AN ABSTRACT OF THE THESIS OF

Todd V. Fletcher for the degree of Doctor of Philosophy in Education presented December 17, 1985.

Title: Comparative Analysis Between the Woodcock Psycho-Educational Battery in Spanish and the Wechsler Intelligence Scale for Children--Revised Mexican Version

Abstract approved: Redacted for Privacy
Don H. Duncan

The purpose of this study was to compare and describe test score differences on the Woodcock Psycho-Educational Battery in Spanish (BATERIA) and the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM). Analyses were performed on the two test batteries, investigating mean score differences on the different clusters as a function of type of subject. The sample consisted of 43 fifth grade subjects classified as normals drawn from public and private schools in Mexico City and a sample of 47 subjects classified as referrals drawn from the same classes.

The research questions addressed included comparability of performance between normal and referrals on: (a) the full scale scores of the two batteries; (b) the Special

Cognitive Ability clusters of the BATERIA; (c) the Scholastic Aptitude clusters of the BATERIA; (d) the Scholastic Achievement clusters of the BATERIA; and (e) the Verbal and Performance scales of the WISC-RM.

The most significant finding was that, after controlling for differences in the norms, referral subjects scored about 5.5 standard score points higher on the Full Scale score of the WISC-RM than on the Full Scale score of the BATERIA. In addition, the results indicated that the Full Scale score of the BATERIA correlated better with the three areas of achievement (mathematics, reading, and written language) than did the the Full Scale score of the WISC-RM.

Normal subjects scored significantly higher than referrals on the three cognitive ability clusters, the three scholastic aptitude clusters, and the three scholastic achievement clusters of the BATERIA. Referral subjects scored significantly higher on the Performance scale of the WISC-RM than on the Verbal scale, while normal subjects exhibited no difference when comparing the two scales.

This study provided insights regarding the impact of the norming procedure on obtained norms, and provides a comparison of performance on the BATERIA and the WISC-RM based on two samples of subjects.

Comparative Analysis Between the Woodcock Psycho-Educational Battery in Spanish and the Wechsler Intelligence Scale for Children--Revised Mexican Version

by

Todd V. Fletcher

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed December 17, 1985

Commencement June 1986

APPROVED:

Redacted for Privacy

Professor of Special Education in charge of major

Redacted for Privacy

Head of department of Special Education

Redacted for Privacy

Dean of Graduate School

Date thesis is presented December 17, 1985

Typed by Bill McMechan for Todd V. Fletcher

DEDICATION

This disseration is dedicated to John McCarthy, a fellow teacher killed during the revolution in Nicaragua, who was dedicated to the education of children and the pursuit of knowledge.

Many people provided assistance in the process of completing this thesis. Dick Woodcock, from the inception of this project was a constant source of support and guidance. His willingness to listen and provide feedback throughout this project was invaluable. Thank you for the innumerable hours that we engaged in fruitful discussion and debate. Don Duncan, my major professor, provided a delicate balance of flexibility and direction and kept me looking at the lighter side of life with his keen sense of humor. Leona Gilliland, a dedicated special education teacher, was an excellent role model and an inspiration for me to aspire to become the best that I possibly could be. Merle Kelley, thank you for your words of encouragement and your strong principles. To my other committee members thank you for your patience and understanding.

I would also like to thank my professors and colleagues at Western Oregon State College, Oregon State University, the University of the Americas, and the University of Arizona for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

COMPARATIVE ANALYSIS BETWEEN THE WOODCOCK PSYCHO-EDUCATIONAL BATTERY IN SPANISH AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED MEXICAN VERSION

CHAPTER 1

INTRODUCTION

An area of critical concern in the field of special education in the United States is the lack of appropriate assessment instruments for evaluating culturally and linguistically different children referred for a psycho-educational evaluation to determine if a handicapping condition is present.

This concern is particularly critical with regard to the past and present abuses in the use and interpretation of standardized educational and psychological tests by professionals when working with culturally and linguistically different groups in American schools. This is supported by the fact that there have been disproportionate numbers of immigrant and minority students, in particular children from a Spanish speaking background, placed in special education classrooms when compared to their anglo counterparts (Cummins, 1984). Studies by Williams (1968) and Mercer (1973)

have documented the over-representation of Mexican-American children in special classrooms based on the proportion of these students in American schools. Mercer (1973) indicates that Mexican-American children were 10 times as likely to be placed in special education programs as their anglo counterparts.

The reason for this over-representation according to Mercer (1979) is the indiscriminate use of psychological and educational tests which embody the cultural and linguistic orientation of American society and values. This pattern of mislabeling and misplacement began with the large influx of immigrants in the late 1800's and early 1900's and has continued until present day (Gelb, 1982). Children of these immigrant families were frequently placed in special classes because they could not meet the expectations and standards as a member of a regular class (Miller, 1916).

This concern increases with the continual and projected influx of immigrant and minority students, particularly Hispanic, into our country. Cummins (1984) reaffirmed the need to examine and reassess the procedures, policies, and practices used to assess these minority language students. As Cummins stated the adjustment to the reality in the educational context is slow "because it is difficult for some educators to acknowledge that these diagnostic assessment tools . . . are inappropriate for an increasing number of students" (p. 2).

In the past 20 years this growing concern has been accompanied by court cases and litigation trying to arrest the wave of misclassification of Mexican-American children into special education programs. The Education for All Handicapped Children Act of 1975 (P.L. 94-142) stands out as landmark legislation with regard to meeting the individual needs of exceptional minority children. Baca and Cervantes (1984) stated that meeting the needs of exceptional students from language minority backgrounds will require the efforts of special educators in improving current assessment instruments and procedures so that discriminatory practices and misplacement of these individuals can be reduced. Bernal (1983), in discussing the current needs in the emerging area of bilingual special education, observed that clinical based research is necessary, especially in the area of assessment, to determine how valid IQ tests are for diagnosing the cognitive exceptionalities of limited English proficient students.

The Woodcock Psycho-Educational Battery in Spanish (BATERIA) and the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM) are two tests currently used by professionals to assess Spanish language dominant students in the diagnosis of learning disabilities. As specified in P.L. 94-142, one criteria for determining whether a student has a specific learning disability is the presence of a discrepancy between aptitude and achievement. Both tests are used as measures of scholastic aptitude and

scores obtained are often compared with achievement scores from other standardized tests. Previous studies comparing the two Full Scale mean scores of the corresponding English versions of the BATERIA and WISC-RM have produced mixed results (Reeve, Hall, and Zakreski, 1979; McGrew, 1983). Research is needed on monolingual Spanish speakers to determine the differences that can be expected between the two measures when assessing dominant language, Spanish speaking students who are suspected of having a learning disability.

Purpose of the Study

The purpose of this study is to provide descriptive and comparative data regarding two major test batteries now available in Spanish that were normed in Latin America. These two batteries are the Woodcock Psycho-Educational Battery (BATERIA) and the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM).

Objectives of the Study

1. To analyze the difference between mean scores when comparing the Broad Cognitive Ability Scale of the Woodcock Psycho-Educational Battery in Spanish (BATERIA) and the Full Scale score on the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM) with a group of normal and a group of referral subjects.

2. To analyze the mean score differences when comparing the three special cognitive ability cluster scores from

the BATERIA with a group of normal and a group of referral subjects.

3. To analyze the mean score differences when comparing the three scholastic aptitude cluster scores from the BATERIA with a group of normal and a group of referral subjects.

4. To analyze the mean score differences when comparing the three scholastic achievement cluster scores from the BATERIA with a group of normal and a group of referral subjects.

5. To analyze the mean score differences when comparing the Verbal and Performance Scale scores on the WISC-RM with a group of normal and a group of referral subjects.

Discussion of Terms

The psychological profession has developed a number of terms that are utilized to describe certain human traits and characteristics in this field. The following terms are described and discussed as a means of containing this study and containing the focus of the research problem.

Intelligence tests. Thorndike and Hagan (1977) suggest that intelligence tests are not pure measures of native potential but rather measures of "developed abilities." Anastasi (1982) defines "intelligence tests" as a heterogeneous sampling of tests that provide a single global score, such as an IQ. She states further that traditional intelligence tests cover a cluster of cognitive skills and knowledge that

are "widely predictive of performance in both academic and occupational settings demanded in modern, advanced technological societies."

Aptitude tests. Aptitude testing has traditionally been employed to "refer to tests measuring relatively homogeneous and clearly defined segments of ability" (p. 15). These tests, according to Anastasi (1982), reflect "the cumulative influence of a multiplicity of experience in daily living" (p. 393). Special aptitude tests are used to determine mechanical, musical, and educational aptitudes in areas such as mathematics, reading, written language, and knowledge. The purpose of differential aptitude tests are to reveal differences in a number of aptitudes by providing a profile of scores, usually one for each aptitude. These scores then serve to predict future performance and how well an individual may expect to benefit from a specified course of training.

Achievement tests. Anastasi (1982) states that these tests are designed to measure the effects of a specific program of instruction or training. She states that they measure the results of a standardized set of experiences, such as classes in arithmetic and reading provided to a student under a controlled (formal) learning environment.

Psycho-educational Assessment. In the area of testing and assessment Stellern (1982) refers to the contribution that psychological and educational theory and tests can make to the understanding of a learner and the consequent

resolution of the learning problem. More recently, the term describes an integrated procedure to selectively correlate the two areas of psychological and educational assessment instruments in a common/unitary assessment battery.

Normals. These are the students in this study selected by a random sample procedure using a table of random numbers from each fifth grade class.

Referrals. These are the students in this study selected from each fifth grade class who the teacher perceived to be achieving the poorest in the scholastic skill areas (any one of three--reading, math, or written language). The assumption was made that these were the children the fifth grade teachers would have referred if services of this type were available.

Overview

This study will provide an analysis of two psychological and educational instruments recently developed in Spanish. There is a need for comparative studies in the Spanish editions to determine the significance and impact these instruments might have on normal as compared to referral language minority student populations. A review of the literature regarding psycho-educational assessment and legal issues and concerns of minority language populations will be presented in the next chapter. The review will also include current comparative research on the two batteries in English and discuss concerns that have been raised with regard to

differences reported on scores obtained when comparing normal and referral samples of subjects.

CHAPTER 2

REVIEW OF RELATED LITERATURE

The review of literature surveyed for this study is divided into two major categories. The first part reviews and summarizes the literature related to the history and current assessment of bilingual exceptional children. The second reviews and summarizes the current literature in the field regarding the research comparing the Broad Cognitive Ability cluster of the Woodcock-Johnson Psycho-Educational Battery and the Full Scale score of the Wechsler Intelligence Scales for Children--Revised in English.

Historical Perspective

With the large influx of immigrants in the late 1800's and early 1900's to the present, and the advent of compulsory education in America, public schools encountered large numbers of students who had few English language skills. It was the fact that these non-English speaking children could not keep up with their companions that necessitated assignment to the "streamer" or "backward" classes (Miller, 1916).

In many cases these new students (non-English speakers) were labeled as "defective" and "feebleminded" because they could not perform as other students their age. As Gelb

(1983) recounts, the attitudes of educators in the early 1900's were strongly influenced by Social Darwinism and the Eugenics movement. The Social Darwinists, by observing their own society, came to the conclusion that the most fit people were those at the top of the socio-economic ladder and the least capable were those at the bottom. The theory of selection was used to explain who the most advanced humans were from an evolutionary standpoint and why. The Eugenicists, in turn, believed that certain races and segments of society were inferior to others. These philosophies provided a basis for the belief that children who spoke foreign languages and were from a different culture were mentally incompetent and needed special education services.

Some of the early leaders in the field of intelligence testing took the Eugenicist position. The Binet test, which was originally developed in France for the purpose of discriminating between slow and normal learners, was used for other purposes by people such as H. H. Goddard and Lewis Terman in the United States. Using his version of the Binet test, Goddard tested newly arriving immigrants in New York City in English to discriminate between persons of normal and borderline intelligence (Gelb, 1983). Lewis Terman, in his book The Measurement of Intelligence (1916), depicted Spanish-speaking Americans and Blacks as being inferior in ability and needing to be separated from other students in schools. As Terman stated:

> Children of this group should be segregated in special classes . . . . They cannot master abstractions, but they can often be made efficient workers . . . . They constitute a grave problem because of their unusually prolific breeding. (1916, p. 6)

Kamin (1975) wrote of the political and social abuses of the intelligence test during the first part of this century.

> Since the introduction to America, the intelligence test has been used more or less consciously as an instrument of oppression against the underprivileged, the poor, the foreign born, and racial minorities. (p. 317)

In 1923, Sweeney, testifying to a House Committee, stated that

> The fact that the immigrants are illiterate or unable to understand the English language is not an obstacle . . . . 'Beta' [an intelligence test] . . . is entirely objective. (January 24, 1923, p. 589-594)

Because of these inequities, culturally and linguistically different students have historically been misplaced in special education programs due to their inability to meet school expectations in English (Mercer, 1973; Tucker, 1980; Zobel, 1980).

George Sanchez (1934) was one of the first to document the way in which IQ tests discriminated against minority populations. He challenged the early hereditarian interpretations of Hispanic's low IQ scores. In 1968, Dunn found that 60-80 percent of the students identified as having mildly handicapping conditions were from low socio-economic backgrounds and were mainly ethnic minorities. Jane

Mercer's study (1973), one of the few that focused on the actual process of assessment, suggested that the over-representation of Mexican-American children in EMR classes was not a result of over-referral from teachers and other professionals but rather, bias in the test instruments being used.

Standardized educational and psychological tests have been developed to meet professional needs in determining academic, behavioral, and social growth of students (Oakland, 1973). But as recounted by Nazzaro, "standard batteries of tests normed on white middle class children were used to determine I.Q. and academic achievement" (1979, p. 1) for all children, regardless of their cultural background.

Recently, with the passage of Public Law (P.L.) 94-142, it was mandated that all students be assessed in their dominant mode of communication. Since that time Spanish language proficiency tests have been developed to determine the relative language proficiency of Spanish speaking students. Many educational and psychological tests have been translated into Spanish while maintaining the same norms by which to compare a student's performance. Use of these tests has met with criticism and concern and has led to classic court cases and litigation which have directly influenced the manner in which assessment procedures must be carried out with minority students. In addition, this litigation has paved

the way for a more comprehensive response to their educational needs.

Legal Issues Concerning the Assessment of Handicapped Language Minorities

During the past two decades in the United States litigation and subsequent federal directives have had significant implications for handicapped minority individuals. Beginning with the Supreme Court's decision in Brown vs. Board of Education, a flood of legislation related to the "disadvantaged" child has ensued. The Civil Rights Act of 1964 was the first piece of major legislation that required school districts receiving federal funds to ensure non-discrimination on the basis of race, color, or national origin. The U.S. Office for Civil Rights (OCR's May 25, 1970, Memorandum, Fed. Regist. 35, 11595) later interpretations guaranteed that school districts receiving federal funds were to use non-discriminatory testing procedures in their institutions. The passage of the Bilingual Education Act (1967) and the enactment of the former led to a large body of research regarding normal children who were limited English proficient (LEP) and those children who were in "double jeopardy," being both LEP and handicapped.

In 1975, with the passing of P.L. 94-142, focus was placed on what constituted an appropriate education for handicapped children and provided guidelines for determining special services to be provided to these individuals.

Within this act, several judicial decisions have issued from court proceedings of the 1960's and 1970's--such as the right to a free public education, individualized education, non-discriminatory evaluation, and due process requirements.

P.L. 94-142 guarantees the rights of handicapped children to an education designed to meet their specific needs. It also requires that local education agencies insure that testing and evaluation materials, as well as the procedures used in the assessment, are administered so as not to be racially or culturally discriminatory. Testing materials or procedures must be provided and administered in the child's primary language as well. This was to insure that testing be conducted in the child's dominant language by a multidisciplinary team of qualified personnel using tests that are valid for their intended purpose.

P.L. 94-142 provides a definition for deciding upon a diagnosis of a learning disability. It states that the determination of a learning disability

> is based on (1) whether a child does not achieve commensurate with his or her age and ability when provided with appropriate educational experiences, and (2) whether the child has a severe discrepancy between achievement and intellectual ability in one or more of seven areas relating to communication skills and mathematical abilities. (U.S. Congress, Federal Register, 1977, 42, p. 65082)

Case law has had a significant impact on clarifying the educational rights of the bilingual handicapped child. State and federal courts defined through a significant

number of cases the process by which handicapped children were to be identified, evaluated, and served by educational agencies. The courts found that discriminatory testing procedures had led to the misplacement of bilingual students in programs for the handicapped.

In 1970, the decision in the case Diana vs. California State Board of Education found that certain IQ tests discriminated against Spanish-speaking children. This case provides a classic example of the delicateness involving the assessment of minority group children. The nine children, ages 8 to 13 years, in this case were from homes where Spanish was spoken in the home. Based on English language IQ tests, the children were placed in classes for the mentally retarded. Retesting of the same students in their native language of Spanish indicated that the children were of normal intelligence. The plaintiffs charged that the initial testing was prejudicial due to the heavy emphasis on verbal skills in the English language and the fact that the tests were standardized on white, native-born Americans.

In this same case U.S. Circuit Court Judge Robert Peckham issued a preliminary injunction on the use of IQ batteries (Wechsler, Stanford-Binet, and Leiter) for placement of children in special education classes. This ban was made permanent in 1979. In condemning these tests as racially discriminatory, he stated that since the tests were developed in the U.S. on white populations and had not been adjusted or examined closely for use with certain groups,

these groups received lower scores. Because of lawsuits and litigation like Diana v. California State Board of Education, many public school systems have become very sensitive to the issues of mislabeling, miscategorizing and misplacing children of culturally and linguistically different backgrounds into special education programs. This litigation, as related to testing and evaluation, was incorporated into P.L. 94-142 so as to insure the rights of minority children. It constituted a landmark settlement in psychological assessment and provided for assessment in the primary language of the student as well as in English (Figueroa, 1983).

In the case of Covarrubias v. San Diego Unified School District (1971) the influence of sociocultural factors in determining a student's learning ability was raised by the plaintiff. This case, as in Diana v. California State Board of Education, was settled out of court with stipulated agreements made involving the reevaluation of assignment of students based on culturally relevant assessment instruments.

The Guadalupe Organization, Inc. v. Tempe Elementary School District No. 3 et. al. (1972) suit resulted from the disproportionate placement of Yaqui Indians and Mexican-American children in classes for the mentally retarded based on IQ tests given in English rather than their native language. As in the other cases a settlement made out-of-court required that the children be reevaluated in their primary language.

The Larry P. v. Riles court case (1974) was another case which brought the concern of appropriate assessment and placement of minority children into the spotlight. This case was concerned with the misplacement of elementary aged children into a class for the mentally retarded. The complaint stated that the students were

> victims of a testing procedure which fails to recognize their unfamiliarity with the white middle class cultural background and which ignores learning experiences which they may have had in their homes.

Retesting of these same students found them all to be of normal intelligence.

Litigation has been crucial in the establishment of rights for language minority children. Other cases such as Arreola v. Board of Education (1968) and Jose P. v. Ambach (1979), extended the consideration of linguistic and cultural factors when providing evaluation for placement and instruction. The court decisions and legislation enacted in the past 25 years have provided procedural safeguards and developed policy regarding the rights of students to an appropriate assessment that is non-biased, and fair and designed to provide appropriate placement.

Assessment Concerns

Even with major legislation and case law being enacted, inappropriate identification, assessment, and placement continue (Ambert and Dew, 1982). As reported by the General Accounting Office (1981) the foremost reasons for creating

this problem are a lack of information on the educational rights of bilingual exceptional students, a shortage of personnel who are competently trained in bilingual special education, and a lack of appropriate assessment instruments, adequately designed programs, and suitable curriculum materials.

One of the reasons, as listed above, is the lack of culturally relevant instruments which are appropriately normed for assessing different student populations. The Mexican-American subculture represents one of the minority groups that have received some of the negative consequences of the use and abuse of psychological and educational standardized tests. Both Hernandez (1973) and Moore (1976) have commented that our educational system has not met the needs of minority subjects, particularly the Mexican-American population. Studies show that an inability to cope with standardized tests is a major obstacle confronting Mexican-American children (Carter, 1970). De Blassie and Franco (1983) stress that the standardized tests presently used are oriented toward and biased in favor of those individuals who are in the mainstream of American society. In further discussion of this bias, Jones (1976) states that bias occurs "at the level of standardization where decisions are made about the population for whom the test is appropriate" (p. 161).

Major complex issues also arise when attempting to identify the handicapped bilingual child. Two possible

dangers are 1) the "false positive" identification, when bilingual children are incorrectly labeled as exceptional and placed in special programs, and 2) the "false negative" identification, when those bilingual children who have exceptional difficulties are not correctly identified and served. These dangers may be further exacerbated by using tests and norms developed in a social and cultural context which are very different from that of the child being assessed.

Even with recent litigation and procedural safeguards provided by P.L. 94-142, current practices of identification, assessment, and placement are questionable at best (Figueroa, 1983). One of the foremost reasons accounting for this is the lack of appropriate assessment tools to correctly identify, assess, and place students. There are few reliable and valid instruments available which are normed on linguistically or culturally diverse populations of children (Archuleta and Cervantes, 1981; Baca, Cervantes and Torres, 1978). As Sattler (1982) has noted, one of the most important tools in the assessment of learning disabled children are reliable and valid intelligence tests and reliable and valid achievement tests. The identification of a bilingual exceptional child with a learning disability is presently complicated by the fact that there are few trained personnel who can carry out an assessment and few adequate assessment instruments and procedures being used in the process (Ortiz and Yates, 1983).

There have been a few attempts to make existing tests more appropriate for the culturally and linguistically different by direct translation of tests. For instance, Spanish translations of the Wechsler Intelligence Scales have been studied and reported (Coyle, 1965; Moran, 1962). However, Mercer (1979) and Schwartz (1962) concluded that the modification of the language of a test or other procedure tends to make the modified scores almost impossible to interpret using any existing normative framework.

Assessment procedures and tests will need to be modified so that exceptional children can be identified in relationship to the cultural and language experience they bring with them to the assessment situation (Pacheco, 1983). Assessment procedures presently used reinforce and legitimize a monocultural value system, since standardized assessment procedures are based on a statistical definition of what is normal. Even though a proportionate number of non-Anglo children may be in the norming sample, the largest subgroup has the greatest influence in establishing the behavioral norms. The other cultural groups which differ from the majority will therefore be defined as abnormal.

The System of Multicultural Pluralistic Assessment (SOMPA), provides an assessment model in which performance is interpreted in relation to a normative framework that was developed with regard to ethnic and socio-economic background of the subject being assessed (Mercer, 1979). The student's test performance is interpreted by looking at the

norms for children of similar backgrounds. Four sociocultural scales are considered in the adjustment of scores in order to eliminate sociocultural bias by comparing the subject to students who have had similar opportunities to learn the test content (Mercer and Ysseldyke, 1977). One criticism of this instrument is that even though the SOMPA may reduce the number of children classified as mentally retarded or learning disabled by adjusting scores upwards based on differential sociocultural norms, the reality is that minority students still must operate in an educational environment that is oriented towards middle-class norms (Cummins, 1984). Plata (1982) also argued that "the SOMPA does not predict how well the student will do in mainstreamed public school instructional programs" (p. 6). Another criticism centers on the fact that the results obtained provide few pedagogical intervention strategies that may be appropriate.

Dean (1979) suggests that, if possible, dual norm comparisons could be made. A non-Anglo's scores could be compared to the norms of the children from the dominant culture. A second evaluation and comparison could be made based on local norms obtained from children with his/her specific subculture background; this could only be done if local norms were available, but as Dean states, this is difficult to do due to the lack of published local norms for different groups. In conclusion Dean states that "any alternative that views clinical measures from a culture-bound

perspective runs the risk of labelling on the basis of culture rather than competence" (p. 14).

One consideration, as Samuda (1975) indicates, would be to depart from traditional testing and develop new measures consistent with the special language characteristics of minority individuals. These new measures would match the language style and vernacular of the individual to enable the subject to respond without anxiety or emotional threat. In a study conducted by Vazquez, Nuttall Associates, Inc. (1984), regarding assessment of limited English proficient LEP populations, it was reported that the authors of the study state that many local education agencies have relied on non-verbal tests and translations of tests in their assessment procedures for minority language students.

Many special educators and psychologists, aware of the shortcomings of many tests when used with minority populations, have used some of the more traditional non-language measures but even then with some reservations (Laosa, 1977). The desire to create tests that would be "culture-free" and "culture-fair" for culturally different subjects has been an ongoing process, but Mercer (1979), Wesman (1968), and Williams (1975) have stated that "the search for a culture-free test is sheer nonsense" (Mercer p. 144).

The importance of developing adequate psychoeducational assessment batteries for culturally, linguistically, and ethnically different students is clear. These measures of cognitive abilities and achievement are usually

norm-referenced measures that examine an individual's performance relative to the performance of other children like him/her. Many authors argue that localized norms of these measures may be the most appropriate standards by which to compare a student's performance (Oakland, 1977; Laosa, 1977; Mercer, 1979; and Ysseldyke, 1973, 1977). They point out that such norms based on the region, state, district, or community could be developed and enhance the meaningfulness of psycho-educational assessment instruments.

Description of Instruments

The Woodcock-Johnson Psycho-Educational Battery (Woodcock and Johnson, 1978) and the Wechsler Intelligence Scales for Children--Revised (Wechsler, 1974) are two English tests currently used in the field of Special Education in the United States to help determine the presence of a learning disability. These two instruments have recently been modified, adapted and normed in Spanish in Latin America. A brief description of the composition and the theoretical model of the Woodcock Psycho-Educational Battery in Spanish (BATERIA) and the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM) will provide background information with regard to what each instrument purports to measure and diagnose.

The Woodcock Psycho-Educational Battery in Spanish (Woodcock, 1982) consists of 17 subtests organized into two parts. Part One--The Tests of Cognitive Ability--contains

10 subtests which are a heterogeneous sampling of intellectual skills ranging from lower order perceptual processing to higher order verbal reasoning and comprehension skills. The 10 subtests provide cluster scores for seven aspects of cognitive functioning: Broad Cognitive Ability; Oral Language; Reasoning; Visual-Perceptual Speed; Reading Aptitude; Mathematics Aptitude; and Written Language Aptitude. The three aptitude clusters provide scores of expected achievement against which the subject's actual achievement may be compared. Part Two--The Tests of Achievement--contains 7 subtests that measure three aspects of scholastic achievement. These are reading, mathematics, and written language. The written language achievement cluster measures the subareas of spelling, capitalization, punctuation, and usage.

The BATERIA was designed and developed to provide information needed in educational decision-making. This psycho-educational assessment model provides information which can differentiate three types of discrepancies in an individual's performance. These are aptitude-achievement discrepancies, intra-cognitive discrepancies, and intra-achievement discrepancies. The model is based on a pragmatic "decision-making" model (Woodcock, 1984) which uses a decision-based test design strategy. Two important questions this design addresses are: 1) What are the most important decisions for which practitioners need psychometric information? and 2) What design or set of tests will provide

the necessary psychometric information in making those decisions?

The Wechsler Intelligence Scales for Children--Revised Mexican version (WISC-RM) contains 12 subtests which are divided into two different scales: Verbal and Performance. The Verbal scale of the test contains 6 subtests and provides information about a subject's verbal comprehension skills. The Performance scale contains 6 subtests which measure perceptual organization skills. Scores obtained from the two scales are a Verbal IQ score, a Performance IQ score, and a Full Scale IQ score, which combines the Verbal and Performance IQ scores into one global score.

The WISC-RM is based on the concept of global intelligence. Wechsler saw intelligence as a multi-determined and multi-faceted entity rather than a single, independent, or unique trait. He avoided singling out one trait, such as verbal comprehension, as more important or crucial than another, but rather conceived of intelligence as "something that is inferred from the way these abilities are manifested under different conditions and circumstances" (Wechsler, 1974, p. 5). Wechsler sought to measure intelligence through as many different "languages" as possible with the underlying construct being that intelligence is not a single unique trait but a composite or global entity.

Comparative Research

A large portion of the research comparing the Woodcock-Johnson Psycho-Educational Battery (WJPEB) in English and the Wechsler Intelligence Scale for Children--Revised (WISC-R) in English have addressed the mean standard score differences on the Broad Cognitive Ability cluster of the Woodcock-Johnson Tests of Cognitive Ability (WJTCA) and the Full Scale scores on the WISC-R. Results of the studies are mixed; some studies reported large mean score differences with the Woodcock-Johnson Tests of Cognitive Ability (WJTCA) being significantly lower (Reeve, Hall, and Zakreski, 1979), and others reporting little or no difference (McGrew, 1983). Since both tests have their parallel editions in Spanish, and, with the exception of modifications discussed previously, the tests are identical conceptual translations based on the original English editions, it is possible to generalize the discussion of issues from the English editions to the Spanish editions.

In one of the first studies reported (Reeve, Hall, and Zakreski, 1979), when comparing the two instruments the investigators noted a significant mean score discrepancy of 12 points on the Full Scale WJTCA and the Full Scale IQ score

from the WISC-R. In their discussion it was pointed out that, based on the results of this study, a child who would be classified as learning disabled based on the WISC-R score would be eligible for services in a class for educable mentally retarded subjects based on the WJTCA results. Finding the correlation between the two measures to be .79, the authors provided an explanation of the difference by stating that the WJTCA may tap an area of cognitive functioning in which children with learning problems seem to have greater difficulty than children included in the standardization sample of the WJTCA. The also explained that the norms for the WJTCA may be in error. The main concern of the authors of this study was that

> "the labeling and placement of children . . . could well be a function of the assessment instrument used rather than the kind and quality of performance assessed" (Reeve et al., 1979, p. 68).

In another early study with 50 learning disabled children (Ysseldyke, Shinn, and Epps, 1981), a discrepancy was reported with the total sample scoring about 7.7 points lower on the WJTCA Broad Cognitive Scale than on the WISC-R Full Scale. In referring to the opinion in Reeve et al. (1979) that the WJTCA taps areas in which learning disabled students perform more poorly, the authors in this study equated these areas of weakness in these subjects to the nature of the WJTCA which they believe to be highly achievement oriented rather than primarily an expected measure of achievement. In further discussion they likened the cognitive

subtests as being product-dominated (Newland, 1971) and a measure of crystallized intelligence (Cattell, 1963). They suggest that these tests sample primarily what has been learned rather than the processes necessary to the acquisition of academic skills which Cattell labeled as fluid intelligence.

There have been a number of other studies performed, both published and unpublished, comparing the two instruments. Woodcock (1984a) summarized the current research comparing the mean scores of the WISC-R and the WJTCA by analyzing each study, type of sample and data bias. Table 1 refers to a summary of 20 studies carried out.

Table 1. Median Difference as a Function of Type of Sample

Type of Sample	No. of Studies	WISC-R/WJTCA Difference
Normals; Gifted	4	1.2
Referrals; Low Achievers	9	4.3
LD (one data bias)	3	7.7
LD (two data biases)	3	11.3
EMR	1	11.7

In the 20 studies examined comparing the mean score difference there was a range of differences from -0.3 to 12.0 standard score points. The mean score difference was

5.8 points with lower mean scores for the WJTCA. In the normal and gifted samples there was a mean score difference from 1.1 to 2.9 standard score points, with a median difference of 1.2 points. In the eight referral samples and one low achieving sample differences ranged from -0.3 to 9.1 standard points, with a median difference of 4.3 points. In summarizing the results from the 20 studies, 2 trends were observed in the differences obtained between the 2 measures. The first was that in the normal samples there were fewer differences in scores, and second, as the studies moved to the more handicapped samples the differences became greater. The data showed that with normal samples of subjects mean score differences between the WISC-R and the WJTCA was about 1 point. In the low-achieving and referral samples the mean score difference was between 4 and 5 points.

In discussion of why there are differences, five different factors have been suggested in the literature: 1) An error in calculating the norms for the WJTCA (Reeve et al., 1979; Ysseldyke et al., 1981a); 2) A bias in the procedure for selecting the norming subjects (Cummings and Moscato, 1984a); 3) An instrument bias in the data causing the mean WISC-R scores to be too high (Woodcock, 1980); 4) A score bias in the data causing the mean WISC-R scores to be even higher (Woodcock, 1980); and 5) A sample bias resulting in lower scores for the WJTCA relative to the

WISC-R, caused by the difference in the correlation of the two tests with success in school (Woodcock, 1980).

In further explanation of the problem related to interpretation of the results of these earlier studies (Woodcock, 1984a), four biases were identified and described in the design and methodology of the investigations which influenced the outcome of these studies:

1) Instrument bias. "One of the two instruments to be compared had been used for the selection or subclassification of the sample" (Woodcock, 1984a). In some studies, WISC-R scores were obtained from learning disabled (LD) students who had already been placed in a special LD program. Since these students previously placed were selected on the basis of their discrepancy between their WISC-R score and their low achievement score, the mean scores were higher. The students selected would tend to have higher WISC-R mean scores since LD classification typically is based upon an ability-achievement discrepancy. Students with lower WISC-R scores were not as likely to be placed in LD programs based on the same criteria for LD eligibility. Therefore, groups of students selected on this basis would tend to have higher mean scores to become eligible for LD classification. It was recommended that to carry out research on LD samples that "referrals," not "selectees," should make up the groups to be sampled.

2) Score bias. "The selection test had not been readministered to obtain new scores if it was to be a part of

the comparison" (Woodcock, 1984a). This is an extension of the instrument bias and states that the selection test used to classify the selected sample needs to be readministered before the comparative analysis is carried out since the WISC-R scores that survived the selection procedure were based "on the most fortuitous score combination of true ability and positive error" (Woodcock, 1984, p. 343).

3) Sample bias. "Selection of the sample had a differential effect on the statistics being compared as a function of each test's correlation with selection criteria" (Woodcock, 1984a). This occurs when a sample is chosen, but not based on random sampling from the general population. What must be examined is the correlation between the tests being used and the selection criteria for the sample. If the correlation of one of the instruments being used is higher than the selection criteria then this will become evident with the mean score being a greater distance from the general mean.

4) Interpretation/reporting bias. "Lack of objectivity in evaluations and reporting these results" (Woodcock, 1984a). This occurs due to the lack of objectivity in the evaluation and reporting of results obtained from studies carried out. This bias has operated in past studies which conclude that the difference in mean scores obtained between the WISC-R and the WJTCA is due to the WJTCA scores being "too low" and the WISC-R scores as "just right." Other considerations, as Woodcock (1984) suggests, may be that the

observed differences are real and can be expected or that the WISC-R scores are too high.

In discussing the possible reasons for the discrepancy issue, Reeve et al. (1979), as stated previously, explained that the WJTCA may be measuring an area of cognition which learning disabled children may have more difficulty with than the children reported in the WJTCA standardization sample. Ysseldyke (1982) also suggested that higher correlations with academic achievement were due to "achievement content" in the subtests. Cummings and Moscato (1984a) stated that

> although the Broad Cognitive Scale appears to exhibit a high degree of shared variance with various measures of academic achievement, an analysis of the subtests' content clearly reveals that those subtests which comprise the scale are not part of the domain traditionally considered academic achievement. (p. 37)

Estabrook (1983), in a canonical correlation study with 152 referral subjects, concluded that a large amount of variance was not shared between the 2 tests and that the difference is "most likely . . . attributable to unique factor structures" (p. 35).

It is significant to note that in a review of the studies done when correlations were calculated between the two instruments, the WJTCA correlated higher than the WISC-R in all areas of achievement with the exception of knowledge. When considering the differences observed, Woodcock (1984) concluded that the major factor causing the differences would appear to be the WJTCA's higher correlation with

achievement. Since the WJTCA model was specifically designed as an educational "decision-making" model highly related to school success, the greater mean score differences for low-achieving and referral subjects could be attributed to these higher correlations with achievement. These greater differences could be seen not as a weakness of the WJTCA, but rather as a more discriminating instrument with regard to measuring skills which are most liked those performed in school.

In further discussion that the WJTCA is primarily a measure of achievement, Woodcock (1984b) discusses the importance of using the four scholastic aptitude cluster scores specifically designed to determine an aptitude-achievement discrepancy rather than the Broad Cognitive Ability score. Cummins and Moscato (1984a) reported correlations that were comparable between the Broad Cognitive Ability scores and the four areas of achievement and the four scholastic aptitude clusters with their corresponding achievement measures. Based on the reported correlations, it is evident that the predictive ability of the scholastic aptitude cluster is equal to the single Broad Cognitive Ability score and more advantageous for four reasons (Woodcock, 1984b): 1) they are not confounded with achievement; 2) they can reduce testing time; 3) they provide differential expectancy information based on norms; and 4) they provide higher clinical validity.

Summary

This study will provide a comparative and descriptive analysis of two psychological and educational instruments recently developed in Spanish. Considering the political and social ramifications of the overselection of children into classes for the mildly retarded (Mercer, 1979), it is imperative to determine, by data collection and research, if these reported differences are also observed in the Spanish editions. Replication of the previous studies in the Spanish editions will provide information with regard to the educational significance and impact these instruments have on normal as compared to referral samples of subjects. The importance of this research was magnified by the Reeve et al. (1979) statement that the

> labeling and placement of children viewed as having learning problems could well be a function of the assessment instrument used rather than the kind and quality of performance assessed. (p. 6)

CHAPTER 3

METHODOLOGY

Introduction

This study compared the performance of normal and referral students using the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM) and the Woodcock Psycho-Educational Battery in Spanish (BATERIA). This chapter describes the research design and methodology used. It includes the design for the study, the sample, selection of the sample, selection and training of examiners, the data gathering procedures, description of instruments, and the statistical analysis used in the study.

Design of the Study

The present study had five objectives:

1) To analyze the mean score difference between the Broad Cognitive Ability score on the BATERIA and the Full Scale score on the WISC-RM as a function of type of subject (normal or referral).

2) To analyze the mean score difference in performance on the Special Cognitive Ability clusters from the BATERIA as a function of type of subject (normal or referral).

3) To analyze the mean score differences when comparing the three Scholastic Aptitude clusters from the BATERIA as a function of type of subject (normal or referral).

4) To analyze the mean score differences when comparing the three Scholastic Achievement clusters from the BATERIA as a function of type of subject (normal or referral).

5) To analyze the mean score differences when comparing the Verbal and Performance scale scores on the WISC-RM as a function of type of subject (normal or referral).

Sample

The subjects for this study were drawn from public and private schools in Mexico City during the 1984-1985 academic school year. All subjects were enrolled in the fifth grade and were tested during January, February, and March of 1985.

Ninety students were selected, with 65 students drawn from 3 public schools and 25 students from 3 private schools. The sample was comprised of 53 females and 37 males ranging in age from 9-11 to 15-1. Twenty students had been retained 1 year, 9 students had been retained 2 years, and 3 students had been retained 3 years.

Forty-three children were included in the normal sample with 32 selected from public schools and 11 from private schools. Forty-seven children were included in the referral group with 33 subjects coming from public schools and 14 from private schools.

Selection of the Sample

Subjects were assigned to the two groups by the following procedure:

1) Referrals. The teacher of each selected fifth grade class was asked to rank order the four students whom he/she perceived to be doing the poorest in one or more of the scholastic areas of reading, mathematics, or written language.

2) Normals. Using a table of random numbers, three normal subjects from each fifth grade class were selected from the class list.

The procedure mentioned above allows for the selection of a subject both as a normal and a referral. If a referral subject were selected as a normal, this is the procedure that was followed: Within each class, referral subjects were selected before normals. Referral subjects subsequently selected as normals were used as normals and were not included in the referral group. Three of the four students identified as the poorest achievers in each fifth grade class were used as the referral subjects. If, in the random selection of normals, one of the referral subjects was selected, the one remaining referral subject was used to fill the quota of the three referrals. If more than one of

the referral students was selected for the normal group, the fourth referral student from another class was used to fill the quota.

Selection and Training of Examiners

Two examiners and the student investigator tested all subjects. The BATERIA was administered by one examiner, a psychologist who was trained and supervised in the administration of the test. The WISC-RM was administered by the other examiner, a student with a master's degree in psychology with prior training in the administration of the test. Both examiners were recommended by the director of the Orientacion Infantil de Rehabilitacion Infantil, an educational evaluation clinic in Mexico City, who had previous knowledge of their skills in the administration of psychological and educational tests. The examiners were monitored frequently during the data gathering to insure that standardized procedures were maintained.

Data Gathering Procedures

The directors of selected schools were contacted and informed of the purpose of the study. The director of each school who participated in the study introduced the examiners to the fifth grade classroom teacher in his/her respective school. The fifth grade teachers were then instructed in the student selection procedure. From a class list, four students (referrals) were selected by the

teacher based on the criteria previously established. These four students selected were listed numerically according to the degree of difficulty they were experiencing in school. Subsequently, three students were randomly chosen (normals) from the class list by using a table of random numbers.

Students selected for the study were administered the BATERIA and the WISC-RM on two consecutive days. Administration time for each test was approximately one and one-half hours with a total testing time of about three to three and one-half hours per subject. These tests were administered in a counterbalanced order with half of each group taking the WISC-RM first and the other half the BATERIA first.

Each subject was administered 17 subtests from the BATERIA and 10 subtests from the WISC-RM. Testing sessions were conducted in empty classrooms or other rooms suitable for individual testing.

Description of Instruments

Within the past few years both the WJPEB and the WISC-R have been developed in Spanish to meet the psycho-educational assessment needs of professionals who seek cognitive functioning and academic achievement levels regarding subjects whose dominant language is Spanish.

The Woodcock Psycho-Educational Battery in Spanish (BATERIA) was validated on 802 subjects both male and female, ages 5-19 from typically urban areas of the Spanish

speaking world: Madrid, Spain; San Jose, Costa Rica; Guadalajara, Mexico; Chihuahua, Mexico; Lima, Peru; and Ponce, Puerto Rico. A representative cross-section of both public and private schools in each community was made by consulting local professionals and school officials as well as the National Ministry of Education in each of these countries. The composite sample was taken primarily from kindergarten and grades 1, 3, 5, 8 (2 secundaria), and grade 11 (2 bachillerato in Mexico, 3 BUP in Spain). The data for the norms were gathered from January to July, 1980.

The General Spanish Norms provided compare a subject's score for the BATERIA with the scores obtained by a representative sample of persons drawn from 5 regions of the Spanish speaking world. Both age and grade norms are provided for ages 5 to 19 and grades K to 12.9. Norms for all parts of the BATERIA are based on data gathered from the same sample of subjects, thus permitting direct comparison among the various measures of intelligence and achievement. The BATERIA in Spanish includes the following components. Part One--The Tests of Cognitive Ability--consists of 10 subtests: Picture Vocabulary, Spatial Relations, Visual-Auditory Learning, Quantitative Concepts, Visual-Matching, Antonyms-Synonyms, Analysis-Synthesis, Numbers Reversed, Concept Formation, and Analogies. Raw scores are obtained for each subtest and combined to obtain the following cluster scores: Broad Cognitive Ability or Full Scale Score; the three Special Cognitive Ability clusters,

Oral Language, Reasoning, and Visual-Perceptual Speed; and the three Aptitude cluster scores, Reading, Mathematics, and Written Language.

The scholastic aptitude clusters provide a measure of a student's expected achievement in three areas: Reading, Mathematics, and Written Language. The clusters were developed using a step-wise multiple regression procedure. Subtests with related achievement were excluded and those subtests that best predicted achievement in the relevant areas were used.

Part Two--The Tests of Achievement--contain seven subtests which make up the three achievement clusters. The Reading clusters contain Letter-Word Identification, Word Attack, and Passage Comprehension. The Mathematics cluster contains Calculation and Applied Problems. The Written Language cluster contains Dictation and Proofing. The derived weights for the subtests are equal.

Score from the BATERIA may be converted into many indices of performance: part scores, grade score equivalents, cluster scores, cluster difference scores, percentile rank scores, and standard scores. A Relative Performance Index provides an estimation of a student's performance on tasks similar to the one tested. A functioning level is then

obtained which provides a verbal interpretation of the subject's performance.

The Wechsler Intelligence Scale for Children--Revised Mexican version is a test of general intelligence. The WISC-RM was published in 1983 and was based on a selection of 1100 public school subjects both male and female in Mexico City, Mexico. The development of the WISC-RM was done in cooperation with the Secretary of Public Education in Mexico City who allowed the selection of the subjects from the elementary and secondary schools in Mexico City. The age range of the test covers students from 6 to 16 years of age. The purpose of the standardization, which began in 1979, was to provide new norms gathered in Mexico City that could be used to compare Mexican children with other children from the same sociological context.

The new adaptation of the WISC-R to Spanish includes three changes from the Standard WISC-R currently used in the United States. Some items on the Verbal scale, specifically in the Information, Comprehension, and Vocabulary subtests, have been modified to be culturally relative. The second change is the exclusion of the final three items on the Picture Completion subtest of the Performance scale. The third change is the reordering of items on the Verbal scale according to the difficulty of the item based on the new standardization carried out in Mexico City.

The WISC-RM includes 12 subtests and provides up to 3 scores based on unweighted composites of subtests. The 12

subtests have been categorized under either the Verbal or Performance Scale. The Verbal Scale includes the subtests of Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span (optional). The Performance Scale includes the subtests of Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes (optional).

Raw scores are obtained for each subtest and then converted to scaled scores that are appropriate for the age of the student. The scaled scores have a normal distribution with a mean of 10 and a standard deviation of 3. The scaled scores of the Verbal and Performance subtests are added to provide Verbal, Performance, and Full Scale scores. These scaled scores are converted to standard scores or intelligence quotients with a mean of 100 and a standard deviation of 15.

Validity and Reliability. As reported, both of these instruments in their English editions have adequate norms, reliability, and validity (Thurlow and Ysseldyke, 1979). WISC-R test-retest reliabilities were reported with all subtests of the test ranging from .63 to .95 (Wechsler, 1974). Three concurrent validity studies done with the WPPSI, WAIS, and the Stanford-Binet indicate that there is satisfactory validity as well (Salvia and Ysseldyke, 1981).

The Woodcock Johnson Psycho-Educational Battery split-half reliabilities for subtest scores and cluster scores are reported in the appendices to the technical manual

(Woodcock, 1978). For subtest scores, the median reliability coefficients exceed .80 except for Punctuation and Capitalization (.78), and Visual Matching (.65). The cluster scores median reliabilities exceed .85 except for the Perceptual Speed cluster (.70). Ten studies are reported in the technical manual providing evidence for the validity of the battery (Salvia and Ysseldyke, 1981).

In their respective Spanish editions no concurrent validity information with other tests is available. Reliability information for the WISC-RM (Wechsler, 1984) reports mean split-half reliabilities ranging from .65 to.82. Split-half median reliabilities of the three scale scores (Verbal, Performance, Full Scale) was .89, .90, and .94 respectively.

Reliabilities reported on the 17 BATERIA subtest scores ranged from .73 to .94. On the 12 cluster scores reliabilities reported ranged from .79 to .95 (Woodcock, 1985).

Data Analyses

An analysis of variance for repeated measures design was used to determine significant differences in performance as a function of type of subject (normal, referral) between the Broad Cognitive Ability Score on the BATERIA and the Full Scale Score on the WISC-RM. The same type of analysis was employed to determine significant differences in performance as a function of type of subject (normal, referral) on the Special Cognitive Ability clusters, Scholastic

Aptitude clusters, Scholastic Achievement clusters, and the Verbal and Performance scales of the WISC-RM.

When comparing the Full Scale scores of the BATERIA and the three scale scores of the WISC-RM, raw scores were converted into standard scores with a mean of 100 and a standard deviation of 15. When significant differences were found among the cluster scores, the Newman-Kuels multiple comparison test was used to determine the differences. The Pearson product moment correlation was used to examine the correlations between the BATERIA cluster scores and the three major IQ scale scores of the WISC-RM. These data were analyzed, using the Biomedical Computer Programs (BMDPV2, Dixon, 1975) and the Systat Program (Wilkenson, 1985).

Limitations of the Study

This study was carried out only in Mexico City. Furthermore, the 90 subjects selected for this study from both public and private schools represents only a microcosm of the student population in Mexico City. Therefore, generalizations from the findings of this study to Hispanic students in other parts of Mexico, other parts of the Spanish-speaking world, or to students in the United States is limited.

The selection procedures used in identifying the low-achieving students in this study was done by each fifth grade teacher based on their own internal norms and experience in working with the selected child during the present

school year. Findings from this study and other studies have shown that there may be some question about the validity and accuracy with which regular classroom teachers make these referrals.

Although low-achieving students were selected based on poor achievement in mathematics, reading, or written language, no attempt was made to divide these students into subcategories for analysis.

CHAPTER 4

ANALYSIS OF DATA

The purpose of this chapter is to describe and compare the results of the study for the subjects classified as normals and the subjects classified as referrals. The following research topics include: a) comparability of full scale scores, b) performance on the Cognitive clusters, c) performance on the Aptitude clusters, d) performance on the Achievement clusters, and e) Verbal and Performance Scale score differences.

Analyses performed

1) The first analysis (Broad Intelligence) examines the mean difference in performance on the BATERIA and WISC-RM as a function of the type of subject.

2) The second analysis (Special Cognitive Abilities) examines the mean difference in performance on the Special Cognitive Ability clusters of the BATERIA as a function of the type of subject.

3) The third analysis (Scholastic Aptitudes) examines the mean difference in performance on the Scholastic Aptitude clusters on the BATERIA as a function of the type of subject.

4) The fourth analysis (Scholastic Achievement) examines the mean difference among the Scholastic Achievement clusters as a function of the type of subject.

5) The fifth analysis (Verbal and Performance Scales) examines the mean difference between the Verbal and Performance Scales of the WISC-RM as a function of the type of subject.

The statistical procedure utilized was an ANOVA repeated measures design. The .05 alpha level was used as the criterion for rejecting the null hypothesis. When a significant *F* ratio was obtained, post-hoc comparisons were conducted using the Newman-Keuls multiple comparison procedure.

Findings

Broad Intelligence. This analysis compared mean Full Scale scores between the BATERIA and WISC-RM for subjects classified as normals and for subjects classified as referrals. For subjects classified as normals, the correlation between the WISC-RM Full Scale and the BATERIA Broad Cognitive Ability was .86. For subjects classified as referrals, the Full Scale correlation was .84. (Appendix A provides correlations between major test scores of the BATERIA and the WISC-RM for each sample.)

The results of the analysis of variance comparing normals and referrals are presented in Table 2. Based on the obtained *F* ratio in Table 2, the main effect difference between groups was significant at the .01 level ($F = 14.58$).

Students from the normal group scored significantly higher than students from the referral group.

Based on the obtained *F* ratio in Table 2 there was no significant difference between tests. Of interest was a trend toward interaction, though not significant, which indicated that the normal group scored relatively higher than the referral group on the Broad Cognitive Ability cluster of the BATERIA than on the Full Scale score of the WISC-RM.

Table 2. ANOVA Tests by Group for BATERIA Broad Cognitive Ability Full Scale vs.WISC-RM Full Scale

Source of Variance	df	Mean Square	F	P
BETWEEN SUBJECTS:				
Normals vs. Referrals (A)	1	24.42	14.58	<.01
$Error_W$	88	1.67	-	-
WITHIN SUBJECTS:				
BATERIA vs. WISC-RM (B)	1	.00	.04	<.01
Test x Classification (A x B)	1	.41	3.39	N.S.
$Error_W$	88	.12	-	-

Mean score differences for the BATERIA and WISC-RM Full Scale measures were compared further. Results are reported in Table 3. The initial finding from the comparison of the Full Scale intelligence scores from the WISC-RM and BATERIA is of particular interest. For each of the two experimental groups, a difference of 6 to 8 points was noted, with the

WISC-RM higher. An examination of information on the development and standardization in the manuals for the two tests indicates a marked difference in the location, distribution, and type of subjects in the norming sample.

Table 3. Full Scale Cognitive Ability Scores Before Standardizing

Type of School	n	BATERIA	WISC-RM	Difference
Normals:				
Combined	43	M 100.5	M 106.9	6.4
		SD 13.6	SD 18.0	4.4
Public School	32	M 97.0	M 102.1	5.1
		SD 13.9	SD 17.3	3.4
Private School	11	M 110.8	M 120.6	9.8
		SD 4.6	SD 11.9	7.3
Referrals:				
Combined	47	M 86.3	M 94.2	7.9
		SD 18.0	SD 19.6	1.6
Public School	33	M 80.0	M 85.4	5.4
		SD 8.9	SD 9.2	0.3
Private School	14	M 93.8	M 114.8	21.0
		SD 9.7	SD 10.7	1.0

The WISC-RM sample was drawn exclusively from students attending public school in Mexico City. The BATERIA, in contrast, was normed on a stratified random sample of students attending both public and private schools in five Spanish-speaking countries. The SES characteristics of the public school population in Mexico City are markedly different from those same characteristics for the private school population. The WISC-RM norming sample is expected to be both

lower and more restrictive than the BATERIA sample because of these differences in the norming sample. It follows, then, that mean scores and standard deviations based on the WISC-RM norm tables will yield inflated scores on a representative sample of the general population, as well as higher scores at the top and lower scores at the bottom of the scale due to restricted range.

A standardizing procedure was performed on the Full Scale scores on the two tests for the normal group to remove the confounding effects of the known norming sample differences on the two measures. The mean for this scale is 100 with a standard deviation of 15. The referral group Full Scale scores for the two measures were then placed on the same scale as the normal group by using the following transformation equation.

$$Y = \left[\left(\frac{X_r - \bar{X}_n}{SD_n} \right) 15 \right] + 100$$

Whereas, Y is a score on the derived scale based on the normal sample,

X_r is the score made by a referral subject,

$\bar{X}_n$ is the mean score for the normal subjects,

SD_n is the standard deviation of the normal sample,

15 sets the standard deviation of the new scale to 15,

100 is a constant that sets the mean of the new scale to 100.

The purpose for this standardizing procedure is to examine the differences between the two tests with referral groups, when differences for the normal group are held at zero. This allows for evaluation of the differences in scores after controlling for the confounding effects of different norms. Table 4 reports means and standard deviations for the transformed scores of the normal and referral groups. A study of the results shows that the referral group differs from the normals with a lower mean on both measures. A comparison of means for the referral group indicates a discrepancy between the WISC-RM and BATERIA of 5.5 standard score points.

Table 4. Full Scale Cognitive Ability Scores After Standardizing

Sample	n	BATERIA	WISC-RM	Difference
Normals	43	M 100.0 SD 15.0	M 100.0 SD 15.0	0.0 0.0
Referrals	47	M 84.3 SD 19.9	M 89.8 SD 16.4	5.5 3.5

Table 5 reports BATERIA and WISC-RM Full Scale score correlations for the normal sample with the three Scholastic Achievement clusters from the BATERIA (Reading, Written Language, and Mathematics). The highest correlations for the Broad Cognitive Ability (BATERIA) were with Written Language

(.77) and Reading (.74). Broad Cognitive Ability and WISC-RM Full Scale scores both correlated .72 with Mathematics achievement. Lowest correlations for all achievement tests were found with the WISC-RM Performance scale.

Table 5. BATERIA and WISC-RM Correlations with BATERIA Scholastic Achievement Clusters Normal Sample (n = 43)

	BATERIA Achievement Measures		
Intelligence Measures	Reading	Written Language	Math
BATERIA			
Broad Cognitive Ability	.74	.77	.72
WISC-RM			
Full Scale	.68	.65	.72
Verbal Scale	.69	.65	.57
Performance Scale	.54	.54	.48

Special Cognitive Ability Clusters. An analysis of variance for repeated measures was computed to determine the significance of differences in performance on the Special Cognitive Ability clusters of the BATERIA as a function of type of subject (normal or referral). The results, as shown in Table 6, based on obtained F ratio revealed that there was a significant difference across tests for normals and referrals at the $p < .01$ level ($F = 20.09$); the normals scored significantly higher than the referrals on the three Special Cognitive Ability clusters.

There was a significant main effect difference at the .01 level among the Special Cognitive Ability cluster scores (F = 8.09). The Newman-Keuls multiple comparison test was performed to determine where the differences were. After the post-hoc comparison, it was found that the Visual-Perceptual Speed cluster was significantly higher than the Reasoning and Oral Language clusters. There was no statistically significant difference between the Reasoning and Oral Language clusters.

Table 6. ANOVA Tests by Group for BATERIA Oral Language, Reasoning, and Visual-Perceptual Speed Clusters

Source of Variance	df	Mean Square	F	P
BETWEEN SUBJECTS:				
Normals vs. Referrals (A)	1	6530.93	20.09	<.01
$Error_W$	88	325.07	-	-
WITHIN SUBJECTS:				
Test (B)	2	778.05	8.09	<.01
Test x Classification (A x B)	2	161.97	1.69	N.S.
$Error_W$	176	96.12	-	-

Table 7 reports mean W scores for normal and referral samples on the Special Cognitive Ability and Scholastic Aptitude clusters from the BATERIA. Results indicate that the normal sample scored at or above the mean for their grade placement (5.5) on all measures. The referral group scored

below the mean for their grade placement on all measures. This suggests that scores for the normal sample are closer to those of the norming group than are the scores for those of the referral group.

Table 7. Mean W Scores on BATERIA Special Cognitive Ability and Scholastic Aptitude Tests for Normal and Referral Samples

	Mean W Score		
BATERIA	Normals n = 43	Mean For Grade 5.5	Referrals n = 47
Special Cognitive Abilities			
Oral Language	503.7	503.0	495.0
Reasoning	504.9	503.0	496.2
Visual-Perceptual Speed	511.6	504.0	498.7
Scholastic Aptitude			
Reading	504.6	504.0	494.2
Written Language	503.3	502.0	493.3
Mathematics	506.8	505.0	496.5

Scholastic Aptitude Clusters. Analysis of variance for repeated measures was computed to determine differences in performance on the three Scholastic Aptitude clusters of the BATERIA as a function of the type of subject (normal or referral). The results shown in Table 8, based on an obtained F ratio, revealed there was a significant main effect for type of subject at the .01 level (F= 21.37). Normals scored

significantly higher than the referrals on the three Scholastic Aptitude clusters.

There was a significant main effect at the .01 level for the three Scholastic Aptitude scores averaged for both groups (F = 27.56). A Newman-Keuls multiple comparison test was performed to determine where the differences were. All three Scholastic Aptitude cluster scores were significantly different from one another. Students scored highest on the Math Aptitude cluster followed by the Reading and Written Language clusters. Further analysis of the mean W scores in Table 7 exhibit the differences for normals and referrals on the three clusters. Normals performed about 10 W score points higher on all clusters.

Table 8. ANOVA Tests by Group for BATERIA Reading, Math, and Written Language Aptitude Clusters

Source of Variance	df	Mean Square	F	P
BETWEEN SUBJECTS:				
Normals vs. Referrals (A)	1	7088.48	21.37	<.01
$Error_W$	88	331.71	-	-
WITHIN SUBJECTS:				
Test (B)	2	264.64	27.56	<.01
Test x Classification (A x B)	2	.88	.09	N.S.
$Error_W$	176	9.60	-	-

Scholastic Achievement Clusters. Analysis of variance for repeated measures were computed to determine differences

in performance on the Scholastic Achievement clusters of the BATERIA as a function of the type of subject (normal or referral). The results shown in Table 9, based on obtained F ratios, reveal that there was a significant main effect difference across tests for normals and referrals at the .01 level (F = 33.17). The normal group scored significantly higher than the referral group on all three Scholastic Achievement clusters.

There was a significant main effect difference at the .01 level among the three Scholastic Achievement cluster scores on the BATERIA (F = 38.29). A Newman-Keuls multiple comparison test was performed and all three Scholastic Achievement cluster scores were found to be statistically different from one another. Mathematics achievement was the highest, followed by Reading and Written Language.

Table 9. ANOVA Tests by Group for BATERIA Reading, Math and Written Language Achievement Clusters

Source of Variance	df	Mean Square	F	P
BETWEEN SUBJECTS:				
Normals vs. Referrals (A)	1	7605.68	33.17	<.01
$Error_W$	88	229.27	-	-
WITHIN SUBJECTS:				
Test (B)	2	1824.90	38.29	<.01
Test x Classification (A x B)	2	12.20	2.54	N.S.
$Error_W$	176	47.66	-	-

Table 10 reports mean W score results of the Scholastic Achievement tests for normals and referrals. The scores indicate that there was about a 13 point difference on Written Language, an 11 point difference for Mathematics, and about an 8 point difference for Reading between the two groups.

Table 10. Mean W Scores for BATERIA Scholastic Achievement Clusters for Normal and Referral Samples

	Mean W Score		
BATERIA-Achievement	Normals n = 43	Mean for Grade 5.5	Referrals n = 47
Reading	506.7	503.0	498.6
Written Language	506.1	504.0	493.5
Mathematics	514.2	503.0	503.0

Note: Mean for BATERIA norming sample

WISC-RM: Verbal and Performance Scales. Analysis of variance for repeated measures were computed to determine differences in performance on the Verbal and Performance scales of the WISC-RM as a function of the type of subject (normal or referral). The results shown in Table 11, based on obtained F ratios, reveal that there was a significant main effect for type of subject ($F = 10.08$, $p < .01$) and a significant main effect for type of test ($F = 13.52$, $p < .01$). A significant interaction effect ($F = 7.69$, $p < .05$) was obtained. The significant interaction is derived from the results that indicate that for normals there

was no statistically significant difference between the Verbal and Performance scales for normals (x = 105.2 vs. 106.7), but for referral subjects there was a significant difference (x = 89.2 vs. 100.0).

Table 11. ANOVA Tests by Group for WISC-RM Verbal Scale vs. Performance Scale

Source of Variance	df	Mean Square	F	P
BETWEEN SUBJECTS:				
Normals vs. Referrals (A)	1	5749.12	10.08	<.01
$Error_W$	88	570.09	-	-
WITHIN SUBJECTS:				
Test (B)	1	1698.34	13.52	<.01
Test x Classification (A x B)	1	966.00	7.69	<.01
$Error_W$	88	125.61	-	-

Table 12 reports mean standard scores for both groups for the three scales of the WISC-RM. As observed, the normal group scored significantly higher on all three scales. The referral group scored 10.8 standard score points lower on the Verbal scale than on the Performance scale. The referral group also scored 15.9 standard score points lower than the normal group on the Verbal scale.

Table 12. Mean Standard Scores for Normals and Referrals on WISC-RM Scales

WISC-RM	Sample	
	Normal n = 43	Referral n = 47
Verbal	105.3	89.2
Performance	106.7	100.0
Full Scale	106.9	94.2

Scholastic Aptitude-Achievement Correlations. Table 13 reports correlations among the Aptitude and Achievement clusters from the BATERIA. Results indicate that for normals the best predictor of Reading, Written Language, and Mathematics achievement is the Full Scale Broad Cognitive Ability cluster. For the referral sample, the Reading and Mathematics Scholastic Aptitude clusters predicted as well or better than the Full Scale Broad Cognitive Ability cluster. This would seem to support the use of these Scholastic Aptitude clusters as predictors of achievement with the referral population.

Table 13. Correlations of BATERIA Scholastic Achievement Clusters with Respective Aptitude Clusters and Full Scale Broad Cognitive Ability for Normal and Referral Groups

	Actual Achievement					
	Reading		Written Language		Mathematics	
BATERIA Aptitude Measure	Normal n = 43	Refer. n = 47	Normal n = 43	Refer. n = 47	Normal n = 43	Refer. n = 47
Reading	.71	.78				
Written Language			.70	.78		
Mathematics					.55	.71
Full Scale Broad Cognitive Ability	.74	.71	.77	.81	.72	.71

Summary of Results

The results of this study may be summarized as follow:

1. Both normal and referral subjects scored higher on the Full Scale score of the WISC-RM than on the BATERIA. Normal subjects scored 6.4 standard score points higher and referral subjects scored 7.9 standard score points higher. These results imply a 1.5 standard score point greater difference between the two tests for the referral sample than for the normal sample. There is reason to question the validity of these observed differences based on the normative sample characteristics of the two tests. Because of the known differences in the selection of the norming sample for the two tests, the BATERIA and WISC-RM Full Scale scores for

the normal group were standardized in order to remove these confounding effects and set the mean for each test at 100 and the standard deviation of 15 for the normal group. The referral group Full Scale scores on both tests were then placed on the same scale using a transformation equation. This procedure in effect allowed for evaluation of the differences in the referral subjects' performance on the two tests when differences for the confounding effects of different norming samples were controlled. After placing the referral subjects on the same scale as normals, the referral subjects scored 5.5 standard score points higher on the WISC-RM than on the BATERIA. This indicates that the referrals had a greater difference, of 5.5 standard score points, between the two tests than did the normals. This is in contrast to only a 1.5 standard score point difference before controlling for norming sample differences by standardizing the scores.

2. Normal subjects scored significantly higher than the referrals on the three Special Cognitive Ability clusters. There was also a significant difference found among the three Special Cognitive Ability clusters. The Visual-Perceptual Speed cluster was significantly higher than the Reasoning and Oral Language clusters across all subjects.

3. Normal subjects scored significantly higher than referrals on the three Scholastic Aptitude clusters. In

addition, all three Scholastic Aptitude clusters were significantly different from one another with Mathematics the highest, followed by Reading and Written Language.

4. Normal subjects scored significantly higher than referrals on the three Scholastic Achievement clusters. There was a significant difference found between each of the Achievement cluster scores across all subjects with Math being the highest followed by Reading and Written Language.

5. Referral subjects scored significantly higher on the Performance Scale than on the Verbal Scale of the WISC-RM. The difference between the Verbal and Performance Scale scores was not significant for the normal subjects.

CHAPTER 5

SUMMARY AND DISCUSSION

The purpose of this chapter is to provide a summary of the study, a discussion of the results including implications for the field of bilingual special education, and suggestions for future research.

Purpose of the Study

The major purpose of this study was to analyze and compare test scores on the Woodcock Psycho-Educational Battery in Spanish (BATERIA) and the Wechsler Intelligence Scale for Children--Revised Mexican version (WISC-RM) for a sample of normal subjects and a sample of referral subjects.

Subjects and Design

Forty-three 5th grade subjects classified as normals and forty-seven 5th grade subjects classified as referrals were selected to participate in this study. Subjects were administered the BATERIA and the WISC-RM. All testing was performed within a three month period. Tests were counterbalanced to avoid an order effect. Both the BATERIA and WISC-RM protocols were scored by the investigator.

Statistical analyses were performed at the University Computer Center using the BMDP2V program. Repeated measures ANOVAs were used to answer the research questions. All descriptive and comparative data were compiled and analyzed by this investigator.

Several important implications can be drawn from the results of this study. Discussion of the findings and implications will be presented topically, following the order of the research questions.

Discussion of the Findings

Broad Intelligence. Significant differences between BATERIA and WISC-RM Full Scale mean scores were observed in both the normal and referral groups. A discrepancy of 6.4 standard score points difference was noted between the normal group and a 7.9 standard score point difference for the referral group. These results are in contrast to the U.S. findings where there was only a 1 point difference between the two tests with normals, but about a 4 point difference with referrals and low-achieving subjects (Woodcock, 1984).

Based on information about the normative samples for the two tests, there are important differences in the location, distribution, and type of subject included. Such factors may have led to substantial differences in the norm tables for the two tests. The BATERIA was normed on a stratified random sample of public and private school subjects from five Spanish speaking countries. In contrast, the

WISC-RM normative data were gathered from public schools only in Mexico City. Since tuition payments are required to attend a private school, there is a selective factor operating that results in children from lower socio-economic level families being enrolled in the public school while children from higher socio-economic level families are more likely enrolled in private schools. Furthermore, private schools in Mexico City are in session daily from 8:00 a.m. to 2:00 p.m., while public schools operate from 8:00 a.m. to 12:30 p.m. Private school subjects receive supplemental curricular offerings in addition to the mandated government curriculum.

If a norming sample is drawn from a population which is lower in ability than the general population, those norms, when applied to the general population, will result in higher mean scores for subjects from the general population. Furthermore, a norming sample from a restricted range will exaggerate the high and low scores in the general population. Evidence of these effects in the norms are seen when contrasting BATERIA and WISC-RM standard scores for normal subjects from private and public school. Mean scores for the Full Scale WISC-RM were 18 points lower for the public school subjects than for the private school subjects. The WISC-RM standard deviation (18.0) was not only higher than that of the BATERIA (13.6) but greater than 15.0. The higher standard deviation of the WISC-RM for the normal

group suggests that the data gathered in the WISC-RM normative study was from a sample with a restricted range and less variable than a representative cross-section from the general population. Because of these known differences in the norming samples for the two tests, it is inappropriate to directly evaluate the obtained differences between scores from the two tests, and to compare these differences to the differences observed in the U.S. studies.

To control for the differences in the norming samples, the Full Scale scores for the normals on both tests were standardized and placed on a common scale with a mean of 100 and a standard deviation of 15. A transformation procedure was then performed and the referral subjects' Full Scale scores on both tests which placed their scores on the same scale as that for the normal group. This procedure, in effect, provided common norms for the two tests, as if they were normed at the same time on a common sample of subjects.

When comparing mean Full Scale scores of referrals after the standardizing procedure there was, of course, a zero point difference between normals on the two tests, but there was a 5.5 standard score point difference for the referral group. This is in contrast to the 6.4 point difference for the normal group and the 7.9 point difference for the referral group before standardizing. These findings, after the standardizing of scores, appear to replicate the results of studies in the United States (Woodcock, 1984). The median

differences in those studies for normals was about 1 point, and for referral and low-achieving subjects 4.3 points.

Regarding the correlation of the BATERIA and the WISC-RM on the three scholastic achievement clusters, results indicated that the Full Scale BATERIA score for the normal group correlated higher in Reading and Written Language than the WISC-RM and the same for Mathematics as the WISC-RM. This pattern of correlation suggests that the BATERIA may be a "more discriminating instrument with respect to measuring cognitive skill relevant to school success and similar life endeavors" (Woodcock, 1984, p. 352). It follows then that groups such as the referral students in this study would be expected to score lower on the BATERIA Cognitive Ability than on the WISC-RM Full Scale score. The results from this study are consistent with those from other studies in the U.S. involving referral or low-achieving groups which compared the two measures.

The measured difference between normals on the two tests may primarily be a result of the non-comparability of norms that were gathered on the two tests (a problem that exists to some degree between any pair of tests which have not been co-normed). The WISC-RM sampled only public schools in Mexico City and the BATERIA sample included both public and private schools representative of students across the Spanish speaking world. In addition, the mean score differences between referral subjects on the two tests may

be attributed in part to unique factor (Estabrook, 1983) structures and their correlation with school achievement.

Special Cognitive Ability Clusters. A specific pattern of performance indicated that normals on the average scored significantly higher on all three Special Cognitive Ability clusters of the BATERIA than did the referrals. The results removed also indicated that all subjects performed significantly higher on the Visual-Perceptual Speed cluster than on either Reasoning or Oral Language. There were no significant differences between the Reasoning and Oral Language clusters. In addition, predictably, the referral group performed below the mean for their grade placement on all clusters.

The higher performance on the Visual-Perceptual Speed cluster and lower performance by all subjects on Reasoning and Oral Language was not consistent with expectations regarding students with higher abilities (Gallagher, 1975). Although significance was obtained with Visual Perceptual Speed higher than the other two Special Cognitive Ability clusters for both groups, actual differences were about 2 to 3 points higher for referrals and 7 to 8 points higher for the normals.

Scholastic Aptitude Clusters. Normals on the average scored significantly higher on all three Scholastic Aptitude clusters from the BATERIA than did the referrals. These results are predictable in light of the fact that the low-achieving students in this study, like learning disabled

students, were actually performing poorly in one or more of the achievement areas of reading, math, or written language. The referral group could be predicted to perform more poorly than normals on tests designed to estimate a subject's expected achievement in academic skills (Woodcock, 1978).

The Scholastic Aptitude clusters place heavy demands in the verbal processing area. Typically, learning disabled and low-achieving individuals tend to perform better on nonverbal cognitive tests than on verbal cognition measures (Hessler, 1982). Results from this study indicated that the referral sample scored significantly higher on the Performance scale of the WISC-RM than on the Verbal scale, thereby confirming these findings.

Results from previous studies using the English version indicate that the Scholastic Aptitude clusters from the BATERIA place heavy emphasis on verbal conceptual processes and correlate highly with other verbal measures (Reeve, Hall, and Zakreski, 1979; Ysseldyke, Shinn, and Epps, 1981a). The correlation coefficients between the BATERIA Scholastic Aptitude clusters and the WISC-RM Full, Verbal, and Performance scales were consistently higher for the Verbal scale than the Performance scale. These results are not surprising, since measures of verbal conceptual processing can be expected to correlate higher with academic achievement than do measures of nonverbal processing (Hessler, 1982).

It is interesting to note that subjects from both groups performed the highest on the Mathematics Aptitude cluster. This cluster primarily measures visual-perceptual and nonverbal conceptual abilities. The Mathematics Aptitude cluster, typically correlates most highly with Reasoning and Visual-Perceptual Speed.

The results of this study are consistent with this pattern, showing higher scores on Visual-Perceptual Speed and Mathematics Aptitude than on other special cognitive and aptitude measures for both groups.

Scholastic Achievement Clusters. As expected, the referral subjects scored significantly lower than the normal subjects on all three academic clusters of the BATERIA. Both referral and normal groups scored significantly better on the Mathematics Achievement cluster than on Reading or Written Language. For both groups, Written Language was the lowest cluster.

Students selected for the referral group in this study were achieving poorly in one or more of the academic achievement areas. Since many students with learning and achievement problems typically exhibit low performance in reading and writing, these findings corresponded with current research and are expected (Kirk and Gallagher, 1983; Hessler, 1982; Lerner, 1981).

WISC-RM: Verbal and Performance Scales. The normal group scored significantly higher on both scales. Means for both groups were higher for the Performance scale than for

the Verbal scale, though not a significant difference for the normals. These results are consistent with previous studies which found that learning disabled and low-achieving individuals typically tend to perform better on nonverbal cognitive tests than on verbal cognition measures (Hessler, 1982). Studies analyzing performance of learning disabled students in the current literature indicate that these students score significantly higher on the Performance scale than on the Verbal scale on the WISC-R (Anderson, Kaufman, and Kaufman, 1976; Smith, 1978; Smith, Coleman, Dokecki, and Davis, 1977; Zingale and Smith, 1978).

Anderson, Kaufman, and Kaufman (1976) reported Verbal-Performance discrepancies which were significantly larger than the average discrepancy for a sample of normal subjects. Even though the results from this study were found to be statistically significant, the practical implications of these differences are not clear. According to Kaufman (1979), a Verbal-Performance score difference of 9 points is required for statistical significance at the $p < .15$ level on the WISC-R. An analysis of standardization data of the WISC-R (Kaufman, 1976c) reveals that the average WISC-R discrepancy (regardless of direction) is 9.7 score points for all children aged 6 to 16 years of age.

Conclusions and Implications

Subjects grouped as normal and referral for this study were observed to score higher on the WISC-RM Full Scale

score than on the BATERIA Broad Cognitive Ability Full Scale score. Observed differences in means and standard deviations for the two groups may be attributable to differences in the norming samples for the two tests. These results demonstrate the effect that the composition of a norming sample may have upon the later evaluation of subjects selected from the general population. If the norming sample is lower in average ability than the general population, the obtained scores will be too high. If the range of ability in the normative sample is more restricted than the range of ability in the general population, then the effect will be to underestimate low scores and overestimate high scores in the general population. The normative scores provided by the WISC-RM could well produce "false positives" and "false negatives," depending upon the ability of the subject being compared to the standardization group on which the table of norms is based. This has important implications in regard to clinical assessment of Spanish speaking students referred for a cognitive diagnostic evaluation. Practitioners need to be aware of the differences that can be expected based upon the findings of this study.

After standardizing scores from the two tests, the referral group scored 5.5 points higher on the WISC-RM than on the BATERIA. These results closely approximate the 4.3 standard score points median difference reported in the literature when comparing the corresponding English versions (Woodcock, 1984). This finding may be explained by the

BATERIA's higher correlation with school achievement, but practitioners should be aware of this trend in performance when evaluating lower achieving and referral students.

For the normal group the BATERIA Full Scale Broad Cognitive Ability correlated higher with actual achievement in Reading and Written Language than Full Scale WISC-RM scores The two measures both correlated .72 with Mathematics achievement. These results suggest that for this population the BATERIA is the better predictor for scholastic achievement.

On BATERIA Special Cognitive and Scholastic Aptitude clusters, both normal and referral groups scored significantly higher on Visual-Perceptual Speed and the Mathematics Aptitude clusters than on the other measures.

Correlations between Scholastic Aptitude and Achievement scores from the BATERIA were lower for the normal group than for the referral sample (.71 for Reading, .70 for Written Language, and .55 for Mathematics). Correlations for the referral group between the same measures were .78 for Reading and Written Language and .71 for Mathematics.

On the three BATERIA achievement measures, the referral group scored significantly below the normal group. Though not significant, a trend was observed with both groups scoring highest on Mathematics and lowest on Written Language.

Concluding Remark. This study examined the performance of a sample of subjects classified as normal and a sample of subjects classified as referral on the BATERIA and the

WISC-RM. The results provided insights regarding test performance and test validity for the two samples studied. The findings from this study point out what significance norming a standardized test on a restricted range of subjects may have in the evaluation and diagnosis of subjects selected from the general population. The analysis and description of performance by the two groups has implications for the application of scores from the two tests in psycho-educational work with Spanish speaking subjects.

Implications for Future Research

Though the results of this study replicated the findings from similar studies in the United States using the Woodcock-Johnson Psycho-Educational Battery and the Wechsler Intelligence Scale for Children--Revised, additional studies would serve to validate the findings of this study and further document the extent of difference between the two tests to be expected and to what extent these differences are a function of type of student.

Although low-achieving students were selected based on poor achievement in mathematics, reading, or written language, no attempt was made to divide these students into subcategories for analysis since the common underlying characteristic of the subjects classified as referrals was poor achievement. Future studies could examine and determine if different results on the two tests would be obtained depending upon the type of student referred.

Based on the findings from this study, it is suggested that the technical adequacy of the Mexican norms of the WISC-RM be further evaluated. Evidence suggests that the mean ability of the WISC-RM norming sample may have been too low, thus, the scores provided by the norm tables would be too high for use with the general population. Furthermore, since the range of the norming sample may have been restricted, the high scores would be exaggerated and the low scores would also be exaggerated when the test is used with the general population. Additional research should be performed to verify or discount the tentative conclusions about the WISC-RM norms presented in this study. An appropriate transformation equation might be developed which could provide an adjustment to norms provided by the present WISC-RM tables.

BIBLIOGRAPHY

Adelman, H. (1970). An interactive view of causality. Academic Therapy, 6, 43-52.

Ambert, A. & Dew, N. (1982). Special education for exceptional bilingual students: A handbook for educators. Milwaukee, Wi.: University of Wisconsin--Milwaukee. Midwest National Origin Desegregation Assessment Center.

Anastasi, A. (1982) Psychological testing (5th ed.). New York: Macmillan Publishing Co.

_____. (1984). The K-ABC in historical and contemporary perspective. Journal of Special Education, 18, 45-55.

Anderson,M., Kaufman, A., & Kaufman, N. (1976). Use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 13, 381-386.

Archuleta, K. & Cervantes, H. T. (1981). The misplaced child: does linguistically different mean learning disabled? In P. C. Gonzales (Ed.), Proceedings of the Eighth Annual International Bilingual Bicultural Education Conference. Rosslyn, Va.: National Clearinghouse for Bilingual Education.

Arreola v. Santa Ana Board of Education. (1968). 160-577 (Orange County, California).

Baca, L. & Cervantes, H. T. (1984). The bilingual special education interface. St. Louis, Missouri: Times Mirror/Mosby College Publishing.

_____. (1978). The use of bilingual testing instruments in Colorado. The Bilingual Journal, 3, 10-14.

Bergin V. (1982). Special education needs in bilingual programs. Rosslyn, Va.: Inter America Research Associates, Inc. National Clearinghouse for Bilingual Education.

Bernal, Ernesto R. (1983). Trends in bilingual special education. Learning Disability Quarterly, 6(4).

Carter, T. D. & Segura, R. D. (1979). Mexican Americans in the school: A decade of change. New York: College Entrance Examination Board.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.

Cervantes, H. T. & Baca, L. (1979). Assessing minority students: The role of adaptive behavior scales. Journal of Non-White Concerns, 7 (3), 122-127.

Covarruvias v. San Diego Unified School District. (1971). No. 70-394-T (San Diego, California).

Coyle, F. A. (1965). Another alternate wording on the WISC. Psychological Reports, 16 (3), 1276-1285.

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

_____. (1982) Tests, achievement and bilingual students. Focus, National Clearinghouse for Bilingual Education, 9, 1-7.

Cummings, J. A. & Moscato, E. M. (1984a). Research on the Woodcock-Johnson Psycho-Educational Battery: Implications for practice and future investigations. School Psychology Review, 13, 13-40.

_____. (1984b). Reply to Thompson and Brassard. School Psychology Review, 13, 45-48.

Davis, C., Haub, C., & Willette, J. (1983). U.S. Hispanics changing the face of America. Population Bulletin, 38 (3).

De Blassie, R. R. & Franco, J. N. (1983). Psychological and educational assessment of bilingual children. In D. Omark & J. D. Erickson (Eds.), The bilingual exceptional child. San Diego: College-Hill Press.

Dean, R. S. (1979). Distinguishing patterns for Mexican-American children on the WISC-R. Journal of Clinical Psychology, 35.

Diana v. Board of Education. (1970). Civil Actions No. C-70-37 (N. C. California).

Dunn, L. (1968). Special education for the mildly retarded--is much of it justifiable? Exceptional Children, 35, 5-22.

Figueroa, R. A. (1983). The Diana visits reveal inequities in assessment of Hispanic students. CASP Today, 4-6.

Estabrook, G. (1983). A canonical correlation analysis of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability in a sample referred for suspected learning disabilities. Manuscript submitted for publication.

Fradd, S.& Hallman, Clemens, L. (1983). Implications of psychological and educational research for assessment and instruction of culturally and linguistically different students. Learning Disability Quarterly, 6, 468-478.

Gaito, J. (1977). Repeated measures design and counter-balancing. In R. O. Collier & T. J. Hummel (Eds.), Experimental design and interpretation. Berkeley, Ca.: McCrutchan Publishing Corporation.

Gallagher, J. J. (1975). Teaching the gifted child (2nd ed.). Boston: Allyn and Bacon, Inc.

Gelb, Steven. (1982). A guide to working with minority language students in special education. Prepared for the Superintendent of Public Instruction. Seattle, Washington.

Hessler, G. L. (1982). Use and interpretation of the Woodcock-Johnson Psychoeducational Battery. Hingham, Ma.: Teaching Resources Corporation.

Jones, R. L. (1976). Mainstreaming and the minority child. Minneapolis, Mn.: Council for Exceptional Children.

Kamin, L. J. (1975). Social and legal consequences of IQ tests as classification instruments: Some warnings from our past. Journal of School Psychology, 13, 317-323.

Kirk, S. A. & Gallagher, J. J. (1983). Educating exceptional children (4th ed.). Boston: Houghton-Mifflin.

Laosa, L. M. (1977). Nonbiased assessment of children's abilities: Historical antecedents and current issues. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Bruner/Mazel.

Larry P. v. Wilson Riles. (1974). 343 F. Supp. 1306, 502F, 2nd 963 (9th Circuit Court).

McGrew, K. S. (1983). Comparison of the WISC-R and Woodcock-Johnson tests of cognitive ability. Journal of School Psychology, 21, 271-276.

Memo directed to special educators and Title VII program directors. (1982). Rosslyn, Va.: National Clearinghouse for Bilingual Education.

Mercer, J. R. (1973). Labeling the mentally retarded. Berkeley: University of California Press.

Mercer, J. R. (1979). System of multicultural pluralistic assessment. New York: The Psychological Corporation.

_____ & Ysseldyke, J. (1977). Designing diagnostic-intervention programs. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Bruner/Mazel.

Miller, M. (1916). The school and the immigrant. Cleveland: Survey Committee of the Cleveland Foundation.

Moore, J. W. (1976). Mexican American (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall Inc.

Moran, R. E. (1962). Observations and recommendations on the Puerto Rican version of the Wechsler Intelligence Scale for Children. Pedagogia, Rio Piedras 10, 89-98.

Nazzaro, J. (1979). Assessment of minority students (fact sheet). Reston, Va.: Eric Clearinghouse on Handicapped and Gifted Children.

Newland, T. E. (1971). Psychological assessment of exceptional children and youth. In W. Cruickshank (Ed.), Psychology of exceptional children and youth. Englewood Cliffs, N.J.: Prentice-Hall.

Nuttall, Ena Vazquez. (1984). Assessing and treating bilingual handicapped children: National state of the art. Paper presented at the annual meeting of the American Psychological Association. Toronto, Canada.

Oakland, T. & Matuszek, P. (1977). Using tests in non-discriminatory assessment. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Brunner/Mazel, Inc.

Ochoa, A. M., Pacheco, R. & Omark, D. R. (1983). Addressing the learning disability needs of limited English proficient students: Beyond language and race issues. Learning Disability Quarterly, 6, 416-423.

Omark, D. R. & Erickson, J. G. (1983). The bilingual exceptional child. San Diego: College-Hill Press.

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

_____. (1984). Staffing and the development of individualized education programs for bilingual exceptional students. In Baca, L. & Cervantes, H. T. (Eds.), The Bilingual Special Education Interface. St. Louis, Mo.: Time Mirror/C. V. Mosby.

Pacheco, R. S. (1983). Bilingual mentally retarded children: language confusion or real deficits? In D. R. Omark & J. G. Erickson (Eds.), The bilingual exceptional child. San Diego: College-Hill Press.

Plata, M. (1982). Assessment, placement, and programming of bilingual-exceptional pupils: A practical approach. Reston, Va.: The Council for Exceptional Children.

Reeve, R., Hall, J., & Zakreski, R. S. (1979). The Woodcock-Johnson tests of cognitive ability: Concurrent validity with the WISC-R. Learning Disability Quarterly, 2, 63-69.

Reynolds, C. R. (1985). Critical measurement issues in learning disabilities. The Journal of Special Education, 18, 450-476.

Rubin, R. A., Krus, P., & Balow, B. (1973). Factors in special education placement. Exceptional Children, 39, 42-50.

Salvia, J. & Ysseldyke, J. E. (1981). Assessment in special and remedial education. Boston: Houghton-Mifflin Company.

Samuda, R. (1975). Psychological testing of American minorities. New York: Harper and Row Publishers, Inc.

Sanchez, G. L. (1934). Bilingualism and mental measures: A word of caution. Journal of Applied Psychology, 18, 756-772.

Sattler, J. (1982). Assessment of children's intelligence and special abilities. Boston: Allyn & Bacon, Inc.

Schwartz, P. A. (1962). Adapting tests to the cultural setting. Educational and Psychological Measurement, 4, 673-686.

Senf, G. M. (1982). Future research needs in learning disabilities. In R. P. Anderson & C. G. Halcomb (Eds.), Learning disability/minimal brain dysfunction. Springfield, Ill.: Charles C. Thomas.

Shinn, M., Marston, D., & Ysseldyke, J. (1980). Review of the Woodcock-Johnson Psycho-Educational Battery. School Psychology International, 1 (4), 20-22.

Smith, M. D., Coleman, J. M., Dokecki, P. R., & Davis, E. E. (1977). Intellectual characteristics of school labeled learning disabled children. Exceptional Children, 43, 352-357.

Smith, M. W. (1974). Alfred Binet's remarkable questions: a cross-national and cross-temporal analysis of the cultural biases built into the Stanford-Binet Intelligence Scale and other Binet tests. Genetic Psychology Monographs, 6 (12), 18-47.

Stellern, J. (1982). Diagnostic-prescriptive teaching. San Juan Capistrano, Ca.: Exceptional Press.

Thompson, P. L. and Brassard, M. R. (1984). Cummings and Moscato soft on Woodcock-Johnson. School Psychology Review, 13, 41-44.

Thorndike, R. L. & Hagan, E. (1977). Measurement and evaluation in psychology and education. New York: John Wiley & Sons.

Tyler, Leona E. (1984). Testing the test: What tests don't measure. Journal of Counselling and Development, 63, 48-50.

U.S. Congress. (1977). Public Law 94-142 (The Education for All Handicapped Children Act of 1975), 20 U.S.C. 1401. P.L. 94-142 Regulations. Federal Register, 42(May 4).

U.S. General Accounting Office. (1981). Disparities still exist in who gets special education: Report to the Chairman, Select-Subcommittee on Education, Committee on Education and Labor, U.S. House of Representatives. Washington, D.C.: U.S. Government Printing Office, Sept. 30, 1981.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1974). Wechsler intelligence scale for children--revised. New York: The Psychological Corporation.

_____. (1984). Escala de inteligencia para nivel escolar Wechsler. Mexico, D.F.: El Manual Moderno.

Wesman, A. G. (1968). Intelligence testing. American Psychologist, 23, 267-274.

Wilkerson, T. (1985). The system for statistics. Evanston, Ill.: Systat Inc.

Williams, R. (1975). Moderator variables as bias in testing black children. Journal of Afro-American Issues, 8, 77-90.

Woodcock, R. W. (1978). Standardization and development of the Woodcock-Johnson Psycho-Educational Battery. Allen, Texas: DLM Teaching Resources.

_____. (1980). A response to some concerns raised about the Woodcock-Johnson Psycho-Educational Battery. Proceedings of the Tenth Annual Invitation Conference on Leadership in Special Education, 10, 107-117.

_____. (1982). Bateria Woodcock psico-educativa en espanol. Allen, Texas: DLM Teaching Resources.

_____. (1984). A response to some questions raised about the Woodcock-Johnson I. The mean score discrepancy issue. School Psychology Review, 13, 342-354.

_____. (1984). A response to some questions raised about the Woodcock-Johnson II. Efficacy of the aptitude clusters. School Psychology Review, 13, 355-362.

_____ & Johnson, M. B. (1977). Woodcock-Johnson Psycho-Educational Battery. Allen, Texas: DLM Teaching Resources.

Ysseldyke, J., Algozzine, B., & Shinn, M. (1981). Validity of the Woodcock-Johnson Psycho-Educational Battery for learning disabled youngsters. Learning Disability Quarterly, 4, 2.

Ysseldyke, J. & Algozzine, B. (1982). Critical issues in special and remedial education. Boston: Houghton-Mifflin.

Zingale, S. A. & Smith, M. D. (1978). WISC-R pattern for learning disabled children at three SES level. Psychology in the Schools, 15, 199-204.

APPENDICES

APPENDIX A

Pearson Product-Moment Correlation Coefficient for BATERIA and WISC-RM Standard Scores for Subjects Classified as Normal (n = 47)

	WISC-RM		
BATERIA	Full Scale	Verbal	Performance
Broad Cognitive Ability	.86	.84	.74
Oral Language	.84	.83	.71
Reasoning	.75	.77	.57
Visual-Perceptual Speed	.61	.50	.66
Reading Aptitude	.78	.75	.72
Written Language Aptitude	.80	.76	.73
Mathematics Aptitude	.75	.72	.67
Reading Achievement	.68	.69	.54
Written Language Achievement	.65	.65	.54
Mathematics Achievement	.59	.57	.48

APPENDIX B

Pearson Product-Moment Correlation Coefficient for BATERIA and WISC-RM Standard Scores for Subjects Classified as Referral (n = 43)

	WISC-RM		
BATERIA	Full Scale	Verbal	Performance
Broad Cognitive Ability	.84	.84	.68
Oral Language	.90	.87	.77
Reasoning	.65	.64	.52
Visual-Perceptual Speed	.61	.59	.51
Reading Aptitude	.87	.84	.72
Written Language Aptitude	.88	.85	.73
Mathematics Aptitude	.84	.82	.68
Reading Achievement	.71	.71	.60
Written Language Achievement	.73	.74	.58
Mathematics Achievement	.67	.69	.49