THE UNIVERSITY OF MANITOBA

The Concurrent Validation of the Developing Cognitive Abilities Test as an Instrument for Identifying Cognitive Entry Behaviors

by

Bernice Duma

A thesis
submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements
for the degree of Master of Education

Department of Educational Psychology
Winnipeg, Manitoba
January, 1989

ISBN 0-315-63322-0

Canada

THE CONCURRENT VALIDATION OF THE DEVELOPING COGNITIVE ABILITIES TEST AS AN INSTRUMENT FOR IDENTIFYING COGNITIVE ENTRY BEHAVIORS

BY

BERNICE DUMA

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF EDUCATION

TABLE OF CONTENTS

Table

LIST OF FIGURES

Figure

LIST OF CHARTS

Chart

ABSTRACT

The main purpose of this research was to examine the concurrent validity of the Developing Cognitive Abilities Test (DCAT) as an edumetric instrument for identifying cognitive entry characteristics of learners. The successful identification of entry characteristics would facilitate improved diagnosis, curriculum design, instructional strategies, and placement decisions. In order to determine the concurrent validity of the DCAT, the relationship of the DCAT was compared to the performance of learners on the Concept Learning and Development Assessment (CLDA), the Stanford Achievement Test (SAT), and the Canadian Test of Basic Skills (CTBS). In order to determine the comparability of the DCAT, this study was conducted using two culturally different school populations. These two population samples were located in Brochet and Winnipeg, Manitoba, Canada. Tests were administered to 14 to 40 grade three students and 17 to 70 grade seven students in Winnipeg and Brochet. Correlations ranging from -.01 to $^{+}$.84 were indicated among the test scores at both grade levels. The DCAT Verbal subtest correlated more highly with the SAT Vocabulary subtest at the grade seven level in Winnipeg than with any other test at any other grade level or location. All other DCAT, CTBS, SAT, and CLDA Noun correlations obtained were within the medium or low range with the majority of correlations lacking significance at the .05 level.

These research findings should prove to be valuable to the classroom teachers that participated in the study, to all educators concerned with diagnosis, curriculum design, instruction, and placement decisions. Also, these research results should prove to be of interest to theorists in the area of cognitive psychology and evaluation.

ACKNOWLEDGEMENTS

I wish to extend sincere appreciation to Dr. W. Rampaul for his assistance, constructive criticism, and patience during this study.

I wish to extend gratitude to Professor J. Belford and Dr. R. Henjum for serving on my thesis committee and for providing valuable suggestions.

I wish to extend sincere thanks to Dr. N. Isler, Mr. D. Mandryk, Mr. S. Didkowski, Mr, Ottawa, and Mr. C. De Landro for permission to conduct this research study.

I wish to extend sincere thanks to Ms. D. Derry, Mr. V. Karpik, Ms. C. Sproule, and Mr. F. Luschak, and the teachers of Brochet school for assisting with the research in Winnipeg and Brochet.

Last, but not least, I wish to thank my husband, Roy, and my mother, Helen Kianski for their support, kindness, and patience during this study.

CHAPTER I

INTRODUCTION

Presently, educators are recommending that classroom teachers should concern themselves with the diagnosis of cognitive entry characteristics which are alterable variables rather than with abstract global characteristics which are unalterable (Hunter, 1980). The identification of cognitive entry characteristics is essential as a basis for diagnostic teaching to accelerate "predictably" and "significantly" the quality and quantity of learning for almost all learners (Hunter, 1980). A knowledge of the specific cognitive entry characteristics of learners can provide the classroom teacher with a concrete foundation for making accurate educational diagnosis as a means for basing prescription and remediation, improving instruction, creating curriculum, furnishing materials, and providing a solid basis for determining appropriate placement of learners. Continuous assessment using standardized or informal tests is an essential ingredient for promoting instruction, learning, and effective schools (Hunter, 1985; Possemato, 1985).

This research was aimed at the DCAT and its potential as a valid edumetric instrument for filling a gap in current educational evaluation practice. The DCAT possesses the potential for providing educators with a solid foundation for making sound educational decisions as the DCAT is constructed on a theoretical framework based on current learning theory. According to the DCAT Technical Manual and Norms (Gage, 1983), the learner's performance is a function of individual cognitive skills and academic achievement. Similarly, according to the DCAT Technical Manual

and Norms (Gage, 1983), the DCAT is based on the underlying assumption that appropriate instruction can modify and improve those cognitive characteristics and abilities that contribute to academic achievement.

The DCAT measures two aptitude dimensions: (1) verbal, quantitative, and spatial abilities which are essential for success in all school-related subjects (DCAT Technical Manual and Norms, 1983); and (2) five of the six cognitive Classes of Bloom's Taxonomy (Bloom, Englehart, Furst, Hill, and Krathwohl, 1956). The assessment of cognitive characteristics which are amenable to alterations as a result of instructional strategies (DCAT Technical Manual and Norms, 1983), makes the DCAT a unique ability test. Likewise, the combination of assessing cognitive abilities and content areas provides the teacher with an exclusive multi-dimensional edumetric instrument. Because the DCAT measures processes that are prerequisite to academic achievement, the DCAT fills a void in present educational evaluation practice. The major intent of this research study was to concurrent by validate the DCAT so that educators will have greater confidence in the DCAT as an edu- metric instrument.

BACKGROUND TO THE STUDY

The DCAT deserves attention and recognition for the following reasons:

1. The concurrent validity of the DCAT is essential for determining cognitive entry behaviors. The successful identification of cognitive entry behaviors fills a gap in current academic evaluation;
2. The successful indentification of cognitive entry behaviours can enhance the quality and quantity of learning so that up to ninety percent or more of all learners can succeed academically (Bloom, 1981a; Bloom, 1985). With an accurate diagnosis of the learner's

cognitive entry behaviors, the classroom teacher has a firm foundation for designing instruction, curriculum, and formulating decisions;

3. The DCAT is a relatively new edumetric instrument. While the DCAT possesses face validity, no educational research data are presently available to substantiate the concurrent validity of the DCAT for diagnostic, curriculular, and placement decision-making purposes;
4. The DCAT was standardized in October, 1980, and April, 1981, and reflects current curricular content (DCAT Technical Manual and Norms, 1983);
5. The DCAT is a unique edumetric instrument because it measures individual cognitive characteristics of learners based on five of the six classes of Bloom's Taxonomy--Knowledge, Comprehension, Application, Analysis, and Synthesis. These are familiar to teachers. In addition, the second area of the DCAT assesses those skills that are associated with academic performance (DCAT Technical and Norms Manual, 1983):
   (a) Verbal skills--related to academic performance in language and reading;
   (b) Quantitative skills--related to academic performance in numbers, algebra, and science;
   (c) Spatial skills--related to academic performance in geography, science, and geometry.

   Because of these two different assessment dimensions, the DCAT is a unique edumetric instrument.
6. Bloom's Taxonomy which forms the cognitive dimension of the DCAT, is very successful (Reilly and Lewis, 1983), is familiar to many edu-

cators as an educational tool (Seifert, 1983), and is extensively used (Wiersma and Jurs, 1985). Classroom teachers can relate to Bloom's Taxonomy both in theory and from classroom practice.

7. According to the DCAT Technical and Norms Manual (1983), the DCAT has many classroom uses:
   (a) Information on individual learners;
   (b) Information for curriculum development and improvement;
   (c) Identification of learners with learning difficulties;
   (d) Identification of learners with low achievement and high aptitude;
   (e) Identification of gifted learners;
   (f) Prediction of achievement if used in combination with the Achievement Series of the Comprehensive Assessment Program. Prediction is reported in plus, minus, and asterisk signs rather than actual grade differentials in order to prevent misinterpretation;
8. The DCAT is based on current cognitive learning theory. The alterability of cognitive characteristics, once assessed in relation to academic performance, has profound implications for diagnosis, instruction, curriculum, and decision-making;
9. The DCAT possesses the potential for providing an edumetric instrument based on sound pedagogical learning theory in lieu of the weaknesses inherent in the psychometric approach to diagnosis, instruction, curriculum, and decision making;
10. The CTBS and the SAT measure academic achievement only. The DCAT, on the other hand, measures cognitive entry characteristics in relation to all academic subject areas. Hence, the DCAT possesses the potential for a wider range of diagnostic data than the CTBS

and the SAT;

11. The DCAT is similar to the CLDA in that the CLDA and the DCAT are based on current learning theory.

SIGNIFICANCE OF THE STUDY

Central to the DCAT and current cognitive learning theory is the concept of cognitive entry characteristics. Cognitive entry characteristics refer to the specific skills, ability, or knowledge that are necessary prerequisites for learning a specific academic subject or task (Bloom, 1981a).

The importance of cognitive entry behaviors is stated succinctly by Ausubel (1978):

> If I had to reduce all educational psychology to one principle it would be this: The most important single factor is what the learner already knows (p.iv).

In cognitive psychological literature, the term "cognitive entry characteristics" has synonymous terms such as "cognitive entry behaviors", "readiness", "prerequisites", and "cognitive structures". While the terminology used may vary, their meanings and implications for education are similar. Cognitive entry characteristics are related to contemporary cognitive learning theory. Cognitive psychologists such as Piaget, Ausubel, Klausmeier, Gagne, Bruner, and Bloom view learning as a developmental, sequential, and information-processing task. Cognitive psychologists stress the importance of how learners acquire and process information. Bruner (1966) has said, "Knowing is a process, not a product" (p.72). In brief, cognitive psychologists possess the following similar views with regard to learning (Reilly and Lewis, 1983):

1. Learning involves meaning and understanding;
2. The substance of new material must be understood so that transfer of learning can take place;

3. New material to be learned must be "anchored" to material existing in the cognitive structure;
4. The learner must learn broad concepts and principles;
5. Material to be learned must be organized in cognitive structures or the underlying structure of the disciplines;
6. Learning must occur at the work-a-day level;
7. An attempt is made to explain events related to the mind rather than external entities;
8. Language is important for thinking and communicating; language is the main tool for learning;
9. Readiness is very important as the starting point for meaningful learning;
10. Classroom learning must be meaningful to the learner.

The role of cognitive entry behaviors in the education process has many implications. The relationship of cognitive entry characteristics to instruction is explained thus:

> While an initial (not exhaustive) diagnosis is essential, the only way to maintain the essential prerequisites for a particular learning is by diagnostic teaching, a process of continual "dip-sticking" that guides the professional decision to reteach, to practice or extend, to move to the next learning or to "abandon ship" because "now's not the time" (Hunter, 1980, p.122) . . . It is possible for the content to remain the same, but those who have achieved the learning can be stretched while those who need more teaching will receive it . . . . Reteaching or remediation should be provided as soon as "dip-sticking" indicates such a need (Hunter, 1985. p.64).

Cognitive entry characteristics can be improved because they are composed of particular content and skills which can be learned if reviewed and relearned (Bloom, 1984). Much of the variation in academic achievement is directly related to variations in learner's cognitive

entry characteristics (Bloom, 1981a). When learners are able to reach sufficient competency levels on necessary cognitive characteristics, the majority of learners may achieve high levels of academic learning with little achievement variation (Bloom, 1981b; Bloom, 1984; Guskey and Gates, 1986). Bruner (1966) states that, providing the learner possesses the appropriate readiness level:

> Any subject can be taught in some intellectually honest form to any child at any stage of development (p. 33).

The learner's readiness for learning is dependent on the presence of a sequence of learning activities (Wilson, 1984). Gagne, Briggs, and Wager (1988) describe seven types of learning that are distinguishable from each other in their degree of complexity and in terms of their prerequisites. Levels of complexity of intellectual skills are indigenous to, but independent of, all types of subject matter (Gagne, Briggs, and Wager, 1988). Curriculum developed on a cognitive framework specifies the sequence of cognitive processes that are a part of learning in addition to defining curricular objectives (Gibson, 1980). Cognitive psychologists suggest adapting curriculum to the learner rather than adapting the learner to the curriculum (Gibson, 1980). In order to develop independent learners, curriculum must possess objectives and goals beyond the knowledge level (Bloom, 1981a; Bloom, 1986). Using sound instructional practices based on learning theory, the rate of cognitive development can be enhanced by educational programming (Woolfolk and McCune-Nicolich, 1984) within certain developmental limits (Ausubel, Sullivan, and Ives, 1980).

The need to identify appropriate cognitive entry behaviors has been expressed in various ways in the educational literature by various scholars:

1. Glasser (1981) states that present test theory and techniques have failed to keep up with recent developments in cognitive and learning psychology;
2. Curtis and Glaser (1985) state that intelligence testing can become more sensitive to current educational and social needs by a study of the types of performance needed for scholastic success;
3. Klausmeier and Sipple (1980) emphasize the need for determining entry behaviors in each subject area;
4. Popham (1983) states that measurement should serve as a catalyst for instruction;
5. Shaha and Wittrock (1983) discuss the need to identify cognitive processes as a means of increasing the potential of learners;
6. Webster (1978) discusses the need for a diagnostic method for determining how well a learner processes information. Webster (1978) also states that this method would reduce cultural and racial bias due to a lack of norm-referenced criteria;
7. Child (1985) suggests that a crucial need exists for the development of tools and skills which can assist teachers in making precise diagnosis and remediation of individual learner's problems;
8. Frase (1980) tells of the need for an analysis of the skills in specified subjects and methods for relating these skills to instructional methodology and instructional outcomes;
9. Messick (1985) declares the need for information about a constellation of abilities and knowledge structures which could serve to make decisions pertinent to learning techniques or organization of content for individual learners;

10. Hunt (1985) claims that a cognitive science approach to measurement can be used as a theoretical base. This theoretical base can generate requirements for individualized intelligence measures which are different from tests used for predicting performance in some poorly defined circumstance. Using measurement of individual mental ability is very different from using measurement which is justifiable in predictive validity terms;
11. Slavin (1988) emphasizes the increased diagnostic usage of standardized tests in the future;
12. Slavin (1988) states that test content must be related to school curriculum to increase the function of tests for purposes of evaluation and to provide for curriculum development;
13. Mehrens and Lehmann (1984) state that evaluation serves four main functions:
    (a) Instructional--evaluating learning outcomes, teaching, and curriculum, diagnosing learning, differentiating class assignments, grading, and motivation;
    (b) Guidance--personal, educational, and occupational decisions;
    (c) Administrative--classification, selection, placement, curriculum evaluation and planning, public relations data, teacher evaluation, information to outside publics, and grading;
    (d) research;
14. Bloom (1981a) emphasizes the need to identify cognitive entry

behaviors as a pre-assessment technique for successful mastery learning.

The identification of cognitive entry behaviors should help to enhance current educational delivery systems such as Individualized Instruction Plan (IEP), Individually Guided Education (IGE), Mastery Learning, Diagnostic Prescriptive Instruction, Direct Teaching, and many other delivery systems which depend on accurate diagnosis as a starting point for prescription, curriculum design, remediation, instruction, and placement decision making. Likewise, when cognitive entry characteristics, which are alterable variables, replace intelligence, which is an unalterable variable, the harmful effects of classification and prediction which characterize decision-making based on standardized intelligence tests would cease to exist (Bloom, 1981a). Cognitive psychologists view learning as an on-going maturational process that proceeds continually irrespective of chronological age (Ausubel, 1978; Gibson, 1980).

This study is aimed at determining the validity of the DCAT as an edumetric instrument for identifying cognitive entry characteristics. Since two of the main purposes of the DCAT are to identify the various cognitive levels of learners in various school subjects in three cognitive dimensions, it possesses the potential for powerful educational utility. Unfortunately, no research currently exists to support or

refute the concurrent validity of the DCAT. Since the CTBS, and the SAT are widely known and widely used standardized, academic achievement tests, these tests have been selected as a basis for evaluating the concurrent validity of the DCAT. Likewise, the CLDA has been selected for helping to determine the concurrent validity of the DCAT as the CLDA is a widely respected and currently used learning theory based test. By administering these four tests in two different localities, the comparability of the DCAT to the CTBS, SAT, and CLDA can be readily determined.

RESEARCH QUESTIONS

1. Based upon the assumption that the CTBS and the DCAT are measures of the cognitive level of learners, what are the degrees of relationship among the scores of the CTBS and DCAT when these tests are administered to grade three and seven learners in Winnipeg and Brochet?
2. Based upon the assumption that the SAT and the DCAT are measures of the cognitive level of learners, what are the degrees of relationship among the scores of the SAT and the DCAT when these tests are administered to grade three and grade seven learners in Winnipeg and Brochet?
3. Based upon the assumption that the CLDA and the DCAT are measures of the cognitive level of learners, what are the degrees of relationship among the scores of the CLDA and the DCAT when these tests are administered to grade three and grade seven learners in Winnipeg and Brochet?

DEFINITIONS

CROSS CULTURAL EVALUATION - The process of determining whether a test measures what it purports to measure when administering to groups of individuals possessing different values, customs, and beliefs.

VALIDITY - A test is considered to possess validity when it measures what it purports to measure.

EDUMETRIC INSTRUMENT - A test used for evaluating academic achievement.

COGNITIVE ENTRY CHARACTERISTICS - Those skills or behaviors which a learner must possess as prerequisites before new concepts, principles, or skills can be attained.

DIAGNOSIS - The identification or measurement of specific abilities, skills, or qualities from among a wide variety possessed by the individual.

CURRICULUM DESIGN - A sequential plan of specific content and skills in a given subject area for instructional purposes.

PLACEMENT DECISIONS - Decisions pertinent for determining appropriate group, grade, instructional, or curricular placement to best suit the needs of a learner or group of learners.

GENERALIZABILITY - The application of a behavior, skill, or an idea to comparable people or situations.

ALTERABLE VARIABLES - Variables which can be altered or changed.

ABSTRACT GLOBAL CHARACTERISTICS - Broad, general, and abstract qualities that describe learner behavior or achievement.

PRESCRIPTION - An educational plan designed to include curriculum and instructional strategies for an individual learner or a group of learners.

REMEDIATION - Curriculum and instructional techniques designed for an individual or a group of learners as a result of education diagnosis.

COGNITIVE SKILLS - Skills for learning, thinking, remembering, and analyzing which assist individuals in processing environmental information.

ACADEMIC ACHIEVEMENT - Achievement attained in a school-related subject field.

BLOOM'S TAXONOMY - A hierarchy of educational objectives which has three major categories: cognitive, affective, and psychomotor motor objectives.

KNOWLEDGE - The first category of Bloom's cognitive educational objectives, knowledge, including rote memory, is the easiest kind of learning. The individual can possess knowledge but does not need to possess understanding or comprehension.

COMPREHENSION - The second category of Bloom's cognitive educational objectives. Comprehension consists of restating or identifying information in a format that is not identical to the original presentation.

APPLICATION - The third category of Bloom's cognitive educational objectives. Application refers to problem solving ability involving problems that are similar but differ from problems previously experienced.

ANALYSIS - The fourth category of Bloom's Taxonomy of Cognitive Objectives. Analysis refers to the breaking down of an entity into its components.

SYNTHESIS - The fifth category of Bloom's Taxonomy of Cognitive Objectives. Synthesis consists of being able to combine skills, ideas, knowledge, experiences, and concepts to formulate original products.

COGNITIVE LEARNING THEORY - Theory which emphasize internal thinking: memory, acquisition, and relationships of information.

PSYCHOMETRIC APPROACH - An approach to diagnosis, curriculum, instruc-

tion, and placement decisions based on test results derived from standardized achievement or intelligence tests by a learner or a group of learners.

DIAGNOSTIC TEST - A test used for diagnosis, or determining areas of academic weaknesses and strengths, as opposed to tests used solely for assigning grades.

FACE VALIDITY - The way in which a test possesses content that seems to be relevant to the knowledge, skill, or ability that it claims to examine.

CONCURRENT VALIDITY - Concurrent validity results from an accumulation of data obtained at the same time from various sources.

CULTURAL BIAS - Inclination toward prejudice or discrimination against an individual or group of individuals who possess values, customs, beliefs, and language which differs from the dominant cultural group.

RACIAL BIAS - Prejudice or discrimination against an individual or group of individuals who possess a different racial origin than the dominant racial group.

MEASUREMENT - A procedure which employs a rule to designate numerical descriptions to some characteristics of an object, event, or person.

EVALUATION - The examination of all available data concerning learner(s), educational program(s), and teacher(s), to determine the amount of change in the learner and to make valid judgement pertinent to the program(s) in use.

FEEDBACK - Knowledge of the accuracy of the learner's response.

LIMITATIONS

Cross-cultural academic achievement. This study is limited to providing individual comparison of learner's test results obtained from the DCAT, CTBS, SAT, and CLDA. Two different populations were selected to determine the comparability of the DCAT. Comparison of test results on a class, individual, or cultural basis is not a consideration in this study.

Academic Achievement. This study will be limited to the DCAT, CTBS, SAT, and CLDA. While other achievement, cognitive abilities, or concept attainment tests exist which may test similar or other achievement, cognitive, or concept areas, consideration to other standardized or informal edumetric instruments were not a concern for this study.

Time. This study is limited to test results obtained by learners in June and December, 1985. While approximately seven to twelve school days would be required to administer the test battery to a classroom of learners, it is beyond the limits of this study to specify the actual number of days in which the tests must be administered.

Examiners. This study is limited to the extraneous and intrinsic variables which each examiner brings to the testing situations. While each examiner will be requested to adhere strictly to the administration procedures outlined in the administration manual or section for each test, time and cost factors prohibit the probability of a higher degree of examiner consistency by using one examiner for test administration in each classroom in this study.

DELIMITATIONS

Examinees. This testing sample will consist of examinees from grades three and seven in Winnipeg and Brochet. While these two groups of examinees have been selected owing to their cultural disparities, the results and implications of the test results may not necessarily be applicable to every classroom of grade three and seven in the world. However, the findings of this study should serve as an indicator of academic achievement of learners from the specific school community which each group of examinees represent.

ORGANIZATION OF THE REMAINDER OF THE STUDY

A review of the literature related to this study will be the subject content discussed in Chapter II. Chapter III will outline the design of the study. The results and an analyses of the results of research findings will be examined in Chapter IV. The implications of the research findings and recommendations for future research will be discussed in Chapter V.

CHAPTER II

REVIEW OF RELATED LITERATURE

The literature review will be segmented into three major sections as stated below. References cited are those bearing the most relevance to the present study.

Cross-Cultural Evaluation

1. Culture fair tests
2. Culture free tests
3. The concept of culture and test bias

1. CULTURE - FAIR TESTS

The historical framework which gave rise to the creation of culture-fair tests is based on the sordid view that had once prevailed in the United States with reference to the superior test performance of white Americans in comparison to black Americans and immigrant minorities (Blum, 1978). After World War II, the following factors helped to stimulate the creation of culture fair-tests (Blum, 1978):

1. The supremacy of the white intellect in comparison to the intellect of Negroes and immigrant minorities as measured by prevailing intellegence tests;
2. The growth of the belief of racial equality;
3. The availability of new knowledge about the effects of experience and environment on intelligence test performance;
4. The growing need for manpower in industrial sectors as the result of demands for more skilled workers, "industrial efficiency", and economic productivity to meet the demands created by World War II;
5. In recent times, the Civil Rights Movement and court cases challenging the discriminatory effects of intelligence tests on minority groups.

Culture-fair tests were designed to question learners from all cultural groups without discriminating against any sub-culture or cultural group (Gibson, 1980). Culture-fair tests are based on two common assumptions (Mehrens and Lehmann, 1984): (1) no genetic difference exists among various subcultures; and (2) tests that measure innate ability will reveal no significant difference among subcultures. Culture-fair tests

possess test items which eliminate the influence of verbal skills on test results (Ebel and Frisbee, 1986), language, motivation, test-taking attitudes, test-wiseness, speed, and competitiveness (Gronlund, 1985). Culture-fair tests also provide opportunities for learning the skills and knowledge which the test measures in an attempt to derive a measurement of ability which is free of most or all of these differences (Gronlund, 1985). One technique for creating a culture-fair test was the design of test items using figures, drawings, and objects which require selection, arrangement, classification, or some type of manipulation (Ebel and Frisbee, 1986). Culture-fair tests which require the elimination of culturally biased items have defeated their own purpose since individual test responses are culturally-loaded for all individuals (Ebel and Frisbee, 1986). Likewise, a totally culture-fair test would not discriminate between individuals (Ebel and Frisbee, 1986). If a test lacks discriminability, no need exists for the test (Ebel and Frisbee, 1986). A test designed to be culture-fair had to meet these five criteria (Oakland, 1982):

1. Test standardization was to be representative of the national population;
2. Standard deviation and mean scores were to be similar for all social classes, and for all racial and ethnic groups;
3. Similar validity and reliability estimations were to be available for all subgroups;
4. Tests demanded a minimum use of language (listening, reading, speaking, and writing);
5. Time limits were to be removed for test completion.

Unfortunately, culture-fair tests have proven to be highly inadequate (Ebel and Frisbee, 1986; Woolfolk, 1987). Likewise, culture-fair

tests lack validity, reliability, and fairness to minority groups (Das, Kirby, and Jarman, 1979; Reschley, 1981; Nenty, 1986). The notion that a pencil and paper culture-fair test can be created to measure innate ability is fallacious (Blum, 1978). Similarly, the practical problems of creating and using culture-fair tests are impossible without research analysis on cultural differences, a definition of intelligence, and more information on the structure of intelligence (Blum, 1978; Cattell, 1979). Numerous measurement authorities hold the position that if culture-fair tests could be created, these tests would have less utility than current tests which are biased due to environmental factors (Mehrens and Lehmann, 1984). Many psychologists hold the position that culture-fair tests possess less predictive validity than achievement and aptitude tests (Mehrens and Lehmann, 1984). If the student's earlier environment is related to success at school, then the use of a test that blots out environmental influences may result in a decline of predictive validity (Mehrens and Lehmann, 1984).

Gay (1985) states:

> One problem with so called culture-fair tests is that they do not do as good a job at predicting "success" as traditional tests. Thorndike has stated that a test is fair only if it predicts success for the same proportion of minority group people that actually achieve success, success being defined as the criterion measure. It has repeatedly been found that culture-fair tests lack empirical validity, specifically predictive validity (Anastasi, 1982). This should not be surprising since basically culture-fair tests are attempting to use a non-language predictor to predict a criterion which is heavily influenced by traditional language proficiency. Thus, while the intent of culture-fair tests is laudable, their usefulness is limited by the institutional environment in which they are used (p.163).

Since culture-fair tests had proven to be so inappropriate in being fair, the use of tests bearing the generic label "culture-fair" fell

into disrepute. However, since the demise of culture-fair tests, test publishers have created new types of tests designed to be culture-fair, which, generally speaking, are more sophisticated in their rationale and design. Secondly, practices have been implemented using conventional standardized tests with the intention of providing cultural "fairness". The remainder of this section will be devoted to discussing the most recent attempts at culture-fair testing as opposed to the earlier culture-fair tests.

One current solution to minority assessment has been the development of pluralistic tests. One of the most recently designed culture-fair tests has been the System of Multicultural Pluralistic Assessment (SOMPA). This test varies from the earlier culture-fair tests in that, in addition to receiving conventional intelligence test scores, the test scores of minority examinees are adjusted upwards and possesses pluralistic norms (Mercer and Lewis, 1978). The "Adusted Intelligence Quotient" uses the examiner's knowledge of neighbourhood, family, school, and community to assess internal control, self-direction, and complexity (Mercer and Lewis, 1978). Individual test results are derived by two methods (Mehrens and Lehmann, 1984): (1) "uncorrected" scores for determining immediate learning needs, and (2) "corrected" scores (based on a correction of the WISC-R) for determining a student's "latent scholastic potential" as a way of avoiding labelling. The "latent scholastic potential" is a technique for considering the sociocultural level of the student (Woolfolk and McCune-Nicolich, 1984). Some inherent weaknesses of the SOMPA are:

1. Test users may come to perceive the "corrected" scores as reflective of reality (Mehrens and Lehmann, 1984). A student who has not

mastered certain capabilities or skills or to disregard them on the premise that, "He's no worse than other children of similar background" (p.396), is of no assistance to the learner whatsoever;

2. Norms are based on California children only (Woolfolk and McCune-Nicolich, 1984);
3. Estimation of learning potential scores do not possess the same degree of accuracy as standard intelligence scores as predictive indicators of academic achievement (Woolfolk and McCune-Nicolich, 1984). Students who have low scores on standardized achievement and intelligence tests require assistance irrespective of reasons for the attainment of low scores (Woolfolk and McCune-Nicolich, 1984). While students who attain low test scores should not be classified as retarded, these students should not be allowed to remain in programs that are leading to academic failure (Woolfolk and McCune-Nicolich, 1984).

Another current solution has been the design of culture-specific tests as a means of altering the assessment process for minorities. The purpose of culture-specific tests is to assess learners from specified ethnic-racial and social classes possessing common, identifiable geographic, and cultural areas (Oakland, 1982). The Black Intelligence Test of Cultural Homogeneity (BITCH) is an example of a culture-specific test. Unfortunately, the BITCH has proven to be biased favoring black middle class learners over white middle class learners and lower class black and white learners (Joseph). The development of culture-specific tests has been hampered by the multiplicity of ethnic-racial, geographic, and socially diverse cultural groups in existence (Oakland, 1982). Furthermore, resources and human willpower are lacking for such

a wide scale undertaking (Oakland, 1982).

In an attempt to make tests more linguistically relevant, conventional standardized tests have been translated into various languages for bilingual and multilingual learners (Oakland, 1982). Unfortunately, cultural differences cannot be readily removed by solely translating test items (Mercer, 1971). Altering the language of a test has a negative effect on the standardization properties and item difficulty of the test (Oakland, 1982). Translating conventional standardized tests has had many deleterious effects (Cabello, 1981; Leiblich and Kugelmass, 1981; Oplesch and Genshaft, 1981; Chavez, 1982; Crawford, 1985). Emphasis should be placed on testing the students in their own dominant language as a real disability should be obvious in both languages.

Minimum Competency Testing (MCT) has been a recent attempt to provide equality of opportunity for minorities by adjusting passing scores to a minimal level so that minorities can attain the same outcomes as white middle class learners (Baratz, 1980). The possibility that the tests may be responsible for the high black drop-out rate has been posited by Serow and Davis, 1982; Catterall, 1986; Catterall, 1987. MCT is fraught with many common problems lacking simple resolution including how to ensure fairness to minority groups (Gronlund, 1985). Some of these problems are:

1. MCT is similar to achievement testing (Reynolds and Bezruczko, 1988);
2. MCT in basic curriculum subjects may result in: (a) teachers teaching for the test, (b) creation of minimum standards at the price of high quality, and (c) weakened creativity and learning transfer (Urzillo, 1987);
3. MCT tests are too easy (Reed, 1987);

4. Test performance data and judgement are more valid measures of competency than the "cardiac approach" [traditional passing standard of 80% (Berk, 1987)];
5. Difficulties in test creation and graduation standards for handicapped students (Wildemuth, 1983);
6. Summer school attendance for failing students (Partridge, 1986);
7. Kindergarten retention and resultant lowering of student's self-esteem; overemphasis on reading competency (Partridge, 1986);
8. MCT influences curricular content (Partridge, 1986);
9. Retention of weaker students as a means of controlling the number of failures (Partridge, 1986);
10. Court challenges regarding three issues (ERIC Clearinghouse, 1984):
    (a) Constitutional concerns pertinent to due process--the time required to implement the testing programs, test reliability, and test validity;
    (b) Equal protection--provisions for fair education to racial minorities, non-English persons, and the handicapped;
    (c) Negligence--the need to document every phase of the performance of the student, certification of teachers, and accountability of the school.
11. Mehrens and Lehmann (1986) list the following five disadvantages of MCT:
    (a) Focuses less attention on harder to measure educational outcomes;
    (b) Creates teaching for the test;
    (c) Ceases to furnish sufficient instructional challenges within the

school as the "minimums" will become "maximums";

(d) Labels students unfairly and results in the retention of the academically weak students;

(e) Increases costs, particularly for remediation and implementation.

Some test publishers have recently implemented methods which were designed to reduce difficulties related to minority testing. Many test publishers presently employ personnel representative of various minority groups to detect test items which may be culturally biased (Gronlund, 1985). Judges are screened out using a modified caution index (Jaeger and Busch, 1986). However, more black professionals should be included in test development and interpretation (Johnson, 1988). Even if minority personnel could remove all bias caused by test items, bias in all likelihood, could emanate from such factors as examinee, examiner, and test interpretation variables (Alford, 1984; Edelsky and Harman, 1988). Secondly, hiring minority group test reviewers has proven to be a costly and complicated process (Popham, 1981). Current attempts at removing culturally biased test items are, at best, very mundane attempts to deal with the problem of item bias. Many issues related to item bias go far beyond hiring minority reviewers to remove culturally biased test items. For example, the disparity between white and black test performance is the result of item complexity rather than cultural rarity (Flynn, 1980). This fact remains true whether the test item involves spatial, verbal, or numerical content (Flynn, 1980).

Statistical analysis has been another solution employed to remove culturally biased test items (Gronlund, 1985). Statistical methods for detecting item bias have met with variable degrees of success (Bleistein, 1986). Various techniques lack agreement as to which test items

possess bias and the quantity of biased test items that exist in a specific test (Hills, 1981). Item bias procedures can be easily manipulated (Linn, 1984; Schueneman, 1985) to provide various interpretations. Current dissatisfaction with various item analysis techniques has created fragmentation. Furthermore, test developers are more concerned with "statistical elegance and scientific methodology" than with minority fairness (Gonzales-Tamayo, 1984). Likewise, test developers are affected in their perceptions by their majority group socioeconomic positions (Gonzales-Tamayo, 1984).

Also, test developers have attempted to reduce cultural bias by including a representative sample of minorities in the normative sample (Popham, 1981). This technique is futile if the proportion of minorities is not parallel to the general population (Popham, 1981). Even if norms would reflect accurate numerical representations of minority groups in the standardization population, minority groups could continue to be screened out since only individuals with the highest scores are selected (Popham, 1981). In addition, standardized achievement tests are invalid for learners whose academic curricula does not parallel the curricula of the standardization sample (Gibson, 1980). Neely and Shaunessy (1984) states that there are six problematic areas in minority testing: (a) unsuitable content, (b) unsuitable standardization samples, (c) language and examiner bias, (d) unacceptable societal consequences, (e) evaluation of differing constructs, and (f) variable predictive validity.

While psychometric discriminatory admissions policies still persist (Zorn, 1983; Willie, 1985), new solutions have been implemented in regard to the admissions policies for institutions of higher learning for minority groups. One solution has been to determine the number of admissions on the basis of the percentage of the national population a particular minority group represents (Popham, 1981). This type of quota system has been referred to as "compensatory justice", "reverse racism" (Popham, 1981), or a "double standard" (Eysenck, 1979). Another common practice has been to have separate cut- off scores or add bonus points to minority test scores for vocational and educational selection (Gronlund, 1985). This method is a very arbitrary approach for this method alters the predictive validity of the test in a very haphazard way. A test can be considered to be unbiased or fair only if the test predicts for minority and majority groups with the same degree of accuracy (Mehrens and Lehmann, 1984; Gronlund, 1985). Similarly, lowering the cutoff scores for minority groups can foster an attitude that minorities are different from the majority, less capable than the majority, and that these differences are fixed (Popham, 1981). Lowering the cutoff scores leads to a spirit of "second-classism" and prevents the creation of educational systems in which ethnicity is not a factor in test interpretation (Popham, 1981).

In addition, adding bonus points to the test scores of minority learners will not solve their learning difficulties but will only result in a misuse of tests (Scarr, 1978). Likewise, low achievement test scores may indicate inadequate mastery of the skills the test is designed to measure irrespective of the minority group tested (Gronlund, 1985). Lowering the cutoff points or adding bonus points can obscure

information pertinent to the knowledge of actual skill acquisition a learner has attained. Williams (1983) describes a method frequently used by employers in selecting employees which is based on a pre-set value judgement as to what criteria constitutes fairness to various cultural groups and setting a predictor cutoff based on the amount of risk an employer is prepared to take in hiring an employee. Determining cutoff points for various subgroups is intended to maximize the employee selection process. However, minority selection may be impeded by unreliable prediction or a large standard error (Williams, 1983).

Reynolds (1980) recommends the demonstration or factorial invariance for all cultural groups for whom the test was designed as a means of easing test interpretation. Reynolds (1980) suggests that whether adjustments are necessary for specific populations is an issue which requires further study. Spencer (1983) discusses some of the inequities that result from using various statistical procedures for comparing group scores. Astin (1979) suggests that alternatives must be found if minorities are to achieve access to higher education. Harman (1980) suggests that university enrollments have declined and community colleges have sprung up owing to the unfair testing practices that currently prevail. Presently, tests are unable to identify gifted minorities.

Fair use of tests for selection is a part of a bigger issue that should be resolved by society as a result of court rulings (Gronlund, 1985). Two new reforms have been Truth-in-Testing legislation and the Golden Rule Principle (Weiss, 1987a). Truth-in-Testing was legislated to furnish more information relating to the validity, accuracy, and cultural bias of test items in tests related to college admissions

(Weiss, 1987a). Similarly, the Golden Rule Principle states that for questions which are of similar validity and difficulty in every content subject, questions that possess the least differences in success levels between minority and majority examinees must be given first priority (Weiss, 1987a). Hence, safeguards must be created to guarantee that tests are measuring important knowledge variations between examinees, rather than cultural specific factors that lack relevance (Weiss, 1987a). The Golden Rule Principle is difficult to apply (Gonzales-Tamayo, 1987), ineffective (Linn and Drasgow, 1987), and is opposed by professional associations (Faggen, 1987; Jaeger, 1987). Support is being mustered to effect new legislation that provides a breakdown of test scores by ethnic group, race, and sex as confidence prevails that bias exists (Jaschik, 1987).

SUMMARY

Many methods have been employed to provide culture-fair tests and many methods have been employed to overcome the weaknesses inherent in culture-fair tests: translating tests into different languages, minimum competency testing, hiring minority test reviewers, statistical analysis techniques, and including a normative sample of minorites in the normative sample. Similarly, attempts to reduce discriminatory selections and admissions policies, easing test interpretation, and legislative reforms have all been attempts to provide culture-fair tests. However, many difficulites persist currently in culture fair-testing which require resolution in terms of fair assessment of academic achievement.

2. CULTURE-FREE TESTS

The main goal of culture-free tests was the development of measures that assessed innate characteristics and eliminated environmental characteristics (Oakland, 1982). Presently, the view exists that neither of these two characteristics can be independently isolated (Oakland, 1982). Like culture-fair tests, culture-free tests have been subjected to much criticism. Some of the criticisms of culture-free tests include:

1. Due to the elimination of verbalized content and reliance on performance based on spatial factors, culture-free tests do not measure intelligence (Feinberg, 1978);
2. The lack of language interaction between examiner and examinee causes the tests to lose credibility (Ekberg, 1979);
3. Controlling for familiarity of materials provided and verbalization requirements are inadequate since experimental variables include not only what a learner knows but also how the examinee thinks (Grover, 1981);
4. Controlling for verbalization and familiarization of materials does not necessarily insure that intellect rather than cultural differences are being assessed (Grover, 1981);
5. Culture-free tests require language interaction between the examiner and the examinee (Roth, 1976). This fact makes the notion that tests are not reliant on language, which is an entity mediated on culture, an incredible impossibility (Feinberg, 1978);
6. The fact that a test is restricted to pictorial representations does not imply that the test is culture-free or culture-fair (Feinberg, 1978);

7. Culture-free (and culture-fair) test tasks are rarely scrutinized to see if they are familiar to the examinee in terms of the actual test material, to the operation required, or to the operation when applied to specific material (Goodenow, 1976; Grover, 1981);
8. Test content which is non-verbal or non-academic may depend on experience for correct responses as the tasks are differentially encoded by various examinees and could trigger various cognitive strategies that depend on the experience of the examinee (Wagner, 1978);
9. Some of the test items used are rote memory achievement test type items (Feinberg, 1978). Similarly, changes in the type of test items occur during the test (Feinberg, 1978);
10. Ambiguous test instructions (Feinberg, 1978);
11. Subtle difficulties exist with respect to test items, responses, and examiner-examinee variables (Ekberg, 1979);
12. Lack of question-answer formats (Feinberg, 1978);
13. Error and ambiguity arising from various interpretations as to what constitutes a correct response (Feinberg, 1978);
14. Culture-free tests lack predictive validity (Harrington, 1979);
15. No test can be culture-free as responses depend on what the examinee has learned in his own culture (Noll, Scannell, and Craig, 1979);
16. Culture-free tests are culture-bound thus making it impossible to design a totally culture-free test (Harrington, 1979);
17. With specific reference to the culture aspect of culture-free tests, Ebel and Frisbee (1986) state:

> Attempts to build "culture-free" tests by eliminating items that discriminate between different cultures have been no more successful. If carried far enough, they result in

> eliminating all the items. There is no difference between individuals in their response to any test item that cannot be attributed to differences in culture, if culture is defined inclusively enough. Each of us lives in a somewhat different culture. Not only Eskimos and Africans, but also Vermonters and Virginians, farmers and city dwellers, boys and girls, even first-born and next-born in the same family live in somewhat different "cultures". The differences are not equally great in all these instances, but they exist as differences in all cases, and they can be used to support the contention that any item that discriminates is unfair. It is logically impossible for a culture-free test to discriminate among individuals, and there is no reason to use a test that does not discriminate between those who have more or less of an ability that is of interest to the user (p.307).

18. No culture-free test has been created which has gained wide usage or wide substitution for conventional intelligence tests (Noll, Scannell, and Craig, 1979).

SUMMARY

Culture-free tests have not been effective for cross-cultural evaluation of academic achievement. Culture-free tests possess many inadequacies related to the lack of language, test items, responses, examiner-examinee variables, predictive validity, and ambiguous instructions. Individual differences in response to test items can be attributed to culture (Ebel and Frisbee, 1986). Since a culture-free test cannot discriminate between individuals, a culture-free test is useless (Ebel and Frisbee, 1986) as one of the main purposes of tests is to discriminate between differences among individuals.

5. THE CONCEPT OF CULTURE AND TEST BIAS

The classical anthropological definition of culture is that culture is (Tylor, 1871):

> that complex whole which includes knowledge, belief, art, morals, law, custom, and other capabilities acquired by man as a member of society (p.1).

While there are various definitions of the word "culture", most of these definitions include beliefs, attitudes, rules, and values that define behavior in a given group of people (Woolfolk, 1987). In addition, groups can be described along religious, ethnic, regional, or other categories (Woolfolk, 1987). There are many cultures within a given nation (Woolfolk, 1987). All the individuals within a given country may share numerous similar values and experiences--especially due to the effect of mass communication; but different facets of living are influenced by differential cultural backgrounds (Woolfolk, 1987). Within a culture, uniformity among the members is fostered (Woolfolk, 1987). This uniformity serves to reinforce the distinctions among the various cultural groups and the diversity of culture in the total population (Woolfolk, 1987).

Every cultural group attempts to teach specific "lessons" in regard to living. Woolfolk (1987) illustrates the "lessons" that are indigeneous to all cultural groups as indicated in Chart 1.

CHART 1

Values and Attitudes: Potential Cultural Conflicts in the Schools

Significant Areas for All Cultures	Majority Cultural Values: The School's Expectations	Minority Cultural Values: Student Expectations
Interpersonal relationships	Competition among individuals; emphasis on individual accomplishment	Many Native American and Hispanic cultures: Mutual assistance; emphasis on group accomplishment
Orientation toward time	Planning for future; individual works for own future	Some Native American groups: Focus on present; cultural group provides for individuals' future
		Some Oriental cultures: Significance of past, tradition, ancestors
Valued personality type	Busy, occupied, efficient	Some Oriental and Hispanic cultures: Methodical, relaxed, meditative
Relationship of humanity to nature	Humans control and improve nature; focus on technology	Native American cultures: Humans at one with nature; mutual support of nature and humanity
Most cherished value	Individual freedom	Some Oriental cultures: Tradition; group loyalty

Adapted from M. L. Maehr (1974), Sociocultural origins of achievement. Monterey, CA: Brooks/Cole.

Due to these essential cultural variations, the values and behaviors a student gains at home or in the community may be unlike teacher or school expectations (Woolfolk, 1987). Generally, schools demand and reward the abilities, attitudes, and behaviors encouraged by the culture of the teachers (Woolfolk, 1987). An examination of the chart provides examples of possible conflict merely as a way of illustrating the part that culture plays in causing differences among people (Woolfolk, 1987). At present, there is a growing interest in preserving and valuing diversities in place of attempting to formulate a national culture (Woolfolk, 1987). However, within every group there are large disparities between persons (Woolfolk, 1987). Even though persons within the same neighbourhood share similar socioeconomic and cultural backgrounds, these persons are likely different from each other in various ways (Woolfolk, 1987). Another very likely source of differentiation is the influence of the family (Woolfolk, 1987).

Culture may include ethnic background, neighbourhood, peers, social milieu, and impacting variables such as language, religion, occupation, income, and values (Harrington, 1979). Cultural experiences differ between groups (Hynd and Garcia, 1979). Nobody has attempted to catalogue all the groups that create bias but the quantity of possibilities can make one wonder if the task is possible--Orientals, Latinos, Indians, and Blacks, people from different parts of a country, individuals from different communities but possessing the same ethnic identity, urban and rural, to name but a few (Hills, 1981).

Cross-cultural evaluation becomes exceedingly problematic when one takes into consideration the interrelationship and inseparability of culture, language, and cognition. There are differences in cognitive processing in various cultures. While "language makes the man" (p.52), language also creates barriers between various groups of men (Farb, 1968). The difficulty is far greater than merely a matter of language translation, for language questions the ways we experience and perceive the world (Farb, 1968). Farb (1968) states:

> Linguistically speaking, man is not born free. Our linguistic minds were made up for us from the day we were born. We have inherited our culture's particular habits of perception and expression, and these particular habits often differ markedly from those inherited by people in different cultures (p.52).

Language, in addition to blinding its speakers to particular perceptions, also directs speakers' attentions into specific habitual thought patterns (Farb, 1968). Alone, vocabulary is the least significant distinction among various languages (Farb, 1968). Rather, the totality of the internal structure and pattern of language is most significant (Farb, 1968). Each culture classifies experiences through its own lan-

guage (Farb, 1968). Mankind also learns unconsciously the method his own cultural group chooses from the sensory bombardments and classifies what it has chosen (Farb, 1968). Farb (1968) states that if one thinks in one language, one will think in a certain way, but if one thinks in a different language, one will think in a different way. The concept that man is a victim of his own language is derived from "the Whorf hypothesis". Whorf claimed that the composition of a given language (Farb, 1968):

> ... is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade ... We dissect nature along the lines laid down by our native languages ... Languages have grammars, which are assumed to be merely norms of conventional and social correctness, but the use of language is supposed to be guided not so much by them as correct, rational or intelligent THINKING (p.53).

In more recent times, Vygotsky (1978) describes the relationship between personal development and sociohistorical evolution. To Vygotsky, individual development is dependent on the individual using the tools of culture for expressing mental powers (Bruner and Haste, 1987). Also, numerous anthropologists are currently examining how language, non-verbal communication, and behavior structures constitute culture, and how meaning cultural frames mold the conceptions and perceptions of the individual (Bruner and Haste, 1987). Recent efforts on culture and language support the belief that cultural subgroups generate frames or schemas identifying some individuals as part of ingroups and others as part of outgroups (Bruner and Haste, 1987). In addition, the role of language in cognition is now recognized for its ability to differentiate concepts and to make these concepts available for transmitting a piece of culture (Bruner and Haste, 1987). Bruner (1987)

states that communication is not only composed of spoken content but also "illocutionary" characteristics through which an individual's intentions are sent and by which an individual clarifies her or his interpretation of the particular context in which the communication occurs. Even youngsters become very skillful in "reading" the intent of an individual's words, even though it may not correspond literally to the traditional content of the spoken message (Bruner, 1987). Culture is first, last, and always learned (Hall, 1986). An individual from China develops different schemata than the individual from New York (Grippin and Peters, 1984). Moreover, different cultures have different perceptions (Harrington, 1979). However, when an individual must make adjustments to, and must compete in a subculture or culture differing from which the individual is familiar, then cultural disparities quickly create cultural disadvantages for the individual (Anastasi, 1971). Attempts to "culturalize" the students, "providing a holding centre", "provide a crash course in ... customs", and "give students one year to adjust to a [new] culture", as well as assimilationist attitudes (Samuda, 1989a, p. 118) are ludicrous in the extreme in view of the complexity of culture.

What is most significant for the educational success for immigrant students is the quality of the communication between children and adults rather than the home language (Cummins, 1984). Parents and peers furnish the development of the interaction of culture, language, and cognition. Parents reject or accept selectively specific stimuli that are presented to the child: parents schedule, frame, filter, organize events, and mediate the relationships of space, time, affection, and causation (Lewis, 1989). Through such experiences, children eventually

construct structures of cognition and attach themselves to their past culture and societal reality (Lewis, 1989). Similarly, each culture furnishes a structure in which the organizing, interpreting, and comprehending of relationships and events occurs as a result of experience and exposure (Lewis, 1989). Organizing of experiences links the person with their culture or society (Lewis, 1989). In addition, it permits the person to be creative, adaptable, flexible, grounding themselves in the past, coping with the present, and looking forward to the future in their contextual culture (Lewis, 1989). Language plays a central role in these processes, as well as in other kinds of sharing and communicating (Lewis, 1989).

The absence of the development of linguistic structure can impede information processing in spite of well-constructed conceptual and experiential structures (Pascual-Leone and Ijez, 1989). This can happen in the instance of linguistic and cultural minorities who lack linguistic and conceptual learning experience in their new environment and in their new language (Pascual-Leone and Ijez, 1989). Hence, these minorities are unable to effectively coordinate the linguistic and conceptual structures in the new culture with experience structures derived from their original culture due to overtaxing individual mental processing capacities (Pascual-Leone and Ijez, 1989). Trueba (1987) claims that minority children who have learned English through a different culture are at a disadvantage in school learning due to a deficiency of social and cognitive skills which presupposes substantial and specific linguistic and cultural knowledge. Heath (1986) describes a comparative study of how two different working-class cultural views of language perform differentially in acquiring and retaining literacy. Similarly,

Lorimer (1986) speaks of the disservice schools provide due to the use of 'massified' or 'generic' culture textbooks and other teaching materials thereby enhancing cultural selection at the expense of allowing for any important differences in the needs of ethnic minorities.

Also, individual adjustment and learning styles are grounded in the culture of the home, the quality of one's interactions with one's peers, and societal interaction in the broadest sense (Samuda, 1989b). Culture also has a huge impact upon the student's cognitive styles (Berry, 1976). The learning styles which may be considered suitable in the culture of the home, may be considered inappropriate in the environment of the school (Das, 1973). Hence, qualitative variations between the learning style of the student and the instructional mode can increase the student's disorientation feelings and therefore impair educational achievement (Samuda, 1989b).

Research results provide credibility to the fact that minority students are grossly overrepresented in educable mentally retarded classes (Mercer, 1973), learning disability classes (Oritz and Yates, 1983), basic programs, vocational and special education classes, while being systematically omitted from educational programs that provide socioeconomic and professional mobility (Samuda, 1989c). Intelligence tests are linguistically and culturally biased in support of white, middle-class students and possess poor validity for students who are different socioculturally (Pascual-Leone and Ijez, 1989). Similarly, due to the powerful emphasis on skills and knowledge that are previously gained "from within a given cultural environment" (p.146), these tests are biased (Pascual-Leone and Ijez, 1989). Cultural bias comes from giving norm-referenced tests to individuals from different

cultural or ethnic backgrounds (Popham, 1988). Many children from these backgrounds have suffered irreparable damage as a consequence of decisions based on norm-referenced testing (Popham, 1988). Also, the technical procedures used for appraising the adequacy of norm-referenced testing instruments are inappropriate for analyzing tests devised mainly to serve evaluation functions (Popham, 1988).

Even though litigation has furnished minority students with the right to assessment in their dominant or primary language (Cummins, 1989), presently controversies continue unceasingly in reference to the extent to which bilingual education is effective and appropriate for fostering the academic development of minority students (Hakuta, 1986). Mastery of the native language and the host language implies complete command of "symbolic culturally determined archetonics" (Camilleri, 1986, p.142). The native language and the host language present an area of conflict which divides both communities (Camilleri, 1986). Two codes of language with conflicting significations provide both attraction and repulsion (Camilleri, 1986). Similarly, although acquisition of bilingualism relies on the way parents place themselves relative to the linguistic controls they put on their children and the kind of relationship they have with their children, parents, too, are victims of this bilingual context (Camilleri, 1986). The language employed in tests varies from the language of disfavourable classes, and, hence, confusion may result with respect to testing directions, and unfairness in testing verbal problems and verbal items (Camilleri, 1986).

In terms of the overrepresentation of minorities in learning disability classes, major junctures between practice and policy in fair testing (Garcia and Yates, 1986; Maldonado-Colon, 1986) and gaps in

knowledge by psychologists who have been isolated regarding such issues as inadequacies of standardized assessment, bilingualism at home, and language development patterns are all causal (Cummins, 1984). In addition, "institutionalized racism" in regard to testing is unchallenged virtually at policy and special provision levels, professional certification and training, and (with only minor exceptions) programs initiated by school boards (Cummins, 1989). In addition, the following causes have been identified as being causal on the overrepresentation of minorities in low educational streams (Samuda, 1989c):

1. Weak teacher expectation;
2. Weak student expectation;
3. Curriculum that is insensitive to multicultural background and non-white contributions to the larger society;
4. Curriculum and system rigidity which fails to utilize cultural diversification as a positive instructional resource;
5. Adverse impressions of heritage language and English as a Second Language (ESL) programs;
6. Indication of systematized racism in the schools;
7. Passive teaching and learning models instead of interaction teaching and learning models;
8. Lack of acceptance and recognition by educators in the mainstream of the various values, learning styles, and behaviors of the different minority groups;
9. Tendency to group classes homogeneously;
10. Penchant to place ESL students in special education or basic programs;
11. Absence of guidance counsellors with ample training in antiracism and non-biased testing;

12. Lack of interface and interaction between principals, counsellors, and teachers;
13. Absence of involvement of parents in assessment and placement decision-making;
14. Unsuitable assessment or no assessment when assessment should have been done.

Barriers that negatively affect assessing and programming opportunity of students due to their culture, ethnicity, race, and language include (Samuda, 1989c):

1. Ethnocentric and monolingual testers and tests;
2. Inaccurate and insufficient knowledge of minorities, testing, and tests;
3. Absence of a clear-cut policy with respect to testing and placement or/and differences between practice and policy;
4. Preservation of the status quo due to change implications (e.g. too costly to buy new tests and materials and the difficulties of altering teachers' beliefs and behaviors);
5. Lack of usage of the student's native language, background, and culture in the curriculum;
6. Lack of awareness and knowledge by teachers pertinent to the different and unique individual and cultural learning modes;
7. Insufficient training and insensitivity to the needs and problems of minorities by school staff, particularly those in counselling and guidance positions;
8. Insufficient understanding by educators about behaviors which are representative of various cultural backgrounds;
9. Unconscious prejudice and stereotyping by educators which reflect

expectation and interaction patterns in the wider community;

10. Insufficient communication between parents and others who are familiar with the student;
11. Incorrect perceptions about what tests measure and faulty notions of the knowledge that tests furnish;
12. Lack of knowledge and information by policy creators in regard to the state of the art of testing;
13. Erratic assessment methods;
14. Absence of continuous assessment and open placement;
15. Unfair and erroneous labelling practices derived from the results of one test or limiting assessment information;
16. Preconceived notions that achievement problems cannot be altered;
17. Psychological testing interpretation which emphasizes performance expectation rather than instructional intervention;
18. Practices and policies that have been contradicted by research (e.g. assessing immigrants after two years of national residency);
19. Absence of senior administrative level resources and policies in regard to testing and placement of minorities;
20. Predilection to homogeneous classroom groupings;
21. Absence of booster or transitional programs helping and encouraging students to move on to various levels and to experience various program choices;
22. Maintenance of a learning environment that is oriented toward transmission of culture;
23. Rigidity in regard to methods of gaining and expressing knowledge and comprehension i.e., the reliance on printed texts;
24. Lack of information provided to students and parents pertaining to

program selection and the consequences of individual choices at critical transitional points;

25. Assessment technology (instruments, software, and hardware) can distract or intimidate minorities;
26. Insufficient utilization and identification of school resource personnel.

Currently, tests provide a present functional level in regard to abilities and skills relevant to the culture to which the tests are designed, but there is misinterpretation frequently of what has actually been tested and the inability to test other, more critical competencies (Samuda, 1989c). Present assessment weaknesses include (Samuda, 1989c): (1) absence of legislative direction to special education with respect to racial and cultural diversity, new immigrants, and visible minorities and their parents, (2) false conceptions by test givers and users that capacity has been measured rather than knowledge acquisition as specified by majority middle-class assessment developers, and (3) inability to recognize test inadequacies and that the present practice of depending on standardized assessment methods is insufficient. No test, whether informal or formal, is adequate in itself since it is fallacious to believe that better and more of similar things will eventually lead to equitable assessment (Samuda, 1989c).

To hope to design a test which will satisfy all cross-cultural groups is foolhardy. Whether culturally biased tests are of any value depends on the definition of culturally biased tests and the purpose for which the particular test is employed (Kelley, 1982). The magnitude of culture as an impinging factor on standardized testing is clearly illu-

strated when one is reminded of Binet's two chief assumptions in regard to intelligence testing: (1) Those individuals being compared by a test should have similar experiences, and (2) the test should be a sample of intelligent behavior for the individual being examined (Scarr, 1978). The many dimensions of the concept of culture make impossible a clear-cut universal definition of culture that can be quantified numerically for psychometric purposes. The issue of cultural bias with specific reference to all standardized tests deserves much caution and attention. Clearly, Humphreys, Kendrick, and Wesman, (1975) define test bias as:

> A test is considered fair for a particular use if the inference drawn from the test score is made with the smallest feasible random error and if there is no constant error in the reference as a function of membership in a particular group.

Standardized tests lack validity for learners whose background and current academic curricula are at variance with that of the standardization sample (Gibson, 1980). Schooling reflects social values (Rosenbach and Mowder, 1981). Rosenbach and Mowder (1981) suggest that resolution to test bias issues require socio-political mobilization rather than psychometric improvement as test validity is continually high.

The curriculum may vary considerably from one culture to the next as the particular needs of a culture specify the direction of curriculum, content, and acculturation as derived from casual modelling (Stone and Neilson, 1982). Likewise, cultural bias may be caused by the criterion (Rosenbach, 1979), or by the nature of the criterion and the predictor (Williams, 1983). The problems of test bias cannot be solved until the testing purposes are clarified (Rosenbach, 1979).

Other cultural biasing factors in cross-cultural evaluation pertain

to instructional mode (Grover, 1981), task, situational, motivational, and personality factors (Dillon and Stevenson-Hicks, 1983). Examinee variables contributing to bias include: locus of control (Neely and Shaunessy, 1984), self-concept (Neely and Shaunessy, 1984), test anxiety (Neely and Shaunessy, 1984), cheating (Stanwyck and Abdelal, 1984), and lack of test-wiseness (Brescia and Fortune, 1988). Examiner variables that contribute to bias include: attribution patterns (Tom and Cooper, 1984), interrater agreement (Rengal, 1986), type of reinforcement (Carlson, 1983), and time and speed variables (Camilleri, 1986). Similarly, communication between the examiner and examinee can contribute to bias (Taylor and Lee, 1987).

Samuda (1989c) suggests that a good testing model must mirror the concepts of educational equality and equality for a "just" society by accounting for students' learning styles, motivation modes, linguistic, and cultural differences, and diversified capacities. Also, a good testing model should place assessment in a more suitable academic context including how testing is related to, and is reflective of instruction, how involving parents interacts with the assessment processes, and consideration of the knowledge foundations, mind sets, role definitions, and the process relationships of students, parents, testing, class teachers, and administrative personnel (Samuda, 1989c).

Hilliard III (1984) states that while no doubt the future will provide prolific evaluation activities, including new practices, materials, processes, applications for processing data, statistical techniques, and processes in observation technology, the fundamental concerns in evaluation are not only technological but theoretical, philosophical, and, possibly, political. In a democracy with democratic

education goals, there are processes of evaluation that are fitted to the broad goal (Hilliard III, 1984). These processes should incorporate attention to the context in which evaluation happens (Hilliard III, 1984). This entails developing systematized attention to theory, culture, history, and pedagogy (Hilliard III, 1984). Similarly, this also includes the notion that evaluation should be directed toward improving the processes of instruction (Hilliard III, 1984). While the ability to achieve these improvements are currently available, practical refinements must continue (Hilliard III, 1984). Test bias and misuse must end and new evaluation techniques must be developed with respect to cultural groups thereby reducing barriers to schools and occupations (Weiss, 1987b). Also, Rhodes (1988) suggests that while much controversy, work, and research on cultural content bias exists, very little effort has been devoted to cultural process bias.

Cross-cultural evaluation experience recommends sensitivity to cultural groups and the culture of the individual (Ginsburg, 1986). However, difficulties prevail due to the variability of culture:

1. The distribution of abilities is far larger within each ethnic or racial group than between groups (Woolfolk, 1987);
2. Culture is multicultural in nature (Hansen, 1979). Culture is an instrument for organizing experience (Hansen, 1979). Culture helps the individual to organize perception, process information, and solve the problems of daily living (Hansen, 1979). The individual capacity to learn and the disposition for learning are under the influence of standards maintained by other individuals or standards maintained by the individual being raised within a specific group (Hansen, 1979). Concurrently, differences in biographic experiences, variations in biogenetic composition, and predictable communication vagaries which

render sent messages, each donate to heterogeneity in communities which share many cultural standards (Hansen, 1979). Not one culture is exclusively shared by everyone in a group (Hansen, 1979). Also, even elements that may be shared will only be shared with some of the members of the group (Hansen, 1979). In that respect each group can be considered to be multicultural (Hansen, 1979).

3. Differences among subcultures may exceed cultural differences among nations (Camilleri, 1986);
4. One can regard children and culture as an analogy; every child has an individual culture or perspective (Ginsburg, 1986). Every child creates his own individual culture through interaction with various other "culture carriers", every individual possesses an individual cultural "resevoir" for daily living (Camilleri, 1986). Education is a type of negotiation between the individual and the educator's "cultural montage" (Camilleri, 1986). Likewise, each individual has subcultural variants (Camilleri, 1986).

Traditional, standardized testing methods often provide absurd data regarding competence (Cole and Scribner, 1974). Test items are frequently misinterpreted and misunderstood with the end result that norm standardization is frequently invalid (Ginsburg, 1986). One solution is to find tasks which possess relevance to the context of the particular culture (Ginsburg, 1986). While one given task can be appropriate for assessing competence in one culture, another task can be appropriate for assessing the same competence in a different culture (Ginsburg, 1986). Even though the tasks are different, the tasks may be equivalent subjectively in assessing the same processes (Ginsburg,

1986). Conversely, the same task may be inequivalent subjectively among cultures (Ginsburg, 1986). The main key for the measurement of competency is "subjective equivalence" rather than "objective identity" (Ginsburg, 1986, p. 173).

Standardizing is nonsensical and destroys the purpose for which standardization was designed (Ginsburg, 1986). For the achievement of "subjective equivalence", "objective identity" frequently has to be discarded and tasks modified to every perspective or culture since recent cross-cultural research indicates that cultural groups on occasion create distinctive cognitive patterns in responding to localized environmental needs (Ginsburg, 1986). Ginsburg (1986) recommends the exploration of nontraditional and flexible assessment methods including the clinical interview technique, introspection, and talking out loud. However, ultimately, cross-cultural or cross-ethnic comparison involves value judgements and social system structural bias (Samuda, 1989d).

SUMMARY

The concept of culture can be discussed from various perspectives with various negative implications for cultural test bias and academic achievement. While various educational authorities make suggestions for improving minority assessment and cross-cultural evaluation, test bias difficulties prevail due to the individuality and variability of culture as well as due to a lack of flexible testing methods. Therefore, at the present time, norm-referenced assessment seems to be a foolhardy and precarious enterprise at best. While much research and literature criticizes the folly of cross-cultural norm-referenced assessment, the DCAT claims fair assessment of minority groups on the basis of bias analysis of test items conducted by the delta method in the development of the DCAT. In view of the intensity and quantity of arguments in support of test bias in cross-cultural and minority group assessment, the concurrent validity of the DCAT is of high interest presently.

SUMMARY OF THE RELATED LITERATURE

A review of the literature reveals that cross-cultural evaluation is fraught with a very large number of variables which have an impact on academic achievement outcomes. A variety of factors have differential effects on academic achievement results so as to make cross-cultural evaluation of academic achievement a difficult, complex task. To date, culture-fair tests have proven to be highly inadequate.

Culture-free tests, like culture-fair tests have proven to be inadequate and senseless as culture-free tests lack the power to discriminate between differences among individuals. Cultural test bias is prevalent due to the variability and individuality of culture. In spite of the fact that cross-cultural evaluation is filled with so many impinging variables that can affect educational achievement outcomes, the CTBS, SAT, and CLDA are highly respected educational assessment tools that have been used for this purpose as these tests are among the best tests available currently. This study investigated the concurrent validity of the DCAT in relation to the CTBS, SAT, and CLDA.

CHAPTER III

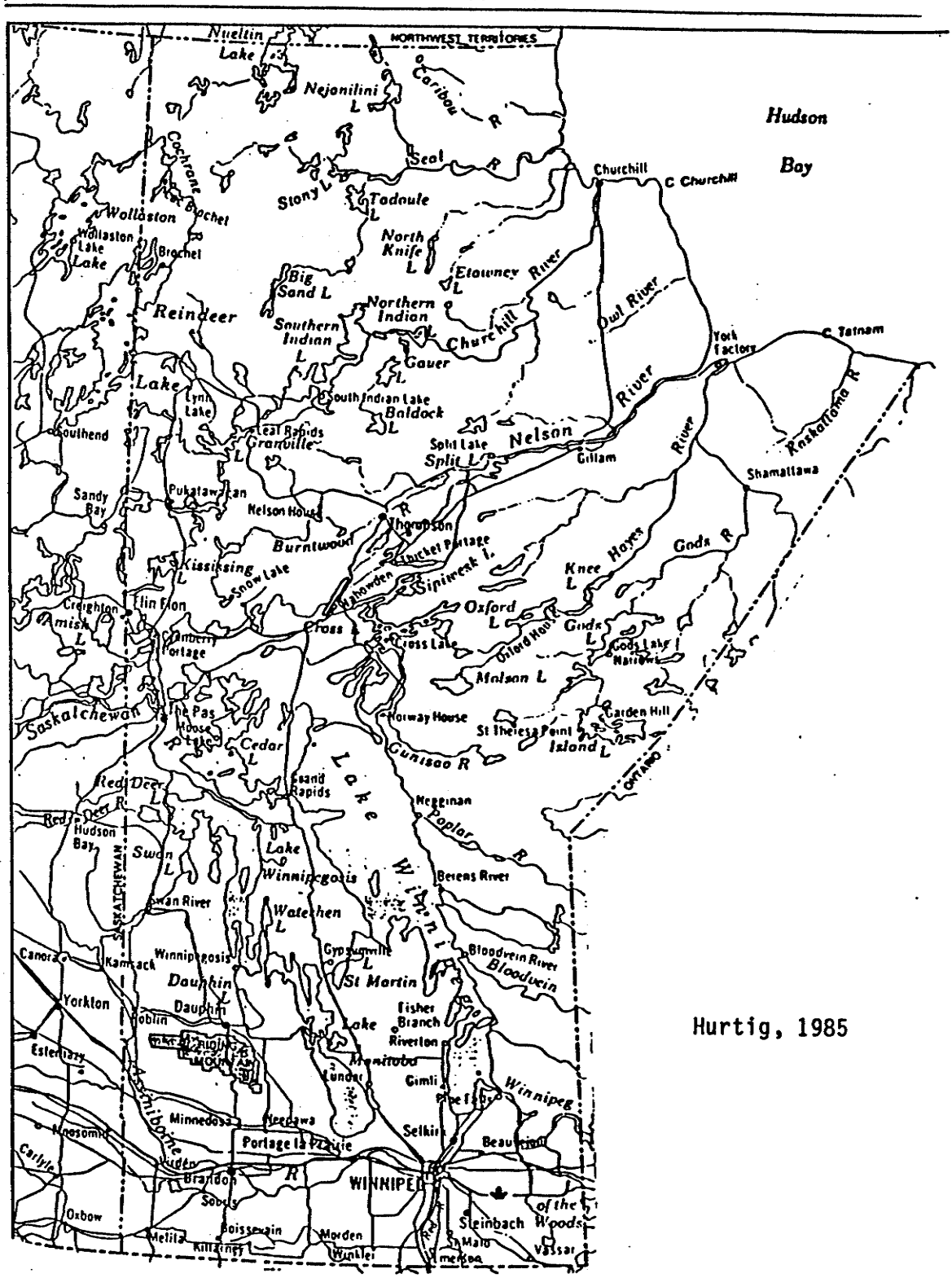
METHODOLOGY

INTRODUCTION

This chapter provides a description of the methods employed in the collection and analysis of data. Information is reported on the following:

1. Description of Brochet, Winnipeg, the schools, and the examinees
2. The test instruments employed
3. Test administration and scoring procedures
4. Scoring responses
5. The research design

FIGURE 1



Hurtig, 1985

1. BROCHET, THE SCHOOL, AND THE EXAMINEES

Brochet is a village located in northern Manitoba. Geographically, Brochet is located at latitude 57 degrees north and longitude 101 degrees west. Brochet was formally recognized as a village in 1906 by Treaty No. 10 (Smith, 1978). Prior to 1906, Brochet was known as Lac Caribou (Smith, 1978). Economically, Brochet served as a mission and trading post centre for Cree and Chipewyan Indians (Singh, 1982). In 1859, the Hudson Bay Company established an outpost at Brochet (Singh, 1982). In 1861, Bishop Grandin visited Brochet for the first time to perform baptismals and confirmations on the local population (Darveau, 1982). In 1973, the population of Brochet, which was approximately 1,500, was reduced by half when the Chipeweyans moved to Lac Brochet, approximately 90 kilometers north of Brochet (Singh, 1982).

Singh (1982) provides the following description of Brochet:

1. Population: 491;
2. Language: Mainly Cree; English literacy is less than fifty percent;
3. Religion: Catholic;
4. Economy: Approximately seventy-five percent of the population relies on social welfare assistance. Some Indians derive a livelihood from hunting, fishing, and trapping;
5. Communication and transportation: telephone, radio and airplane;
6. Expensive airfares, unpredictable weather, isolation, remoteness, and the absence of recreational facilities, makes Brochet a disadvantaged community for its inhabitants and for educational opportunity.

Brochet School is located in Frontier School Division No. 48 in the Province of Manitoba. Singh (1982) provides the following description of Brochet School:

1. The school enrollment consists of approximately 200 students, fourteen teachers, and two teacher aides;
2. The mother tongue for 97% of the school population is Cree;
3. The school staff, except for the two teacher aides, are all non-Native;
4. Educational services include nursery school to grade nine;
5. Teacher turnover rates are high: 1980--70%; 1981--36%; 1982--50%. Most new teachers have had no prior teaching experience and stay at Brochet for only one or two years.

WINNIPEG, THE SCHOOLS, AND THE PARTICIPATING EXAMINEES

Winnipeg is located at latitude 49 degrees and longitude 97 degrees (Energy, Mines, and Resources Canada, 1981), sixty miles north of the United States Border (Grolier, 1958). Metropolitan Winnipeg ranks fifth in size of Canadian cities after Toronto, Montreal, Vancouver, and Calgary (Gale, 1987). Winnipeg is the largest city as well as the capital city of the province of Manitoba (Hurtig, 1988). Situated midway between the Pacific and Atlantic Oceans, Winnipeg has been called the "Bull's Eye of the Dominion" (Hurtig, 1988). Also, Winnipeg has been called the "Gateway to the West" since Winnipeg is located where the Canadian Shield yields to the Canadian prairies (Hurtig, 1988). Winnipeg was formally incorporated as a city in 1874 with a population of 3,700 (Hurtig, 1988). Winnipeg has a land area of 3,394.82 square

kilometres and currently has a population of 625,304 (Statistics, Canada, 1988).

Winnipeg contains half the population and 68% of the employees of Manitoba (Hurtig, 1988). Also, Winnipeg produces 83% of manufactured items, and accounts for 62% of the retail sales for Manitoba (Hurtig, 1988). Winnipeg is well-known as a centre of transportation (Hurtig, 1988). The key to continued expansion is secondary manufacturing (Hurtig, 1988). Substantial increases in employment are evident in the private and public service industries (Hurtig, 1988). In addition, Winnipeg is also an insurance and financial centre (Hurtig, 1988). In 1979, the Winnipeg Development Incorporation was created for attracting high-technology industries to Winnipeg (Hurtig, 1988). Gale (1987) states that Winnipeg's population is composed of 40% British origin with a strong Scottish element. Other nationalities include: Ukrainian, German, French, Italian, Dutch, Phillipinos, Chinese, and Vietnamese (Gale, 1987). The economic, cultural, and social background of Winnipeg is similar to neighbouring United States areas (Gale, 1987).

Metropolitan Winnipeg consists of eleven school divisions providing educational services for 119,057 students (Weir and Wai Lai, 1978). Seven Oaks School Division, one of these eleven Metropolitan school divisions, has a total school enrollment of 7,332 students (Weir and Wai Lai, 1978) in seventeen elementary, junior high, and high schools. The Maples area possesses a total population of 13,975 with the ethnic distribution (Statistics Canada, 1988) as illustrated in the population summary chart. Elwick School and James Nisbet schools are elementary schools (grades kindergarten through grade six inclusive) within the Maples area of Seven Oaks School Division having total

CHART 2 - Maples Area

TOTAL POPULATION BY ETHNIC ORIGIN(29) BY SEX(3) - 1986 CENSUS
20% SAMPLE DATA
17 NOVEMBER 1988

FUNCTION: T1=COUNT

AREA: PA00483317 WIN. AR. (KILO) THE MAPLES

	TOTAL - BOTH SEXES	MALE	FEMALE
TOTAL - ETHNIC ORIGIN	13,975	6,725	7,255
BRITISH AND FRENCH	225	95	130
BRITISH AND OTHER	1,660	735	920
FRENCH AND OTHER	240	110	130
NATIVE AND OTHER	265	130	135
ALLOTHER MULTIPLE RESPONSES	1,840	880	960
MULTIPLE RESPONSES	4,220	1,950	2,275
BRITISH	1,725	800	920
FRENCH	240	120	125
DUTCH	25	-	15
GERMAN	765	385	375
ITALIAN	545	245	295
PORTUGUESE	550	285	265
JEWISH	910	485	420
UKRANIAN	1,390	645	745
OTHER EASTERN EUROPEAN	755	350	405
SCANDINAVIAN	110	45	65
ABORIGINAL PEOPLE	70	35	40
PACIFIC ISLAND ORIGINSINCL FILIPINO	1,145	560	585
OTHER EAST-SOUTHEAST ASIAN	200	90	105
SOUTH ASIAN	675	360	310
LATINCENTRAL AND SOUTH AMERICAN	30	25	-
CARIBBEAN ORIGINS	65	25	40
BLACK ORIGINS	130	75	55
ALLOTHER SINGLE RESPONSES	425	235	185
SINGLE RESPONSES	9,750	4,775	4,980

school populations of approximately 350 students each drawn from the above ethnic origins.

Edmund Partridge Junior High School (grades seven through nine inclusive) is a school within the Seven Oaks School Division having a total school population of approximately 325 students. The Edmund Partridge School area possesses a total population of 4,715 with the ethnic distribution as stated in the population summary chart (Statistics Canada, 1988). English is the language of instruction in these three schools, with French taught as a second language in grade four and up. Provision is made for heritage language instruction by those ethnic groups desiring to provide such instruction to their children after the end of the school day in the schools. The school year is from September to June. The pupil-teacher ratio is 1:23 at the elementary level.

CHART 3

Edmund Partridge Area

TOTAL POPULATION BY ETHNIC ORIGIN
1986 CENSUS NOVEMBER 1986
CENSUS CANADA, 1988

TOTAL - Ethnic Origin	4,715
British	820
French	100
Ukrainian	945
German	310
Polish	400
Filipino	90
Jewish	375
Other Single Origins	450
Multiple Origins	1,220

average number of years of teacher experience is 12 years in Seven Oaks School Division (Weir, 1983). Similarly, teachers qualifications are quite high with approximately 60% of teachers possessing a bachelor's degree (or equivalent) plus one year of teacher training (Weir, 1983), approximately 20% with two degrees or more, and approximately 20% without a Bachelor of Arts degree or equivalent (Weir, 1983).

2. THE TEST INSTRUMENTS EMPLOYED

The instruments used in this research study consisted of the following four tests:

A. The Canadian Test of Basic Skills (CTBS);

B. The Stanford Achievement Test (SAT);

C. The Concept Learning and Development Assessment (CLDA);

D. The Developing Cognitive Abilities Test (DCAT).

A. THE CANADIAN TEST OF BASIC SKILLS (CTBS), form 5 & 6 levels, levels 9-14 inclusive 1982 edition. The first edition of the CTBS was published in Canada in 1966. The CTBS is the Canadian version of the Iowa Test of Basic Skills. The nature and purpose of the CTBS according to the CTBS teacher's guide is to (King, 1982, p.3):

1. determine the developmental level of each pupil in order to adapt materials and instructional procedures more precisely to individual needs and abilities;

2. diagnose specific, qualitative strengths and weaknesses in a pupil's educational development;

3. indicate the extent to which individual pupils have the specific readiness skills and abilities needed to begin instruction or to proceed to the next step in a planned instructional sequence;

4. provide information useful in making administrative decisions in grouping or programming to accomodate individual differences;

5. diagnose strengths and weaknesses in group performance (class, buildings, or system) which have implications for change in curriculum or instructional procedures or emphasis;

6. provide a behavioral model to show what is expected of each pupil and to provide feedback which will indicate progress toward suitable individual goals;

7. report progress in learning the basic skills to parents in objective, meaningful terms.

The CTBS assesses reading, mathematics, and study skills (King, 1982). Measurement test items require a knowledge of metric (King, 1982). The CTBS was intended for use in grades kindergarten to twelve inclusive (King, 1982). The arrangement of the CTBS into multilevels permits the same test materials to be used by all

learners; less capable learners can begin at earlier items while no ceiling exist for the more capable learners (King, 1982). However, starting and stopping at various points in the test booklet may confuse some learners, hence, contributing to administration problems (Gronlund, 1981; Mehrens and Lehmann, 1986). The CTBS provides the following scores: norm-referenced scores, criterion-referenced scores, grade equivalent scores, percentile ranks, and stanines (King, 1982). The CTBS was originally standardized in 1966 using 30,000 English speaking students from 225 Canadian schools (Buros, 1972); the 1973 standardization consisted of 1.25% of Canadian schools or 139 Catholic and non-Catholic schools and 74,689 students in all ten provinces (Nelson, 1975).

The Canadian Test of Basic Skills reports the following validity data (King, 1982, p.2):

> Content specifications are based on over forty years of continuous research in curriculum, measurement procedures, and interpretation and use of test results. The skills 207 objectives represented in the tests were determined through systematic consideration of courses of study, statements of authorities in method, and recommendations of national curriculum groups. The item selection process involved a combination of empirical and judgemental procedures, including evaluation by representative professionals from diverse cultural groups.

Reliability varies per grade and per test (King, 1982). For the Multilevel Edition, grades three to eight the Teacher's Guide (King, 1982) reports internal consistency reliability coefficients for five of the eleven subtests only. Reported internal consistency reliability coefficients range from .87 to .96 for these five areas (King, 1982). Total reliability across grades three to eight is .97 to .98 (King, 1982). In summary, Mehrens and Lehmann (1986) state:

> ... the current ITBS (CTBS) has been carefully constructed. The 1978 ITBS (CTBS) was carefully normed on a representative sample. The Multilevel Edition is attractively packaged in a re-usable, spiral-bound booklet. The illustrations are clear and the type is easy to read. To accommodate more "tailor-made" individualized testing, new reporting services were developed. To assist teachers whose pupils are mostly "out-of-level", the tests were prepared and packaged by age rather than grade levels (p.290).

The CTBS can be machine or hand-scored (King, 1982). The total test battery requires approximately five hours for administration of which four hours and four minutes is actual student working time (See Table 1). The CTBS subtests administered in this research study at the grade three level and grade seven levels are: Vocabulary and Reading. At the grade three and seven levels, total actual student working time is 57 minutes. The CTBS subtests used in this research study are indicated with an asterisk (*) on Table 1.

B. THE STANFORD ACHIEVEMENT TEST (SAT), form E, Primary 3, and form E Advanced, Basic Battery. These editions of the SAT were published in the United States in 1982. The Stanford Achievement Tests, Primary 3, and Advanced tests assess listening comprehension, reading, arithmetic, spelling, and arithmetic skills. The Stanford Achievement Tests are designed for reflecting instruction in Canadian schools (Gardner, Rudman, Karlsen, and Merwin, 1982). In an attempt to ensure content validity, current curricular materials, textbooks, guidelines, syllabuses, research information, and curriculum specialists were consulted in order to prepare instructional objectives and to furnish well-proportioned curriculum coverage (Gardner, Rudman, Karlsen, and Merwin, 1982). Item construction principles were carefully adhered to (Gardner, Rudman, Karlsen, and Merwin, 1982). Every subtest and every test item were edited and reviewed for suitability for measure-

TABLE 1

Number of Subtest Test Items for Levels
Nine and Thirteen of the C.T.B.S. **

	Subtest	No. of items grade 3, level 9	No. of items grade 7, level 13	Time Limits
*	V: Vocabulary	30	43	15
*	R: Reading	44	57	42
	L-1: Spelling	30	43	12
	L-2: Capital-ization	28	31	12
	L-3: Punctuation	28	31	14
	L-4: Usage	27	31	14
	W-1: Visual Materials	36	52	40
	W-2: Reference Materials	37	47	25
	M-1: Math Concepts	28	42	25
	M-2: Math Problems	23	30	25
	M-3: Math Computation	39	45	20
	Total Test	350	452	244
*	Total Research Study	* 74	* 100	* 57

ment of the curricular objective, style, and content (Gardner, Rudman, Karlsen, and Merwin, 1982). Furthermore, each test item was edited and reviewed for sexual, racial, cultural, and ethnic bias (Gardner, Rudman, Karlsen, and Merwin, 1982). The Stanford Achievement tests were examined by curriculum and measurement experts, and participating teachers in local and national tryout programs (Gardner, Rudman, Karlsen, and Merwin, 1982). A panel of eight independent minority educators reviewed the SAT for elimination of sexual, ethnic, cultural, and racial bias (Gardner, Rudman, Karlsen, and Merwin, 1982).

The SAT was standardized in 1981 (Gardner, Rudman, Karlsen, and Merwin, 1982). The Fall Standardization sample consisted of approximately 250,000 students in 300 school districts whereas the Spring Standardization consisted of 200,000 students (Gardner, Rudman, Karlsen, and Merwin, 1982). Each student was also administered the Otis-Lennon School Ability Test in order to define the measurement ability of the SAT (Gardner, Rudman, Karlsen, and Merwin, 1982). Alternate form reliability was established by administering Form E and Form F to 20,000 pupils in grades two to eleven (Gardner, Rudman, Karlsen, and Merwin, 1982). These 20,000 pupils were also administered the Otis-Lennon School Ability Test (Gardner, Rudman, Karlsen, and Merwin, 1982). In order to equate the various levels of the SAT, scores on subtests were equated to scores on adjacent level subtests and a continuous scale of scores was developed to permit score interpretation across the test levels (Gardner, Rudman, Karlsen, and Merwin, 1982). In addition, in order to equate the various levels of the test, 20,000 pupils from grades one to eight, and grade ten were administered two adjacent test levels of the Stanford Achievement tests, using a random

assignment technique (Gardner, Rudman, Karlsen, and Merwin, 1982).

The SAT provides the following scores: norm-referenced scores, percentiles, stanines, grade equivalents, and scale scores, and normal curve equivalents (Gardner, Rudman, Karlsen, and Merwin, 1982).

Construct validity was ascertained by having a larger ratio of pupils passing test items at higher grade levels (Gardner, Rudman, Karlsen, and Merwin, 1982). The test authors claim that test validity can be determined by careful evaluation of the test content which is furnished in the Stanford Index of Instructional Objectives and comparing this information to your own curricular instructional objectives as a means of judging the validity of the test for your own needs (Gardner, Rudman, Karlsen, and Merwin, 1982). Internal consistency reliability, alternate form reliability, and standard error of measurement data as well as intercorrelations between subtests and the Otis-Lennon School Ability Test are provided (Gardner, Rudman, Karlsen, and Merwin, 1982). The Form E Primary 3 level of the SAT for grades 3 and 4 provides reliability coefficients for the subtests between .84 and .96 for grade 4, Fall Standardization Sample, alternate-form subtest reliability coefficients between .73 and .90 for grade three, and intercorrelations between .45 and .80 with the Otis-Lennon School Ability Test at the beginning of grade four (Gardner, Rudman, Karlsen, and Merwin, 1982). No information is provided re: reliability coefficients of subtests for grade 3, or inter-correlations with the Otis-Lennon School Ability Test (Gardner, Rudman, Karlsen, and Merwin, 1982).

For the Form E Advanced Stanford Achievement Test, reliability coefficients between .80 and .94 are reported for grade 7, Fall Standardization Sample only. Alternate forms reliability coefficients between

.74 and .95 for the subtests, and intercorrelation coefficients between .47 and .85 with the Otis-Lennon School Ability Test at the beginning of grade eight are reported (Gardner, Rudman, Karlsen, and Merwin, 1982).

With respect to the SAT, Mehrens and Lehmann (1986) state:

> ... one of the most popular and useful standardized achievement batteries used in our schools (p.265-266) ... it represents one of the better test batteries for surveying school achievement from kindergarten to high school (p.266) ... The Stanford is quite valid for evaluating pupil status and progress. For teachers who frequently like to obtain a cumulative index of their pupils' progress, the Stanford ... provides a cumulative assessment of pupil knowledge with an articulated series of tests from grades K. to 13. Many instances arise when a teacher is interested in knowing whether their pupils are working at their capacity. The Stanford, because it was standardized with the Otis-Lennon Ability Test, provides for such information. ... Despite some minor criticisms of the Stanford ... we recommend it highly. The Stanford series were meticulously constructed and standardized (p.286).

The SAT can be machine or hand-scored. The total test battery requires approximately 275 minutes of actual student working time at the grade three level and approximately 220 minutes of actual student working time at the grade seven level (see Table 2). The SAT subtests administered in this research study at the grade three and seven levels are: Reading Comprehension, Vocabulary, Concepts of Number, Mathematics Computation, and Mathematics Applications. The SAT grade three level subtests used in this study require 140 minutes approximately at the grade three level and 145 minutes approximately at the grade seven level. The SAT subtests used in this study are indicated in Table 2 with an asterisk (*).

C. CONCEPTUAL LEARNING AND DEVELOPMENT ASSESSMENT (CLDA) (Wisconsin Research and Development Centre For Cognitive Learning, 1977).

The purpose of the CLDA is to test concept attainment mastery. This test is designed on the Conceptual Learning and Development Model

TABLE 2

Number of Subtest Test Items for Primary 3, and Advanced Level, SAT **

	Subtest	No. of items Primary 3 Level	Time limits Primary 3 Level	No. of items Advanced Level gr. 7	Time limits Advanced Level
*	Reading Comprehension	60	30	60	30
*	Concepts of Number	34	20	34	20
*	Mathematics Computation	42	35	44	40
*	Mathematics Applications	38	35	40	35
	Spelling	36	15	50	15
	Language	46	30	59	30
*	Vocabulary (Teacher-dictated)	38	20	40	20
	Listening Comprehension (Teacher-dictated)	40	Approx. 30	40	Approx. 30
	Word Study Skills	54	30	--	--
	Total Test	388	Approx. 245	367	Approx. 280
*	Total Research Study	* 212	Approx. * 140	* 218	Approx. * 145

(CLDM) by H. J. Klausmeier.

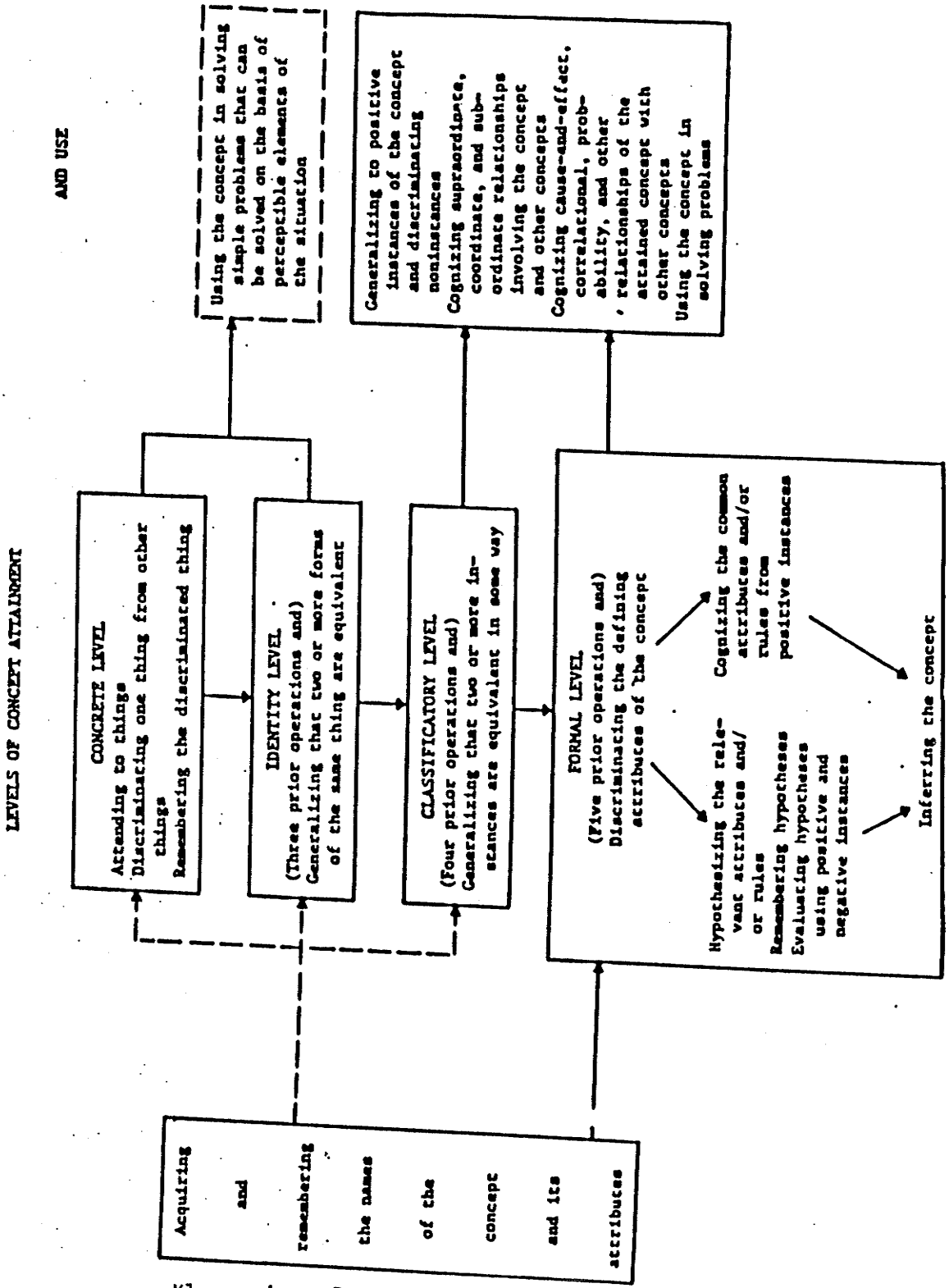
The CLDM is a descriptive, analytical model which defines four levels of concept attainment and likely extensions and uses of concept mastery, describes the cognitive processes needed in learning new concepts at the four levels, and suggests external and internal learning conditions related to the specific levels. The concept attainment levels plus the conditions and processes of learning have been isolated through research and analyses in schools and laboratories. Similarly, guidelines have been created using the CLDM and school-based research. According to the CLDM, a concept is defined as (Klausmeier, Bernard, Katzenmayer, and Sipple, 1977):

> ... ordered information about the properties of one or more things -- objects, events, or processes -- that enables any particular thing or class of things to be differentiated from, and also related to, other things or classes of things. The word concept is used by Klausmeier, Ghatala, and Frayer (1972) to designate mental constructs of individuals as well as identifiable public entities that comprise part of the substance of the various disciplines. Thus, the term concept is used appropriately in two different contexts just as many other English words are ... Klausmeier, Ghatala, and Frayer (1972) carried the definition further by specifying eight attributes of concepts: learnability, usability, validity, generality, power, structure, instance numerousness, and instance perceptibility (p.3).

The CLDM uses concepts symbolized by words which are definable by attributes. Since one is unable to locate definitions for every word, researchers and developers of curricular materials have to specify the clarifying attributes cooperatively or independently with representatives from the subject disciplines.

The Levels of Concept Attainment diagram illustrates the operations that are a part of attaining the various concept levels. These opera-

CHART 4



Klausmeier, Bernard, Katzemeyer & Sipple, 1977

tions provide a context to explain short-term phenomena and to identify changes occurring over time as newer operations surface and increasingly higher levels of attainment are made possible. The CLDM uses the term operations as Guilford (1967) did. Guilford describes memory, productive thinking, cognition, and evaluation operationally and formally in test performance terms. Guilford claims that cognition and cognized patterns should be related. Guilford (1967) defines cognition thus:

> Cognition is awareness, immediate discovery or rediscovery, or recognition of information in various forms; comprehension or understanding .... The most general term, awareness, emphasizes having active information at the moment or in the present ... the term, recognition, is applied to knowing the same particular on a second encounter ... if cognition is practically instantaneous, call it recognition; if it comes with a slight delay, call it "immediate discovery" (p.203-204).

Classes and information units, which are two lower levels of Guilford's taxonomy, include instant discovery, recognition, and awareness. However, comprehension, equated with understanding, is applicable to systems and relations which are products of a higher level. Hence, understanding of patterns, structures, sequences, or principles requires comprehension. The four levels of Concept Attainment are illustrated as a means of describing the four levels for which learners may master an identical concept, the processes involved at every level, concept use and extension, and acquiring names and related attributes.

An exclusive model feature is the specification of four different levels of one concept instead of suggesting an end mastery level as soon as a concept is learned for the first time. Hence, the CLDM describes in a developmental, long term context the changes that happen in concept mastery for the same learner.

At the concrete level, concept attainment occurs when the learner recognizes an object with which he or she has had prior experience. At this time, the object that he or she has experienced initially is experienced in the same fashion precisely in the second and later situations. The individual attends to and discriminates the object from other objects (Gagne, 1970) and represents it internally as an image of an object directly experienced by the senses (Woodruff, 1961). At the identity level, concept attainment occurs when the individual recognizes the same object as at the concrete level but the object is now presented in a varying modality or a different spatial-temporal perception. Generalization of two or more representations of the same objects as being equivalent is the most significant feature at this level.

Classificatory level concept attainment occurs when the individual deals with two examples, at minimum, of similar sets of objects as being equal in some manner. However, individuals may group objects and still remain unable to describe the reasons for grouping (Henley, cited in Deese, 1967). At the formal level, one is able to name and discriminate the concept, name defining values and attributes, and evaluate accurately situations as being placed or not being placed in a set due to the absence or presence of definitional attributes. Concept attainment at the formal or classificatory level can be used in four different ways as illustrated previously: (1) generalization to new situations, (2) cognition of supraordinate-subordinate relationships, (3) cognition of numerous other relationships among concepts, and (4) generalization to problem-solving instances. Horizontal and vertical transfer occurs through extension and utilization of attained concepts.

The CLDA battery is designed for usage from kindergarten through grade twelve. However, it is most likely undesirable or unnecessary to administer all subtests or items at every grade level, as intermediate students will get all test items in booklet A correct and primary children will not be able to do some items in booklets C and D correctly. The amount of test items that should be administered must be decided by taking into account research goals. The test battery is given to whole classes of students at higher primary levels and upper grades, and to small groups of five to seven students at the lower primary levels.

The total test battery including directions are read to the students at all grade levels. Student responses are marked on the test booklet pages. Students are not permitted to alter responses on subtests previously taken. For this research study only the complete Noun test, which consists of 62 items and requires approximately 50 minutes to administer, was given to grades three and seven. Table 3 illustrates the Noun test information pertinent to this research study. The low Reliability coefficients obtained in this study indicate that the Noun Test test items possess a low degree of consistency for grade seven Winnipeg and grade three and seven Winnipeg groups. Hence, the Noun Test is not as consistent a measure of language skills for grade three Winnipeg and grade three and seven Brochet groups as it is for grade seven Winnipeg groups.

D. DEVELOPING COGNITIVE ABILITIES TEST (DCAT), forms A and B, 1980.

The DCAT measures those abilities and characteristics which contribute to academic achievement (DCAT Technical Manual and Norms, 1983). The DCAT measures two aptitude dimensions (DCAT Technical Manual and Norms, 1983). The first dimension of the DCAT measures verbal, spatial, and quantitative abilities (DCAT Technical Manual and Norms, 1983). Whereas verbal skills are the foundation for achievement in areas con-

TABLE 3

Number of Subtest Test Items for
CLDA Noun ** for Grades 3 and 7

No. of Items	Time Limits
62	50 min. approx.

sisting of language and reading, quantitative skills are the foundation for achievement in numerical, algebraic, and science subjects. Spatial skills are required in geography, science, and geometry (DCAT Technical Manual and Norms, 1983). The second dimension of the DCAT furnishes data on five of the six classes of Bloom's Taxonomy: Knowledge, Comprehension, Application, Analysis, and Synthesis (DCAT Technical Manual and Norms, 1983). Achievement results from these subtests provide information about the degree of individual abstract thinking (DCAT Technical Manual and Norms, 1983). For the grade two level only, the content domain is composed of verbal, quantitative, and spatial levels only. The cognitive dimension was omitted as the test designers felt that a group test would not measure the five categories with precision at such a young age (DCAT Technical Manual and Norms, 1983). The DCAT is a unique assessment instrument tool for the measurement of student ability (DCAT Technical Manual and Norms, 1983). The combining of these two aptitude dimensions (cognitive level and content domain) provides the test user with a unique instrument for assessing examinee ability (DCAT Technical Manual and Norms, 1983). The DCAT is based on the premise that suitable instruction can improve and change those abilities and characteristics (DCAT Technical Manual and Norms, 1983). The data gained from the DCAT can provide the

TABLE 4

Number of Items For Each of the DCAT Subtests
Levels Three and Seven/Eight A **

	Subtests	No. of items Level 3 Grade 3	Time Limits	No. of items Levels 7/8A Grade 7 & 8	Time Limits
*	Verbal	40		31	
*	Quantitative	26		29	
	Spatial	14	14	20	22
	Total Test	80	50	80	50
*	Total Research Study	* 66	* 42	* 60	* 38

foundation for altering instruction for meeting individual needs as maturity and appropriate instruction can create positive changes in cognitive and academic skills (DCAT Technical Manual and Norms, 1983).

The DCAT measures aptitude from grades two through twelve using six different tests (DCAT Technical Manual and Norms, 1983). While level two consists of nine subtests composed of 80 test items, levels three to 9/12 consist of a single test composed of eighty test items (DCAT Technical Manual and Norms, 1983). In addition, while working time for level two is paced by the examiner, 50 minutes is suggested as working time for the entire test (DCAT Technical Manual and Norms, 1983). For level two, items were chosen for eighty-four various tasks. Two criteria were important for item selection: (1) each task was to contain three easy items that 90 percent of the students could pass, and (2)

items were to be chosen that would enhance test reliability. The data from this second tryout were used to choose items for the final edition of level two (DCAT Technical Manual and Norms, 1983). Traditional and Rasch latent trait analyses were employed to help determine final test items (DCAT Technical Manual and Norms, 1983). For levels three to 9/12 inclusive, test items were created to meet the precise criteria for both test dimensions (DCAT Technical Manual and Norms, 1983). Verbal test items were controlled by selecting words and terms from The Living World Vocabulary (Dale and O'Rourke, 1979), quantitative test items were selected from the most used mathematics programs, and spatial test items were designed without standards to ascertain level of difficulty other than indices of statistical difficulty (DCAT Technical Manual and Norms, 1983). Two tryouts were administered (DCAT Technical Manual and Norms, 1983).

The DCAT was standardized during the same year as the Comprehensive Assessment Program Achievement Series and the School Attitude Measure (DCAT Technical Manual and Norms, 1983). These three tests comprise the Canadian Comprehensive Assessment Program (CCAP) (DCAT Technical Manual and Norms, 1983). The DCAT was standardized on 13,047 students from grades two through twelve from 45 Canadian school districts and 10 separate school systems. The standardization sample was derived by using a stratified random sampling technique which is based on proportionate numbers of students from Western, Central, and Eastern Canada (DCAT Technical Manual and Norms, 1983).

Validation research studies were done during and after test standardization to furnish data necessary for validating all forms and all levels of the DCAT. Test content was formed, examined, and refined on commonly used objectives by qualified groups. Test items were designed from these objectives and evaluation was conducted by reviewers and a minority-group panel. Similarly, classroom teachers examined item

content during field assessment. Psychometric analysis was also employed (DCAT Technical Manual and Norms, 1983). Item analysis is currently being done (DCAT Technical Manual and Norms, 1983). The Technical Manual and Norms (DCAT Technical Manual and Norms, 1983) furnishes little data on test validity. Correlation tables with the Comprehensive Assessment Program Achievement Series (CAPAS) are furnished. Verbal subtests have high correlations with the verbal CAPAS subtests, especially Reading (Fox, 1985). Modest correlations between CAPAS Mathematics and DCAT Quantitative and Spatial are evident (Fox, 1985). Examination of Quantitative and Verbal subtests suggest that the DCAT is very similar to the Scholastic Aptitude Test (SAT) and the School and College Abilites Test (SCAT) (Fox, 1985). However, no studies are documented which compare the DCAT with the same type of measures (Fox, 1985). Research is required, especially for the Spatial Component (Fox, 1985).

Reliability is "fairly high" for all DCAT forms and levels for Spatial and Verbal subtests while estimations for Quantitative subtests tend to be somewhat lower (i.e. from .60 to .68) for level 5/6 for grades 5, 6, and 7 (Fox, 1985). In addition, parallel test forms correlate highly with composite scores but provide lower correlations for Spatial subtests (Fox, 1985). Kuder-Richardson Formula 20 estimates, which tend to be higher than Kuder-Richardson Formula 21 estimates, are being calculated currently for DCAT test items (DCAT Technical Manual and Norms, 1983). Statistical item bias analysis was also conducted using the delta method so that bias for women and minority groups is within a reasonable range (DCAT Technical Manual and Norms, 1983).

The DCAT provides test scores for predicting anticipated achievement levels on the CCAP (DCAT Technical Manual and Norms, 1983). Similarly, the DCAT provides the following scores for the content dimension: raw scores, percentage correct, local stanines, national

stanines, and equal interval scores and the following scores for the cognitive dimension: percentage correct and average percentage correct (DCAT Technical Manual and Norms, 1983). The DCAT may be machine or hand-scored. Table 4 illustrates test subtests pertinent to this research study with an asterisk (*).

In reviewing the DCAT, Fox (1985) states:

> A great deal of time, effort, and thought appears to have gone into the construction of the DCAT. In time it may well become a widely used and respected measure of developing intellectual abilities. Its potential for diagnostic-prescriptive instructional planning in relationship to the taxonomy could indeed make it more generally useful than other more traditional aptitude measures. At present the reliability and predictive validity studies are not as extensive and impressive as for some other tests as the SAT and SCAT. This does not negate its potential for immediate use, but more research is needed to establish whether or not the test has predictive powers equivalent to, greater than, or less than more traditional measures. For now, the user should be cautioned about the ways in which this test may be closer to an achievement measure or different from other aptitude measures. In this case different may indeed eventually prove better (p.461).

3. TEST ADMINISTRATION AND SCORING PROCEDURES

The tests were administered according to the directions supplied by the tests' administration manuals with respect to procedures, time limits, and materials. The entire test battery was administered in both locations within a two week period.

4. SCORING OF RESPONSES

All the tests were marked in accordance with the directions given in the test manuals. Raw scores were computed for each of the subtests administered for this study.

5. RESEARCH DESIGN

Data collected from the administration of the four tests to classrooms of grade three and grade seven learners were organized into the following eight groups:

1. Winnipeg - Grade 3 - Verbal Component
2. Winnipeg - Grade 3 - Quantitative Component
3. Winnipeg - Grade 7 - Verbal Component
4. Winnipeg - Grade 7 - Quantitative Component
5. Brochet - Grade 3 - Verbal Component
6. Brochet - Grade 3 - Quantitative Component
7. Brochet - Grade 7 - Verbal Component
8. Brochet - Grade 7 - Quantitative Component

Each group of data was analyzed to provide the means, standard deviations, sample sizes, Kuder-Richardson Reliability coefficients (KR20), and the correlations between the DCAT and the other three tests for each group. The Cronbach Apha was used to estimate the internal consistency of each test. The Pearson Product Moment Correlation coefficients between the DCAT and the other three tests were computed to identify the relationships between the DCAT and the other three tests.

CHAPTER IV

RESULTS, ANALYSES, AND DISCUSSION

The purpose of this chapter is to present the descriptive and inferential statistics and to discuss the results of this study.

The means, standard deviations, and the correlation coefficients among the variables were computed on the 286 IBM microcomputer using the StatPac Gold program. Table 5 presents the means, standard deviations, Kuder-Richardson (KR20) Reliability coefficients, and sample sizes for each of the subtests. Table 6 presents a summary of the Cronbach Alpha Reliability coefficients for each test used in this study. Tables 7 - 14 present the correlation matrices for the sets of variables in each location and grade. The Pearson Product Correlation coefficient was used to measure the relationship between pairs of subtests. The Cronbach Alpha Reliability coefficient was used to examine the average intercorrelations of the subtests within a group (i.e. verbal and quantitative versus grade versus location). To compare the equality of correlation coefficients across both locations, Bonferoni at the .01 level of significance was used.

PRESENTATION OF DATA

Descriptive Statistics

Table 5 contains the results of the descriptive statistics of the ten variables. The means (M), standard deviations (SD), sample sizes (N) and the reliability coefficients (KR20) are grouped by grade levels for each location.

Verbal Component

DCAT Verbal The range of scores for grade three Winnipeg was 13 to 37. The grade three Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 28.38, 5.71, and 0.83 respectively. A raw score of 29 is equivalent to the 48th percentile (DCAT Technical Manual and Norms, 1983). The grade three Winnipeg mean of 28.39 is equivalent to the 43rd percentile which is seven points below the 50th percentile. Twenty-three students performed at or below the 48th percentile and 17 students performed above the 48th percentile.

The range of scores for grade three Brochet was seven to 30. The grade three Brochet mean, standard deviation, and KR20 Reliability coefficient were 19.61, 7.98, and 0.90 respectively. A raw score of 29 is equivalent to the 48th percentile. The grade three Brochet mean of 19.61 is equivalent to the 12th percentile which is 30 points below the 50th percentile. Twenty-eight students performed below the 48th percentile and three students performed above the 48th percentile.

The range of scores for grade seven Winnipeg was six to 26. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 18.59, 5.06, and 0.82 respectively. A raw score of 19 is

TABLE 5

Means, Standard Deviations, KR20 Reliability Coefficients, and Sample Sizes For Grade 3 and Grade 7 Winnipeg and Brochet

T E S T S	G R A D E			
	3 Winnipeg	3 Brochet	7 Winnipeg	7 Brochet
DCAT-Verbal				
Mean	28.38	19.61	18.59	11.00
S.D.	5.71	7.98	5.06	2.85
KR20	0.83	0.90	0.82	0.34
N	40	31	17	30
CTBS-Vocabulary				
Mean	20.43		27.20	
S.D.	5.67		8.95	
KR20	0.85		0.91	
N	40		49	
CTBS-Reading				
Mean	24.98		27.55	
S.D.	7.55		11.68	
KR20	0.85		0.92	
N	40		49	
SAT-Rdg. Comp.				
Mean	36.77	23.36	33.12	23.77
S.D.	10.66	5.81	14.14	7.11
KR20	0.94	0.65	0.95	0.77
N	39	25	42	30
SAT-Vocabulary				
Mean	18.74	11.00	23.81	11.29
S.D.	4.77	2.69	7.21	2.95
KR20	0.78	0.05	0.86	0.23
N	38	25	42	28
CLDA-Noun				
Mean	33.92	31.00	42.91	35.20
S.D.	4.51	2.97	7.58	4.03
KR20	0.25	0.37	0.86	0.46
N	21	14	44	30
DCAT-Quantitative				
Mean	11.35	7.45	10.71	9.00
S.D.	3.10	2.73	3.69	3.59
KR20	0.48	0.38	0.57	0.61
N	40	31	17	30
SAT-Concepts of No.				
Mean	19.33	10.44	19.02	11.47
S.D.	5.55	3.31	7.47	2.80
KR20	0.85	0.40	0.89	0.21
N	39	27	46	30
SAT-Math Comput.				
Mean	23.76	16.55	20.96	14.53
S.D.	8.95	8.48	9.68	5.93
KR20	0.91	0.90	0.92	0.81
N	40	27	46	30
SAT-Math Applics.				
Mean	22.03	11.31	19.74	11.40
S.D.	7.30	5.21	9.04	4.78
KR20	0.87	0.77	0.91	0.69
N	40	26	46	30

equivalent to the 47th percentile. The grade 7 Winnipeg mean of 18.59 is equivalent to the 47th percentile which is three points below the 50th percentile. Ten students performed at or below the 47th percentile and seven students performed above the 47th percentile.

The range of scores for grade seven Brochet was five to 15. The grade seven Brochet mean, standard deviation, and KR20 Reliability coefficient were 11.00, 2.85, and 0.39 respectively. A raw score of 19 is equivalent to the 47th percentile. The grade 7 Brochet mean of 11.00 is equivalent to the 6th percentile which is 44 points below the 50th percentile. Twenty-eight students performed below the 47th percentile and two students performed above the 47th percentile.

The mean DCAT Verbal scores obtained by all groups were variable. While the mean DCAT Verbal score obtained by grade seven Winnipeg was average, the mean DCAT Verbal scores were slightly below average for grade three Winnipeg and below average for grade three and grade seven Brochet. The mean score obtained by the grade three Winnipeg group indicated that the grade three Winnipeg verbal cognitive abilities were average. However, the mean score obtained by grade seven Winnipeg groups was slightly below average. The mean scores obtained by grade three and seven Brochet groups indicated that cognitive abilities were below average for these groups.

Likewise, the KR20 Reliability coefficients obtained for DCAT Verbal were high for grade three Winnipeg, grade three Brochet, and grade seven Winnipeg. However, the KR20 Reliability coefficient for grade seven Brochet was low. The low Reliability coefficient obtained by grade seven Brochet indicated that the DCAT Verbal test items were not a consistent measure of verbal abilities for grade seven Brochet

students. The higher reliability coefficients obtained by grade three Winnipeg, grade three Brochet, and grade seven Winnipeg indicated that the DCAT Verbal test items were more consistent for grade three Winnipeg, grade three Brochet, and grade seven Winnipeg than for grade seven Brochet. Hence, the DCAT Verbal subtest was not as reliable a measure of verbal abilities for grade seven Brochet as it was for grade three Winnipeg, grade three Brochet, and grade seven Winnipeg.

CTBS Vocabulary The range of scores for grade three Winnipeg was nine to 30. The grade three Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 20.43, 5.67, and 0.85 respectively. The raw score of 23 is equivalent to a grade score of 3.9. The mean raw score of 20.43 is equivalent to a 3.5 grade average or 0.4 below grade average. The raw score of nine is equivalent to a grade 2.4 or 1.5 below grade average while the raw score of 30 is equivalent to a 6.1 grade average of 2.3 above grade average. Twenty-five students performed below grade average and 15 students performed at or above grade average.

The range of scores for grade seven Winnipeg was eight to 41. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 27.20, 8.95, and 0.82 respectively. The raw scores of 32 and 33 are equivalent to grade scores of 7.8 and 8.0. The mean raw score of 27.20 is equivalent to a grade average of 7.1 or 0.8 below grade average. The raw score of eight is equivalent to 2.2 or 5.7 below grade average. The raw score of 41 is equivalent to a 9.7 grade average or 1.8 above grade average. Thirty-two students performed below grade average and 17 students performed above grade average.

The mean CTBS Vocabulary scores obtained by grade three and seven Winnipeg groups were slightly below average. The mean scores obtained

in both grades three and seven Winnipeg indicated that the vocabulary skills of grade three and seven Winnipeg groups were slightly below average. The KR20 Reliability coefficients obtained for CTBS Vocabulary were high for both grade three and grade seven Winnipeg. The high reliability coefficients obtained in grade three and seven Winnipeg indicated that the CTBS Vocabulary test items were very consistent for these two groups of students. Hence, the CTBS Vocabulary subtest was a reliable measure of vocabulary skills for Winnipeg.

CTBS Reading The range of scores for grade three Winnipeg was 13 to 38. The grade three mean, standard deviation, and KR20 Reliability coefficient were 24.98, 7.55, and 0.85 respectively. A raw score of 28 is equivalent to a grade score of 3.9. The mean raw score of 24.95 is equivalent to a grade 3.7 average or 0.2 below grade level. The raw score of 13 is equivalent to a grade three score of 2.4 or 1.5 below grade average. The mean raw score of 38 is equivalent to a grade average of 5.0 or 1.1 above grade average. Twenty-three students performed below grade average and 17 students performed at or above grade average.

The range of scores for grade seven Winnipeg was 12 to 54. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 27.55, 11.65 and 0.91 respectively. The raw score of 37 is equivalent to a grade score of 7.9. The mean raw score of 27.55 is equivalent to a 6.9 grade average or 1.0 below grade level. The raw score of 12 is equivalent to a 4.1 grade average or 3.8 below grade average while the raw score of 54 is equivalent to a 10.4 grade average or 2.5 above grade average. Thirty-nine students performed below grade average and 11 students performed above grade average.

The mean CTBS Reading scores obtained by grade three and seven

Winnipeg groups were slightly below average. The mean scores obtained in both grade three and seven Winnipeg groups indicated that the reading skills of grade three and seven Winnipeg groups were slightly below average. The KR20 Reliability coefficients obtained for CTBS Reading were high for both grade three and grade seven Winnipeg. The high reliability coefficients obtained in grade three and seven Winnipeg indicate that the CTBS Reading test items were very consistent for these two groups of students. Hence, the CTBS Reading subtest was a reliable measure of reading skills for Winnipeg.

SAT Reading Comprehension The range of scores for grade three Winnipeg was 11 to 54. The grade three Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 36.77, 10.66, and 0.94 respectively. The raw score of 43 is equivalent to a Spring grade score of 3.9. The mean score of 36.77 is equivalent to a grade score of 3.2 or 0.7 below grade average. The raw score of 11 is equivalent to a 1.7 grade average or 2.2 below grade average. The raw score of 54 is equivalent to a grade average of 7.6 or 3.7 above grade average. Twenty-two students performed below grade average and 17 students performed above grade average.

The range of scores for grade three Brochet was 13 to 38. The grade three Brochet mean, standard deviation, and the KR20 Reliability coefficient were 23.36, 5.81, and 0.65 respectively. The raw score of 43 is equivalent to a Spring grade score of 3.9. The mean raw score of 23.36 is equivalent to a grade average of 2.3 or 1.6 below grade average. The raw score of 13 is equivalent to a 1.8 grade average or 2.1 below grade average. The raw score of 38 is equivalent to a grade average of 3.3 or 0.6 below grade level. Twenty-five students performed below grade average and no students performed above grade average.

The range of scores for grade seven Winnipeg was 10 to 58. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 33.12, 14.14, and 0.85 respectively. The raw scores of 37 and 38 are equivalent to Spring grade scores of 7.7 and 8.0. The mean raw score of 32.57 is equivalent to a grade average of 6.6 or 1.3 below grade average. The raw score of 10 is equivalent to a 3.0 grade average while the raw score of 58 is equivalent to a PHS (Post High School) grade average. Fourteen students performed below grade average and 28 students performed above grade average.

The range of scores for grade seven Brochet was eight to 40. The grade seven Brochet mean, standard deviation, and KR20 Reliability coefficient were 23.77, 7.11, and 0.77 respectively. The raw scores of 37 and 38 are equivalent to Spring grade scores of 7.7 and 8.0. The mean raw score of 23.77 is equivalent to a grade average of 4.7 or 3.2 below grade average. The raw score of eight is equivalent to 2.8 or 5.1 below grade average. Twenty-nine students scored below average and one student scored above average.

The mean SAT Reading Comprehension scores were below average for all grade three and seven Winnipeg groups. The grade three Winnipeg score was the least below average, followed by grade seven Winnipeg score. The grade three Brochet score ranked third below average. Grade seven Brochet had the lowest score of all the four groups. The mean scores obtained by all the groups indicated that reading comprehension skills were below average for all grade three and seven Winnipeg and Brochet groups. The KR20 Reliability coefficients obtained for SAT Reading Comprehension were high for both grades three and seven Winnipeg. The SAT Reading Comprehension reliability coefficients were substantially lower for grades three and seven Brochet. The higher relia-

bility coefficients obtained in both grades three and seven Winnipeg indicated that SAT Reading Comprehension test items were more consistent for grade three and seven Winnipeg than for grade three and seven Brochet. Hence, the SAT Reading Comprehension subtest was not as reliable a measure of reading comprehension for the grade three and seven Brochet groups as it was for grade three and seven Winnipeg groups.

SAT Vocabulary The range of scores for grade three Winnipeg was seven to 32. The grade three Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 18.74, 4.77, and 0.78 respectively. A Spring raw score of 25 is equivalent to a 3.9 grade average. The mean raw score of 18.26 is equivalent to a Spring grade three score of 2.6 or 1.3 below grade average. The raw score of seven is equivalent to Kindergarten five months level or 3.4 below grade average. The raw score of 32 is equivalent to a grade average of 6.5 or 1.4 below grade level. Thirty-six students performed below grade level and three students performed at or above grade average.

The range of scores for grade three Brochet was five to 15. The grade three Brochet mean, standard deviation, and KR20 Reliability coefficient were 11.00, 2.69, and 0.05 respectively. The mean raw score of 11.00 is equivalent to a Spring grade three score of 1.4 or 2.5 below grade average. The raw score of five is equivalent to a Pre-Kingdergarten grade level or 3.9 below grade level. The raw score of 15 is equivalent to a grade average of 2.1 or 1.8 below grade level. Twenty-five students performed below grade average and no students performed above grade average.

The range of scores for grade seven Winnipeg was six to 40. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability

coefficient were 23.81, 7.21, and 0.86 respectively. The mean raw score of 24.36 is equivalent to a grade score of 3.8 or 0.1 below grade level. The raw scores of 25 and 26 are equivalent to Spring grade scores of 7.8 and 8.2. The raw score of six is equivalent to a 2.5 grade average or 5.4 below grade average. The raw score of 40 is equivalent to a grade average of 8.6 or 0.7 above grade average. Twenty-four students performed below a 7.8 grade average and 18 students performed above an 8.2 grade average.

The range of scores for grade seven Brochet was four to 19. The grade seven Brochet mean, standard deviation, and KR20 Reliability coefficient were 11.29, 2.96, and 0.23 respectively. The raw scores of 25 and 26 are equivalent to Spring grade scores of 7.8 and 8.2. The mean raw score of 11.29 is equivalent to a 3.6 grade level or 4.3 below grade average. The raw score of four is equivalent to a 1.4 grade level or 6.5 below grade average. The raw score of 19 is equivalent to a grade average of 5.8 or 2.1 below grade average. Twenty-eight students performed below grade average and no students performed above grade average. The very low reliability coefficient (0.25) indicated that the Sat Vocabulary test items were an inconsistent measure of vocabulary skills for grade seven Brochet students.

The mean SAT Vocabulary scores obtained by grade three and seven Winnipeg groups in Winnipeg and Brochet varied from very slightly below average to highly below average. The grade seven Winnipeg scores were closest to average, followed by grade three Winnipeg, grade three Brochet, and grade seven Brochet. The mean raw scores obtained indicated that vocabulary skills are below average for all groups with Brochet grade seven being the lowest. The KR20 Reliability coefficients

were high for grade seven Winnipeg, and low for grades three and seven Brochet. The higher reliability coefficient obtained in grade seven Winnipeg indicated that the SAT Vocabulary test items were more consistent for grade seven Winnipeg than for grade seven Brochet. The moderate reliability coefficient obtained for grade three Winnipeg indicated that the SAT Vocabulary test items were of moderate consistency for grade three Winnipeg. The low reliability coefficients obtained for grades three and seven Brochet signified that the SAT Vocabulary subtest was not as reliable a measure of vocabulary skills for grades three and seven Brochet as it was for grades three and seven Winnipeg.

CLDA Noun The range of scores for grade three Winnipeg was 22 to 42. The grade three mean, standard deviation, and KR20 Reliability coefficient were 33.92, 4.51, and 0.25 respectively. Twenty-two students performed below the mean and 17 students performed above the mean. The range of scores for grade three Brochet was 27 to 36. The grade three Brochet mean, standard deviation, and KR20 Reliability coefficient scores were 31.00, 2.97, and 0.39 respectively. Five students performed below the mean, and nine students performed at or above the mean.

The range of scores for grade seven Winnipeg was 28 to 54. The grade seven Winnipeg mean, standard deviation, and KR20 Reliability coefficient were 42.91, 7.58, and 0.86 respectively. Twenty-five students performed below the mean and 19 students performed above the mean.

The range of scores for grade seven Brochet was 29 to 47. The grade seven mean, standard deviation, and KR20 Reliability coefficient were 35.20, 4.03, and 0.46 respectively. Twelve students performed below the mean and 18 students performed at or above the mean.

The CLDA Noun mean scores obtained by grade three and seven Winnipeg groups were higher than the scores obtained by grade three and seven

Brochet groups. However, the Brochet group sample size was almost too small to make a credible evaluation. The mean scores indicated that noun skills were higher for grade three and seven Winnipeg groups than for grade three and seven Winnipeg groups. The KR20 Reliability coefficient obtained for grade seven Winnipeg was high whereas the reliability coefficients obtained for grade three Winnipeg, grade three and grade seven Brochet were low. The high reliability coefficient obtained in grade seven Winnipeg indicated that the CLDA Noun test items were more consistent for grade seven Winnipeg than for grade three Winnipeg, grade three Brochet, or grade seven Brochet. Therefore, the CLDA Noun test was not as reliable a measure of language skills for grade 3 Winnipeg, grade 3 Brochet, or grade 7 Brochet as it was for grade 7 Winnipeg.

Quantitative Component

DCAT Quantitative The range of raw scores for grade three Winnipeg was five to 18. The grade three Winnipeg mean, standard deviation, and the KR20 Reliability coefficient were 11.35, 3.10, and 0.48 respectively. A raw score of 14 is equivalent to the 46th percentile while a raw score of 15 is equivalent to the 52nd percentile. The grade three Winnipeg mean of 11.35 is equivalent to the 32nd percentile which is 18 points below the 50th percentile. Nineteen students performed below the 46th percentile. Twenty-three students performed below the 52nd percentile. Eleven students performed at or above the 46th percentile and seven students performed at or above the 52nd percentile.

The range of scores for grade three Brochet was two to 13. The grade three Brochet mean, standard deviation, and KR20 Reliability coefficient were 7.45, 2.73, and 0.38 respectively. A raw score of 14 is equivalent to the 46th percentile while a raw score of 15 is

equivalent to the 52nd percentile. The grade three Brochet mean of 7.45 is equivalent to the 15th percentile or 35 points below the 50th percentile. Thirty-one students performed below the 46th percentile and no students performed above the 46th percentile. Thirty-one students performed below the 52nd percentile and no students performed above the 52nd percentile.

The range of raw scores for grade seven Winnipeg was 6 to 18. The grade three Winnipeg mean, standard deviation, and the KR20 Reliability coefficient were 10.71, 3.69, and 0.57 respectively. A raw score of 13 is equivalent to the 47th percentile while a raw score of 14 is equivalent to the 54th percentile. The grade seven Winnipeg mean of 10.71 is equivalent to the 28th percentile or 22 points below the 50th percentile. Twelve students performed below and five students above the 47th percentile. Thirteen students performed below and four students performed above the 54th percentile.

The range of raw scores for grade seven Brochet was three to 17. The grade seven Brochet mean, standard deviation, and the KR20 Reliability coefficient were 9.00, 3.59, and 0.61 respectively. A raw score of 13 is equivalent to the 47th percentile while a raw score of 14 is equivalent to the 54th percentile. The grade seven Brochet mean of 9.00 is equivalent to the 13th percentile or 37 points below the 50th percentile. Twenty-three students performed below and seven students performed above the 47th percentile. Twenty-six students performed below and four students performed above the 54th percentile.

The mean DCAT Quantitative scores obtained by all groups were below average. Grade three Winnipeg performed best, followed by grade seven Winnipeg, grade seven Brochet, and grade three Brochet. The mean scores obtained by all grade three Winnipeg and Brochet groups indicated that

quantitative skills were below average for all grades and sites. The reliability coefficients were moderate for grade seven Winnipeg and grade seven Brochet. However, the reliability coefficients for grade three Winnipeg and grade three Brochet were low. The moderate reliability coefficients obtained for grade seven Winnipeg and grade seven Brochet indicated that the DCAT Quantitative test items were moderately consistent for grade seven Winnipeg but they were of low consistency as a measure of quantitative skills for grade three Winnipeg and grade seven Brochet.

SAT Mathematics Concepts of Number (M-1) The range of scores for grade three Winnipeg was 6 to 28. The mean, standard deviation, and KR20 Reliability coefficient were 18.85, 6.27, and 0.85 respectively. The raw scores of 22 and 23 are equivalent to Spring grade scores of 3.8 and 4.0. The mean raw score of 18.85 is equivalent to a 3.3 grade average or 0.6 below grade level. The raw score of 6 is equivalent to a grade average of 1.3 or 2.6 below grade average while the raw score of 28 is equivalent to a 5.4 grade average or 1.5 above grade average. 26 students performed below grade average and 13 students performed above grade average.

The range of scores for grade three Brochet was five to 17. The mean, standard deviation, and KR20 Reliability coefficient were 10.44, 3.31, and 0.40 respectively. The raw scores of 22 and 23 are equivalent to Spring grade scores of 3.8 and 4.0. The mean raw score of 10.44 is equivalent to a 1.9 grade average or 2.0 below grade level. Raw scores of five and 17 are equivalent to grade scores of 1.1 and 3.0. Twenty-seven students performed below a 3.8 grade average and no students performed above a 3.8 grade level. Twenty-seven students performed below a 4.0 grade level and no students performed above a 4.0 grade level.

The range of scores for grade seven Winnipeg was seven to 33. The mean, standard deviation, and KR20 Reliability coefficient were 19.02, 7.47, and 0.40 respectively. The raw scores of 19 and 20 are equivalent to Spring scores of 7.7 and 8.0. The mean raw score of 19.02 is equivalent to a 7.7 grade average or 0.2 below grade level. The raw score of eight is equivalent to a 4.9 grade average or 3.0 below grade level. The raw score of 33 is equivalent to a PHS (Post High School) grade level or 5.0^{+} above grade level. Twenty-six students performed at or below the 7.8 grade level and 21 students performed at or above the 8.0 grade level.

The range of scores for grade seven Brochet was seven to 19. The mean, standard deviation, and KR20 Reliability coefficient were 11.47, 2.80, and 0.61 respectively. The raw scores of 19 and 20 are equivalent to Spring scores of 7.8 and 8.0. The mean raw score of 11.47 is equivalent to a grade level of 5.7 or 2.2 below grade level. The raw score of seven is equivalent to a 4.6 grade level or 3.1 above grade level. The raw score of 19 is equivalent to a 7.7 grade level or 0.2 below grade level. Thirty students performed below grade level and no students performed above grade level.

The mean SAT M-1 scores obtained by all groups were below average in varying degrees. While the grade seven Winnipeg mean score was very slightly below average, the mean score obtained by grade three Winnipeg was lower than grade seven Winnipeg. Grades three and seven Brochet obtained the lowest mean scores of the four groups. The mean scores earned by all groups indicated that concept of number skills for all students in all four groups were below average. The KR20 Reliablity coefficients were high for grade three and seven Winnipeg but were low for grade three and seven Brochet. The higher reliability coefficients

obtained in both Winnipeg groups indicated that SAT M-1 test items were more consistent for grade three and seven Winnipeg groups than for grade three and grade seven Brochet groups. Therefore, the SAT M-1 subtest was not as reliable as a measure of concepts of number skills for Brochet as it was for Winnipeg.

SAT Mathematics Computation (M-2) The range of scores for grade three Winnipeg was nine to 39. The mean, standard deviation, and KR20 Reliability coefficient were 23.76, 8.95, and 0.91 respectively. The raw scores of 25 and 26 are equivalent to Spring grade scores of 3.8 and 4.0. The mean raw score of 23.76 is equivalent to a 3.7 grade level or 0.2 below grade level. The raw score of nine is equivalent to a grade level of 1.5 or 2.4 below grade level while the raw score of 39 is equivalent to a 3.5 grade level or 0.4 below grade level. Thirty-four students performed below grade level and six students performed above grade level.

The range of scores for grade three Brochet was seven to 33. The mean, standard deviation, and KR20 Reliability coefficient were 16.55, 8.48, and 0.90 respectively. The raw scores of 25 and 26 are equivalent to Spring grade scores of 3.8 and 4.0. The mean score of 16.55 is equivalent to a 2.7 grade level or 1.2 below grade level. The mean raw score of seven is equivalent to a 1.2 grade level or 2.7 below grade level. The raw score of 16.55 is equivalent to a 5.2 grade level or 2.7 below grade level. Twenty-two students performed below grade level and five students performed above grade level.

The range of scores for grade seven Winnipeg was eight to 36. The mean, standard deviation, and KR20 Reliability coefficient were 20.96, 9.68, and 0.92 respectively. The raw scores of 22 and 23 are equivalent to Spring grade scores of 7.8 and 8.0. The mean raw score of 20.96 is

equivalent to a 7.5 grade level or .4 below grade level. The raw score of eight is equivalent to a 4.8 grade score and the raw score of 36 is equivalent to a PHS grade level. Twenty-eight students performed below grade level and 18 students performed above grade level.

The range of scores for grade seven Brochet was three to 27. The mean, standard deviation, and the KR20 Reliability coefficient were 14.53, 5.93, and 0.81 respectively. The raw scores of 22 and 23 are equivalent to Spring grade scores of 7.8 and 8.0. The mean raw score of 14.53 is equivalent to a 6.2 grade level or 1.7 below grade level. The raw score of three is equivalent to a 3.3 grade score or 4.6 below grade level. The raw score of 27 is equivalent to a 9.2 grade level or 1.3 above grade level. Twenty-nine students performed below grade level and one student performed above grade level.

The raw mean scores obtained by all groups were below average. However, these mean raw scores were not as drastically low as many of the other subtests previously described. The below average SAT M-2 scores indicated that mathematics computation skills are somewhat below average for all Winnipeg and Brochet groups. The KR20 Reliability coefficients obtained for SAT M-2 were high for all groups for both sites. The high correlation coefficients obtained by all groups in all sites signified that the SAT M-2 test items were consistent as a measure of mathematics computation skills.

SAT Mathematics Applications (M-3) The range of scores for grade three Winnipeg was nine to 36. The mean, standard deviation, and KR20 Reliability coefficient were 22.03, 7.30, and 0.87 respectively. The raw score of 25 is equivalent to a Spring grade score of 3.9. The mean raw score of 22.03 is equivalent to a 3.5 grade level or 0.4 below grade level. The raw score of nine is equivalent to a grade average of 1.6

while the raw score of 36 is equivalent to a 8.0 grade level. Twenty-seven students performed below grade level and 13 students performed at or above grade level.

The range of scores for grade three Brochet was five to 19. The mean, standard deviation, and KR20 Reliability coefficient were 11.31, 5.21, and 0.77 respectively. The raw score of 25 is equivalent to a Spring grade score of 3.9. The mean raw score of 11.31 is equivalent to a 1.9 grade level or 2.0 below grade level. The raw score of five is equivalent to a 1.0 grade average while the raw score of 19 is equivalent to a grade level of 3.1. 26 students performed below grade level and no students performed above grade level.

The range of scores for grade seven Winnipeg is two to 34. The mean, standard deviation, and KR20 Reliability coefficient were 19.74, 9.04, and 0.91 respectively. The raw scores of 33 and 34 are equivalent to Spring grade scores of 7.8 and 8.0. The mean raw score of 19.74 is equivalent to a 7.8 grade average or 0.1 below grade level. The raw score of two is equivalent to a 3.0 grade average or 4.9 below grade level. A raw score of 34 is equivalent to a PHS grade level or 9.0^{+} above grade level. Forty-four students performed below grade level and four students performed above grade level.

The range of scores for grade seven Brochet is three to 21. The mean, standard deviation, and KR20 Reliability coefficient were 11.40, 4.78, and 0.69 respectively. The mean raw score of 11.40 is equivalent to a 5.7 grade level or 2.2 below grade level. The raw score of three is equivalent to a 3.5 grade level while the raw score of 21 is equivalent to a grade level of 8.0. Thirty students performed below grade level and no students performed above grade level. The raw mean scores obtained for all groups at both sides were below average in

varying degrees. While the grade three and seven Winnipeg mean scores were very slightly below average, the raw mean scores for grade three and grade seven Brochet were somewhat higher.

The below average SAT M-3 raw mean scores obtained by all groups at both sites indicated that all groups at both sites were slightly or somewhat below average in mathematics application skills.

SUMMARY

For all the groups and sites studied, 36 sets of scores were obtained. Thirty-two of these 36 sets of scores possessed standardized criteria for evaluating raw mean scores. Only the CLDA Noun test possessed no standardized criteria for evaluating raw mean scores. Hence, of the 32 sets of scores evaluated using standardized evaluation criteria, students in grade three and grade seven Winnipeg performed below grade level in 31 of the 32 subtests. The only subtest and group in which achievement was at grade level was DCAT Verbal for grade seven Winnipeg. In none of the groups or sites did the raw means achievement scores exceed grade level or the 50th percentile. Hence, the mean raw scores obtained in this research study indicated that verbal and quantitative skills were below average for all grade three and seven Winnipeg and Brochet groups with the exception of the grade seven Winnipeg group which achieved average performance on the DCAT Verbal subtest. In addition, the raw mean scores obtained by all Brochet groups were lower than the scores obtained for all Winnipeg groups. Therefore, verbal and quantitative skills were more below average for all Brochet groups than for all Winnipeg groups.

The larger differences in the below average raw mean scores obtained by all Brochet groups as compared to all Winnipeg groups, as well as the fact that all groups in both sites, except for grade seven DCAT Verbal, achieved below average, may be attributable to many factors

Some of the factors that may influence low test scores may include language, motivation, test-taking attitudes, test-wiseness, speed, and competitiveness (Gronlund, 1981). Other factors may include language interaction between examiner and examinee (Camilleri, 1986) as well as countless other intrinsic and extrinsic factors related to examinee, examiner, and testing situation variables. Giving norm-referenced tests to individuals from different cultural or ethnic backgrounds creates cultural bias (Popham, 1988) and leads to low test scores is an inadequate explanation for the low test scores. No culture is shared by all members of a group (Hansen, 1979). Low test scores may be due to inadequate mastery of skills the test is designed to measure irrespective of the minority group tested (Gronlund, 1985). Possibly, standardized tests lack validity for learners whose background and current academic curricula are at variance with that of the standardization sample (Gibson, 1980). Moreover, current trends in evaluation are concerned with the culture of the individual. Possibly, standardized tests lack "subjective equivalence" even though the tests may possess "objective identity" (Ginsburg, 1986). Weiss (1987b) suggests that safeguards must be created to guarantee important knowledge variations between examinees rather than culture-specific factors that lack relevance. While some, none, or all of these possibilities mentioned may have influenced the test results, conjecture without evidence is unfair, unjust, and cruel.

TABLE 6
Summary of the Reliability Coefficients
For Each Test

T E S T S	Grade 3 Winnipeg	Grade 3 Brochet	Grade 7 Winnipeg	Grade 7 Brochet
DCAT	.82	.81	.82	.47
CTBS	.91		.95	
SAT	.95	.78	.97	.45
CLDA	.25	.37	.86	.46

The Cronbach Alpha Reliability coefficients for the total of each test in this research study were indicated in Table 6. Of the 14 reliability coefficients obtained, eight were high, one was medium, and five were low. Seven of the eight high reliability coefficients were obtained by the Winnipeg groups whereas four of the five low reliability coefficients were obtained by the Brochet groups. These results indicated that this test battery possessed a high degree of consistency for grade three and seven Winnipeg groups as test measures of verbal and quantitative skills. The low Cronbach Alpha Reliability coefficient obtained by grade seven Brochet was of particular concern as the three tests administered to that group possessed a low degree of consistency for those students. Hence, this test battery possessed a low degree of consistency as measures of verbal and quantitative skills for grade seven Brochet students.

The DCAT obtained high reliability coefficients for all grades and sites except for grade seven Brochet. The conclusion that can be drawn here is that the DCAT lacked consistency as an evaluation instrument for

grade seven Brochet students. On the other hand, the CTBS possessed a high degree of consistency for grade three and seven Winnipeg. The CTBS was not administered to grades three and seven Brochet. The SAT possessed high reliability for grade three and seven Winnipeg but medium consistency for grade three Brochet and low consistency for grade seven Brochet. The CLDA Noun test obtained a high reliability coefficient for grade seven Winnipeg only. However, the CLDA Noun test obtained low reliability coefficients for grade three Winnipeg and Brochet and grade seven Brochet. These results indicated that the CLDA Noun test is not a consistent measure of language skills for the majority of students in grade three Winnipeg and Brochet and grade seven Brochet groups. Low reliability coefficients may have stemmed from a number of causes at which one may only speculate. These causes may be created by test, examinee, examiner, or other intrinisic or extrinsic variables. On the other hand, causation may be due to none of these factors or may be due to other unknown factors.

Of the 36 KR20 Reliability coefficients obtained for the DCAT, CTBS, SAT, and CLDA subtests in this research study, 19 subtests had high, seven subtests had medium, and 10 subtests had low reliability coefficients. In other words, 19 subtests (53%) had high reliability coefficients, seven subtests (19%) had medium, and 10 subtests (28%) had low reliability coefficients. Similarly, the majority of the high reliability coefficients were obtained by the Winnipeg groups whereas the majority of the low coefficients were obtained by the Brochet groups. Also, the Brochet groups had a larger number of low reliability coefficients in spite of the fact that the CTBS was not administered to grade three and seven Brochet groups. The higher number of reliability

coefficients obtained in grades three and seven Winnipeg and Brochet indicated that the test items were more consistent for some groups in one or both locations than for some groups in both locations as reliable measures of verbal and quantitative skills. Likewise, the medium and low reliability coefficients obtained in grades three and seven Winnipeg and Brochet were less consistent for some groups in one or both locations as reliable measures of verbal and quantitative skills.

The fact that 17 subtests or 47% of the subtests had medium or low reliability coefficients for the test items for all groups and sites creates serious concern about the test items as consistent measures of verbal and quantitative skills for all groups and sites. Closer examination of the reliability coefficients revealed that the Winnipeg groups had a far larger number of high and medium reliability coefficients and far fewer low reliability coefficients than the Brochet groups. This indicated that the verbal and quantitative test items were more consistent measures of these skills for Winnipeg than Brochet. The medium and low KR20 Reliability coefficients caused one to speculate as to the causes of these results. Ginsburg (1986) has suggested that test items are frequently misinterpreted and misunderstood with the end result that norm standardization is frequently invalid. Other possibilities may include intrinsic and extrinsic examinee variables, poor match of test content to local curriculum, or any other unknown factor.

INFERENTIAL STATISTICS

Verbal Component

The correlation matrix of the verbal variables for all grade three Winnipeg students is presented in Table 7. Correlations ranged from .29 to .62. Significant medium correlations at the .05 level were obtained for DCAT Verbal and CTBS Vocabulary, DCAT Verbal and CTBS Reading, and DCAT Verbal and SAT Reading Comprehension. These correlations indicated that DCAT Verbal seemed to measure moderately constructs similar to CTBS Vocabulary, CTBS Reading, and SAT Reading Comprehension for grade three Winnipeg students. A low correlation was obtained between DCAT Verbal and SAT Vocabulary. This low correlation indicated that DCAT Verbal and SAT Vocabulary seemed to measure similar constructs to a low degree. On the other hand, no significant correlations at the .05 level were obtained between DCAT Verbal and SAT Vocabulary or between DCAT Verbal and CLDA Noun. These correlations indicated that DCAT Verbal did not seem to measure constructs similar to SAT Vocabulary and CLDA Noun for grade three Winnipeg students.

The correlation matrix of all the verbal variables for all grade three Brochet students is presented in Table 8. Correlations ranged from .06 to .32. No significant correlations at the .05 level were obtained between DCAT Verbal and SAT Reading Comprehension, DCAT Verbal and SAT Vocabulary, and DCAT Verbal and CLDA Noun. These correlations indicated that DCAT Verbal, SAT Reading, and CLDA Noun did not seem to measure the same constructs for grade three Brochet students.

The correlation matrix of all the verbal variables for all grade

TABLE 7

Simple Correlation Coefficients Among the Variables DCAT Verbal, CTBS Vocabulary, CTBS Reading, SAT Reading Comprehension, SAT Vocabulary, and CLDA Noun for Grade 3 Winnipeg

VARIABLES	DCAT Verbal	CTBS Voc.	CTBS Rdg.	SAT Rdg. Comp.	SAT Voc.	CLDA Noun
DCAT Verbal	1.00 (40)					
CTBS Voc.	.6188 (40)	1.00 (40)				
CTBS Rdg.	.6243 (40)	.7604 (40)	1.00 (40)			
SAT Rdg. Comp.	.5068 (39)	.5020 (39)	.6945 (39)	1.00 (39)		
SAT Voc.	.3363 (38)	.4508 (38)	.5304 (38)	.4620 (37)	1.00 (38)	
CLDA Noun	.2910 (39)	.3847 (49)	.3662 (39)	.2105 (38)	.3545 (37)	1.00 (39)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .94

TABLE 8

Simple Correlation Coefficients Among the Variables DCAT Verbal, SAT Reading Comprehension, SAT Vocabulary, and CLDA Noun for Grade 3 Brochet

VARIABLES	DCAT Verbal	SAT Rdg. Comp.	SAT Voc.	CLDA Noun
DCAT Verbal	1.00 (31)			
SAT Rdg. Comp.	.0610 (25)	1.00 (25)		
SAT Voc.	.3242 (25)	-.0984 (22)	1.00 (25)	
CLDA Noun	.2309 (14)	.1070 (14)	.0643 (13)	1.00 (14)

Correlation coefficients indicate no correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .41

seven Winnipeg students is presented in Table 9. Correlations ranged from .34 to .84. Significant correlations at the .05 level were obtained for DCAT Verbal and CTBS Vocabulary, DCAT Verbal and SAT Vocabulary, and DCAT Verbal and CLDA Noun. The correlation between DCAT Verbal and SAT Vocabulary was high while the correlations between DCAT Verbal and CTBS Vocabulary and between DCAT Verbal and CLDA Noun were moderate. These correlations indicated that while DCAT Verbal and SAT Vocabulary seemed to measure similar constructs at a high level for grade seven Winnipeg students, DCAT Verbal, CTBS Vocabulary, and CLDA Noun measured similar constructs at a moderate level for grade seven Winnipeg students. No significant correlations were found between DCAT Verbal and CTBS Reading and between DCAT Verbal and SAT Reading Comprehension. This lack of significant correlation between these subtests indicated that DCAT Verbal did not seem to measure the same constructs as CTBS Reading and SAT Reading Comprehension for grade seven Winnipeg students.

The correlation matrix of the verbal variables for all grade seven Brochet students is presented in Table 10. Correlations ranged from .14 to .43. Significant low correlations at the .05 level were obtained for DCAT Verbal and SAT Vocabulary and DCAT Verbal and CLDA Noun. These correlations indicated that DCAT Verbal, SAT Vocabulary, and CLDA Noun seemed to measure similar constructs at a low level for grade seven Brochet students. No significant correlation at the .05 level was obtained between DCAT Verbal and SAT Reading Comprehension. The lack of significant correlation at the .05 level between DCAT Verbal and SAT Reading Comprehension indicated that DCAT Verbal and SAT Reading Comprehension did not seem to measure similar constructs for the grade seven Brochet students.

TABLE 9

Simple Correlation Coefficients Among the Variables DCAT Verbal, CTBS Vocabulary, CTBS Reading, SAT Reading Comprehension, SAT Vocabulary, and CLDA Noun for Grade 7 Winnipeg

VARIABLES	DCAT Verbal	CTBS Voc.	CTBS Rdg.	SAT Rdg. Comp.	SAT Voc.	CLDA Noun
DCAT Verbal	1.00 (17)					
CTBS Voc.	.6177 (17)	1.00 (49)				
CTBS Rdg.	.3537 (17)	.7804 (49)	1.00 (49)			
SAT Rdg. Comp.	.3409 (13)	.7796 (41)	.8839 (41)	1.00 (42)		
SAT Voc.	.8445 (13)	.8378 (41)	.7735 (41)	.6925 (42)	1.00 (42)	
CLDA Noun	.5609 (14)	.6534 (43)	.6643 (43)	.6709 (37)	.5981 (37)	1.00 (44)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .92

TABLE 10

Simple Correlation Coefficients Among the Variables DCAT Verbal, SAT Reading Comprehension, SAT Vocabulary and CLDA Noun for Grade 7 Brochet

VARIABLES	DCAT Verbal	SAT Rdg. Comp.	SAT Voc.	CLDA Noun
DCAT Verbal	1.00 (30)			
SAT Rdg. Comp.	.1377 (30)	1.00 (30)		
SAT Voc.	.1394 (28)	.5075 (28)	1.00 (28)	
CLDA Noun	.4290 (30)	.0679 (30)	-.0563 (28)	1.00 (30)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .53

Quantitative Component

The correlation matrix of the quantitative variables for all grade three Winnipeg students is presented in Table 11. Correlations ranged from .25 to .54. A significant medium correlation at the .05 level was obtained for DCAT Quantitative and SAT Concepts of Number (M-1), a low significant correlation was obtained between DCAT Quantitative and SAT Mathematics Computation, and no significant correlation between DCAT Quantitative and SAT Mathematics Applications. Hence, SAT Concepts of Number (M-1) measured concepts similar to the DCAT Quantitative in a moderate way, SAT Mathematics Computation measured constructs similar to the DCAT to a low degree, and SAT Mathematics did not measure constructs similar to the constructs measured by DCAT Quantitative.

The correlation matrix of the quantitative variables for all grade three Brochet students is presented in Table 12. Correlations range from 0.3 to .42. The significant low correlation at the .05 level obtained for DCAT Quantitative and SAT Mathematics Applications (M-3) indicated that DCAT Quantitative and SAT Mathematics Applications (M-3) seemed to measure similar concepts at a low level for grade three Brochet students. In addition, no significant correlations at the .05 level were obtained for DCAT Quantitative and SAT Concepts of Number (M-1) and between DCAT Quantitative and SAT Mathematics Computation (M-2). This lack of significant correlations between DCAT Quantitative and SAT Concepts of Number (M-1) and between DCAT Quantitative and SAT Mathematics Computation (M-2) indicated that DCAT Quantitative seemed to measure different constructs than SAT Concepts of Number (M-1) and SAT Mathematics (M-2) for grade three Brochet students.

TABLE 11

Simple Correlation Coefficients Among the Variables DCAT Quantitative, Sat Concepts of Number (M-1), SAT Mathematics Computation (M-2), and Sat Mathematics Applications (M-3) for Grade 3 Winnipeg

VARIABLES	DCAT Quant.	SAT Concepts of No.	SAT Math Comput.	SAT Math Applics.
DCAT Quant.	1.00 (40)			
SAT-Concepts of No.	.5428 (39)	1.00 (40)		
SAT-Math Comput.	.4151 (40)	.5393 (39)	1.00 (40)	
SAT-Math Applics.	.2534 (40)	.5896 (39)	.5217 (40)	1.00 (40)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .78

TABLE 12

Simple Correlation Coefficients Among the Variables DCAT Quantitative, Sat Concepts of Number (M-1), SAT Mathematics Computation (M-2), and Sat Mathematics Applications (M-3) for Grade 3 Brochet

VARIABLES	DCAT Quant.	SAT Concepts of No.	SAT Math Comput.	SAT Math Applics.
DCAT Quant.	1.00 (31)			
SAT-Concepts of No.	.0306 (27)	1.00 (27)		
SAT-Math Comput.	-.0351 (27)	.5034 (27)	1.00 (27)	
SAT-Math Applics.	.4177 (26)	-.2469 (26)	-.3626 (26)	1.00 (26)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .59

The correlation matrix of the variables for all grade seven Winnipeg students is presented in Table 13. Correlations ranged from -.50 to .62. Significant moderate correlations at the .05 level were obtained between DCAT Quantitative and SAT Concepts of Number, DCAT Quantitative and SAT Mathematics Computation (M-2), and between DCAT Quantitative and SAT Mathematics Applications (M-3). The moderate correlations obtained for DCAT Quantitative, SAT Concepts of Number (M-1), SAT Mathematics Computation (M-2), and SAT Mathematics Applications (M-3) indicated that DCAT Quantitative seemed to measure constructs similar to the constructs measured by SAT Concept of Number (M-1), SAT Mathematics Computation (M-2), and SAT Mathematics Applications (M-3) for grade seven Winnipeg students. The significant moderate negative correlation at the .05 level obtained for DCAT Quantitative and SAT Mathematics Computation (M-2) indicated that as the raw scores of DCAT Quantitative increased, the raw scores for SAT Mathematics Computation (M-2) decreased for grade seven Winnipeg students.

The correlation matrix of the variables for all grade seven Brochet students is presented in Table 14. Correlations ranged from -.01 to -.19 No significant correlations at the .05 level were obtained for correlations between DCAT Quantitative and SAT Concepts of Number (M-1), DCAT Quantitative and SAT Mathematics Computation (M-2), and between DCAT Quantitative and SAT Mathematics Applications (M-3). This lack of significant correlations at the .05 level indicated that DCAT Quantitative does not seem to measure constructs similar to the constructs measured by SAT Concepts of Number (M-1), SAT Mathematics Computations (M-2), and SAT Mathematics Applications (M-3).

TABLE 13

Simple Correlation Coefficients Among the Variables DCAT Quantitative, Sat Concepts of Number (M-1), SAT Mathematics Computation (M-2), and Sat Mathematics Applications (M-3) for Grade 7 Winnipeg

VARIABLES	DCAT Quant.	SAT Concepts of No.	SAT Math Comput.	SAT Math Applics.
DCAT Quant.	1.00 (17)			
SAT-Concepts of No.	.6195 (16)	1.00 (46)		
SAT-Math Comput.	-.5023 (16)	.7824 (45)	1.00 (46)	
SAT-Math Applics.	.5897 (16)	.8367 (45)	.8038 (46)	1.00 (46)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .90

TABLE 14

Simple Correlation Coefficients Among the Variables DCAT Quantitative, Sat Concepts of Number (M-1), SAT Mathematics Computation (M-2), and Sat Mathematics Applications (M-3) for Grade 7 Brochet

VARIABLES	DCAT Quant.	SAT Concepts of No.	SAT Math Comput.	SAT Math Applics.
DCAT Quant.	1.00 (30)			
SAT-Concepts of No.	-.1851 (30)	1.00 (30)		
SAT-Math Comput.	-.0081 (30)	.5924 (30)	1.00 (30)	
SAT- Math Applics.	-.1285 (30)	.3872 (30)	.5657 (30)	1.00 (30)

Correlation coefficients underlined indicate correlations significant at the .05 level.

Sample size for each correlation is indicated in the brackets.

Cronbach Alpha = .64

SUMMARY

The inferential statistics revealed that of the 28 DCAT correlations, the only high significant correlation at the .05 level was obtained by DCAT Verbal and SAT Vocabulary for grade seven Winnipeg. This high significant correlation between DCAT Verbal and SAT Vocabulary indicated that DCAT Verbal and SAT Vocabulary seemed to measure similar constructs for grade three and seven students. Of the remaining 15 DCAT Verbal correlations with the other CTBS, SAT, and CLDA subtests, five correlations were moderate, three were low, and seven possessed no significant correlations. Of the 12 DCAT Quantitative correlations, a moderate negative correlation was obtained for DCAT Quantitative and SAT Mathematics Computation (M-2) for grade seven Winnipeg. In total, there were four moderate DCAT Quantitative correlations. In addition, two DCAT Quantitative correlations were low. The six remaining DCAT Quantitative correlations possessed no significance at the .05 level. Twenty-eight correlations were done between the DCAT subtests and the CTBS, SAT, and CLDA Noun subtests. Of these 28 correlations, one correlation was high, nine were medium, and five were of low significance at the .05 level. The remaining 13 correlations lacked significance at the .05 level. Hence, the DCAT Verbal and Quantitative subtests did not seem to measure constructs similar to the CTBS, SAT, and CLDA Noun subtests in a high or medium level of significance in 18 out of 28 instances.

DISCUSSION

Question One

Question One sought to determine the degrees of relationship among the scores of the CTBS and DCAT when these tests were administered to grade three and seven learners in Winnipeg and Brochet. The results of this research study indicated that a moderate degree of relationship exists between the scores for DCAT Verbal, CTBS Vocabulary, CTBS Reading and SAT Vocabulary for grade three Winnipeg groups. However, while a moderate degree of relationship was obtained for DCAT and CTBS Vocabulary scores, no significant relationship was obtained between DCAT Verbal and CTBS Reading Comprehension scores for grade seven Winnipeg groups. Therefore, DCAT Verbal and CTBS Vocabulary and Reading Comprehension possessed a medium degree of relationship for grade three and seven Winnipeg scores. However, for grade seven Winnipeg, DCAT Verbal and CTBS Reading Comprehension scores possessed no significant relationship. The CTBS subtests were not administered to grade three and grade seven Brochet groups.

Question Two

Question Two sought to determine the degrees of relationship among the scores of the SAT and the DCAT when these tests were administered to grade three and seven learners in Winnipeg and Brochet. The results of this research study indicated that a moderate degree of relationship existed between DCAT Verbal and SAT Reading Comprehension scores for grade three Winnipeg groups only. No relationship existed betwen DCAT Verbal and SAT Reading Comprehension scores for grade seven Winnipeg or grade three and seven Brochet scores. The results of this

research study indicated that varying degrees of relationships existed between DCAT Verbal and SAT Vocabulary. While a high relationship existed between DCAT Verbal and SAT Vocabulary scores for grade seven Winnipeg, only a low relationship existed between DCAT Verbal and SAT Vocabulary for grade seven Brochet scores and grade three Winnipeg scores. No relationship existed between DCAT Quantitative and SAT Vocabulary scores for grade three Brochet students.

The research results also revealed inconsistent results for relationships among the quantitative subtests. DCAT Quantitative and SAT Concepts of Number (M-1) scores indicated a moderate relationship for grade three and seven Winnipeg. No relationship existed between DCAT Quantitative and SAT Concepts of Number (M-1) scores for grades three and seven Brochet. Similarly, results of this research study revealed that both DCAT Quantitative and SAT Mathematics Computation (M-2) possessed a medium relationship for grade seven Winnipeg scores only. On the other hand, DCAT Quantitative and SAT Mathematics Computation (M-2) scores possessed a low relationship only for grade three Winnipeg scores, and no relationship whatsoever for grades three and seven Brochet scores. In addition, DCAT Quantitative and SAT Mathematics Applications (M-3) scores possessed a moderate relationship for grade seven Winnipeg scores only. A low degree of relationship existed between DCAT Mathematics Applications (M-3) for grade three Brochet scores and no relationship existed whatsoever for grade three Winnipeg and grade seven Brochet scores.

Question Three

Question Three sought to determine the degree of relationship among the scores of the CLDA and the DCAT when both of these tests were administered to grade three and seven learners in Winnipeg and Brochet. The results of this research study indicated that a medium degree of relationship existed between the scores obtained for the CLDA Noun Test and DCAT Verbal for grade seven Winnipeg scores. A low degree of relationship existed between CLDA Noun and DCAT Verbal for grade seven Brochet. No significant relationship was found between the DCAT and the CLDA for grade three Winnipeg and Brochet.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

This study started with a discussion about the current educational need to identify the cognitive entry characteristics of learners to help enhance the quality and quantity of learning for all learners. Knowledge about cognitive entry characteristics could be helpful in providing a foundation for diagnosis, curriculum design, and evaluation. The need to identify cognitive entry characteristics is well documented in the educational literature. The Developing Cognitive Abilities Test, a new test, attempts to assess cognitive entry characteristics of learners to enhance instruction and evaluation. The results of this research study indicated that the DCAT is of questionable utility for assessing the cognitive entry characteristics of learners, and, hence, for assisting in curriculum design, instructional strategies, and placement decisions.

The descriptive analyses of this research study indicated that: (1) The DCAT possessed a high degree of consistency for all groups except grade seven Brochet, (2) the CTBS possessed a high degree of consistency for Winnipeg groups, (3) the SAT possessed a high degree of consistency for grade three Brochet, and low consistency for grade seven Brochet, and (4) the CLDA possessed a high degree of consistency for grade seven Winnipeg and a low degree of consistency for the other three groups. Also, the descriptive analyses of the research study data indicated that students in grade three and seven Winnipeg and Brochet performed below grade level or the 50th percentile in 31 of the 32

norm-referenced subtests administered in this study. Likewise, all the Brochet groups obtained lower mean raw scores than all of the Winnipeg groups. Similarly, an analyses of the KR20 Reliability coefficients revealed that only 53% or 19 of the 36 subtests used in this study had high reliability coefficients while the remaining 17 subtests had medium or low reliability coefficients. This data created concern about the DCAT test items as consistent measures of verbal and quantitative skills for all groups and sites. Of similar concern is the fact that the Winnipeg groups had a far larger number of high and medium reliability coefficients than the Brochet groups.

The inferential statistics revealed that of 28 correlations the only high significant correlation at the .05 level was obtained by DCAT Verbal and SAT Vocabulary for grade seven Winnipeg. These results indicated that only DCAT Verbal and SAT Vocabulary were measuring similar constructs for grade seven Winnipeg. Also, the inferential statistics revealed that of the 28 DCAT correlations, the only high significant correlation at the .05 level was obtained by DCAT Verbal and SAT Vocabulary for grade seven Winnipeg. Similarly, the DCAT correlations possessed nine medium correlations, five low correlation coefficients, and 13 DCAT correlations possessed no significant correlation. Also this data revealed that the majority of DCAT Verbal and Quantitative subtests obtained low or no significant correlations at the .05 level. These results reveal that the DCAT did not seem to measure the same constructs as the majority of other subtests used in this research study.

The conclusion drawn from this research study was that since the majority of DCAT subtests did not seem to measure the same constructs as

the CTBS, SAT, and CLDA Noun test for all grade three and grade seven groups in Winnipeg and Brochet, the DCAT subtests lack reliability for all grade three and seven Winnipeg groups. Furthermore, since the DCAT subtests lacked reliability for all grade three and grade seven Winnipeg groups, the DCAT also lacked validity for these students.

It is recommended that further research be conducted to replicate this study to obtain additional data to assist in refuting or defending the concurrent validity of the DCAT. To make judgements on the DCAT on the basis of one research study would be unfair. Accumulated evidence is required to verify this research study. Likewise, the concurrent validation of the DCAT in other student populations should be conducted to help determine whether there is a large discrepancy in results among various groups in various locations. With respect to the need to identify cognitive entry characteristics as a means of improving diagnosis, curriculum design, instructional strategies, and placement decisions, it is recommended that grade three and seven Winnipeg groups use the CTBS and the SAT. In addition, the SAT is recommended for grade three Brochet. Evaluation instruments suitable to the needs of grade seven Brochet students should be found or created. Presently, the DCAT, CTBS, SAT, and CLDA cannot fulfill these needs adequately.

REFERENCES

Alford, David W. (1984). IQ test controversy: Past, present, and future trends. (ERIC Document Reproduction Service No. ED 259026)

Anastasi, Anne. (1971). Psychological testing. (5th ed.). Toronto: MacMillan.

Anastasi, A. (1982). Psychological testing. (5th ed.). Toronto: MacMillan.

Astin, Alexander. (1979). Testing in the post "Bakke" period. (ERIC Document Reproduction Service ED 171820)

Ausubel D. (1978). Education psychology: A cognitive view. (2nd ed.). New York: Holt, Rinehart, & Winston.

Ausubel, David P. & Sullivan Edmond V., & Ives, William S. (1980). Theory and problems of child development. (3rd ed.). New York: Gaune and Stratton.

Baratz, J. (1980). Policy implications of minimum competency testing. In R. Jaeger & C. Tittle (Eds.), Minimum competency achievement testing. (pp. 49-68). Berkely: McCutchan.

Berk, Ronald A. (1987). Setting passing scores on competency tests. NASSP Bulletin, 71, (496), 69-76.

Berry, J. W. (1976). Human ecology and cognitive style. Beverley Hills: Sage.

Bleistein, Carole A. (1986). Application of item response theory to the study of differential item characteristics: A review of the literature. (ERIC Document Reproduction Service No. ED 268160)

Bloom, B.S., Englehart, M.B., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). Taxonomy of educational objectives, Handbook I: Cognitive

domain. New York: McKay.

Bloom, Benjamin S. (1981a). Evaluation to improve learning. New York: McGraw-Hill.

Bloom, Benjamin S. (1981b). All of our children learning. New York: McGraw-Hill.

Bloom, Benjamin S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 6, 4-16.

Bloom, Benjamin S. (1985). Developing talent in young people. New York: Balantine Books.

Bloom, Benjamin S. (1986). Interview with Thomas F. Kerner, editor of NASSP: A discussion about instruction and learning, teachers, and schools: An interview with Benjamin S. Bloom. NASSP Bulletin, 70, (493), 53-56, 58-69.

Blum, Jeffrey M. (1978). Ethnic minorities and IQ tests: The problem of cultural bias in tests (Reprinted from Pseudoscience and mental ability (1978), pp. 98-109. In Prentice H. Babptiste, Jr., & Mira Lanier Babptiste (Eds.) (1979), Developing the multicultural process in classroom instruction competencies for teachers. New York: University Press of America. (Reprinted from Pseudoscience and mental ability, 1978, pp 98-109.

Brescia, William & Fortune, Jim C. (1988). Standardized testing of American Indian students. ERIC Digest. (ERIC Document Reproduction Service No. ED 296813)

Bruner, J.S. (1966). Toward a theory of instruction. Cambridge: Harvard University Press.

Bruner, Jerome, S. (1987). Actual minds, possible worlds. Cambridge: Harvard University Press.

Bruner, Jerome S., & Haste, Helen (Eds.). (1987). Making sense. Metheun: New York.

Buros, O.K. (Ed.). (1972). The Seventh mental measurements yearbook. (Vol. 1.) New Jersey: The Gryphon Press.

Cabello, Beverley. (1981). Potential source of bias in dual language achievement tests. Beverley, California: California University, Centre for the Study of Evaluation (ERIC Document Reproduction Service No. ED 218320)

Camilleri, C. (1986). Cultural anthropolgy and education. London: Kagan Page in association with UNESCO.

Carlson, Jerry S. (1983). Applications of dynamic assessment to cognitive and perceptual functioning of three ethnic groups. Riverside, Calif: California University (ERIC Document Reproduction Service No. ED 233040)

Cattell, Raymond B. (1979). Are culture-fair intelligence tests possible and necessary? Journal of Research and Development in Education, 12, (2), 3-13.

Catterall, James S. (1986). Dropping out of school as a process: Implications for assessing the effects of competency tests required for graduation. Los Angeles, Calif: Los Angeles Centre for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 293879)

Catterall, James S. (1987). Toward researching the connections between tests required for high school graduation and the inclination to

drop out of school. Project: Effects of testing reforms and standards. Los Angeles: Calif: Centre for the Study of Evaluation. (ERIC Reproduction Service No. ED 293886)

Census Tracts Winnipeg: Part I. Census Recensement Canada. (1988). Ottawa: Statistics Canada.

Chavez, Ernest L. (1982). Analysis of a Spanish translation of the Peabody Picture Vocabulary Test. Perceptual and Motor Skills, 54, (3), 1335-38.

Child, Dennis. (1985). Educational psychology: Past, present, and future. In Noel Entwistle (Ed.), New directions in educational psychology learning and teaching (pp. 9-23). Philadelphia: The Falmer Press.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.

Cole, M., & Scribner, S. (1974). Culture and thought. New York: Wiley.

Crawford, Alan N. (1985). Test review: Preuba del desarrollo inicial del language. Reading Teacher, 38, (4), 428-31.

CTBS manual for administrators, supervisors, and counsellors. (1975). Canada: Nelson.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. San Diego: College-Hill Press.

Cummins, Jim. (1989). Institutionalized racism and the assessment of minority children: A comparison of policies and programs in the Unites States and Canada. In Ronald J. Samuda, Shiu L. Kong, Jim

Cummins, Juan Pascual - Leone & John Lewis (Eds.), Assessment and placement of minority students. (pp. 95-107). Toronto: C.F. Hogrefe.

Curtis, Mary E. & Glaser, Robert. (1985). Intelligence testing, cognition, and instruction. Pittsburg, PA: Learning Research and Development Centre. (ERIC Reproduction Service No. ED 263163)

Dale, Edgar & O'Rourke, Joseph. (1979). The living world vocabulary. Chicago: Field.

Das, J. (1973). Cultural deprivation and cognitive competence. In Ellis, W. (Ed.), International review of research in mental retardation. New York: Academic Press, 1973.

Das, J.F., Kirby R., & Jarman, Ronald F. (1979). Simultaneous and successive cognitive processes. New York: Academic Press.

Darveau, A. (1982). Mission St. Pierre Du Lac Caribou - St. Peter Mission, Brochet, Manitoba. Unpublished manuscript.

Deese, J. (1967). Meaning and change of meaning. American Psychologist, 22, 641-651.

Developing cognitive abilities test technical manual and norms. (1983). Toronto: Gage.

Dillon, Ronna F. & Stevenson-Hicks, Randy. (1983). Competence vs. performance and recent approaches to cognitive assessment. Psychology in the Schools, 20, (2), 142-145.

Ebel, Robert L., & Frisbee, David A. (1986). Essentials of educational measurement. (4th ed.). Prentice-Hall: Englewood Cliffs.

Edelsky, Carole & Harman, Susan. (1988). One more critique of reading tests--with two differences. English Education, 20, (3), 157-171.

Eckberg, Douglas Lee. (1979). Intelligence and race. New York: Praeger.

ERIC Clearinghouse on tests, measurement, and evaluation. (1984). Princetown, N.J.: Educational Testing Service. (ERIC Document Reproduction Service No. ED 286948)

Eysenck, Hans J. (1979). The structure and measurement of intelligence. New York: Springer-Verlag.

Faggen, Jane. (1987). Golden rule revisited: Introduction. Educational Measurement: Issues and Practice, 6, (2), pp. 5-8.

Farb, Peter. (1968). How do I know what you mean? Horizon, 10, (4), pp. 52-59.

Feinberg, Walter. (1978). IQ tests, intelligence, and the distribution of knowledge. Unpublished. Urbana: University of Illinois.

Flynn, James R. (1980). Race, IQ, and Jensen. London: Routledge and Kegan Paul.

Fox, Lynn H. (1985). Review of developing cognitive abilities test. In The ninth mental measurements yearbook. (Vol. 1.) (pp. 460-461). The University of Nebraska: The University of Nebraska Press.

Frase, Lawrence T. (1980). The demise of generality in measurement and research methodology. In Eva L. Baker & Edys Quellmalz. Educational testing and evaluation (pp. 460-461). Beverly Hills: Sage.

Gagne, Robert M., Briggs, Leslie J., & Wagner, Walter W. (1988). Principles of instructional design. (3rd ed.). New York: Holt, Rinehart, & Winston.

Gardner, Eric E.F., Rudman, Herbert C., Karlsen, Bjorn, & Merwin, Jack C.O. Stanford achievement test national norms booklet: Advanced form E/F. New York: The Psychological Corporation Harcourt, Brace, & Jovanovich.

Gardner, Eric E.J., Rudman, Herbert C., Karlsen, Bjorn, & Merwin, Jack C. (1983). Stanford achievement test national norms booklet. Primary 3 forms E/F. New York: The Psychological Corporation Harcourt, Brace, & Jovanovich.

Gay, L.R. (1985). Educational evaluation and measurement competencies for analysis and application. Columbus: Merrill.

Gibson, Janice T. (1980). Psychology for the classroom. (2nd ed.). Englewood Cliffs: Prentice-Hall.

Ginsburg, Herbert P. (1986). The myth of the deprived child: New thoughts on poor children. In Ulrich Neisser (Ed.), The school achievement of minority children: New perspectives. (pp. 169-89). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Glaser, Robert. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. American Psychologist, 36, (9), 923-36.

Goodnow, J.J. (1976). The nature of intelligence behavior: Questions raised by cross-cultural studies. In L.B. Resnick (Ed.), The Nature of intelligence. (pp. 1-10). New York: John Wiley and Sons.

Gonzales-Tamayo, Eulogio. (1984). Aptitude testing controversy: Beliefs, not values, are on trial. Information analyses-viewpoints (ERIC Document Reproduction Service No. 248286)

Gonzales-Tamayo, Eulogia. (1987). The golden rule agreement is psychometrically defensible. (Evaluation/Feasibility Report No. 142). US: Department of Education Office of Educational Research and Improvement, Educational Resource Centre (ERIC Document Reproduction Service No. ED 283868)

Grippin, Pauline & Peters, Sean. (1984). Learning theory and learning outcomes: The connection. New York: University of America Press.

Gronlund, Norman E. (1981). Measurement and evaluation in education. New York: MacMillan.

Gronlund, Norman E. (1985). Measurement and evaluation in teaching. New York: MacMillan.

Grover, Sonja C. (1981). The cognitive basis of the intellect: A response to Jensen's "bias in mental testing". Washington: University Press of America.

Guilford, J.P. (1967). The nature of human intelligence. New York: McGraw-Hill.

Guskey, Thomas R. & Gates, Sally L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. Educational Leadership, 43, (8), 73-80.

Hall, Edward T. (1986). Unstated features of the cultural context of learning. In Alan Thomas & Edward W. Ploman (Eds.), Learning and development: A global perspective (pp. 239-252). Toronto: The Ontario Institute for Studies in Education.

Hakuta, K. (1986). Mirror of language. New York: Basic Books.

Hansen, Judith Friedman. (1979). Sociocultural perspectives on human learning: An introduction to educational anthropology. Englewood

Cliffs, New Jersey: Prentice-Hall.

Harman, David. (1980). On traditional testing. In Eva L. Baker & Edys S., Educational testing and evaluation. Beverley Hills, Sage.

Harrington, Charles. (1979). Psychological anthropology and education: A delineation of a field of inquiry. New York: A M S Press.

Heath, Shirley Brice. (1986). Critical factors in literacy development. In Suzanne de Castell, Allan Luke & Kieran Egan (Eds), Literacy, Society, and Schooling (pp. 209-229). New York: Cambridge University Press.

Hilliard III, Asa G. (1984). Democracy in education: The evolution of an art-science in context. In Philip Hosford (Ed.), Using what we know about teaching (pp. 113-130). Alexandria, Virginia: Association For Supervision and Curriculum Development: Alexandria, Virginia.

Hills, John R. (1981). Measurement and evaluation in the classroom. Columbus: Merrill.

Hunt, Earl. (1985). Science, technology, and intelligence. Technical Report 9. Lincoln, NE: Paper presented at the Annual Meeting of the Buros Institute (ERIC Reproduction Service No. ED 264258)

Hunter, Madeline. (1980). Altering the alterable variables. The Educational Forum, 45 (1), 121-122.

Hunter, Madeline. (1985). Building effective elementary schools. In J. William Johnson, Education on trial strategies for the future (pp. 53-65). San Francisco: Institute For Contemporary Studies.

Hynd, George W. & Garcia, William I. (1979). Intellectual assessment of the native American student. School Psychologist, 8, (4), 44-54.

Jaeger, Richard M. (1987). NCME opposition to proposed golden rule legislation. Educational Measurement: Issues and Practice, 4, (2), 21-22.

Jaeger, Richard M. & Busch, John-Christian. (1986). The use and effect of caution indices in detecting aberrant patterns of standard-setting recommendations. San Francisco, CA: Paper presented at the 70th Annual Meeting of the American Educational Research Association (ERIC Document Reproduction Service No. ED 269436)

Jaschik, Scott. (1987). Legislature votes to require testers to divulge details of students' responses. Chronicle of Higher Education, (33), 41, 17-22.

Johnson, Sylvia T. (1988). Test fairness and bias: Measuring academic achievement among black youth. Urban League Review, (7) 1-2, 76-92.

Joseph, Andre. Bicultural socialization and the measurement of intelligence. (ERIC Reproduction Service No. ED 138616) (no date provided).

Kelley, Paul H. (1982). Are culturally biased tests useful? New directions for testing and measurement. Academic Testing and the Consumer, 16, 125-133.

King, Ethel. (ed.). (1982). Canadian test of basic skills: Multilevel edition, teacher's guide. Canada: Nelson.

Klausmeier, H.J., Ghatala, E.S. & Frayer, D.A. (1972). Levels of concept attainment and the related cognitive operations.

Klausmeier, H.J., Ghatala, E.S. & Frayer, D.A. (1972). Levels of

concept attainment and the related cognitive operations. Theoretical Paper from the Wisconsin Research and Development Centre for Cognitive Learning. Madison, Wisconsin: The University of Wisconsin.

Klausmeier, Herbert J., Ingison, Linda J., Sipple, Thomas S. & Katzenmeyer, Conrad G. (1977). Development of conceptual learning and development assessment series I: Equilateral triangle. (Technical Report No. 430). Madison, Wisconsin: Wisconsin Research and Development Center For Cognitive Learning, The University of Wisconsin.

Klausmeier, Herbert J. & Sipple S. (Eds.) (1980), Learning and teaching concepts: A strategy for testing applications theory. New York: Academic Press.

Lewis, John. (1989). Innovative approaches in assessment. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, John Lewis, & Juan Pascual-Leone (Eds.), Assessment and placement of minority students (pp. 123-142). Toronto: Hogrefe.

Lieblich, A. & Kugelmass, S. (1983). Patterns of intellectual ability of school children in Israel. Intelligence, 5, (4), 311-313.

Linn, Robert L. (1984). Selection bias: Multiple meanings. Journal of Educational Measurement, 21, (1), 33-48.

Linn, Robert L. & Drasgow, Fritz. (1987). Implications of the golden rule: Settlement for test construction, Educational Measurement: Issues and Practice, 6, (2), 13-17.

Lorimer, Rowland. (1986). The business of literacy: The making of the educational textbook. In Suzanne de Castell, Allan Luke & Kieran

Egan (Ed.), Literacy, society, & schooling (pp. 132-142). New York: Cambridge University.

Maldonado-Colon, E. (1986). Assessment: Considerations upon interpreting data of linguistically/culturally different students referred for disabilities or disorders. In A.C. Willig & H.F. Greenberg (Eds.), Bilingualism and learning disabilities: Policy and practice for teachers and administrators (pp. 69-77). New York: American Library Publishing.

Mehrens, William A. & Lehmann, Irvin J. (1984). Measurement and evaluation in education and psychology. (3rd ed.). Toronto: Holt, Rinehart, and Winston.

Mehrens, William A. & Lehman, Irvin J. (1986). Using standardized tests in education (4th ed.). New York: Longman.

Mercer, J. (1971). Instutionalized anglocentrism: Labelling mental retardates in the public schools. In P. Orleans & W. Russell, Jr. (Eds.), Race, change, and urban society. Urban Affairs Review. (Vol. 5). Los Angeles: Russell Sage.

Mercer, J. (1973). Labelling the mentally retarded. Los Angeles: The University of California Press.

Mercer, J. & Lewis, J. (1978). System of multicultural pluralistic assessment. New York: Psychological Corporation.

Messick, Samuel. (1985). Style in the interplay of structure and process. In Noel Entwistle. (Ed.), New directions in educational psychology learning and teaching, (pp. 83-98). Philadelphia: The Falmer Press.

Neely, Renee & Shaunessy, Michael F. (1984). Assessment and the native

American. Portales, New Mexico: New Mexico University. (ERIC Document Reproduction Service ED 273889)

Nenty, Johnson H. (1986). Cross cultural bias analysis of Cattell culture fair intelligence test. San Francisco, California: 67th Annual Meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 274668)

Noll, Victor H., Scannell, Dale P., & Craig, Robert C. (1979). Introduction to educational measurement. Boston: Houghton-Mifflin.

Oakland, Thomas. (1982). Nonbiased assessment in counselling: Issues and guidelines. Measurement and evaluation in guidance, 15, (1), 107-116.

Oplesch, Marie & Genshaft, Judy. (1981). Comparison of bilingual children on the wisc-r and the escala de intelligencia wechsler para-ninos. Psychology in the Schools, 18, (2), 159-163.

Oritz, A.A. & Yates, J.R. (1983). Incidence of exceptionality among Hispanics: Implications for manpower planning. National Association Black Education Journal, 7, 41-54.

Partridge, Susan. (1986). Negative aspects of minimum competency testing to surface: Implications. (ERIC Document Reproduction Service No. 276748)

Pascual-Leone, Juan, & Ijaz, Helene. (1989). Mental capacity as a form of intellectual-developmental assessment. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, John Lewis (Eds.) & Juan Pascual- Leone. (1989), Assessment and placement of minority students (pp. 143-121). Toronto: C.F. Hogrefe.

Popham, James W. (1981). Modern educational measurement. Englewood

Cliffs: Prentice-Hall.

Popham, James W. (1983). Measurement as an instructional catalyst. New Directions for Testing and Measurement, (Measurement, technology, and individuality in education: Proceedings of the 1982 ETS Invitational Conference), (7), 19-30.

Popham, James W. (1988). Educational evaluation. Englewood Cliffs: Prentice-Hall.

Possemato, Paul M. (1985). Secondary education. In J. Johnston (Ed.), Education on trial strategies for the future (pp. 63-78). San Francisco. Institute for Comtemporary Studies.

Reed, Rodney J. (1987). School and college competency testing programs: Perceptions and effects on the black students in Louisiana and North Carolina. New York: Ford Foundation & Atlanta: Southern Educational Foundation. (ERIC Reproduction Service No. 286960)

Reilly, Robert R. & Lewis, Ernest L. (1983). Educational psychology: Applications for the classroom. New York: MacMillan.

Rengal, Elizabeth. (1986). Agreement between statistical and judgemental item bias methods. Washington, DC: Paper presented at the Annual Meeting of the American Psychological Association (ERIC Reproduction Service No. ED 289890)

Reschley, D.J. (1981). Psychological testing in educational classification and placement. American Psychologist, 36, (19), 1094-1102.

Reynolds, Cecil R. (1980). An examination for bias in a preschool test battery across race and sex. Journal of Educational Measurement,

17, (2), 137-146.

Reynolds, Arthur J. & Bezruczko, Nickolaus. (1988). Assessing the construct validity of a life skills competency test. Paper presented at the Annual Meeting of the National Council of Measurement in Education. (ERIC Document Reproduction Service No. ED 294897)

Rhodes, Robert W. (1988). Standardized testing of minority students: Navajo and hopi Examples. St. Louis, MO: Paper presented at the Annual Meeting of the National Council of Teachers of English. (ERIC Document Reproduction Service ED 299587)

Rosenbach, John H. (1979). Instrumentation for assessing the culturally different: Some conceptual issues. San Francisco, Calif: Annual Meeting of the National Council on Measurement in Education. (ERIC Reproduction Service No. 177210)

Rosenbach, John H. & Mowder, Barbara A. (1981). Test bias: The other side of the coin. Psychology in the Schools, October, 18, (4), 450-54.

Roth, David R. (1976). Reconsidering the distinction between verbal and non-verbal I.Q. tests: A sociological perspective. Research Paper No. 1. Spencer Foundation Grant No. B-229. Mimeographed. Austin: University of Texas.

Samuda, Ronald J. (1989a). The new challenge of student assessment and placement. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, Juan Pascuale-Leone & John Lewis (Eds.), Assessment and placement of minority students (pp. 15-23). Toronto: Hogrefe.

Samuda, Ronald J. (1989b). Student assessment and placement in Ontario

Schools. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, Juan Pascuale-Leone & John Lewis (Eds.), Assessment and placement of Minority Students (pp. 109-122). Toronto: Hogrefe.

Samuda, Ronald J. (1989c). Towards nondiscriminatory assessment: Principles and application. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, Juan Pascuale-Leone & John Lewis (Eds.), Assessment and placement of minority students (pp. 173-189). Toronto: Hogrefe.

Samuda, Ronald J. (1989d). Nature versus nurture. In Ronald J. Samuda, Shiu L. Kong, Jim Cummins, Juan Pascuale-Leone & John Lewis (Eds.), Assessment and placement of minority students (pp. 41-68). Toronto: Hogrefe.

Scarr, Sandra. (1978). Testing minority children: Why, how, and with what effects. Bossone (Ed.). Proceedings of the National Conference on Testing: Major Issues. New York: Centre for Advanced Study in Education. In Sandra Scarr (Ed.) (1981), Race, social class, and individualized differences in I.Q. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schueneman, Janice. (1985). Exploration of causes of bias in test items. Educational Testing Service N.J: Graduate Record Exam Board. (ERIC Reproduction Service No. ED 268157)

Seifert, Kelvin. (1983). Educational psychology. Boston: Houghton Mifflin.

Serow, Robert C. & Davis, James, J. (1982). Resources and outcomes of minimum competency testing as measures of equality of educational opportunity. American Educational Research Journal, 19, (4), 529-539.

Shaha, Steven H. & Wittrock, Merlin C. (1983). Cognitive & affective processes related to school achievement: Implications for assessment. California University: Centre for the Study of Evaluation. (ERIC Reproduction Service No. 228272)

Singh, Moti. (1982). The relationships among academic achievement, self-concept, creativity, and teacher expectation of cree children in a northern community. Master's dissertation, University of Manitoba, Winnipeg.

Slavin, Robert E. (1988). Educational psychology: Theory into practice. (2nd ed.). Englewood Cliffs: Prentice Hall.

Smith, J.G.E. (1978). The emergence of the micro-urban village among the caribou-eater Chipewayan. Human Organization. 37, (1), 38-49.

Spencer, Bruce D. (1983). On interpreting test scores as social indicators: statistical considerations. Journal of Educational Measurement, 20, (4), 317-333.

Stanywck, Douglas & Abdelal, Phyllis. (1984). Attitudes toward cheating behavior in the ESL classroom. West Palm Beach, FL: Paper presented at the Annual Meeting of the Eastern Educational Research Association (ERIC Document Reproduction Service No. ED 250927)

Stone, David R. & Neilson, Edwin C. (1982). Educational psychology: The development of teaching skills. New York: Harper and Row.

Taylor, Orlando L. & Lee, Dorian Latham. (1987). Standardized tests and African-American children; Communication and language issues. Negro Educational Review, 38, (2-3), 67-80.

The canadian encyclopedia. (1985). (Vol. 2) Hurtig: Edmonton.

Tom, David & Cooper, Harris. (1984). Academic attributions for success and failure among Asian Americans. New Orleans, LA: Paper presented at the Annual Meeting of the American Educational Research. (ERIC Document Reproduction Service No. ED 246145)

Trueba, Henry T. (1987). Success or failure? Learning and the language minority student. In D. Deyhle (Ed.), Learning failure: Tests as gatekeepers and the culturally different child. (ERIC Document Reproduction Service No. 286959)

Tylor, E.B. (1871). Primitive culture. London: John Murray.

Urzillo, Robert L. (1987). Competency testing: blessing or bane. Contemporary Education, 59, (1), 13-14.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, E. (Eds.). Cambridge: Harvard University Press.

Wagner, D.A. (1978). Memories of Morocco: The influences of age, schooling, and environment on memory. Cognitive Psychology, 10, 1-28.

Webster, Raymond E. (1978). Information processing variables in the assessment of school-related problems. Ontario, Canada: 86th Annual Meeting of the American Psychological Association. (ERIC Reproduction Service No. ED 173356)

Weir, Thomas R. & Wai Lai, Ngok. (Eds.). (1978). Atlas of Winnipeg. Toronto: University of Toronto Press.

Weir, Robert R. (Ed.). (1983). Manitoba Atlas. Province of Manitoba: Surveys and Mapping Branch Department of Natural Resources.

Weiss, John G. (1987a). Its time to examine the examiners. Negro

Educational Review, 38, (2-3), 107-24.

Weiss, John. (1987b). Truth-in-testing & the golden rule principle: Two practical reforms. Washington, DC: Paper presented at the Annual Meeting of the National Council on Measurement in Education. (ERIC Document Reproduction Service No. ED 283826)

Wiersma, William & Jurs, Stephen G. (1985). Educational measurement and testing. Newton: Allyn and Bacon.

Wildemuth, Barbara M. (1983). Minimum competency testing and the handicapped. ERIC Digest published in the ERIC/TME Update Series (ERIC Reproduction Service No. ED 289886)

Williams, Terence. (1983). Some issues in standardized testing of minority students. (ERIC Document Reproduction No. ED 264258)

Willie, Charles V. (1985). The problem of standardized testing in a free and pluralistic society, Phi Delta Kappan, 6, 626-28.

Wilson, Buford, E. (1984). Knowledge and its acquisition. An introduction and overview. In Ronald K. Bass & Charles R. Dills (Eds.), Instructional development: The state of the art, II (pp. 126-138). Dubuque Kendall/Hunt.

Woolfolk, Anita E., & McCune-Nicolich, Lorraine. (1984). Educational psychology for teachers. (2nd ed.). Englewood Cliffs: Prentice-Hall.

Woolfolk, Anita E. (1987). Educational psychology. (3rd ed.). Prentice-Hall: Englewood Cliffs.

Woodruff, A.D. (1961). Basic concepts of teaching. Concise Edition. San Francisco: Chandler.

Zorn, Jeffrey L. (1983). Possible sources of culture bias in the

validation of ETS language tests. Detroit, MI: Paper presented at the Annual Meeting of the Conference on College Composition and Communication. (ED Reproduction Service No. ED 235504)

(ED Reproduction Service No. ED 235504)