schools were not at expected levels, and that the disparity increased in a progression through the grades (Conry & Conry, 1973a). The cumulative age-grade retardation among elementary Indian pupils in Indian Affairs and Northern schools during the years 1969-73 is evidence of the failure of schooling. In each grade, comparisons were made between actual student chronological age and the usual age of students for that grade level. Beginning in kindergarten, seven percent of the children were at least one year older than anticipated. This small disparity increased markedly in the transition to grade 1 where 53.5 per cent were older; only 46.5 percent of the children were at their expected age. An explanation for this change may be that failures and holdbacks begin in grade 1 in response to lack of academic advancement. Conry and Conry charted the trend through grade eight where only five percent of the children were at the expected age. Therefore 95 percent of all Indian students in grade 8 in Saskatchewan Indian schools were older than their non-Indian cohorts during the period 1969-73 (Ibid., p.158). Another disturbing feature noted was that at some grade levels (specifically 6 and 7) a decrease in age-grade retardation appears. This seeming benefit may actually reflect another problem. The report authors concluded that
this anomaly resulted not from an improvement in schooling, but rather the fact that many students reach the age of 15 or 16 years in these grades, and the drop-out rate increases markedly.

In a different phase of the same project comparisons of achievement were made with students in white schools. Differences were found between Indian and white pupils in all content areas. Again an age-related disparity was noted; by grade 7 the difference was much greater for language skills than at the grade 3 level. The extent of the difference was such that Conry and Conry (1973b) concluded that average Indian students in grade 7 were two to three years below grade level in reading. Since much of school learning is dependent on reading ability, the expectation is that such a disparity might lead to lower achievement in other subject areas, an inference confirmed by the comparisons.

6. Factors Affecting District Achievement

From a Provincial assessment in British Columbia, Greer (1978) reported that in school districts with a high percentage of Native Indians in the student population group performance was low. This was also the case in districts with a high percentage of Native Indian heads of households. At the grade 12 level these differences were not found to be statistically significant, a fact no doubt influenced by the enormous drop-out rate of Native Indian students, with the resultant effect that although the district percentage of Native Indian students may be high, the proportion in grade 12 is actually quite small.

An interesting comparison found no similar effect for districts with high percentages of ESL students or ESL heads of households. Language effects seem to be limited for those districts with a high proportion of Native Indians (Greer,
1978, p. 24). Other factors noted in poor achieving districts were a less well-educated populace, fewer persons employed in professional or managerial positions, and large rural districts with small student populations (p. 19). These districts also tended to have more transient and less-experienced teaching staffs, and a higher rate of secondary dropouts (p. 27). At the grade 4 level, there was a strong negative correlation between district performance and the percentage of Native Indian heads of households for Math and Reading subtests (correlations ranged from -.62 to -.76). Similar results were found for districts with high percentages of Native Indian students ($r = -.57$ to $-.70$). Correlations greater than .29 were significant at the .01 level. Therefore, the incidence of a non-dominant language and/or cultural indices of Native Indians were found to be predictive of lower performance in reading and mathematics.

7. Indian Education: Lillooet

Matthew and More (1987), in an evaluation of schools in the Lillooet district of British Columbia, surveyed elementary and secondary teachers to ask their opinions of concerns for Native Indian learning. The teachers replied that language arts skills are major learning problems for Native Indian students. Undeveloped language skills are an impediment to academic success in all areas since so much of learning requires reading, especially after the primary grades. Primary teachers reported that pre-reading skills are few when Indian students first arrive at school.

Matthew and More tested achievement using the Peabody Picture Vocabulary Test and the Canadian Test of Basic Skills. When the placement grade was compared to expected grade for chronological age, 59% of grade 3
students were at least one year behind. This lag was found at other grades and was progressive ranging from a low of 54% in grade 1, to 78% in grade 8.

Achievement comparisons were made to first and third grade white children, who scored near the age norms on the PPVT, whereas the Indians were generally one and a half years below the norms. Matthews and More suggest that the white students begin school with broader vocabulary knowledge than Indian students. Essentially the white students begin with a head start on Indian children in terms of vocabulary. The net effect of cultural and economic differences is that Indian students are at a disadvantage. They have the ability to learn, but they start school with less vocabulary knowledge than white children, and this is reflected by lower scores on the PPVT.

Similarly, Matthew and More found that Indians' scores on the CTBS were below those of white cohorts for every subtest of the CTBS in every grade tested, 3 through 8. The poor Reading and Vocabulary subtest scores confirm findings from the PPVT.

The teachers felt that the greatest problem faced was the lack of support from parents. They also agreed with comments by Chrisjohn et al (1987) that the readings available were generally culturally inappropriate. To improve schooling, teachers advocated greater Indian community involvement to develop structures more suited to Indian learning styles, and more relevant curricular materials.

An interesting finding came from a questionnaire that asked teachers and the Indian people to rank order the goals of schooling. Both groups placed intellectual development at the top of the priority list. The second most important objective as seen by the Indian people was cultural learning. However, teachers
ranked this goal eighth of nine. This is evidence of the different perceptions about education held by Indians and the persons who teach their children.

8. Indian Education: Okanagan-Nicola

Achievement of Indian students in the Okanagan-Nicola region of British Columbia was found to be lower than whites (More, 1984a). Age-grade placement comparisons found 41% of the Indian students to be at least one year behind. As in the Lillooet district, CTBS results were poorer in the upper grades, with 52% of eighth grade students at least one year behind, compared to 23% for first graders. One trend noted is expressed by the comment that an achievement "nose-dive" occurs in reading at the grade 4/5 level (p. 11). Teachers in Okanagan-Nicola had many of the same concerns as those in the Lillooet district. Teachers reported that Indian achievement was not adequate, especially in language arts. The teachers stated concern for the lack of self-esteem shown by the students, and also the lack of family support in the schools. Teachers recommended inclusion of more information about Indian culture in the curriculum, and the involvement of the members of the Indian community.

9. Quality of Indian Education in Canada

More, (1984) in his review of the quality of education of Native Indian students in Canada, found that "reading, writing and attendance problems seriously affect the accuracy of attempts to measure student achievement" (p. 106).

There are many problems with the accuracy of measuring Indian achievement, but on the basis of reports available More concluded that the
quality of Indian education appears to be low (p. 108). The expectation of Indian parents and educators is that Indian performance on standardized tests should be roughly equal to the national norms. More cautions against the belief in norms as standards to be attained, especially when one is measuring across cultures. However, the norms are useful as indicators of relative achievement. More notes that in all studies reviewed the Indian students were behind non-Indians. The average Canadian student from the population as a whole attains a higher score on the CTBS or other standardized tests. The effect of schooling and testing is more likely to be beneficial in terms of educational success for the non-Indian than it is for the Indian student, which is the reason for More’s conclusion that the quality of Indian education in Canada is poor.

This poor quality was reflected in the inadequate preparation and inservice instruction of teachers in the adaptation and/or development of curricula suited to the cultural background and educational needs of Indian students (More, 1984b).

10. Teaching Cognitive Skills

There are three constellations of skills relevant to sociocultural intelligence: social competence, verbal ability and problem-solving ability (Sternberg, 1985). If such skills are teachable, then perhaps intelligence can be taught. Some people learn the ability to perform well at some aspects of intellectual functioning, whereas others may not. Each skill constellation is a component of the general intelligence factor that affects information processing ability in many intellectual tasks (Sternberg, 1985). If instructional intervention can produce an improvement in any single constellation of skills, then improvement in the general intelligence factor may also be achieved. Remedial measures are needed to teach cognitive
skills to culturally disadvantaged children.

Such training procedures have been described as "cognitive therapy" (Whimbey, 1975). Cognitive therapy seeks to provide instruction in areas of school-related learning typical of crystallized abilities. Much of school learning is predicated on white middle-class verbal constructions. Therefore, in order to be successful in such a system, teachers must ensure that these verbal constructions are understood by the Indian students. McShane (1984) cites study findings that American Indian children in New Mexico had three times the number of language impairments of the general population. Interventions that consist of verbal stimuli may be required if teachers determine that level of crystallized verbal ability attained is insufficient to succeed in the school tasks which rely on inquiry, analysis, explanation, and deduction (Whimbey, 1975). Stimuli presented could also include instruction in basic reasoning patterns that teach analytical reasoning and sequential tasks of problem-solving (Bereiter & Engelmann, 1966). Declarative knowledge would thus be supplemented by the teaching of general learning strategy techniques varied in complexity according to the maturation of the child (Weinstein & Mayer, 1985). This instruction would be beneficial to Indian students who tend to be not active problem-solvers, but more passive receivers of information in an attitude typical of the low-aptitude student. On the other hand, high-aptitude students often use sequential analyses to solve problems, drawing on information already in their possession to clarify the question and proceed through a chain of steps to the solution (Whimbey, 1975). High-aptitude students treat such ill-structured problems by simplifying them into a series of small well-structured subproblems (Simon, 1973). This problem-solving procedure depends on the existence in the student's mental make-up of a framework for
the solution of well-structured problems, a framework unlikely to exist in the memory of an Indian child. The development of a problem-solving skill begins with the acquisition of domain-specific propositions and the relevant problem-solving procedures (Fredericksen, 1984). These procedures are applied to various problems and the results are tested for the applicability to the problem. The problem must be represented accurately, and a solution strategy selected. Fredericksen suggests there is little transfer from one domain of knowledge to another, propositions from each subject area must be acquired separately along with the requisite frame for problem-solving.

This entails a serial processing model of intellectual functioning which may differ from that of the Indian child, who is more likely to process problems simultaneously, and seek to integrate the situation globally, rather than decompose it analytically in a series of steps. The well-structured subproblems are in concrete schemata. This too, differs for the Indian child who may rely on imagery to learn concepts which is the learning style employed in the cultural transmission of legends (More, 1984b).

A solution strategy that might be more successful with Indian children would concentrate on holistic learning of concepts, and an integration of problems (Ibid., p. 12). The desire to seek a solution to a problem as a whole sheds light on the above comment that Indian students act as passive receivers of information.
11. Tests of Intelligence

Evidence of impaired performance on standardized tests by Indian children has been documented (Mueller et al, 1986; Seyfort et al, 1980). Analysis of scores on the Verbal Scale of the WISC-R (Wechsler, 1974) have indicated that fully one-third of the items were in the extreme deciles of relative difficulty for the Inuit children tested.

Low scores on intelligence tests by children from different cultural groups result primarily from difficult test items that depend on knowledge of concepts and skills relevant to schooling (Cleary et al, 1975). However, on items that measure raw learning and memory capacity, scores are much closer to national norms (Bereiter & Engelmann, 1966).

Intelligence is a multi-faceted concept, and assessments of intelligence sample these separate facets to produce a combined score indicative of a general intelligence factor. Impaired performance on any component measure will yield a lower cumulative score. A test samples the universe of skills that exists on the date of the test, and yields a total score. Binet did not represent this score as a unitary entity. This score was meant to be a composite of individual mental skills that could be taught, which could result in a concomitant improvement in test scores (Haney, 1984).

Much of the attitude toward Indian intelligence is based on poorly-developed theory, and the imposition of models that are inappropriate or ill-conceived. The observation that Indian students are passive learners has led to conclusions that they must be learning disabled. This assumption rests on the faulted syllogism that learning disabled students are passive in learning situations, Indian students are passive, therefore Indian students are learning disabled. This
logic is flawed, ignoring findings that Indians are more likely to seek global solutions (More, 1984b).

Chrisjohn and Peters (1986) debunk the model of the right-brained Indian stating that the evidence of Indian intellectual functioning as being based on the right-brain is weak. The benefits of such a curriculum for Indians might exist if Indians were indeed "right-brained", but Chrisjohn and Peters (1986) make the observation that the theory may be a passing educational fad, with little relevance to Indians or anyone else (p. 5).

12. Test Frustration

If an Indian student has limited exposure to Anglo culture, chances of success on a standardized test are reduced. Failure is a concept experienced during schooling by Navajo children and reinforced by test-taking (Deyhle, 1986). Failure has a negative effect on ability formation because the student comes to regard low ability as a structure of his intellectual make-up, beyond his capacity to change. Once the perception of low ability is fixed, learned helplessness may result. The student no longer attempts to change his levels of performance because he sees intellectual ability as being separate from his control (Rosenholtz & Simpson, 1984).

Deyhle (1986) showed this process by comparing Navajo and Anglo students in the primary grades. She found that young Navajo students did not distinguish test-taking from other classroom activities. They were excited by the tests which they judged to be game-like and free from anxiety. However, comparable white children approached the testing process with considerable trepidation. The Anglo children were aware of the judgmental nature of testing
and became nervous in anticipation and relieved when the test was completed. When older Navajo children came to the realization that tests were events used to judge personal academic performance, anxiety became much more prevalent. The anxiety inhibited performance and caused frustration. When the results turned out to be poor, the students became apathetic in response to their lack of achievement, a symptom of learned helplessness.

The onset of test anxiety is sudden and swift, at about the third or fourth year in school for Navajo children. Anglo students understand the purpose of tests when they enter the school system, and are therefore in a position to adapt more gradually to test anxiety. Taking a test is a cultural experience concerned with two values: that one should display competence during the test; and that individual achievement is valuable. Without these values test-taking is a "culturally incongruent activity" (Deyhle, 1986). Poor performance on tests may have negative social impact if the child is placed in special classes or labelled as being mentally retarded, which could have devastating effects on self-respect and social adjustment (More & Oldridge, 1980).

Mercer (1979) suggests that such mislabelling is a function of the "distance between the child and the core culture" (p. 14). The distance may result from factors other than race, such as socioeconomic status, which is another main determinant of cultural disadvantage. Mercer developed the System of Multicultural Pluralistic Assessment (SOMPA) to address this problem, but Jirsa (1983) pointed out weaknesses and suggested that the SOMPA contributes little to our understanding of multicultural differences in intellectual abilities. Moreover, Jirsa downplayed the need for alternate instruments, having seen little evidence of bias associated with race in the construction of the WISC-R and
stated that a "competent administration should yield valid results" (p. 15). This statement is supported by evidence that the same constructs are measured by the WISC-R regardless of grouping by socioeconomic status (Hale, 1983; Carlson, Reynolds, & Gutkin, 1983).

13. Review of Indian Intelligence Research

In a discussion of research on testing Indian intelligence, Chrisjohn and Lanigan (1984) noted five problems that cast doubt on the conclusions and recommendations that have been drawn. First, the assumption of Pan-Indianism, is the belief that conclusions drawn are necessarily valid across Indian nations. The second problem is the use of small sample sizes. Third, in many instances, improper instruments have been used to test Indian children. Fourth, there is the difficulty that psychometric research has not dealt with fundamental assumptions about Indian abilities. The fifth problem noted by Chrisjohn and Lanigan may be the most important and that is the lack of an overarching theory of Indian intelligence.

There has not been enough research on Indian cognitive structure, and the research that has been done is usually conducted within specific nations. There is little evidence of homogeneity of cognitive structure among different national and linguistic populations. Thus, Chrisjohn and Lanigan suggest that the results from an investigation, no matter how valid, cannot be defensibly extended beyond the nation involved (p. 51).

The validity of many conclusions is suspect because of limited sample sizes which rarely exceed 200 in Indian studies. Another flaw in studies with seemingly large samples, is that the number of subjects has been obtained by
aggregating across a number of age groups, so that conclusions are reached about students assuming homogeneity across several developmental levels.

Based on the poor performance of Native Indians on standardized tests, some researchers have turned away from tests such as the WISC-R in favor of others presumably less biased. Chrisjohn and Lanigan suggest that these efforts have been largely naive, since the bias "is a function of the interpretation of the results obtained" from testing (p. 52). Lower scores do not necessarily provide evidence of bias, but unsubstantiated conclusions drawn from the test results would. Tests such as the WISC-R have been rigorously validated and examined in over 100 studies with Native Indians; Chrisjohn and Lanigan recommended the retention of tests that have been well-examined rather than to turn to other tests which may not be acceptable according to psychometric standards (p. 52).

The fourth problem deals with this paucity of psychometric research about factors affecting test results. Indian students' exposure to test content, both curricular and cultural, is likely to be different from other groups. For this and other reasons Chrisjohn and Lanigan contend that "construct equivalence" has not been established across ethnic groups, for most tests including the WISC-R. A truly equivalent test would reliably measure the same nominal dimension with comparable intervals between groups (p. 52).

Without adequate research, an adequate theory of Indian intelligence has not been constructed. Imposition of tests devised for other ethnic groups, by administrators from other ethnic groups, is a poor substitute for testing by Indians with instruments developed by Indians, addressing concerns relevant to both their cognition and culture. Chrisjohn and Lanigan have called for the development of tests based on "an articulated theory of intelligence relevant to
the population to be served" (p. 55).

14. Cultural Bias in Tests

Mean test score differences between racial groups have been cited by critics of psychological testing as evidence of inherent racial bias. Mercer (1979) cites the contention "that it is differences in the mean scores which are central to the determination of test bias" (p. 14). Following this logic, a comparison of mean scores between ethnic groups is required. When it has been established that differences in mean scores on tests do exist, the results are subject to interpretation. It may be that prediction of poor academic performance by Indian children on the basis of lower test scores is an unbiased assessment because in the majority of cases these people are relatively unsuccessful in school and in the workplace. A question arises, however, as to the concern that an assessment becomes a self-fulfilling prophecy when teachers, school officials, parents, and the child all believe that test results are accurate assessments of educational potential and that surety of failure is immutable. Mercer cautions against the imposition of labels to avoid creating a self-fulfilling prophecy. An assumption of assessment materials is equality of access to knowledge (Lawton, 1975). However, if access to knowledge is not a fact, then there is a possibility of bias in tests. As the cultural milieu varies from that of the white middle class, access to knowledge decreases. The opportunity to learn may exist as a formal setting, but the ultimate source of inequality may be the cultural context in which the child lives (Keniston, 1977; Wilensky, 1975). Mercer (1972) has argued that what intelligence tests measure, to a large extent, is exposure to Anglo culture.
15. Testing Native Indian Students

A number of factors that contribute to impaired performance on tests by native Indian students have been discussed. These factors are components of a cultural disadvantage which constrains the learning process. Development in an essentially nonverbal environment is inadequate preparation for school tasks. Native families do not concentrate on the sequential transfer of abstract skill concepts. Low socioeconomic status limits the availability of enriching experiences. Some native societies lack a concept or appreciation of the purposes of testing.

The environmental context for the development of intelligence in Indian children is one of an experiential environment that is less rich than it is for white children.

For a multitude of reasons cultural and linguistic minorities in North America are at a greater risk for educational failure, mental illness, and economic disability. This also means, therefore, that they are...at very high risk for experiencing misclassification, misdiagnosis and misplacement...Historically standardized norm-referenced tests of "intelligence" have led to a disproportionately large number of Black, Hispanic and Indian children being...placed in less challenging educational programs (McShane, 1984, p. 83).

Instructional interventions designed to increase educational success have been discussed, including cognitive therapy. Proponents of these treatments recommend instruction in verbal ability and problem-solving, in addition to providing stimuli and relevant feedback, in the belief that early intervention could reduce cultural disadvantage.

Studies that investigate the results of 'strategy training' have shown considerable success with learning disabled students, by teaching metacognitive strategies, and by presenting specific strategy content to facilitate the transfer of learning (Brown & Campione, 1980; Whyte, 1985).
Test performance reflects an interaction between innate intellectual capacity and prior knowledge with current test demands (Cress, 1974). The best assessment of intellectual capacity would thus be the one in which these interactions are controlled, better to examine the main effect of intellectual ability. Cress and O'Donnell (1974) suggested that test validity cannot be assumed when the sample includes Indian students. However, they did conclude that cognitive tests are accurate predictors of success in the dominant culture.

B. SUMMARY OF CHAPTER

This chapter has reviewed different aspects of the problems and considerations of educating and testing Canadian Native Indian students. A consistent conclusion drawn has been that these students are capable of learning, but that their scores on standardized tests of achievement and ability may be lower than those for the majority of the population (MacArthur, 1968).

The reasons for lower test scores are numerous. Beginning with learning style, Native Indians are not used to the concentration on verbal learning typical in schools (Greenbaum, 1985). Neither are they used to having communication controlled by a central figure such as a teacher (Philips, 1972). The nonoral basis of Native Indian learning and communication puts students at a disadvantage at testing time. There is also evidence that the purpose of testing is a foreign concept: children from this ethnic minority group are not prepared to apply all of their intellectual skills to tests, because they are unaware of the necessity to do so (Deyhle, 1986).

Tests may not be constructed to fairly assess the knowledge acquired by ethnic minority groups. Claims have been made of inherent bias (Mercer, 1979)
or that test content is a sample of an unfamiliar domain (Lawton, 1975). Bias has been inferred as a result of differential performance on standardized tests and a concomitant lack of educational success. Performance deficits have been noted (Conry and Conry, 1973b; MacArthur, 1968: Mueller et al, 1986; Seyfort et al, 1980). However the extrapolation that poor test performance is evidence of bias ignores the differences that groups bring to the testing environment.

That tests are standardized means that the procedures of testing, administration, and interpretation have been regularized, and that distributions of results have been reported based on testing a representative sample of the national population. Test content will be more or less relevant to subsets of the population, depending on how similar their educational and developmental experience is to that of the standardization sample. If Canadian Native Indian children are shown to score lower than the general population, a number of cautions must be exercised. First, there should be no confusion that the norms of standardized tests are a 'standard' to be attained (More, 1984b). The norms are simply portrayals of a test's performance in the norming group. Second, a score should not be taken as an immutable measure of mental or educational competence. Content on tests can be taught, and instruction in the goals and meaning of testing will make students better equipped to cope with tests (Whimbey, 1975). Third, teaching to Indian student strengths, and the use of cognitive strategy intervention can aid student adaptation to the school environment (Sternberg, 1985). Fourth, the validity of test use will lie in the identification of disparities in the development of school-related crystallized abilities, reflected in content knowledge between Indian and non-Indian students (Chrisjohn and Lanigan, 1984). If intervention is required to remediate a gap in experience
between Canadian Native Indian students and the majority of the population, then standardized tests may have a useful role in the identification and calibration of these gaps, and in framing prescriptive answers for closing them.
CHAPTER III. METHODOLOGY

The present chapter begins with a description of two methodologies: the techniques used by the evaluation team to collect and analyze data, and the procedures followed in this thesis in the presentation of a second order analysis of test appropriateness for an ethnic minority group. Under the assessment team methodology, there is a description of the sample, the standardized tests, and the testing procedures. Under the thesis methodology there is a description of a judgment model used to evaluate the appropriateness of the standardized tests, and the data analysis.

A. METHODOLOGY OF THE SURVEY OF LEARNING PROBLEMS

1. Sample

The assessment team tested the entire population (n=76) of a reserve school (K-7) for Native Indians in British Columbia. Table 1 gives the distributions of gender and grade level for these children. In Table 2 the range, median, and mean ages are given for students in each grade.

2. Instruments

The assessment team employed a multiple assessment model, including several standardized tests: for Kindergarten the Metropolitan Test of Reading Readiness Level I, Form P (Nurss & McGauvran, 1974); for kindergarten through Grade 5 the Developmental Test of Visual-Motor Integration (Beery & Buktenica, 1976); for kindergarten through Grade 7 the Peabody Picture Vocabulary Test-Revised, Form L (Dunn & Dunn, 1981); for Grades 1 through 7 the
Table 1

Distribution of Sample by Gender and Grade

<table>
<thead>
<tr>
<th>GRADE</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>76</td>
</tr>
</tbody>
</table>

\(^{1}n = 76\)

Table 2

Age of NIRS Students by Year and Month

<table>
<thead>
<tr>
<th>Age</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youngest</td>
<td>5-5</td>
<td>6-5</td>
<td>7-7</td>
<td>8-6</td>
<td>9-9</td>
<td>10-10</td>
<td>11-10</td>
<td>13-4</td>
</tr>
<tr>
<td>Oldest</td>
<td>6-5</td>
<td>8-1</td>
<td>9-6</td>
<td>9-6</td>
<td>13-7</td>
<td>12-2</td>
<td>15-2</td>
<td>15-3</td>
</tr>
<tr>
<td>Median</td>
<td>5-11</td>
<td>7-1</td>
<td>8-8</td>
<td>9-2</td>
<td>11-2</td>
<td>11-2</td>
<td>12-7</td>
<td>14-2</td>
</tr>
<tr>
<td>Mean</td>
<td>5-10</td>
<td>7-3</td>
<td>8-6</td>
<td>9-0</td>
<td>11-3</td>
<td>11-4</td>
<td>13-0</td>
<td>14-2</td>
</tr>
</tbody>
</table>

\(^{1}n = 76\)

Canadian Cognitive Abilities Test (Thorndike, Hagen, Lorge, & Wright, 1974); and also for Grades 1 through 7 the Canadian Test of Basic Skills, Primary and Multilevel Editions, Form 5 (King, 1982b). In order to determine if perceptual problems were present, two additional tests were administered: one of
telebinocular vision; and the other an audiometer rating of hearing ability. Students in Grades K-7 were given tests of hearing acuity by qualified technicians using Zenith and Beltone audiometers. Students in Grades 1-7 were also given vision screening tests using the Keystone Telebinocular Test.

The Metropolitan Readiness Test

The Metropolitan Readiness Test (Nurss & McGauvran, 1976) Level 1 Form P (MRT) was selected to provide measures of developmental skills necessary for readiness to begin school learning. Level 1, is recommended for use from the beginning through the middle of kindergarten. The six subtests used in the assessment cover four pre-reading abilities: Auditory Memory; Rhyming; Visual skills, and Language skills. Visual skills consist of Letter Recognition and Visual Matching while Language skills consist of School Language, Listening, and Quantitative Language. The subtest scores are combined to achieve a seventh score, the Pre-Reading Composite. The MRT is a hand-scorable booklet that yields performance ratings of Low, Average, and High on each of the six skill areas.

The split-half reliability coefficients are reported in Appendix 1 for each of the subtests. The data reported by the test authors has low errors of measurement and relatively high reliability coefficients, typically in the vicinity of 0.8. Validity evidence is derived from correlations between MRT and the Metropolitan Achievement Test. Since the MRT is designed as an assessment of skills necessary for learning, correlations with skill scores from the related Metropolitan Achievement Test are indicative of a valid assessment. The correlations are generally high, approximating 0.6.
Developmental Test Of Visual-Motor Integration

The VMI (Beery, 1982) is a form-copy test that presents children with 24 forms graduated according to developmental age and characteristics. The test was chosen because of its utility as a screening device that aids early identification of learning difficulties. Designed for pre-school and early grades, the VMI may be used with subjects as young as two years or with adults. In this project the test was administered to children in kindergarten through grade five, but not beyond because of a demonstrated tendency of decline in predictive correlations with maturation after age 11 (Klein, 1978; Tucker, 1976).

Reliability information reported by the test authors (p. 15) mentions test-retest reliability estimates ranging from 0.63 to 0.92 and split half reliabilities ranging from 0.66 to 0.93. Concurrent validity of the VMI has been established for academic skills (ranging from 0.51 to 0.73 for reading and math) and chronological age (0.89). Analysis of low socioeconomic groups' test scores has shown the VMI to be a good predictor of academic achievement (Bray, 1974; Buktenica, 1966). Researchers have found that when used as a component of a test battery, the VMI has proved to be a reliable predictor of academic achievement, especially in the primary grades of kindergarten and grade one (University City, 1969). Early identification of potential learning difficulties provides the best opportunity to institute appropriate remediation.

The Peabody Picture Vocabulary Test - Revised

The PPVT-R (Dunn & Dunn, 1981) was administered to students in this sample from kindergarten through Grade seven. The PPVT-R is a measure of receptive vocabulary that is commonly used as a correlate of IQ to indicate student level of achievement in language. Parallel forms L & M are tests
consisting of items arrayed in order of ascending vocabulary difficulty. Norm-referenced basal and ceiling points permit administration of the test to a wide range of ages without having to administer the entire test. This avoids the problem of performance decrements due to fatigue. The test has been standardized on a population ranging from ages 2 1/2 through 18. The test is designed to be administered to children individually by a trained examiner.

Split-half reliability coefficients are reported in Appendix 2 for students aged 5 through 15. Correlations between odd and even items are generally high, approximating 0.8. Split-half reliability coefficients for odd/even items on alternate forms L & M are found in Appendix 3 for ages 5 through 15. Again, reliabilities are high, near 0.8. The reliability coefficients in this table are for both raw scores and standard score equivalents. Table 6 shows the alternate form reliability coefficients raw scores of Forms L & M for children in kindergarten through grade seven. The average reliability coefficients between the two presentation orders range from 0.64 to 0.90 for the grade levels with a high median coefficient of 0.84. Content validity was established by searching Webster's New Collegiate Dictionary (p. 58). Construct validity has been garnered from over 300 studies. Fifty-five concurrent validity studies between the PPVT-R and other vocabulary tests yielded a median correlation of 0.71. In 72 studies correlating the PPVT studies with intelligence tests the median correlation was 0.62; and with four achievement tests the correlations ranged from 0.29 to 0.68.

**The Canadian Test Of Basic Skills**

The complete battery of the CTBS Form 5 Multilevel (King, 1982a) of 11 subtests, comprising 15 scores based on six skill areas, was administered to the sample. Depending on age and grade, the test level ranged from Level 9 in
Grade three to Level 13 in grade 7. The eleven subtests are: Vocabulary; Reading; four language scores: Spelling; Capitalization; Punctuation; and Usage; two Work-Study scores: Visual Materials and Reference Materials; and three Mathematics scores: Concepts; Problem Solving; and Computation. Since the sample also included primary students the CTBS Multilevel Edition Form 5 was administered — Level 6 in Grades 1 and Level 8 in Grade 2. The Primary Edition consists of all the subtests of the Multilevel Edition plus Listening and Word Analysis. Including composites, the battery is comprised of 17 scores.

The internal consistency (KR-20) reliability coefficients (King, 1984, p. 67-69) are reported in Appendix 4 for each of the subtests. Reliability coefficients are consistently high. Test validity evidence for the CTBS is based on over forty years of development, including comparing test content with curriculum used in Canada to ensure content validity.

**Canadian Cognitive Abilities Test**

The Multilevel Version of the CCAT (Thorndike, Hagen, Lorge, & Wright, 1974) is a series of tests of cognitive abilities designed for children in the primary grades, Kindergarten through Grade 7. Canadian Cognitive Abilities Test Primary Version, Level 1 was administered to the students in grade One; and Level 2, Form 3 to grade Two. The Multilevel Edition was administered to the students in grades three through seven. The CCAT has three components: a verbal score; a quantitative score; and a nonverbal score. The verbal battery has four subtests: Vocabulary; Sentence Completion; Verbal Classification; and Verbal Analogies.

KR-20 reliability coefficients are reported in Appendix 5 for grade 1 through 7. Reliability is uniformly high in the vicinity of 0.90. The CCAT was
standardized using the same norming sample as the CTBS. Criterion-related validity evidence as reported by the test authors is that the three components of the CCAT correlated well with the CTBS composite, ranging from 0.61 on the noverbal to 0.86 on the verbal.

3. Testing Procedures

Testing was conducted in the late spring by a team of graduate students in school psychology under the supervision of a registered psychologist, hereafter called the assessment team. The late spring was deemed to be the optimal time for audiometric screening, because of the lower likelihood of respiratory infections which might affect hearing acuity. Also, it was the right time for testing because of the program decisions which motivated the evaluation contract.

4. Assessment Team Data Analysis

The evaluation team scored tests and generated standard scores according to the procedures as outlined in the respective manuals. The scores were presented to the board of trustees as part of a confidential report. The primary focus of this report was to identify children with learning problems to aid the school staff in the planning of appropriate educational programs. The report also included individual results of vision and hearing screening.
B. THESIS METHODOLOGY

1. Judgment Model

To assess whether the instruments used in the screening battery are appropriate for a uniracial ethnic minority sample, a judgment model was created; it is illustrated in Figure 1.

The judgment model was inspired by an evaluation model for organizing data developed by Stake (1967). This thesis employs modifications of the descriptive and judgmental components of the Stake model. Stake’s matrix presentation was utilized to aid comparison of results with a standard. The evaluation questions are listed under two headings: Research Questions and Subsidiary Research Questions; the second group addresses more specific aspects of the first. The research questions are intended to identify the individual performance of the NIRS students on the battery of standardized tests. The third column, headed Standards, gives the anticipated outcomes. Comparison of the outcomes with observations provides the basis for a judgment of appropriateness of the standardized tests along the dimensions defined by the research questions. The fourth column, Judgment of Appropriateness, is blank in the model as presented. Based on the results given in Chapter 4, assessments of appropriateness are developed in Chapter 5.

2. Thesis Data Analysis

The data produced by the assessment team was re-examined to provide answers to the subsidiary research questions, "What were the NIRS student test scores on the battery of tests relative to the population norms?" and "What was
<table>
<thead>
<tr>
<th>RESEARCH QUESTIONS</th>
<th>SUBSIDIARY QUESTIONS</th>
<th>STANDARDS</th>
<th>JUDGMENT OF APPROPRIATENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does the battery achieve its goal of identifying children who are in need of educational assistance.</td>
<td>1a. What were the NIRS student scores on the battery of tests relative to population norms.</td>
<td>1a. Instruments should indicate range of achievement and ability scores.</td>
<td>Judgments of appropriateness will be based on the results in Chapter 4; conclusions drawn are presented in Chapter 5.</td>
</tr>
<tr>
<td></td>
<td>b. What was the prevalence of hearing and vision problems?</td>
<td>b. Incidence of hearing and/or vision problems should be higher than the population.</td>
<td></td>
</tr>
<tr>
<td>2. Do results of the battery reveal the type and magnitude of assistance required?</td>
<td>2a. Compared to test norms, did NIRS scores differ?</td>
<td>2a. Comparisons should show grade levels and test areas where NIRS means are above/below population means.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Did achievement decrease in higher grades? What was the incidence of age grade retardation?</td>
<td>b. Battery should indicate grades and test areas where achievement decreases, and where age is greater than expected for grade placement.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Did scores vary with the indicators of ability level? Was achievement for higher ability students lower than expected?</td>
<td>c. Instruments should identify students achieving below, at, or above ability level.</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 1. A JUDGMENT MODEL FOR EVALUATING THE APPROPRIATENESS OF STANDARDIZED TESTS IN A NATIVE INDIAN RESERVE SCHOOL**
<table>
<thead>
<tr>
<th>RESEARCH QUESTIONS</th>
<th>SUBSIDIARY QUESTIONS</th>
<th>STANDARDS</th>
<th>JUDGMENT OF APPROPRIATENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2d. Were language scores relatively lower than other component test scores?</td>
<td>2d. Instruments should identify students with language scores below other skill area scores.</td>
<td>The judgment of appropriateness will be arrived at based on the results in Chapter 4. The conclusions drawn will be presented in Chapter 5.</td>
<td></td>
</tr>
<tr>
<td>e. Did perceptual problems affect test scores?</td>
<td>e. Instruments should identify students whose performance was affected by perceptual problems.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 3. Does use of a multi-dimensional battery lead to biased assessment when used in such a sample? | 3a. Did the PPVT-R have the same test difficulty characteristics for the NIRS sample as the norming sample? | 3a. Item difficulty rankings should be the same for the sample and the population. | Same as above. |
| 3b. Are Canadian Native Indians underrepresented in the norm reference group for the tests? | b. Proportion of Indians in the norming sample should equal that of the population. |
| 3c. Was there examiner or language bias? | c. Canadian Indians should be the qualified examiners. Students should understand the instructions. |
| 1. Were the examiners Canadian Native Indian? | | |
| 2. Did examiners accurately communicate their intents and purposes to the children? | |

FIGURE 1 (CONTINUED)
The next five subsidiary questions (2A to 2E) address the research question "Do results of the battery reveal the type and magnitude of assistance required?" The first consideration regarding the NIRS sample was "Compared to test norms, did NIRS scores differ?" (2A). The single sample to whom the test was administered had no control counterpart; no appropriate group was available. The only comparison available was the national norms for appropriate grades. Therefore, confidence intervals were constructed around the calculated means for the group and the population. The confidence interval around the population parameter represents the sampling distribution of the mean, an estimate of the dispersion of group means for samples equal in size to that of the NIRS group. Standard deviations and means for the norming sample were obtained from the technical manuals for the respective tests. Assuming that the mean and standard deviation of the norm group are population parameter estimates, the estimate of the standard error was calculated by dividing the population standard deviation by the square root of the NIRS sample size. Mean scores and standard error estimates for the NIRS sample were computed using the Statistical Package for the Social Sciences (SPSS Inc., 1988). From these results, 95% confidence intervals were constructed around the sample mean and the mean of the norming population, using the formula \( X \pm 2.56(Se) \). Thus a one-tailed test of significant differences between means was made possible. The confidence interval constructed around the norm sample mean represents the sampling distribution of the means with sample size equal to the NIRS sample. If the mean of the NIRS sample was found to lie outside this confidence interval, the null hypothesis would be rejected.
The next three subsidiary questions were also intended to address the question "Does use of a multi-dimensional battery lead to biased assessment when used in such a sample?" Score differences were examined to determine if differential results appeared in grade or ability levels and also to examine the influence of language and perceptual problems.

The subsidiary question "Did the PPVT-R have the same test difficulty characteristics for the NIRS students as the norming sample?" (3A) was examined by comparison of rank order item difficulties on the PPVT-R. On tests like the PPVT-R, which are administered to students of wide ranges in age and ability, the item presentation arrangement is intended to be in order of ascending difficulty. Entry and exit points are defined by "basal" and "ceiling" rules. The 'basal rule' assumes that all items prior to the entry point would be passed by the individual tested. In scoring the PPVT-R, the individual is given credit for all items with difficulty lower than that of the basal item. Similarly, the 'ceiling rule' assumes that all items beyond the ceiling item would be failed. In order for these assumptions to be valid, the difficulty-based ordering of items should hold true for different samples (Nunnally, 1978). Different rank order of items by difficulty for a sample is an indication that the items may represent inappropriate content for that group. Jensen (1982) claimed that if items measure the same ability, then the items should maintain the same rank order across groups. Rank order was evaluated in terms of relative item position change between groups. Misranking was defined as a placement change of more than eight rank positions between the two samples. The eight-rank-position movement reflects the 'ceiling rule' used for the PPVT of six incorrect responses in a series of eight items. Rank position of item difficulty for the PPVT was defined
as the placement order in the test (based on ascending order of difficulty). Rank position of item difficulty on the PPVT-R for the sample group was defined by the calculated item difficulty values for the group.

Another comparison of rank order item difficulty was made by calculating a correlation coefficient between the difficulty rankings for the two groups. The coefficient used was Spearman's \( \rho \) which ranges from 0.0 to 1.0. A correlation coefficient of 0.0 reflects no association between the two rankings, whereas a coefficient of 1.0 would reflect that rankings correspond perfectly between the two groups. Jensen (1982) suggested that a correlation coefficient above 0.95 reflects a similar and desirable level of correspondence in item difficulty.

The index of discrimination is the correlation between the item and an external criterion, in this case the total score on the items answered in the range. The appropriate coefficient is the point-biserial \( r_{pbis} \) when comparing a dichotomously scored (right/wrong) item with a continuous variable such as total score. Test reliability is enhanced when correlations between items and the total test score are greater than 0.30 (Nunnally, 1978).

The subsidiary research question "Are Canadian Native Indians underrepresented in the norm reference group for the tests?" (3B), was answered by an inspection of test manuals. The final subsidiary question, "Was there any examiner or language bias?" (3C) was assessed by two further questions, "Were the examiners Canadian Native Indians?" and "Did examiners accurately communicate their intents to the children?"
C. CHAPTER SUMMARY

This chapter has provided a description of the procedures used to analyze data. A judgmental model was also presented showing the research questions arrayed in a framework permitting examination of evidence with respect to the research questions. The evidence is reported in Chapter 4 under two headings, Results of the Survey of Learning Problems and Results of the Thesis.
CHAPTER IV. RESULTS

The results presented in this chapter are divided into two major sections: ‘Results of the Survey of Learning Problems’ and ‘Results of the Thesis’. Under each of the two sections, results are organized according to the research questions. The results constitute the third column headed ‘Evidence’ of the judgment model described in Chapter 3.

A. RESULTS OF THE SURVEY OF LEARNING PROBLEMS

1. Sample Scores on the Battery of Tests

Results are given here for each of the individual standardized tests. The logic is to present the information from the survey of learning problems to provide the reader with an understanding of student achievement relative to age and grade placement. The results reported in this section are descriptive. The purpose of a separate examination of the scores from the survey of learning problems is to detect trends, in this case whether age and grade equivalent scores on the standardized tests were lower for the NIRS students. Lower scores would indicate the presence of learning problems. More detailed analysis of results is presented in Section B, Results of the Thesis.

1.1 Metropolitan Readiness Test

Only kindergarten pupils were given the MRT, which is composed of a pre-reading composite and six subtests. The composite scores are discussed in this section. The evaluation team reported MRT results according to three categories: high, average, and low. Thirteen of 20 pupils’ stanines were in the low range on the pre-reading composite. One student was in the high range and the
remainder in the average range.

### 1.2 **Canadian Test of Basic Skills**

Table 3 contains summary scores on the four skill areas of the CTBS: Vocabulary, Reading, Language Total, and Math Total. For each skill area the range, median, and mean scores are reported, followed by the expected achievement level, and the difference (if any) between placement at time of testing and achieved grade equivalent scores. A simple discrepancy is not evidence of differences in means. Fluctuations of a few months around placement are to be expected. It is only when the differences become large, approximately one year below placement that there is cause for concern.

For both vocabulary and language the mean grade equivalent scores achieved were below the mean grade placement at every grade level. With the exception of grade 1, the same statement is true for reading and math scores. However, except for one subtest, in grades 1 through 5 the difference was less than one and a half years. Using the one year criterion, scores were commensurate with placement from grades 1 through 4. The difference is larger in grades 6 and 7. All but one mean subtest score at these grade levels was more than two years below placement. The significance of the result is difficult to determine since the samples were very small. It is also likely that the results were influenced by the disproportionate numbers of slow learners. Three of four grade 6 students tested on the CCAT and 2 of five grade 7 students were categorized as slow learners (Table 4).

### 1.3 **Canadian Cognitive Abilities Test**

Table 4 is a list of student frequencies categorized by ability level on the CCAT. This table shows the number of students at each grade and ability level.
Table 3

Differences in Mean Scores Between Grade Placement and CTBS Grade Equivalent Scores

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRADE PLACEMENT</strong></td>
<td><strong>LEVEL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest</td>
<td>0.9</td>
<td>1.1</td>
<td>2.5</td>
<td>2.0</td>
<td>3.3</td>
<td>2.6</td>
<td>3.9</td>
</tr>
<tr>
<td>Highest</td>
<td>1.4</td>
<td>4.1</td>
<td>3.4</td>
<td>5.6</td>
<td>5.4</td>
<td>4.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Median</td>
<td>1.1</td>
<td>2.5</td>
<td>3.1</td>
<td>2.9</td>
<td>4.4</td>
<td>3.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Mean</td>
<td>1.1</td>
<td>2.4</td>
<td>3.1</td>
<td>3.3</td>
<td>4.4</td>
<td>3.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Expected*</td>
<td>1.7</td>
<td>2.7</td>
<td>3.7</td>
<td>4.7</td>
<td>5.7</td>
<td>6.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Difference**</td>
<td>-.6</td>
<td>-.3</td>
<td>-.6</td>
<td>-1.4</td>
<td>-1.3</td>
<td>-3.1</td>
<td>-2.8</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest</td>
<td>1.1</td>
<td>1.6</td>
<td>2.4</td>
<td>2.4</td>
<td>2.9</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Highest</td>
<td>3.0</td>
<td>3.4</td>
<td>4.3</td>
<td>5.3</td>
<td>5.2</td>
<td>5.4</td>
<td>6.8</td>
</tr>
<tr>
<td>Median</td>
<td>2.0</td>
<td>2.3</td>
<td>3.3</td>
<td>3.8</td>
<td>4.0</td>
<td>4.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Mean</td>
<td>2.1</td>
<td>2.3</td>
<td>3.3</td>
<td>3.8</td>
<td>4.0</td>
<td>4.6</td>
<td>5.4</td>
</tr>
<tr>
<td>Expected*</td>
<td>1.7</td>
<td>2.7</td>
<td>3.7</td>
<td>4.7</td>
<td>5.7</td>
<td>6.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Difference**</td>
<td>+.4</td>
<td>-.4</td>
<td>-.4</td>
<td>-.9</td>
<td>-1.7</td>
<td>-2.1</td>
<td>-2.3</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest</td>
<td>0.6</td>
<td>1.3</td>
<td>2.2</td>
<td>2.0</td>
<td>2.5</td>
<td>3.8</td>
<td>4.2</td>
</tr>
<tr>
<td>Highest</td>
<td>2.8</td>
<td>3.2</td>
<td>4.3</td>
<td>5.7</td>
<td>5.4</td>
<td>5.2</td>
<td>6.6</td>
</tr>
<tr>
<td>Median</td>
<td>1.6</td>
<td>2.2</td>
<td>3.4</td>
<td>3.1</td>
<td>4.9</td>
<td>4.6</td>
<td>6.0</td>
</tr>
<tr>
<td>Mean</td>
<td>1.5</td>
<td>2.1</td>
<td>3.3</td>
<td>3.6</td>
<td>4.5</td>
<td>4.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Expected*</td>
<td>1.7</td>
<td>2.7</td>
<td>3.7</td>
<td>4.7</td>
<td>5.7</td>
<td>6.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Difference**</td>
<td>-.2</td>
<td>-.6</td>
<td>-.4</td>
<td>-1.1</td>
<td>-1.2</td>
<td>-2.2</td>
<td>-2.2</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest</td>
<td>0.8</td>
<td>1.4</td>
<td>2.3</td>
<td>3.5</td>
<td>4.6</td>
<td>4.7</td>
<td>4.6</td>
</tr>
<tr>
<td>Highest</td>
<td>2.8</td>
<td>4.3</td>
<td>4.9</td>
<td>5.6</td>
<td>5.7</td>
<td>5.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Median</td>
<td>1.5</td>
<td>2.9</td>
<td>3.4</td>
<td>4.3</td>
<td>5.1</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Mean</td>
<td>1.7</td>
<td>2.0</td>
<td>3.5</td>
<td>4.0</td>
<td>5.2</td>
<td>5.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Expected*</td>
<td>1.7</td>
<td>2.7</td>
<td>3.7</td>
<td>4.7</td>
<td>5.7</td>
<td>6.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Difference**</td>
<td>0.0</td>
<td>-.7</td>
<td>-.2</td>
<td>-.7</td>
<td>-.5</td>
<td>-1.5</td>
<td>-2.2</td>
</tr>
</tbody>
</table>

* The "expected" grade equivalent score was derived from an estimate of the time during the school year when assessment took place. For example, 1.7 would be the expected grade equivalent score for students having completed seven months of grade 1.

** The "difference" reported is the mean grade equivalent score subtracted from the expected mean grade equivalent score.
Table 4

Student Frequencies in each CCAT Ability Category

<table>
<thead>
<tr>
<th>Grade</th>
<th>n</th>
<th>Slow Learner (68-78)</th>
<th>Low Average (79-88)</th>
<th>Average (89-110)</th>
<th>High Average (111-120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>7</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>42%</td>
<td>23%</td>
<td>33%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Note. The n's in this table do not always agree with those in Table 1 as some students were absent at time of testing.

Categorization involved examining standard age scores ("normalized standard scores with a mean of 100 and a standard deviation of 16" (Thorndike et al, 1974, p. 46) according to the following levels: slow learner (68-78), low average (79-88), average (89-110), and high average (111-120). As there is only one CCAT score at grades 1 and 2, but three component scores for grades 3 through 7, comparisons between grades are problematic. Therefore, the three CCAT component scores were averaged in grades 3 to 7, so as to yield one summary standard age score for each student.

In grade 1 the majority of students were rated average, whereas most in grade 2 were rated as slow learners. In grade 3 one student was assessed as
being in the high average category, three were average, one low average, and
two slow learners. The majority of students in grade 4 fell into the slow learner
category. Grade 5 students were distributed in the three categories: slow learner,
low average, and average. In grade 6, all students but one were rated as slow
learners. Finally, in grade 7 three of the five students were low average and
the other two were categorized as slow learners. The CCAT scores of the NIRS
sample are unevenly distributed, since approximately 2% are above average and
60% below average. That 42% of the sample was categorized as slow learners
suggests a different distribution compared to the population. From the published
norms it is expected that only 9% of the population would score below the Low

1.4 Developmental Test of Visual-Motor Integration

Number of students, chronological age ranges, age equivalent score ranges,
and the frequency and percentage of students below chronological age are shown
in Table 5 for kindergarten through grade 7 on the VMI. Chronological ages are
reported in years and months (Y-M). This age is compared to the age equivalent
scores from the test results. For comparison purposes the "expected" age score
attained was deemed to be within one year of the student's actual chronological
age at time of testing. This is a rather crude determination given the fluctuation
in the range of scores that could occur in testing. However, the intention was to
detect trends. Previous studies had found Indian students to score below
placement (Conry & Conry, 1973a; Matthew & More, 1987; & More, 1984a). It
was thought that the age equivalent scores for the NIRS students might tend to
be below their respective chronological or "expected" ages. This was found to be
the case. Forty-seven percent of Kindergarten students' age equivalent scores
Table 5

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>n</th>
<th>Chronological Age Range</th>
<th>Age Equivalent Score Range</th>
<th>Number Below</th>
<th>Percentage Below</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>19</td>
<td>5.5 - 6.5</td>
<td>3.1 - 5.4</td>
<td>9</td>
<td>47.3</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>6.5 - 8.1</td>
<td>5.7 - 7.3</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>7.7 - 9.6</td>
<td>5.4 - 8.5</td>
<td>8</td>
<td>72.7</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8.6 - 9.6</td>
<td>5.7 - 9.11</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>9.9 - 13.7</td>
<td>7.3 - 12.7</td>
<td>8</td>
<td>66.7</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>10.10 - 12.2</td>
<td>6.9 - 10.5</td>
<td>5</td>
<td>83.3</td>
</tr>
</tbody>
</table>

*Assessment of age equivalent scores below expected was made on the basis of an attained age score one year below chronological age. Note 1. Total n = 63.

were one year below chronological age, three of eight (38%) grade 1 students were below their expected age equivalent. Only three grade 2 students obtained age equivalent scores comparable to chronological age, whereas the remaining eight (73%) students’ scores were one year below the expected age equivalent. In grade 3 (n = 7) only one student’s age equivalent was within one year of chronological age, while in grade 4 (n = 12) four students had satisfactory and eight (67%) had scores one year below chronological age. Finally, in grade 5, five of the six (83%) students’ scores were below chronological age.

1.5 Peabody Picture Vocabulary Test-Revised

The data reported in Table 6 for the PPVT-R is the same information that was provided in Table 5 for the VMI. Again a simple comparison was made between the age equivalent score obtained from PPVT-R results and each student’s actual chronological age. As with the VMI the "expected" age equivalent
Table 6

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>n</th>
<th>Chronological Age Range</th>
<th>Age Equivalent Score Range</th>
<th>Number Below</th>
<th>Percentage Below</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>19</td>
<td>5,5 - 6,5</td>
<td>3,1 - 6,6</td>
<td>16</td>
<td>84.2</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>6,5 - 8,1</td>
<td>5,4 - 7,4</td>
<td>1</td>
<td>12.5</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>7,7 - 9,6</td>
<td>4,3 - 9,4</td>
<td>7</td>
<td>63.6</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8,6 - 9,6</td>
<td>6,0 - 7,10</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>9,9 -13,7</td>
<td>6,10-11,3</td>
<td>11</td>
<td>91.7</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>10,10-12,2</td>
<td>7,9 -11,6</td>
<td>5</td>
<td>83.3</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>11,10-15,2</td>
<td>5,6 - 9,10</td>
<td>6</td>
<td>100.0</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>13,4 -15,3</td>
<td>7,11-10,2</td>
<td>6</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note. Total n=75. Age is in years and months.

score was defined as being within one year of the student’s age at time of testing. Using this standard, 16 of 19 (84%) Kindergarten students scored below their expected age equivalent. In grade 1 only one student scored lower than expected for chronological age. Seven of eleven grade 2 pupils had age equivalents below chronological age. Nearly all grade 3 pupils (6 of 7) scored lower than their expected age equivalents. In each of grades 4 and 5 there was one student whose reported age equivalent score exceeded chronological age with the remainder one year below. No student in either grade 6 or 7 attained the expected standard, that is an age equivalent score within one year of
chronological age at time of testing.

2. Vision and Hearing Problems

Nearly 40 percent (22/56) of students tested were identified as needing professional ophthalmic testing. An additional 11 students required subsequent teacher monitoring for vision problems. Sixteen students (21%) were found to require treatment for noticeable hearing loss.

B. RESULTS OF THE THESIS

1. Difference Between Sample Scores and the Norming Population

Scores for NIRS students were compared to those of the norming sample for each of the standardized tests. Score results are reported en masse by each test, rather than for each grade. Figures 2 to 6 include 95 percent confidence intervals for both the norming population and the Native Indian sample in terms of performance on four of five tests. Confidence intervals were not constructed for the VMI because an inspection of the technical manual failed to reveal means and standard deviations for each grade level.

The logic for calculating confidence intervals was to establish whether there were differences in mean scores between the NIRS sample and the population. The confidence interval around the population mean (designated by ‘P’ in the figures) was constructed by using a sampling distribution of the mean. That is, instead of showing a 95% confidence interval based on the population spread, the interval was based on a sample size equal to that of the NIRS sample. The interval thus represents the distribution of scores around the mean
that would occur with the typically small sample sizes for the NIRS students. The test for a significant difference between means requires a determination of whether the sample mean (designated by ‘S’ in the figures) falls within the confidence interval around the population.

In a similar fashion 95% confidence intervals were constructed around the sample means. These intervals are indicators of the possible distribution of means for samples of this size from the Native population.

1.1 Metropolitan Readiness Test

Figure 2 contains 95 percent confidence intervals of kindergarten students’ performance on the Metropolitan Readiness Reading Test. The calculations are based on raw scores of the pre-reading composites. The mean score for the NIRS group is much lower than that of the norming sample (46.5 vs. 63.7) and is outside the confidence interval, evidence of a statistically significant difference.

1.2 Peabody Picture Vocabulary Test

In Figure 3, 95 percent confidence intervals are displayed for the performance of kindergarten students through grade 7 for both the norming sample and the NIRS sample on the PPVT-R. Again, calculations are based on raw scores.

In kindergarten the NIRS mean score (47.79) was significantly lower than the norming sample (69.3) as shown in Figure 3. However, the NIRS mean for Grade 1 (76.0) is not significantly lower than the norm (79.3), since the estimate of the true sample mean falls within the norm sample confidence interval. For all grades 2 through 7, the NIRS means lie beyond the norming sample confidence interval, evidence of significant difference between means.

The results for the PPVT-R indicate that receptive vocabulary knowledge
Kindergarten Students

P - Population Mean
p - Confidence Interval for the Norming Population
S - Sample Mean
s - Confidence Interval for the Native Indian Sample

FIGURE 2. NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE METROPOLITAN READINESS TEST.
FIGURE 3 CONTINUED.
NINETY-FIVE PERCENT CONFIDENCE INTERVALS

P - Population Mean
p - Confidence Interval for the Norming Population
S - Sample Mean
s - Confidence Interval for the Native Indian Sample
for the NIRS students is substantially lower than that of the students in the norm sample. In only one instance (grade 1) did the NIRS mean fall within the confidence range of the norming sample. For the other seven grades the hypotheses of no between group differences must be rejected.

1.3 The Canadian Test of Basic Skills

Results of calculations for the Canadian Test of Basic Skills are shown in Figure 4. Comparison of confidence intervals between the NIRS and norm sample statistics are reported for grades 1 to 7 on the CTBS. The unit score reported is the total composite which is composed of the results of each subtest combined to give a total grade equivalent achievement score for each student. Student grade equivalent values were summed and averaged to produce a mean grade equivalent score.

For the first five grades there is no evidence of a statistically significant difference between groups. At every level the mean of the NIRS sample fell within the confidence interval of the norming sample. Moreover, for grades 1, 2, and 3 the NIRS mean score was higher than the mean reported in the test norms. The NIRS scores in grades 4 through 7 were lower than the norms, but only in grades 6 and 7 were the differences statistically significant. In these two instances the intervals were widely disparate and the NIRS sample mean was in excess of one and one half grade equivalents below the reported norm.

The trend detected in the preceding analysis is that of a decline in student achievement scores through the grades on the CTBS. Results from the lower grades were satisfactory with the mean grade equivalents in grades 1, 2, and 3 all higher than the norming sample means. However, a downward trend is evident in the intermediate grades, leading to a differential in excess of one
FIGURE 4. NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE CTBS COMPOSITE SCORE.
P - Population Mean
p - Confidence Interval for the Norming Population
S - Sample Mean
s - Confidence Interval for the Native Indian Sample

FIGURE 4 CONTINUED.
NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE CTBS COMPOSITE SCORE.
grade level by grade seven.

1.4 Canadian Cognitive Abilities Test (Primary)

Results for the CCAT-P are reported in Figure 5 along with population norm comparisons and calculated confidence intervals. The CCAT-P was administered to pupils in grade 1 and 2. There is only one mean score reported for each grade, since the primary level is not divided into subtests. Values are for raw scores, not deviation IQs. The mean for grade 1 was 54.9 which is close to the comparison mean (55.6) and within the range of the confidence interval for the norm sample, evidence that the difference was not statistically significant. However, the grade 2 results were evidence of a statistically significant difference. The mean for grade 2 student scores (47.3) was lower than the population mean (57.2) and outside the confidence range constructed around the norm sample mean.

1.5 Canadian Cognitive Abilities Test (Multilevel)

Mean scores from the administration of the CCAT-M to pupils in grades three through seven at NIRS are reported in Figure 6 along with the relevant norm sample means and calculated confidence intervals. The multilevel form of the CCAT is comprised of three components: Verbal, Nonverbal, and Quantitative. Comparisons are reported for all three subtests at each grade level.

In grade 3 the NIRS means for Verbal and Nonverbal subtests were below the norm sample means (49.4 vs. 55.0; and 44.6 vs. 56.0), whereas the NIRS mean exceeded that of the population on the Quantitative component (39.2 vs. 34.6). For each subtest the norm sample mean was contained within the range of the NIRS sample confidence interval. No inference may be made of a significant difference between groups at this grade level.
FIGURE 5. NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE CCAT (GRADES 1 AND 2).

P - Population Mean
p - Confidence Interval for the Norming Population
S - Sample Mean
s - Confidence Interval for the Native Indian Sample
FIGURE 6. NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE CCAT (GRADES 3 TO 7).
FIGURE 6 CONTINUED.
NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE NORMING POPULATION AND THE NATIVE INDIAN SAMPLE ON THE CCAT (GRADES 3 TO 7).
FIGURE 6 CONTINUED.
NINETY-FIVE PERCENT CONFIDENCE INTERVALS
FOR THE NORMING POPULATION AND THE NATIVE INDIAN
SAMPLE ON THE CCAT (GRADES 3 TO 7).

P - Population Mean
p - Confidence Interval for the Norming Population
S - Sample Mean
s - Confidence Interval for the Native Indian Sample
Quantitative scores were again the best subtest results in grade 4. The NIRS sample mean was contained within the norming sample confidence interval, although the NIRS mean was lower (35.5 vs. 37.4). Verbal and Nonverbal scores too were lower (44.1 vs. 58.8; and 50.4 vs. 57.8) and beyond the range of the confidence interval, evidence of a statistically significant difference.

In grade 5 the NIRS mean was contained within the norming sample confidence interval for both the Quantitative and Verbal subtests, evidence that NIRS sample results were not significantly different. However, on the Nonverbal subtest the NIRS mean (43.5) was lower than the norming sample mean (58.0) and below the range of the confidence interval, suggestive of a statistically significant difference. Native Indian nonverbal performance measures are commonly equal or superior to other groups. McShane and Plas (1984) discussed the nonverbal developmental backgrounds of Indian, from which one might expect higher performance on the nonverbal components of tests. This was not found for NIRS students.

The mean scores for the NIRS group were lower than the norm means for all three subtests in grade 6, and the Verbal and Nonverbal NIRS mean scores were below the confidence interval of the norm indicating a statistically significant difference. However, the Quantitative mean score for the NIRS group fell within the confidence interval.

In grade 7 the NIRS mean on the Nonverbal subtest (47.6) was lower than the norming sample mean (56.5) but not below the confidence interval. However, the lower NIRS means on the Verbal and Quantitative subtests were both significantly different from the norms.
2. Score Decrements in Higher Grades

To assess achievement trends, scores for all NIRS subjects were compared on the PPVT-R, VMI, and CTBS. Figures 7, 8, and 9 are graphic representations of students in each grade who attained achievement scores commensurate with chronological age or grade placement. The PPVT-R and the VMI age equivalent scores were compared to chronological age, and the CTBS composite grade equivalent scores were compared to placement. Equivalent scores were readily available from the results of the survey of learning problems. These scores were used to increase comparability across tests. Grade and age norms fit well with the graduated progression in school and in intellectual development. The age and grade equivalents give the teacher an indication of student progress and are therefore of use when it is necessary to interpret test scores (Nunnally, 1978). "Achievement commensurate with chronological age" was defined as an age equivalent score within one year of actual age; "achievement commensurate with grade placement" was defined as a grade equivalent score within one school year of placement at time of testing. The one year criterion was chosen for the following reason. It was necessary to compare student performances between tests. This was made possible by a rationale developed from the Peabody Picture Vocabulary Test-Revised. The 95% confidence interval around a standard score equivalent for the PPVT-R is ± 14, based on a standard error of measurement of 7 (Dunn & Dunn, 1981, p. 40). Translated to age equivalent scores this interval would correspond to ± one year. That is, for a student in grade 5.5 an acceptable range of his true score would be from a grade equivalent of 4.5 to 6.5. Therefore, to be confident that an obtained PPVT-R age equivalent score was below chronological age, the difference should be one year. Since this
FIGURE 7. PERCENTAGES OF STUDENTS ATTAINING
ACHIEVEMENT SCORES COMMENSURATE WITH CHRONOLOGICAL
AGE ON THE PPVT-R
FIGURE 8. PERCENTAGES OF STUDENTS ATTAINING ACHIEVEMENT SCORES COMMENSURATE WITH CHRONOLOGICAL AGE ON THE VMI
FIGURE 9. PERCENTAGES OF STUDENTS ATTAINING ACHIEVEMENT SCORES COMMENSURATE WITH GRADE PLACEMENT ON THE CTBS
criterion is available for the PPVT-R, it was adopted also for the other tests in order to have a similar criterion for comparison. Figures 7, 8, and 9 show the percentages of students at each grade level whose achievement is within one year of age or placement.

2.1 PPVT-R vs. Chronological Age

Using the standard of an expected age equivalent score being not less than one years below chronological age, approximately 22% (16/74) of NIRS pupils met or exceeded their expected performance according to age on the PPVT-R. As shown in Figure 7, the largest percentage of these pupils were in grade 1 (88%). Three of these pupils were in kindergarten (16%), three in grade 2 (27%), one in grade 3 (14%), one in grade 4 (8%), and one in grade 5 (17%). No pupil in grades 6 or 7 attained a PPVT-R age equivalent within one year of chronological age. This represents a dramatic trend of increasing performance decrements across the eight year span in the schools.

2.2 VMI vs. Chronological Age

Following the one-year comparison standard used with the PPVT-R, 24 of 62 pupils (39%) tested on the VMI met or exceeded their expected age equivalent scores: 10 of these children were in kindergarten (53%); five in grade 1 (63%); three in grade 2 (27%); one in grade 3 (14%); four in grade 4 (33%); and one in grade 5 (17%). Figure 8 shows that the majority of students with satisfactory performance are in the lower grades. The percentages of students with lower scores increased through the grades. Four of the fourteen pupils who met or exceeded expected age equivalent scores also met or exceeded age equivalent scores on the PPVT-R. There was at least one child in each grade achieving an age equivalent equal to chronological age.
2.3 CTBS vs. Grade Level

As shown in Figure 9, the percentage of students attaining commensurate achievement scores is higher for the CTBS than the other two tests in the primary grades. Again, for comparison purposes, one year is the criterion. Of the 54 pupils tested on the CTBS, 25 met the one year criterion. Although most pupils in the primary grades (1 through 3) met this standard, the proportions dropped in the intermediate grades. The decrease is sharp after grade 3 with fewer and fewer students reaching expected achievement levels until in grades 6 and 7, the CTBS percentages are similar to those of the PPVT-R. The standard was attained by all pupils in grade 1, nine (82%) in grade 2, six (86%) in grade 3, four (33%) in grade 4, two (33%) in grade 5, and none in grades 6 or 7. The satisfactory primary performance is offset by less adequate results in the higher grades.

All students identified as having adequate attainment on the PPVT-R and VMI also demonstrated adequate performance on the CTBS. There was one seeming inconsistency: a grade 4 student whose CTBS grade equivalent of 3-1 was more than one and one half years below placement showed good performance on the VMI (nine months above chronological age). The generally low CTBS composite result might be explained by the fact that language ability is a major constituent of this score. On the PPVT-R -- which assessed receptive vocabulary -- this pupil earned an age equivalent score four years below chronological age. Thus, the test results lend corroborative support to one another in the identification of students whose performance was not commensurate with grade placement and/or chronological age.
3. Age-Grade Retardation

NIRS student ages at time of testing were analyzed as a key to whether they were behind in grades relative to others. Using the premise that student age upon entering kindergarten is between 4 years 6 months and 5 years 6 months, the expectation is that the range of ages would be five to six half-way through the year. In each successive year, the corresponding ages would raise by one year, with a range of six to seven years in grade 1 and seven to eight years in grade 2. Note, that the timing is stated as half-way through the school year. This is a slightly conservative estimate since testing actually occurred somewhat later, in the spring.

An inspection of students’ records revealed that students in all grades were older than the anticipated range. For the students as a group, 55.4% were older than the expected age range for the grade in which they were placed. This is an indication that grades have been repeated, or that the student performance is in some way lower than expected for age and/or grade.

In grade by grade analysis the percentages of students older than the range for placement were: kindergarten, 21%; grade 1, 63%; grade 2, 91%; grade 3, 17%; grade 4, 92%; grade 5, 33%; grade 6, 50%; and grade 7, 83%. With the exception of students in kindergarten and grades 3 and 5, the majority of students were older than the usual cohort for the grade. The high percentage in grade 4 (92) is indicative of special problems appearing at this grade. The transition form primary to intermediate grades where instruction relies more heavily on reading, may be responsible.
4. Ability Level Differences in Achievement

To answer the question "Did scores vary with indicators of ability level?", age or grade equivalent scores on the PPVT-R, VMI, CTBS were compared for students in each of the CCAT ability levels. The graphs in Figure 10 chart the percentage of students in each CCAT ability category whose test scores were not commensurate with their age and/or their grade. Figure 10 contains bar graphs representing the percentages of students in each ability category who did not achieve equivalent scores on the VMI, PPVT-R, or CTBS within one year of age or placement. The CCAT ability levels used in this figure are slow learner (68-78), low average (79-88), average (89-110), and high average (111-120).

In grades 1 and 2 the CCAT score is a single measure, whereas from grades 3 to 7 it is a composite of three subtests. Students were included in the comparison only if they had been tested on at least two of the measures. The total number of students in the sample is 51, since kindergarten pupils were not tested with this instrument.

4.1 Equivalent Scores Within One Year Of Age or Placement

From Figure 10 it can be seen that, twelve percent of slow learners obtained age equivalents within one year of chronological age on the VMI, whereas none did on the PPVT-R. More than half of the slow learners' grade equivalent scores on the CTBS were within one year of placement.

Students' equivalent scores tended to correspond with their ability groupings. The average students were most likely to achieve scores near their age or placement. Conversely, the slow learners were least likely to achieve scores within one year of age or placement. One consistent finding across ability levels was better performance on the CTBS. Students from all CCAT categories
RESULTS / 80

SLOW LEARNERS (68-78)

LOW AVERAGE (79-88)

AVERAGE (89-110)

HIGH AVERAGE (111-120)

FIGURE 10. PERCENTAGE OF STUDENTS 1 YEAR BELOW EXPECTED AGE OR GRADE LEVEL ON THE VMI, PPVT-R, AND CTBS FOR EACH CCAT ABILITY
were more likely to attain satisfactory levels of performance on this test than the VMI or PPVT-R.

5. Achievement of Higher Ability Students

"Was achievement for higher ability students lower than expected?" This is one of the components of the subsidiary research question 2C. For the NIRS sample, students categorized as either average or high average on the CCAT were considered the higher ability group. Table 7 includes the numbers and percentages of students categorized as high average or average who attained scores more than one year below grade level on the CTBS. The percentage of students with satisfactory performance (within one year of placement) in each grade is also shown.

The primary grades (1 to 3) had greater proportions of students categorized as average or high average than the intermediate grades. Only 4 of the 17 higher ability students (24%) were in the grade range 4 to 5. No student in grade 6 or 7 was categorized as higher ability. There was little evidence of achievement deficiency in grades 1 or 2, with only one student in each grade scoring one year below grade level. None of the higher ability students in grades 3, 4, and 5 had CTBS scores one year below expected grade level.

In summary, two of 17 average or high average students achieved CTBS grade equivalent scores more than one year below placement. Therefore, according to the CTBS results, the achievement of the higher ability students was satisfactory.
### Table 7

CTBS Achievement of Students with Average or High Average CCAT Ability Ratings

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>n</th>
<th>f</th>
<th>n Below Expected</th>
<th>% Below Expected</th>
<th>% At Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The expected grade equivalent score is based on when testing occurred during the school year. For example, 1.7 would be the expected grade equivalent score for students having completed seven months of grade 1. To be rated as below the grade equivalent achieved must be more than one year below placement.

### 6. Language Scores Relative to Other CTBS Components

The results of the CTBS, (both composite and subtest scores) were examined to answer question 2D "Were language scores relatively lower than other component test scores?" Since the CTBS composite is comprised of subtest scores, systematically lower scores in one subject could result in somewhat lower composites. Evidence of the effect of language scores would be lower achievement on the Language subtests, according to expected scores. A high percentage of scores below the composite, would constitute evidence of impairment. Of specific
interest is the percentage of Language scores below the composite, and to a lesser extent, the percentage of Reading scores. However, with the one year criterion no student had any CTBS subtest scores one year below the composite. Since reading is a component of language ability, lower scores on the reading subtest may also be a reflection of language difficulty. However, none of the students' Reading subtest scores were more than one year below the composite. Therefore, the conclusion reached is that language subtest scores were not lower relative to the other CTBS components.

7. Effects of Perceptual Problems on Test Scores

Grade level, number of students in each grade, number of students with perceptual problems, and frequencies and percentages of students with perceptual problems who scored more than one year below age or placement on the VMI, PPVT-R, and CTBS are listed in Table 8. In the following discussion the results of this comparison is reported for each grade level.

Nearly 32 percent of the pupils in kindergarten, (six of nineteen) were diagnosed with hearing difficulties. Two of these children's equivalent scores were below chronological age on the VMI, three on the PPVT-R. On both the VMI and PPVT-R 33% of the students with perceptual problems were at least two years behind. Compared to kindergarten results as a whole, where most children attained satisfactory scores, this performance is lower. For kindergarten students, there is evidence that hearing problems do affect test scores.

Results are less conclusive for the four grade 1 pupils with hearing problems, one of whom was also found to have impaired vision. On the CTBS all of these students obtained satisfactory grade equivalent scores. One of four
### Table 8

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>n</th>
<th>Incidence Perceptual Problems</th>
<th>VMI</th>
<th>PPVT-R</th>
<th>CTBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>K</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>83</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>5</td>
<td>**</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td>**</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>76</td>
<td>41*</td>
<td>28</td>
<td>34</td>
<td>83</td>
</tr>
</tbody>
</table>

* Total number of students with hearing and/or vision problems.
** Not tested.

The performance for grade one students with perceptual problems was better than in other grades.

None of the grade 2 children had vision problems, but seven of eleven, or nearly 64 percent were diagnosed with hearing difficulty. On the VMI, 86% were one year below and 71% were two years below age. On the PPVT-R 75% were one year and 43% were two years below age. There is evidence here of a
The incidence of hearing problems was high for grade 3 (5 of 7; 71%), and two of the five were also diagnosed with vision difficulty. However, none of the CTBS composite scores was more than one year below placement for these students. On the VMI 100% were at least one year below age level and 60% were two years below. The corresponding figures for the PPVT-R were 80% and 20%. These lower scores may be attributable to the perceptual difficulties of these children.

Test score results were low on all three tests for the grade 4 students. There was also a high incidence of problems with perception (eight of twelve pupils). On the VMI, 75% of the students with perceptual problems were one year behind and 63% were two years behind. On the PPVT-R all students' age equivalent scores were at least two years below actual age.

Assessment identified one pupil with hearing impairment, and two with vision problems in Grade 5. All three students' VMI equivalents were more than one year below chronological age and two of the three were two years behind. Two students scored one year below placement on the CTBS. On the PPVT-R, both students' scores were two years below their chronological age. The combined incidence of perception problems was 50 percent (3 of 6) for this grade. This is suggestive of impaired performance related to perception.

Five of six students in grade 6 (83%) had hearing or vision problems. None of them achieved at a grade 5 level on the CTBS, one was below grade 4. All were at least two years below their chronological age on PPVT-R age equivalents. These results constitute evidence of perception-related score impairment.
Three children had perceptual problems in grade 7. Two had both types of difficulties. On the CTBS, grade equivalent scores were 1, 2, and 3 years below placement. On the VMI age equivalents were 4, 5, and 6 years below chronological age. Again, lower test performance can be related to hearing and vision difficulties.

The incidence of vision and hearing problems for the total group was 58% and 21%, respectively. This is similar to rates reported by McShane and Plas (1982) who found hearing loss ranging from 20 to 76 percent in Native Indian samples. The comparable figure for the general population is five percent. Thus, in the NIRS sample, hearing loss was found to be four times that of the general population. This, combined with lower test results, especially in the higher grades, is evidence that test performance may well be affected by hearing or vision problems.

8. PPVT-R Item Difficulty Rankings for the Sample

NIRS students' responses were analyzed to calculate rank order item difficulties. The logic of this determination is that two different groups should produce similar item difficulty rankings, if the same construct is being measured (Jensen, 1982). This comparison could be made only for the PPVT-R, because it is a test that contains items arranged in order of ascending difficulty. The expectation was that if the PPVT-R items were found to be sequenced in ascending difficulty, an inference could be made that the test was appropriate for the NIRS sample. Conversely, evidence that PPVT-R items were not sequenced in ascending difficulty, an inference could be made that the content was inappropriate for the NIRS sample.

Item and test analysis (LERTAP: Nelson, 1974) were used on students'
responses to the PPVT. In order to obtain a set of items that had been answered by a sizeable portion of the students, 60 sample items (41 - 100) were selected. These items would be the main constituents of the test for students in middle grades and there would also be overlap from students in the lower and higher grades. The total number of items in the PPVT-R is 175. In this administration no student responded to an item past 114. Therefore, the number of items excluded from the analysis was 54 including items one to 40. Every actual student response was included. Following this procedure the number of responses per item ranged from a low of 23 to a high of 53 on the 60 items selected.

The procedure described in Chapter 3 to identify misranked items was to compare, the test’s presentation order with ranked item difficulty indices. A discrepancy of more than eight positions was said to constitute misranking. An examination of Table 9 will show the PPVT-R item number, calculated difficulty index, presentation rank, and calculated rank. In the presentation ranks, the numeral 1 indicates the first item in the series, which should be the easiest, and therefore should have the highest index of difficulty. The numerals in the computed rank column follow this same procedure. For instance item 41 is presented first, but on the basis of the item difficulty index of 0.77, the calculated rank was twenty-three; this therefore was the 23rd easiest item. Asterisks beside numbers in the computed rank column were used to indicate misranking.

From Table 9, it can be seen that 39 of 60 items or 65% displayed were misranked. This constitutes a marked difference from the rankings determined by item difficulties in the norming population. There was a low
Table 9

PPVT-R Form L Item Difficulty Indices for Items 41 to 100

<table>
<thead>
<tr>
<th>Item</th>
<th>n</th>
<th>Difficulty Index</th>
<th>Norms Rank</th>
<th>NIRS Rank</th>
<th>Item</th>
<th>n</th>
<th>Difficulty Index</th>
<th>Norms Rank</th>
<th>NIRS Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>30</td>
<td>.77</td>
<td>1</td>
<td>23*</td>
<td>71</td>
<td>45</td>
<td>.11</td>
<td>31</td>
<td>59*</td>
</tr>
<tr>
<td>42</td>
<td>30</td>
<td>.57</td>
<td>2</td>
<td>43*</td>
<td>72</td>
<td>44</td>
<td>.88</td>
<td>32</td>
<td>3*</td>
</tr>
<tr>
<td>43</td>
<td>30</td>
<td>.80</td>
<td>3</td>
<td>18*</td>
<td>73</td>
<td>44</td>
<td>.86</td>
<td>33</td>
<td>8*</td>
</tr>
<tr>
<td>44</td>
<td>29</td>
<td>.76</td>
<td>4</td>
<td>26*</td>
<td>74</td>
<td>44</td>
<td>.81</td>
<td>34</td>
<td>16*</td>
</tr>
<tr>
<td>45</td>
<td>29</td>
<td>.79</td>
<td>5</td>
<td>20*</td>
<td>75</td>
<td>44</td>
<td>.75</td>
<td>35</td>
<td>28</td>
</tr>
<tr>
<td>46</td>
<td>32</td>
<td>.78</td>
<td>6</td>
<td>21*</td>
<td>76</td>
<td>43</td>
<td>.80</td>
<td>36</td>
<td>17*</td>
</tr>
<tr>
<td>47</td>
<td>32</td>
<td>.40</td>
<td>7</td>
<td>56*</td>
<td>77</td>
<td>43</td>
<td>.86</td>
<td>37</td>
<td>7*</td>
</tr>
<tr>
<td>48</td>
<td>32</td>
<td>.75</td>
<td>8</td>
<td>27*</td>
<td>78</td>
<td>45</td>
<td>.72</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td>49</td>
<td>31</td>
<td>.26</td>
<td>9</td>
<td>58*</td>
<td>79</td>
<td>51</td>
<td>.60</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td>50</td>
<td>34</td>
<td>.85</td>
<td>10</td>
<td>9</td>
<td>80</td>
<td>51</td>
<td>.84</td>
<td>40</td>
<td>12*</td>
</tr>
<tr>
<td>51</td>
<td>36</td>
<td>.61</td>
<td>11</td>
<td>39*</td>
<td>81</td>
<td>53</td>
<td>.61</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>52</td>
<td>36</td>
<td>.69</td>
<td>12</td>
<td>34*</td>
<td>82</td>
<td>52</td>
<td>.94</td>
<td>42</td>
<td>1*</td>
</tr>
<tr>
<td>53</td>
<td>36</td>
<td>.78</td>
<td>13</td>
<td>22*</td>
<td>83</td>
<td>51</td>
<td>.63</td>
<td>43</td>
<td>38</td>
</tr>
<tr>
<td>54</td>
<td>37</td>
<td>.57</td>
<td>14</td>
<td>44*</td>
<td>84</td>
<td>52</td>
<td>.88</td>
<td>44</td>
<td>2*</td>
</tr>
<tr>
<td>55</td>
<td>36</td>
<td>.67</td>
<td>15</td>
<td>35*</td>
<td>85</td>
<td>51</td>
<td>.73</td>
<td>45</td>
<td>29*</td>
</tr>
<tr>
<td>56</td>
<td>35</td>
<td>.63</td>
<td>16</td>
<td>37*</td>
<td>86</td>
<td>51</td>
<td>.86</td>
<td>46</td>
<td>6*</td>
</tr>
<tr>
<td>57</td>
<td>34</td>
<td>.76</td>
<td>17</td>
<td>25</td>
<td>87</td>
<td>51</td>
<td>.53</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>58</td>
<td>34</td>
<td>.41</td>
<td>18</td>
<td>55*</td>
<td>88</td>
<td>50</td>
<td>.65</td>
<td>48</td>
<td>36*</td>
</tr>
<tr>
<td>59</td>
<td>35</td>
<td>.43</td>
<td>19</td>
<td>54*</td>
<td>89</td>
<td>49</td>
<td>.62</td>
<td>49</td>
<td>41</td>
</tr>
<tr>
<td>60</td>
<td>35</td>
<td>.77</td>
<td>20</td>
<td>24</td>
<td>90</td>
<td>48</td>
<td>.49</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>61</td>
<td>36</td>
<td>.69</td>
<td>21</td>
<td>33*</td>
<td>91</td>
<td>48</td>
<td>.56</td>
<td>51</td>
<td>45</td>
</tr>
<tr>
<td>62</td>
<td>36</td>
<td>.56</td>
<td>22</td>
<td>46*</td>
<td>92</td>
<td>46</td>
<td>.44</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>63</td>
<td>47</td>
<td>.81</td>
<td>23</td>
<td>15</td>
<td>93</td>
<td>43</td>
<td>.48</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>64</td>
<td>49</td>
<td>.88</td>
<td>24</td>
<td>4*</td>
<td>94</td>
<td>39</td>
<td>.49</td>
<td>54</td>
<td>49</td>
</tr>
<tr>
<td>65</td>
<td>49</td>
<td>.84</td>
<td>25</td>
<td>11*</td>
<td>95</td>
<td>36</td>
<td>.46</td>
<td>55</td>
<td>52</td>
</tr>
<tr>
<td>66</td>
<td>49</td>
<td>.86</td>
<td>26</td>
<td>5*</td>
<td>96</td>
<td>34</td>
<td>.08</td>
<td>56</td>
<td>60</td>
</tr>
<tr>
<td>67</td>
<td>48</td>
<td>.79</td>
<td>27</td>
<td>19</td>
<td>97</td>
<td>31</td>
<td>.85</td>
<td>57</td>
<td>10*</td>
</tr>
<tr>
<td>68</td>
<td>47</td>
<td>.83</td>
<td>28</td>
<td>13*</td>
<td>98</td>
<td>29</td>
<td>.21</td>
<td>58</td>
<td>57</td>
</tr>
<tr>
<td>69</td>
<td>46</td>
<td>.70</td>
<td>29</td>
<td>32</td>
<td>99</td>
<td>26</td>
<td>.72</td>
<td>59</td>
<td>30*</td>
</tr>
<tr>
<td>70</td>
<td>47</td>
<td>.81</td>
<td>30</td>
<td>14*</td>
<td>100</td>
<td>23</td>
<td>.46</td>
<td>60</td>
<td>51*</td>
</tr>
</tbody>
</table>

* Items misranked: frequency = 39 percentage = 65

r = .16 (Spearman r)
The correlation between the sequence of difficulty for the NIRS students and the presentation order (Spearman’s $r = 0.13$). The difficulty of these vocabulary items was much different for this aggregated group of native Indian students than for the population.

Subsequent computations were conducted for item difficulty rankings on the PPVT-R for students in each grade. Table 10 is a presentation of item difficulty indices by grade for items 41 through 70 on the PPVT-R. By inspection of Table 10, it can be seen that initial items were the ones with the highest indices of difficulty. This is the case for each grade. Item difficulty indices were highest for the first items answered by the majority of the students in each grade. Item difficulty levels tended to decline from that point indicative of increasing difficulty for students in the sample. This would conform with the PPVT-R presentation order. These results would cast doubt on the initial assessment of misranking. The suggestion is that the misranking evident in the group analysis is more of a function of aggregating the results of heterogeneous age groups, rather than truly anomalous difficulty rankings.

Point biserial correlations between items and the total score correct are evidence of a reliable test when the value of the correlation exceeds 0.30 (Nunnally, 1978). The number of items presented in Table 10 with coefficients of 0.30 or higher was 31 of 60 (51.7%). Therefore, more than half of the items had acceptable levels of reliability for this test. In addition, 18 of the items had point biserial correlations with the total score correction excess of 0.50.

Discrimination indices of 0.30 or lower were found for 29 of the 60 items (48.3%). Items of extreme difficulty identified by indices of 0.15 or lower were found for only two of 60 items (3.0%). The 10 items that were the most
Table 10

PPVT-R Item Difficulty Indices for Items 41 to 70 by Grade

<table>
<thead>
<tr>
<th>Item</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>0.63</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>42</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>43</td>
<td>0.68</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>44</td>
<td>0.67</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>45</td>
<td>0.72</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>46</td>
<td>0.66</td>
<td>0.83</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>47</td>
<td>0.22</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>48</td>
<td>0.55</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>49</td>
<td>0.18</td>
<td>0.17</td>
<td>0.25</td>
<td>0.67</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>0.65</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>51</td>
<td>0.47</td>
<td>0.67</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>52</td>
<td>0.35</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>53</td>
<td>0.81</td>
<td>1.00</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>54</td>
<td>0.37</td>
<td>0.86</td>
<td>0.75</td>
<td>0.00</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>55</td>
<td>0.64</td>
<td>0.57</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>56</td>
<td>0.36</td>
<td>0.80</td>
<td>0.75</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>57</td>
<td>0.47</td>
<td>0.86</td>
<td>0.75</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>58</td>
<td>0.31</td>
<td>0.86</td>
<td>0.25</td>
<td>0.00</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>59</td>
<td>0.38</td>
<td>0.42</td>
<td>0.50</td>
<td>0.80</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>60</td>
<td>0.36</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>61</td>
<td>0.50</td>
<td>0.86</td>
<td>0.60</td>
<td>0.71</td>
<td>0.75</td>
<td>0.50</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>62</td>
<td>0.67</td>
<td>0.86</td>
<td>0.33</td>
<td>0.71</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>63</td>
<td>0.44</td>
<td>0.75</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>64</td>
<td>0.38</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>65</td>
<td>0.50</td>
<td>0.63</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>66</td>
<td>0.63</td>
<td>0.75</td>
<td>0.91</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>67</td>
<td>0.43</td>
<td>0.88</td>
<td>0.73</td>
<td>0.86</td>
<td>0.86</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>68</td>
<td>0.43</td>
<td>0.75</td>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>69</td>
<td>0.33</td>
<td>1.00</td>
<td>0.80</td>
<td>0.57</td>
<td>0.71</td>
<td>0.67</td>
<td>0.33</td>
<td>1.00</td>
</tr>
<tr>
<td>70</td>
<td>0.20</td>
<td>1.00</td>
<td>0.70</td>
<td>0.71</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

difficult for the NIRS sample follow: (the number in brackets is the test item number)

1. frame (47)
2. faucet (49)
3. tambourine (58)
4. disappointment (59)
5. casserole (71)
6. isolation (92)
7. adjustable (95)
8. fragile (96)
9. appliance (98)
10. blazing (100)

An interesting note is that the two most difficult words for the NIRS students were "fragile" and "faucet".

9. Underrepresentation of Canadian Native Indians in the Norm Reference Groups

Test manuals were examined to answer the question, "Are Canadian Native Indians underrepresented in the norm reference group for the five tests?"

The PPVT-R manual reports a U.S. norming group of 4,200, of which 1.2 percent were categorized as "Other". This group consisted of "Indian, Japanese, Chinese, Filipino, and all other races not categorized as black, white, or Hispanic" (Dunn & Dunn, 1981, p. 18). Therefore, the number of American Indians included in the norming sample was very small. Similarly, the number of native American Indians reported in the MRT manual was only half of one percent, or 96 out of 17,852 in the norming group (Nurss & McGauvran, 1976, p. 23).

The CTBS and CCAT tests were normed at the same time. The group used was a stratified random sample of the English-speaking Canadian school
RESULTS / 92

population (King, 1974, p. 45). Attempts were made to ensure that the proportions of backgrounds in the norming groups were representative of the general population. Breakdowns by race or ethnic group were not reported (Ibid., p. 45). From the results of the 1981 norming study of the VMI for 3,090 children, less than one percent of the score variance was deemed to be attributable to ethnic background (Beery, 1982, p. 17). This absence of cultural bias was confirmed by Price (1980), who found no significant difference between American Indian and non-Indian Kindergarten students on VMI test scores.

10. Potential Examiner and/or Language Bias

The fact that the NIRS assessment team were not Native Indian is a factor to be considered. However, since there are no comparison studies, it is difficult to determine whether or not this is a serious limiting factor. Ideally, examiners would be of the same linguistic and cultural background as the students to be tested. However, there is also the necessity that these examiners be competent in test administration procedures. The best match in cultural background between the evaluation team and the NIRS students would have been well trained Canadian Indian examiners. Unfortunately, there are currently very few of these, and if students are to be tested, it is important to use trained examiners.

It is not possible to determine the effect of the cultural difference between the examiners and the students tested. If a student gives an incorrect response to a question, then it is possible that language may have impeded understanding. The students in this sample all spoke English as their first language, but their dialect was non-standard. They may not have understood the problems. They
may have misunderstood directions. There is also the possibility the students did not assert themselves to report instances when they failed to comprehend.

C. CHAPTER SUMMARY

In this chapter, analyses of test scores for the NIRS pupils were presented in relation to comparable population parameters. These results constitute evidence that the native Indian pupils performance was generally below expected according to the norms established by test authors. Note was also made of trends in performance according to subject area at different grade and ability levels. The most important result of the analysis of this chapter is the finding that for every three years of elementary schooling, the NIRS sample fell one year behind the population as represented by the norm sample group. Another notable finding concerns the ability ratings as determined by CCAT assessments. The CCAT was found to be an accurate predictor of achievement in school. A third finding related to schooling is the prevalence of hearing problems. The incidence of hearing difficulty among the NIRS students was found to be four times the national average. Finally, preliminary analysis of the PPVT-R which is a widely-used language ability measure, found a different order of item difficulty for this sample. However, this evidence was found to be unreliable because of the procedure used, which aggregated results of all students across age levels. Analysis of grade grouping results seemed to refute this finding. When item difficulty indices were calculated for homogeneous grade groups, the difficulty progression intended by the test authors became apparent. The first items answered by students in each grade had high difficulty indices. The inference is therefore, that this sample of students found the first portion of the test to be
the easiest. Confirmation of increasing difficulty level leads to an inference that
the PPVT-R functioned as intended for these students.
CHAPTER V. CONCLUSIONS

Chapter 5 begins with a summary of the thesis. Next the completed judgment matrix is presented. In it the previously mentioned research questions and criteria are compared with the results from Chapter 4.

The next section of the chapter is a discussion of judgments derived from the research questions addressing the problems of using standardized tests for an ethnic minority group and the particular appropriateness of the tests used with the NIRS students. In the last two sections of the chapter the limitations of the thesis are described along with implications for future study.

A. SUMMARY OF THE THESIS

The Native Indian Reserve School (NIRS) is the fictitious name of an Indian reserve school on the southwest of Vancouver Island in British Columbia operated under the aegis of a board of trustees comprised of Native Indians. The purpose of the school is to serve the educational needs of the local children. To this end, a comprehensive assessment of student ability and achievement was contracted to a team of university-based evaluators in 1986. Assessment results were to be used to aid educational planning including the creation of individualized education programs.

1. Purpose

This thesis is an outgrowth of that assessment and was designed to evaluate the appropriateness of a multiple assessment battery administered to a uniracial ethnic minority.

The three following secondary research questions were designed to deal
with various aspects of the problem:

1. Does the battery achieve its goal of identifying children who are in need of educational assistance?

2. Does the battery reveal the type and magnitude of assistance required?

3. Has bias had an effect on the measurement of scores for this sample? Does the use of a multi-dimensional battery lead to a biased assessment when used in such a sample?

In order to answer these research questions, 11 subsidiary questions were derived as outlined in the judgment model of Chapter 3.

2. Methodology

The research methodology of this thesis involved the manipulation of the data obtained from the administration of five standardized tests and two tests of perceptual acuity. The evaluation team reported standardized test results for 76 NIRS students from grades kindergarten through grade 7.

To answer the research questions, quantitative data was analyzed for individual students and also for the group. Descriptive group statistics were compared to parameters derived from published norms. Obtained information was summarized in terms of frequencies, percentages, means, and standard errors of measurement.
B. JUDGMENTS

1. Does the Battery Achieve Its Goal of Identifying Children Who are in Need of Educational Assistance?

1a. What Were The NIRS Student Scores On The Battery Of Tests Relative To The Population Norms?

Group mean scores were below expected values when age equivalent scores were compared to chronological age, and grade equivalent scores compared to grade placement. These results are consistent with the findings of More (1984) in his review of Indian education. More commented that in all studies reviewed Indian students were behind non-Indians. In particular, a comparison between grade equivalent scores obtained by NIRS students on four skill areas of the CTBS and actual grade placement showed that the mean sample scores were low (see Table 3). Reading and math achievement were below placement in every grade after one. Vocabulary and language grade equivalent scores were below placement in all grades. All children in the sample, except for, one were classified in the CCAT ranges average or below.

1b. What Was The Prevalence Of Vision And Hearing Problems?

Approximately 40% of the students were diagnosed with vision problems and an additional 14% were recommended as being in need of subsequent teacher monitoring for possible vision problems. Twenty-one percent of the sample suffered from hearing defects. These findings are in line with reports by McShane and Plas (1982) that the incidence of hearing loss due to middle ear disease ranges between 20 and 76% for Native populations.

Judgment -- The battery of instruments was useful in specifying the range
of student ability and achievement. From the test scores it was possible to determine the children with satisfactory performance and those in need of educational assistance. It was also possible to determine the subject areas in which achievement was below expected for the corresponding grade level on the CTBS. However, research has shown (Deyhle, 1986) that Native Indian students may not realize that their best effort is required on a test. At the very least, the assessment was able to identify students in need of assistance in responding to tests. A determination with profound overtones for learning, was the utility of the screening for hearing and vision problems. Identifying perceptual difficulties is the first step in a plan of educational assistance.

2. Does the Battery Determine the Type and Magnitude of Assistance Required?

2a. Compared To Test Norms, Did NIRS Scores Differ?

The mean scores for the NIRS sample were lower than the population norms and for the majority of the test measures. The NIRS mean scores were below the confidence interval constructed around the population norms. The conclusion drawn was that the difference between the NIRS sample and the norming sample was statistically significant.

However, comparison with norms is not the best means of comparisons. More (1984b) suggests that norms from standardized tests may be misunderstood as standards to be achieved. Published norms represent the average score of the student in the standardization sample which is usually underrepresentative of Native Indians. Ideally, norm comparisons would be with other Native Indian groups, preferably from the same nation. Therefore, some caution should be used
in the interpretation of these lower mean scores for NIRS students.

2b. Did Achievement Decrease In Higher Grades? What Was The Incidence Of Age-Grade Retardation?

Equivalent scores on the PPVT-R, VMI, and CTBS were compared across grade levels. In grade 1 the percentage of students obtaining scores comparable to their age and grade ranged from 63 to 100 per cent. However, a decline begins immediately. In grade 6 the percentage is zero on the PPVT-R and CTBS. The VMI was not given in these grades. In the Lillooet area evaluation (Matthew & More, 1987) 59% of grade 3 students were at least one year behind. The corresponding figure for grade 8 was 78%.

More than half (55.4%) of all students were older than the expected range of ages for the grade in which they were placed. This finding is supported by More (1984a) who found 45% of Okanagan Nicola students at least one year behind, and Conry and Conry (1973a) who found age-grade retardation to progress from lower to higher grades. Seven per cent of kindergarten students were older than expected rising to 95% in grade 8. An explanation of the school selection procedure is necessary to explain, in part, the poorer performance of students in higher grades. At the school, it has been a fairly common occurrence to send students to the local public school if they have been identified as having higher ability. Thus the upper grades of NIRS become a catchment for the students of lower ability, and their performance may not be representative of other Native Indian students in these grades. Also, some of the students in grades 6 and 7 could be those who have transferred back to NIRS after having experienced difficulties in the public school system. Again, the result is that the NIRS classes tend to retain those students whose achievement has been lower.
The resulting affect is that the age/grade retardation noted here may seem worse than it is for Native Indian students as a whole.

2c. Did Scores Vary With The Indicators Of Ability Level?

Was Achievement For Ability Students Lower Than Expected?

The achievement scores did tend to vary according to ability indicators. Students with average ability ratings on the CCAT were more likely to achieve below average scores on the PPVT-R and the VMI. Slow learners and low average students performed less well on the tests. All ability groups scored relatively higher on the CTBS than the PPVT-R or VMI.

Only one student was categorized as being of high average ability, and that student attained above average scores on the CTBS and VMI, but not the PPVT-R. No further conclusion could be drawn.

2d. Were Language Scores Relatively Lower Than Other Component Test Scores?

Language scores were not found to be lower than other component scores. Undeveloped language skills have been cited as impediments to education success in other instances (More, 1984a; Matthew & More, 1987) by teachers commenting on common problems for their Indian students, but this evidence was not found in the NIRS results.

2e. Did Perceptual Problems Affect Test Scores?

Students diagnosed as having either hearing or vision problems, or both, showed poor performance on the achievement measures. The results varied according to the different grade levels. By grade 6 none of the students diagnosed with hearing or vision problems achieved a CTBS grade equivalent even one grade level below placement.
Judgment -- First, the results of the tests in the battery showed that the NIRS students' achievement was below the norming population. Second, the difference was shown to increase through the grades. This cumulative performance decrement was found to be quite small, or non-existent in grade 1, but in grades 6 and 7 the development was very pronounced extending to two years below placement. Third, regardless of ability level, all students' performed poorly on the PPVT-R, evidence of a vocabulary problem. The assessment results showed that students of average ability were having difficulty with both the VMI and PPVT-R. Even the high average student had difficulty with the PPVT-R. Since the PPVT-R is a test of receptive vocabulary, the results prescribe an intervention in vocabulary instruction. Lower language scores on the CTBS were not found to be lower than the composite. The inference is that achievement in language was commensurate with achievement in other skill areas. Fourth, the hearing and vision screening results showed that the students with perceptual problems were likely to have achievement difficulties. Therefore, the battery was effective in determining the type and magnitude of the learning problems of the students.

The evidence suggests that early intervention, especially in language development, would narrow or eliminate the differences that were found between the NIRS group and the norming population.

3. Does Use of a Multi-Dimensional Battery Lead To Biased Assessment when used in such a sample?

3a. Did The PPVT-R Have The Same Test Difficulty Characteristics For The NIRS Sample As It Did For The Norming Sample?
Grade by grade analysis of item difficulty indices revealed that the presentation order of the items was that of ascending difficulty. This is the order intended by the test authors, and is therefore evidence that the difficulty characteristics of the test were similar for NIRS students and the norming sample.

3b. Are Canadian Native Indians Underrepresented In The Norm Reference Group For The Test?

The PPVT-R, MRT, and VMI were developed in the United States, which means that Native Indians represented in the norming samples would not be from Canada. The number of American Indians included in the sample was small. However, it was reported in VMI validity studies that little score variance was attributable to ethnic background. The CTBS and CCAT did not report ethnic group membership of individuals tested in their standardization sample. Based on the information available it would seem that Canadian Native Indians are underrepresented.

3c. Was There Examiner Or Language Bias?

The examiners were not Canadian Native Indians. The evaluation team sought to deliver a competent assessment. However, this does not rule out unintended bias intruding because of differences in the ethnic, cultural, and socioeconomic background of the examiners from the testing group. Similarly, communication might not have been as clear as could be. The students may not have understood instructions. The language problem may have also introduced bias in measurement because of student were unfamiliarity with vocabulary.

Judgment -- The evidence is inconclusive whether bias had an effect on the measurement of scores for NIRS students. The standard required that
PPVT-R item difficulty rankings should be the same for the sample and the population. This standard was met. In terms of the standards established (see Figure 1), two criteria were not met: the proportion of Native Indians in the standardization samples did not match the proportion in the general population and the examiners were not themselves Native Indians. Students may not have understood instructions. It is possible that a problem with communication may have existed, but was not voiced. In terms of the norming samples, Canadian Native Indians were distinctly underrepresented.

C. EVALUATION OF APPROPRIATENESS

In summary, the assessment battery did achieve its goal of identifying children with perceptual problems. In terms of these children, the specific assistance was suggested to ensure that vision or hearing difficulties did not act as an impediment to education. The battery was also successful in the identification of students in terms of their test performance; that is those students who were performing below, at, or above their ability levels. The general trend was that NIRS students’ scores were lower than the population norms. This confirms results of other studies that have found lower test scores for Indian students (MacArthur, 1968; Conry & Conry, 1973a; Matthew & More, 1987; and More 1984a). However, the possibility exists that a failure to understand the tests or the testing experience may have contributed to the lower scores. The results showed that a need for educational assistance.

The battery provided results from which it was possible to identify the type of assistance required and the kinds of problems such as, poor language performance, poor performance due to perceptual problems, poor performance in
higher grades, below level performance of average ability students, and lower performance in comparison to norms.

There is a possibility of bias in the difficulty of the PPVT-R. From the norming sample information available, the evidence is that Canadian Native Indians were underrepresented. The fact that the examiners were not Native Indians gives rise to the possibility that some bias may have been introduced into the testing.

For the most part, the multiple assessment battery was appropriate to be administered to a uniracial ethnic minority. The PPVT-R difficulty characteristics for the grade groupings of the NIRS students were similar to those of the norming sample. This lends credence to the finding that the lower PPVT-R results are indicative of lower vocabulary knowledge for the NIRS students. The PPVT-R results were useful in determining the specific items that are unfamiliar. The CTBS was found to be the test with which the students had the least difficulty. Perhaps, the vocabulary and content of this test is closer to the experience of the NIRS students. The suggestion therefore is that the CTBS is useful in locating skill areas weaknesses for Indian students. It was useful in determining the difficulty with language skills for NIRS students.

One final comment about appropriateness of testing. Deyhle (1986) has suggested that testing, its rationale and implications, may be unfamiliar to Indian students. This was certainly true in this case. When the university-based team was contracted for the original evaluation, testing was included as an adjunct to the evaluation of the reserve school. Although the students were familiar with teacher-made tests, this was the first time that NIRS students had ever been tested in a formal setting using standardized tests and procedures. This complete
lack of experience is an illustration of the differential experience that Indians have with testing, especially relative to non-Indians in public schools.

D. GENERALIZABILITY

The results of this study were drawn from a sample of students at a Native Indian reserve school on the southeast coast of Vancouver Island. Chrisjohn and Lanigan (1986) caution against generalizing findings across nations or linguistic groups. However, it would seem appropriate that the conclusions drawn from this study would apply to other Indian bands, in the coastal region of southwestern British Columbia. Since the findings replicate results found in other parts of British Columbia, Alberta, and Saskatchewan, it also would be reasonable to apply conclusions to other elementary Indian students in Western Canada.

E. LIMITATIONS OF THE STUDY

1. This study was limited to an analysis of scores on a single administration of a battery of standardized tests. Inferences drawn were related only to this single testing. Since there was no prior or subsequent testing, determinations of academic and developmental gain were not appropriate.

2. Generalizability of findings beyond this uniracial group in a non-urban reserve are not justified. The original intention of the evaluation team was to match NIRS pupils by age and grade with Native Indian pupils at a public school in an adjoining locale. Unfortunately approval was not given to study these other students leaving the NIRS students as a group for which no control comparisons was available. Therefore, desirable biodemographic
controls could not be applied.

3. The team of evaluators were all caucasian. There is a small possibility that bias may have been introduced because of the difference in race, language, and socioeconomic status.

4. School and home environmental factors could not be analyzed. Circumstances of the *milieu* in which each child lives may have contributed to poor test performance. Other factors such as socioeconomic level and nutrition were not examined.

5. The sample size was small. Although the total group was 76, most comparisons required summary statistics by grade, and grade groups were as small as six. With such small numbers, the reliability and validity of results may be marginal.

**F. IMPLICATIONS FOR FUTURE STUDIES**

At most grade levels the NIRS sample size was small, as few as six students. The possibility exists that larger samples might have produced different results. Large sample studies of homogeneous age cohorts are required to reliably determine item difficulty rankings of the PPVT-R for Native Indians.

Mean standardized test scores for reserve children are typically well below the national averages represented by norms. Since test norms are largely based on non-Indian populations, it is possible that cultural differences have depressed the NIRS students' scores (Matthew & More, 1987). In addition to cultural and academic factors, there are other variables that may affect test performance such as socio-economic status, school attendance, and stability within the district (Boloz & Varrati, 1983). If low socio-economic status is a variable that the reservation
student brings to the school and testing situations, then consideration needs to be made of remedial steps to take to change this deficit. Similarly, attendance may be a factor in academic achievement. Examination of attendance records was not included as part of the study. Neither was an analysis made of the stability of students within the school system. As mentioned, it is common for high-achieving students to leave NIRS students for public schools. Conversely, many students return to NIRS after experiencing the academic, social, or cultural problems in the public schools. These and other variables may have contributed to the noted progressive decline in performance through the elementary grades. These factors deserve closer examination. Results of such studies could have implications for trustees, the decision-makers within native Indian districts. These and other factors may be subsequently determined as causative factors of declining performance.

Regardless of the causes, the results of this thesis show that impaired performance exists on all indicators of achievement in the intermediate grades. A particular problem is undeveloped language arts skills. The fact that achievement of students in the primary grades is often satisfactory, suggests that some factors other than ability are contributing to impaired performance of Native Indian students on standardized tests.
REFERENCES


APPENDIX 1

SPLIT-HALF AND KR-20 RELIABILITY COEFFICIENTS FOR LEVEL 1, FORMS P AND Q, OF THE METROPOLITAN READINESS TEST, SKILL AREA AND PRE-READING SKILLS COMPOSITE SCORES
### Appendix 1

#### Split-Half and KR-20 Reliability Coefficients for Level I, Forms P and Q, of the Metropolitan Readiness Test, Skill Area and Pre-Reading Skills Composite Scores

<table>
<thead>
<tr>
<th>Test/Skill Area/Composite</th>
<th>FORM P</th>
<th></th>
<th>FORM Q</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory Memory</td>
<td>.73</td>
<td>.74</td>
<td>7.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Rhyming</td>
<td>.80</td>
<td>.77</td>
<td>9.0</td>
<td>3.1</td>
</tr>
<tr>
<td>Letter Recognition</td>
<td>.88</td>
<td>.88</td>
<td>7.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Visual Matching</td>
<td>.79</td>
<td>.79</td>
<td>10.6</td>
<td>3.0</td>
</tr>
<tr>
<td>School Language &amp; Listening</td>
<td>.66</td>
<td>.66</td>
<td>11.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Quantitative Language</td>
<td>.75</td>
<td>.67</td>
<td>7.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Visual Skill Area</td>
<td>.85</td>
<td>.88</td>
<td>18.4</td>
<td>5.3</td>
</tr>
<tr>
<td>Language Skill Area</td>
<td>.83</td>
<td>.80</td>
<td>18.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Pre-reading Skills Composite</td>
<td>.93</td>
<td>.92</td>
<td>53.8</td>
<td>12.4</td>
</tr>
</tbody>
</table>

(MRT Manual, p. 25)
APPENDIX 2

SPLIT-HALF RELIABILITY COEFFICIENTS FOR FORM L OF THE PPVT, BY AGE
Appendix 2

Split-Half Reliability Coefficients for Form L of the PPVT, by Age

<table>
<thead>
<tr>
<th>Age Group</th>
<th>n</th>
<th>Odd/Even r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-0 - 5-5</td>
<td>100</td>
<td>.79</td>
</tr>
<tr>
<td>5-6 - 5-11</td>
<td>98</td>
<td>.73</td>
</tr>
<tr>
<td>6-0 - 6-5</td>
<td>94</td>
<td>.84</td>
</tr>
<tr>
<td>6-6 - 6-11</td>
<td>99</td>
<td>.77</td>
</tr>
<tr>
<td>7-0 - 7-11</td>
<td>96</td>
<td>.83</td>
</tr>
<tr>
<td>8-0 - 8-11</td>
<td>113</td>
<td>.77</td>
</tr>
<tr>
<td>9-0 - 9-11</td>
<td>94</td>
<td>.83</td>
</tr>
<tr>
<td>10-0 - 10-11</td>
<td>109</td>
<td>.82</td>
</tr>
<tr>
<td>11-0 - 11-11</td>
<td>97</td>
<td>.77</td>
</tr>
<tr>
<td>12-0 - 12-11</td>
<td>90</td>
<td>.86</td>
</tr>
<tr>
<td>13-0 - 13-11</td>
<td>91</td>
<td>.84</td>
</tr>
<tr>
<td>14-0 - 14-11</td>
<td>112</td>
<td>.87</td>
</tr>
<tr>
<td>15-0 - 15-11</td>
<td>107</td>
<td>.86</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>.80</td>
</tr>
</tbody>
</table>

¹(Robertson & Eisenberg, 1981, p. 34)
APPENDIX 3

ALTERNATE FORMS RELIABILITY COEFFICIENTS FOR PPVT BASED ON RAW SCORES OF IMMEDIATE RETEST SAMPLE, BY GRADE
Appendix 3

Alternate Forms Reliability Coefficients for PPVT Based on Raw Scores of Immediate Retest Sample, by Grade

<table>
<thead>
<tr>
<th>Grade</th>
<th>n</th>
<th>Testing Sequence</th>
<th>r</th>
<th>Average r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>25</td>
<td>L-M</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>M-L</td>
<td>.56</td>
<td>.89</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>L-M</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>M-L</td>
<td>.87</td>
<td>.70</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>L-M</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>M-L</td>
<td>.66</td>
<td>.85</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>L-M</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>M-L</td>
<td>.86</td>
<td>.64</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>L-M</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>M-L</td>
<td>.79</td>
<td>.83</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>L-M</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>M-L</td>
<td>.93</td>
<td>.84</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>L-M</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>M-L</td>
<td>.88</td>
<td>.90</td>
</tr>
<tr>
<td>median</td>
<td>15</td>
<td>L-M</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>M-L</td>
<td>.82</td>
<td>.89</td>
</tr>
</tbody>
</table>

¹(Robertson & Eisenberg, 1981, p. 36)
APPENDIX 4

CTBS INTERNAL-CONSISTENCY RELIABILITY COEFFICIENTS (KR-20)
## Appendix 4

### CTBS Internal-Consistency Reliability Coefficients (KR-20)

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Level</th>
<th>Grade</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>K</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Listening</td>
<td>.75</td>
<td>.76</td>
<td>.71</td>
<td>.67</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.80</td>
<td>.64</td>
<td>.87</td>
<td>.83</td>
<td>.91</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
<td>.92</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Word Analysis</td>
<td>.81</td>
<td>.83</td>
<td>.84</td>
<td>.84</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Reading</td>
<td>---</td>
<td>.92</td>
<td>.95</td>
<td>.90</td>
<td>.91</td>
<td>.90</td>
<td>.90</td>
<td>.90</td>
<td>.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling (L1)</td>
<td>---</td>
<td>---</td>
<td>.86</td>
<td>.83</td>
<td>.91</td>
<td>.90</td>
<td>.91</td>
<td>.91</td>
<td>.91</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>Capitalization (L2)</td>
<td>---</td>
<td>---</td>
<td>.83</td>
<td>.89</td>
<td>.88</td>
<td>.86</td>
<td>.82</td>
<td>.81</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punctuation (L3)</td>
<td>---</td>
<td>---</td>
<td>.74</td>
<td>.76</td>
<td>.78</td>
<td>.81</td>
<td>.80</td>
<td>.80</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage (L4)</td>
<td>---</td>
<td>---</td>
<td>.79</td>
<td>.77</td>
<td>.87</td>
<td>.86</td>
<td>.92</td>
<td>.86</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Total</td>
<td>.83</td>
<td>.69</td>
<td>.92</td>
<td>.94</td>
<td>.95</td>
<td>.95</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Materials (W1)</td>
<td>---</td>
<td>---</td>
<td>.84</td>
<td>.80</td>
<td>.85</td>
<td>.84</td>
<td>.82</td>
<td>.75</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference Materials (W2)</td>
<td>---</td>
<td>---</td>
<td>.83</td>
<td>.77</td>
<td>.86</td>
<td>.86</td>
<td>.89</td>
<td>.85</td>
<td>.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work-Study Total</td>
<td>---</td>
<td>---</td>
<td>.91</td>
<td>.87</td>
<td>.92</td>
<td>.91</td>
<td>.92</td>
<td>.87</td>
<td>.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Concepts (M1)</td>
<td>---</td>
<td>---</td>
<td>.77</td>
<td>.78</td>
<td>.84</td>
<td>.79</td>
<td>.82</td>
<td>.82</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems (M2)</td>
<td>---</td>
<td>---</td>
<td>.76</td>
<td>.69</td>
<td>.79</td>
<td>.81</td>
<td>.82</td>
<td>.87</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computation (M3)</td>
<td>---</td>
<td>---</td>
<td>.79</td>
<td>.80</td>
<td>.87</td>
<td>.88</td>
<td>.86</td>
<td>.83</td>
<td>.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Total</td>
<td>.83</td>
<td>.74</td>
<td>.89</td>
<td>.89</td>
<td>.92</td>
<td>.92</td>
<td>.93</td>
<td>.93</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composite</td>
<td>.93</td>
<td>.93</td>
<td>.97</td>
<td>.97</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1*(Nelson, 1984, pp. 67-69)*
APPENDIX 5
INTERNAL CONSISTENCY RELIABILITY DATA FOR PRIMARY AND MULTILEVEL EDITIONS OF THE CCAT
Appendix 5

Internal Consistency Reliability Data for Primary and Multilevel Editions of the CCAT

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Verbal</th>
<th>Quantitative</th>
<th>Nonverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2591</td>
<td></td>
<td></td>
<td>.88</td>
</tr>
<tr>
<td>2</td>
<td>2532</td>
<td></td>
<td></td>
<td>.86</td>
</tr>
<tr>
<td>3</td>
<td>2274</td>
<td>.95</td>
<td>.89</td>
<td>.92</td>
</tr>
<tr>
<td>4</td>
<td>2627</td>
<td>.93</td>
<td>.89</td>
<td>.92</td>
</tr>
<tr>
<td>5</td>
<td>2870</td>
<td>.92</td>
<td>.89</td>
<td>.91</td>
</tr>
<tr>
<td>6</td>
<td>3099</td>
<td>.91</td>
<td>.88</td>
<td>.91</td>
</tr>
<tr>
<td>7</td>
<td>2973</td>
<td>.91</td>
<td>.89</td>
<td>.90</td>
</tr>
</tbody>
</table>

Data from Canadian standardization sample.