1985

Frequency of Use as a Measure of Word Difficulty in Bilingual Vocabulary Test Construction and Translation

José Miguel Tamayo
Loyola University Chicago

FREQUENCY OF USE AS A MEASURE OF WORD DIFFICULTY

IN BILINGUAL VOCABULARY TEST

CONSTRUCTION AND TRANSLATION

by

JOSE MIGUEL TAMAYO

A Dissertation Submitted to the Faculty of the Graduate

School of Loyola University of Chicago in Partial

Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

MAY

1985

ACKNOWLEDGMENTS

VITA

The investigator, Jose Miguel Tamayo, is the son of Jose Miguel Tamayo and Mercedes Gonzalez. He was born in April 28, 1936 in Bilbao, Spain.

He received the Baccalaureate degree in Classical Studies, the licenciate in Philosophy in 1956, and the Master in Biblical Studies in 1962. He was the editor of Ad Majora, the college magazine, in which he published two articles on biblical studies.

From 1956 to 1958 he was a teacher in grammar school. From 1962 to 1965 he was the principal of a primary school in his native Bilbao, and from 1965 to 1967 he was a counselor in high school. He served in the Spanish army for two years, from June 1967 to June 1969. While in the Army he started to pursue graduate studies in the University of Valencia, Spain. In September 1969 he went to Madrid to complete graduate studies in clinical psychology at the University of Madrid. He obtained thirteen honorific mentions and graduated with the degree of Licenciado in Clinical Psychology in 1972.

During the same period he studied modern languages --English and French-- in the Escuela Superior de Idiomas in Madrid and during three consecutive summers he studied French at L'Alliance Francaise in Paris. He also studied music at Escuela Superior de Musica in Madrid.

In september 1972 he came to the United States, studied English at Roosevelt University in Chicago, and started graduate studies in guidance and counseling at Loyola University of Chicago. In February 1976 he was granted the degree of Master of Education in Guidance and Counseling. In September 1976 he started the Master program in clinical psychology at Roosevelt University in Chicago and graduated in May 1978.

While at Loyola and at Roosevelt he assisted some Hispanic communities in Chicago, was the director of the Young Christian Workers, and served as a counselor for Spanish speaking persons. In 1982 he published the book El Conflicto Matrimonial en los Matrimonios Latinos (Marriage conflict in Latino marriages).

From september 1978 to the present he has worked as a bilingual school psychologist at the Bureau of Child Study, Chicago Board of Education, in which he is a member of the research division and a supervisor of interns.

In 1983 he became an American citizen. He has taught philosophy and psychology courses at Saint Agustine College and currently at North Park College.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Psychologists frequently face the problem of administering to non-English-speaking children in their native languages tests constructed in English and normed with English-speaking samples. As the number of children whose languages are other than English increases in American schools, this procedure is becoming more frequent. As a consequence, the probability of erroneous conclusions when interpreting the test results obtained with careless and uncontroled translations is increasing. Careless, uncontrolled translations refer to the common and widespread practice among bilingual psychologists of using the English version of the test manuals and translating directions and items in an ad hoc manner. A more detrimental practice used by some English speaking psychologists testing linguistic minorities consists of giving the English items to any bilingual personnel available in the school --teachers, secretaries, lunchroom attendants, and the like-- for their translation, and then using this translation for testing purposes. The procedure of using these types of translations or published unvalidated (Spanish) translations of tests written in English was criticized many years ago by Sanchez (1934a). This procedure leads to variations from one administration to another, and from one psychologist to another. The extent to which these variations produce easier or more difficult items is not known, nor is the influence of such variations on test

scores, and on their interpretation or on the conclusions reached from them (Samuda, 1975). Olmedo (1981) indicates that

> Translating existing tests into other languages presents methodological problems that often are not treated properly. For example, direct translations do not ordinarily yield technically equivalent forms because the domains sampled by the different language versions may have little overlap, and the translated items may exhibit psychometric properties substantially different from those of the original English items. (p. 1083)

The most basic principle in the application of any standardized test, particularly intelligence tests, is the necessity to follow, in the most rigorous manner, the procedural directions given in the manual (Wechsler, 1974; Terman and Merrill, 1973; Cronbach, 1960). The very validity of the test results depends greatly upon adherence to these standard procedures. If they are not carefully followed, the test results may be seriously compromised. When a test is given in a language other than that in which the test was standardized, a significant departure from the standard procedures exists. If precautions are not taken, it is impossible to gauge the extent of such departure and the effects on the test results.

The controversy regarding the testing of linguistic minorities that has been going on for four decades (Olmedo, 1977) has emphasized linguistic, cultural, environmental, and social differences. Little systematic attention, however, has been paid to the linguistic aspect of similarity of verbal items and item difficulty. Since the items of most intelligence tests are arranged in order of increasing difficulty (Glasser and Zimmerman, 1967), it is the difficulty of the translated items that must be of primary concern in cases in which a language other than

the original language is used in the test administration. Although most test manuals do not indicate the procedure for ordering items according to difficulty, the most common method used has been the passing percentage of the people in the preliminary studies conducted or in the standardization samples. In nonverbal tests the tasks are the same for children of different languages, consequently the difficulty of test items is not a problem because it could be assumed that the items are equally difficult for both groups. The problem of cultural differences in regard to nonverbal tests and the possibility of bias is not addressed in the present study.

The problem of different levels of difficulty is crucial in verbal tests, particularly in vocabulary tests. Vocabulary, one of the best measures of what is referred to as general intelligence (Cohen, 1959; Glasser and Zimmerman, 1967; Matarazzo, 1972; Zimmerman and Woo-Sam, 1973; Kaufman, 1975; and Jensen, 1980) or verbal fluency (see Oller, 1983), must be carefully translated, accommodated, or constructed if the difficulty of test items is to be the same in both languages and if the results obtained are to be considered valid. This is particularly true when the original tables for calculating scores are used, because local or ethnic norms are not available. It is not enough to translate the English words using non-English words considered to have the same meaning, because even though the translated meaning might be accurate, the pattern of difficulty might be different.

Moreover, while a word by word translation could provide the same results with both English and non-English speaking samples, nevertheless these results might not indicate the same level of intelligence or ver-

bal fluency in both samples. One of the samples might be significantly higher in intelligence or verbal fluency than the other, but results might equal those of the less· intelligent or fluent sample because its list of words is more difficult. While the scores are the same, the intelligence or fluency level necessary to obtain them is different, and so the results are not comparable. When the English verbal items are going to be used in other than the English language, the passing percentage, although it could be a good measure of item difficulty similar to the original one, is not a practical method of measuring item difficulty. Ordering the translated words by their passing percentages would require administration of the words to samples large enough to obtain reliable rank orders for every language in which the tests were going to be administered. This procedure would be very time consuming and impossible to follow in cases in which large samples are not available. An objective and reliable measure of word difficulty that can be used as a reliable procedure for construction or translation and one that does not require administration of words to samples is needed. Such a measure once proved valid, could be used to construct or to translate lists of words just by following the established criteria of similarity and difficulty.

For the purpose of the present study, vocabulary item difficulty or word difficulty refers to those characteristics which indicate whether or not a word is known or unknown to the subject. This knowledge is ascertained from the accuracy of the verbal explanation of the meaning of words presented orally by the examiner. The word characteristics as indicators of word difficulty have to have been clearly identified and

measured in some way by linguists. Word characteristics not fully explained and quantified by linguists would be of no usefulness as an objective measure for translation or construction of vocabulary tests. Indeed, this is not the way test manuals usually measure word difficulty.

Before going further, it is necessary to make clear that this study is primarily concerned with the procedure of translating or constructing vocabulary tests to be used in countries in which children of different languages live and attend school together. However, the implications of this study are directly related to the issue of bias in mental testing, particularly bias with linguistic minorities. Most of the criticisms of mental testing and test items have been based on emotional reactions, social implications of testing, and armchair speculations, with little or no empirical support. Such emotionally laden criticisms are not acceptable in any scientific field. Any assertion in favor or against mental testing has to be validated by empirical evidence or be disregarded as inappropriate. The controversial and emotional atmosphere surrounding the administration of ability tests has generated much confusion and little light to clarify the issue (Carroll and Horn, 1981). The excessive attention that has been devoted to the suspected bias of measurement instruments has prevented a careful and conscientious analysis of the social structures that might be the real sources and maintaining factors of bias in society. In actuality, what the tests have done is just to detect the results of inequality in society rather than causing such disturbing inequalities. While the consequences of the controversy are social in nature, and consequently have to be dealt with in court,

the issue as such is not entirely a social justice one. Rather it is also a scientific truth issue that must be empirically addressed in an academic setting. As Joanne Condas (1980), the Deputy Attorney General of California, indicated in her comment related to the Larry P. case:

> I can only hope that whatever changes are required in California practice as a result of the court order, those who have a professional interest in and commitment to the field of school psychology will continue to study the important validity questions and the educational outcome questions until something definitive emerges in the scholarly realm. It is there, not in the courtroom, where the answers are really to be found. (p. 158)

The same position has been recently advocated by Reynolds and Brown (1984).

From the foregoing one can conclude that the common practice of administering English tests to non-English-speaking children in the children's native language, without accurate translation and validation, is inappropriate for the following reasons: a) there is no guarantee that the level of difficulty of test items is the same in the original and in the translated versions of the tests; b) the basic test administration principle of adherence to test directions in the most rigorous manner is not followed, because directions and items are not given in the original language but rather in the translated version; c) an uncontrolled translation could introduce test bias in an unknown direction, toward either easier or more difficult items.

Purpose of the Study

This study consists of two parts, a theoretical component and an empirical investigation.

Theoretical component

The theoretical component of this study focused on linguistic variables which could account for word difficulty in vocabulary test construction and translation in order to have the assurance that the word items in both tests, English and non English, are similar in the word characteristics related to meaning recall and recognition. It is necessary to emphasize here that the focus of this study is on word characteristics rather than on internal individual processes or strategies for word encoding and retrieval. It is probable that some individuals will rely upon visual or auditory clues or some other mnemonic devices to remember the meaning of words while others will rely upon semantic strategies. These internal processes do not account for word similarity (meaning and difficulty) and they are in the domain of individual differences. It is the similarity in meaning and level of difficulty that is important in the test items from the perspective of test construction and measurement. With tests so constructed or translated it would be possible to obtain comparable results at least with the subjects of the present study: English and non-English speaking children living and attending school in the United States.

Two word characteristics were considered. In order to make both lists equal in the most basic characteristic, the content or general meaning of the word was incorporated in the procedure. Although there is no consensus concerning the concept of meaning (Creelman, 1966), four aspects of word meaning have been pointed out as most important: referential, denotative, associative, and affective or connotative (Glucksberg and Danks, 1975). For the purpose of this study, which presents words out of any context, only the denotative meaning or the generic idea or concept represented by the word was considered.

The second characteristic under study was the level of word difficulty as measured by frequency of use. It is unfortunate that linguists have not generally been concerned with the linguistic aspect of word difficulty. This aspect is crucial in this study which explores some word characteristics in order to find one that could serve as a reliable, objective, and quantifiable measure of word difficulty. Although it is possible that there are many more characteristics, frequency of use has been chosen. This linguistic characteristic can be objectively quantified and statistically analyzed. It is questionable whether some other word characteristics, linguistically more important perhaps, could be equally objective and equally quantifiable. Consequently, frequency of use serves as an operational definition of word difficulty in the present study.

Basic assumptions

The assumptions underlying the selection of frequency of use as criterion of difficulty need clarification. Frequency of use, as a measure of word difficulty, assumes that the words most commonly used are learned faster and remembered better. As a consequence frequency of use could be considered as one valid and practical measure of word difficulty. This first assumption is supported by some linguistic studies. Werner and Kaplan (1952) indicated that children acquire meaning or learn to understand verbal symbols through the adult's direct naming of objects, through verbal definitions, or by grasping the meaning of words in the course of conversation inferring meaning from context. Brown (1958) suggested that "the sequence in which words are acquired is not determined by the cognitive preferences of children so much as by the naming practices of adults." One important aspect of the "naming practices of adults" is obviously the frequency of use of the words in their sentences. A similar position is adopted by Clark (1983) when she indicates that in the process of acquiring meanings, children look for consistencies in adult use of words to pick out particular conceptual categories through the conventions and contrasts that adults observe in their use of language. It is assumed in the present study that this process will be carried out more easily in regard to words of high frequency of use than to low frequency words.

Sternberg and Powell's (1983) theory of learning from context provides some support for this claim as well. This theory explains in detail how people infer the meaning of unfamiliar words from context.

The theory distinguishes between external or contextual cues and internal or mediating variables. Context cues are hints in the passage in which the word is used that facilitate deciphering the meaning of an unknown word. Mediating variables specify those variables that affect the application of the contextual cues present in a given situation. Although the theory indicates that either the context cues and the mediating variables can in some instances impede the process of vocabulary acquisition, it is not unsafe to infer that the more frequently a word is used, the better the context cues and mediating variables will operate. Sternberg and Powell present some initial empirical results that support the facilitative effect of both context and mediating variables. Solomon and Howes (1951) and Howes and Solomon (1951) demonstrated that the ease of recognition of words presented auditorily or visually was a function of their frequency of use in the language. Some theories of learning and retention, particularly the Total Time Hypothesis theory of Bugelski (1962), also provide some support for this assumption. This hypothesis claims that in any learning task the most important condition of learning is the total amount of time devoted to the task. It is assumed in this study that the more frequently a word is used in daily language the more time is devoted to learn the word. Another condition related to learning and retention, similar to the Total Time Hypothesis, is overlearning, when practice occurs beyond the point where immediate and complete recall is possible (Travers, 1982). In vocabulary acquisition overlearning is easier to achieve with the words that are most commonly used in everyday language.

A second assumption concerns the generalizability of the procedure

used in the present study. The procedure proposed in the present study is offered as a valid method of vocabulary test construction or translation for use with individuals of different languages and cultural backgrounds living in the United States. This claim is based on the assumption that vocabulary building, with regard to the concepts used in vocabulary tests, is similar across linguistic and ethnic groups. Although each ethnic group has some concepts that are peculiar to its particular culture, these concepts are not, and should not be, the type of concepts found in vocabulary tests. Vocabulary test concepts usually represent concepts that are, to a certain extent, shared by all the cultures and linguistic groups that live in the United States. This seems to be particularly true for children. Clark (1977a) claims that "the earliest vocabulary of young children seems to be very similar across children and across languages" (p. 25). There appears to be a surprising uniformity in early vocabularies (Clark, 1977b; Clark, 1979; Nelson, 1973). What is important from the point of view of test construction and translation is that these common concepts are represented by words which are equally difficult.

A third assumption, similar to the second, refers to age. The procedure proposed in the present study is offered as a valid method of vocabulary test construction or translation for use with individuals of different ages. Since the frequency of use lists have been compiled with samples of adult language, the question is whether or not these lists are equally valid to construct or to translate vocabulary tests for children as well as for adults. The words with high frequency of use usually represent objects or concepts which are most necessary in

the daily life, and the human needs are basically the same for children and adults. For these reasons it is assumed in the present study that the structure of children's lexicon, in regard to the words found in vocabulary tests, is similar to the structure of adult lexicon. It is obvious that the children's vocabulary includes names of toys and of other familiar objects whose corresponding words, although not frequently used in the language, are very easy for children. These words, however, do not appear to alter significantly the structure of the children's lexicon. Considering the opposite end of the spectrum, words with low frequency of use are so because they represent objects or concepts of little value in daily life and consequently it is assumed that they will be unusual for children as well as for adults.

Finally, the fourth assumption concerns the problem of the actual validity of the obtained measures of frequency of use. Most of them were obtained in the twenties and thirties. Eaton's lists are used in this study. The frequency of use of her English words is based on Thorndike's 1932 Teacher's Word Book of 20,000 Words and her Spanish words are based on Buchanan's 1927 Graded Spanish Word Book. An obvious question is whether or not these frequencies of use are still valid. Although new words have been introduced frequently in the vocabularies of modern languages, it is assumed in the present study that the frequency of use of most words remains constant for long periods of time. Although it is probable that addition of new words into modern vocabularies alters in some way the frequency of use of old words, it is doubtful that significant disruptions will occur as a result of neologisms. The probable outcome of inclusion of neologisms in modern vocabu-

laries is the displacement of some old words from one group of frequency of use to the next rather than significant alterations in the structure of the lexicon. There is substantial evidence that at least the early vocabularies of today's children are the same as the vocabularies of children of fifty years ago (Nelson, 1973; Clark, 1979). Since, as was previously indicated, the words with high frequency of use usually represent objects or concepts which are most necessary in the daily life, and since human needs rarely change, assuming stability in the frequency of use of most words appears to be safe.

The empirical investigation

The empirical component of the present study was designed to determine whether or not frequency of use is a good measure of word difficulty -as defined in this study- that can be used to make vocabulary items -words- equally difficult in both languages. It was predicted that frequency of use will provide a reliable measure of word difficulty.

Summary

This chapter has presented an overview of the purpose intended in this study. The theoretical part intends to select word characteristics that can account for similarity in content and difficulty in bilingual vocabulary test construction and translation. Two characteristics have been selected as a measure of word difficulty: meaning and frequency of

use. In regard to these characteristics four assumptions can be formulated:

1. The frequency of use is one valid and practical measure of word difficulty,

2. Words found in vocabulary tests usually represent and should represent concepts that are common to all cultural and linguistic groups living in the United States,

3. Children's vocabulary is similar to adult vocabulary,

4. The frequency of use of most words remains constant for long periods of time.

The empirical part of this study investigates to what extent frequency of use is one good measure of word difficulty.

CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature is divided into four sections: a) the controversy of test bias; b) the problems of translation; c) theoretical aspects of linguistic and semantic development; and d) review of the theories of semantic development.

1. The controversy of test bias

The problem of test bias has been frequently confused with the issue of the etiology of obtained racial group differences on test performance, with the inappropriate administration and use of tests, and with the social consequences of testing. For these reasons much confusion has been generated, particularly outside the scientific community, and many emotional and political aspects have been included in the controversy. As a scientific issue, the concept of bias is concerned with the statistical properties of the tests, not with their political and social ramifications.

As a social issue bias is not concerned exclusively with the use of tests but with all aspects of the assessment process, from the referral

to the placemennt and its consequences. In this long process bias can
be easily introduced if every single step is not closely scrutinized and
monitored. Different authors· have emphasized single aspects of this
process disregarding others. As a consequence the confusion has
increased. Bias in assessment procedures could be introduced if verbal
tests are mostly used when assessing economically disadvantaged and
bilingual students without tapping other areas of intellectual function-
ing. Bias could be introduced in test use when emphasis is put on the
test results obtained by children who have not been sufficiently exposed
to test taking situations and consequently lack the test taking skills
needed to succeed in the test. In the decision making process, bias
could be introduced when conclusions about placement in special educa-
tion programs are reached based on test results ignoring the adaptive
behavior skills of the student out of the school. Bias could be intro-
duced when the examiner belongs to a race or cultural background other
than the one of the student and fails to establish adequate rapport dur-
ing the testing session. Bias could be introduced when using tests
whose item content is from a particular culture and the student is not
fully a member of this cultural group. Finally, bias could be intro-
duced when the criterion used in·decision making is irrelevant to the
task to be performed. All these types of bias are very important and
are directly related to the social consequences of testing.

As a scientific construct bias is more directly related to the sta-
tistical properties of tests and how they influence test performance of
different ethnic groups. Test bias has been conceptualized and defined
in many ways. Flaugher (1978) has identified eight definitions of test

bias: mean differences, overinterpretation, sexism, differential validity, content, selection model, wrong criterion, and atmosphere or situational factors. Mercer (1978) points out five characteristics as indicative of test bias: test items from a single cultural heritage, differences in average scores among different racial and cultural groups, sociocultural differences within and between cultural and racial groups with these differences accounting for a significant proportion of the variance in test performance, experimental studies demonstrating the effects on test performance of early interventions with culturally different children, and the effects of adoption of minority children into core culture homes. Jones (1978) has suggested four areas in which bias can be introduced: in the content level of test items, in the standardization procedures in which decisions are made concerning the population for whom the test is appropriate, in the administration of the test in cases in which the examiner is unfamiliar with the culture of the child, and in the validation where research may not be conducted concerning test validity for culturally different persons. Jensen (1980) lists ten statistical properties of tests and test characteristics to be examined in order to detect test bias: temporal stability, internal consistency, groups x items interactions, reliabilities and correlations from ANOVA, transformation of the scale of item difficulty, rank order of item difficulty and Delta decrements, the item characteristic curve, item correlation methods, factor analysis criteria of bias, and matched groups and pseudogroups. From a psychometric point of view, bias should not be confused with prejudice, but rather should be considered as "a statistical term referring to a constant error of measure in one specific direction as opposed to random error" (Reynolds and Brown, 1984, p. 2).

In order to have a clear picture of the results of the present mental testing controversy, it is necessary to examine research findings rather than to accept unsupported criticisms or armchair speculations. Vernon (1979) in his extensive review of empirical literature concluded that the claim of bias in mental tests cannot be substantiated. Another extensive review of the empirical studies in education that finds over-whelming evidence against bias has been published by Gordon and Rudert (1979). Reschly (1980) has provided a summary table comparing the common definitions of test bias along with the results of many studies. He concludes: "On most criteria, conventional intelligence tests are not found to be biased" (p. 8). The same conclusion has been reached by Reynolds (1982) in his comprehensive updated review of the literature. Jensen (1980) provided a comprehensive review of the literature concerning the predictive validity on different external criteria and the internal criteria or psychometric features of the mental tests most commonly used. Jensen's conclusions in both areas reveal an overwhelming consistency in support of the "non bias" position. More recently Hunter, Schmidt, and Rauschenberger (1984) have reviewed the empirical findings about tests as estimates of the ability of majority and minority groups to predict job performance. They, concluded that "massive empirical evidence has now accumulated showing that tests are fair to minority members" (p. 93).

A similar conclusion can be drawn from studies of the factorial structure of the Wechsler Scales (Cohen, 1957a; Cohen, 1957b; Cohen, 1959; Silverstein, 1977; Kaufman, 1975; Kaufman, 1979; Reschly, 1978; Vance and Wallbrown, 1978; and Van Hagen and Kaufman, 1975). These stud-

ies indicate that the factorial structure found in Anglo and several minority groups are the same, which demonstrates that these tests are basically measuring the same intellectual characteristics in all groups investigated. Recently, Johnson and Bolen (1984), comparing the factor structure of the WISC-R for blacks and whites, found that this test measures the same theoretical abilities in both ethnic groups. If the factor analytic structure of tests is evidence of their construct validity (Anastasi, 1976), the evidence from these studies clearly supports the construct validity of the Wechsler Scales with minority groups.

In addition to the findings of the factorial structure studies and the findings summarized by Vernon, Gordon and Rudert, Reschly, Jensen, Reynolds, and Hunter et al., recent research findings continue to support the non-bias position. Reynolds (1983) evaluated bias in construct validity and criterion-related validity of intelligence tests and of personality scales and concluded that little evidence exists to substantiate claims of bias for well constructed, properly standardized tests. Oakland (1983) in a study with 1st to 8th graders stratified by racial-ethnic groups (White, Black, and Mexican-American) and social class (middle and lower), concluded that there was little evidence of racial-ethnic or social class bias. Sandoval et al. (1983) found that the item difficulty curves for Anglo, Black, Chicano, and Bermudian compared on each of the verbal subtests of the WISC-R, were remarkably parallel. The results obtained by Ross and Reschly (1983) using several statistical indices, indicated no or negligible bias against Blacks and Mexican-Americans on the Information, Arithmetic, Similarities, Vocabulary, Comprehension, and Picture Completion of the WISC-R.

The non-bias position has been well documented recently with linguistic minorities, particularly with Mexican-Americans. Studies with the Peabody Picture Vocabulary Test - Revised (PPVT-R) comparing Anglo-American and Mexican-American groups, suggested that bias in content , as defined by item-ethnicity interactions of the PPVT-R, was minimal (Argulewicz and Abel, 1984) and that, although the Mexican-American subjects scored almost a standard deviation below the norm, there was no reliable evidence of test bias (Argulewicz et al., 1983). Reynolds and Piersel (1983) examined the reliability of the Boehm Test of Basic Concepts (BTBC) across groups (White and Mexican-American), across group factorial congruence, and equivalence of regression systems in the prediction of early school achievement. They concluded that their findings do not support claims of cultural bias in the BTBC. The results of the investigation of cultural bias in the 46 verbal items of the McCarthy Scales of Children's Abilities (MSCA) indicated that the majority of the items were free from cultural bias, only 3 items reflected systematic cultural bias (Murray and Mishra, 1983).

All things considered, it can be concluded that the tests predict equally well for middle class, for economically disadvantaged, and for linguistic minority students. The psychometric properties of the tests seem to be equal for different ethnic groups, this means that the tests basically behave in the same manner with all ethnic groups. However, the empirical evidence does not support the claim that social justice has been well served by the use of the tests. On the other hand, there is no empirical evidence indicating that not using tests in the decision making process will result in better and more just decisions. Finally,

the present evidence does not indicate the nature of the intelligence being assessed, academic or general intelligence, and less whether intelligence is genetically or environmentally determined. These issues go far beyond the scope of the present study and consequently are not addressed here.

2. The problems of translation

Translating tests, in addition to the question of norms, poses special difficulties. The following are some problems which have been identified: "equivalence of concepts and organized ideas" (Sanchez, 1934b), the presence of idiomatic expressions in the one (language) without exact counterparts in the other, variations in meaning of apparently equivalent words, and problems of dialect variations (William, 1971). Nevertheless these alleged difficulties may be reduced when the intended purpose is to translate the words of vocabulary tests. The aim of this translation is not to transmit the exact feelings, but rather to construct a list of words as similar as possible to the original one in both content and difficulty as measured by the frequency of use. From this point of view, since there is a relationship between the frequency of use and recall (Borude, 1971; Macnamara, et al., 1972), the main concern has to be the level of difficulty of the words as measured by their frequency of use. Moreover, when the intent is to translate isolated words out of context, the associative meaning of a word is reduced to a minimum and only the denotative or direct meaning is of interest. As a

consequence, "transfer of denotative meaning, though difficult and at times impossible, can, as a rule, be exacted in a more or less satisfactory manner" (Winter, 1961).

The review of the literature on test translation shows that the problem of translating vocabularies of intelligence tests into Spanish, the language chosen for this study, has been subject to incomplete solutions. Boder (1925) in his translation into Spanish of Binet's scale in Mexico, compiled a list with new words, but he took into consideration neither the level of difficulty nor the content of both lists. He ordered the Spanish words according to their own difficulty as measured by the passing percentage on the items. The passing percentage is an acceptable criterion in cases in which new norms and tables are compiled or the previous norms are readjusted. When the original tables continue to be used, the passing percentage is not an acceptable criterion (see discussion above). Rodriguez Bou (1950) in his study of the Vocabularies of the Inter American Test of General Ability and the Inter American Test of Language from English into Spanish, took into consideration the content and in some way the level of difficulty of the words as measured by their frequency of use, but he did not compare two samples in order to see whether similar results would be obtained with both. This could be an indication that his method of translation was adequate. Wechsler (Manual del Departamento de Instruccion Publica, 1959) translated the original words into Spanish (Puerto Rico) without considering the level of difficulty, and then ordered the words according to the passing percentage. Cerda (1960) translated the Wechsler tests W-B II, WISC, and WAIS into Spanish (Spain). He provided three vocabularies without any

explanation of the bases for choosing the words and for ordering them in the manner he did. In his translation of the Stanford-Binet Test in Spain, Germain (1970) tried several words choosing them according to a better or more normal distribution of the results and a greater facility in the scoring, but he did not consider their level of difficulty. A similar solution to the one in Puerto Rico was attempted by Ramos Lopez (1970) in Peru. She adopted the translation made in Puerto Rico and ordered the words according to the percentage passing the items, but she did not take into consideration the level of difficulty of the words nor did she take into consideration the several meanings of some words.

3. Theoretical aspects of linguistic and semantic development

Frequency of use has been pointed out by linguists as one of the most important and significant characteristics of words because of the existing relationships between frequency of use and phonic, morphologic, semantic, etymological, and other aspects of words (Zipf, 1949a; Zipf, 1949b; Guiraud, 1954). These relationships have been found in languages as different as the Peipingese Chinese, the American Indian languages of Nootka, Dakota, and Plains Cree, and the Western European languages among others (Zipf, 1949a; 1949b). It seems that some general characteristics of words, such as frequency of use, and phonic, morphologic, semantic, and etymological aspects, are common to most languages. It would be safe, then, to assume that the procedures and conclusions of

this study, which basically relies upon word characteristics, are, to
some extent, applicable to languages other than English and Spanish.
Linguists do not indicate, however, the relationship between frequency
and difficulty. This appears to be due to the fact that their studies
attempt to analyze behavioral or literary aspects rather than to present
guideliness for test construction or item difficulty.

That there is, however, a relationship between frequency and diffi-
culty and that frequency is a useful word characteristic to evaluate age
levels and intelligence levels, was recognized, although not proven, by
Robert Herndon, Chairman of the Committee of Modern Languages of the
American Council on Education, in his foreword to Helen Eaton's (1967)
Word Frequency Dictionary. He claimed that

> The educationist and the teacher of English and the modern
> foreign languages have in this comparative semantic list a
> guide for selecting vocabularies graded to meet the various
> age levels and intelligence levels. This is of particular
> importance for those working with pupils in two languages,
> where it often happens that a concept which is quite usual in
> one of them may find expression in the other only by means of
> a word of lower frequency and therefore less likely to be
> familiar to the learner. (p. viii)

4. Theories of semantic development

Most of the semantic research on words has been done within the con-
text of semantic functions of words in utterances. The focus of this
study is on words as used to refer to or to represent external objects

and events, or word meaning. It will be helpful to review the theories that have attempted to explain how meaning is acquired as an important component of the process of language development. The most relevant theories of semantic development have been summarized by Clark (1973) in the following four hypotheses:

a. The grammatical relations hypothesis (Mc Neill, 1970)

At the beginning of language development, the child has some kind of sentence-meaning dictionary. In this dictionary the lexical entries are tagged with all the grammatical relations that are used in the one-word stage. There are not semantic features at this stage, but only information about grammatical relations. There is, consequently, a primacy of grammatical relations during the earliest stages of language acquisition. Later, at the stage of two-word sentences, the child begins to recognize his dictionary according to a word-meaning rather than sentence-meaning. The child's lexicon develops in two ways, horizontally and vertically. In the horizontal development, when a word enters the child's dictionary, only some semantic features associated with the word enter with it. The child will complete the dictionary entries by adding -horizontally- some other features. The vertical development of the lexicon implies that when a word enters the child's dictionary, all semantic features of a word enter with it at the same time. Initially the features remain separated from each other because they are not recognized as being the same in different entries within the lexicon. The

semantic development consists of collecting separate occurrences of the same features into a unified group.

b. The generalization hypothesis

(Anglin, 1970)

Anglin indicates that semantic development is a generalization process which goes from the perception of concrete semantic relations between words (e. g., a boy and a horse both eat) to the most general or abstract relations (e. g., a boy and a horse may both be animals). Initially children perceive specific concrete relations between words; they perceive words as bound by concrete relations. Therefore the lexicon of children will reflect these concrete relations by means of concrete words. As their semantic knowledge increases, they perceive the more abstract features that relate words and consequently they are able to make generalizations which are expressed by using superordinately more general or abstract lexical items.

c. The universal primitives hypothesis

(Postal, 1966; Bierwisch, 1970)

This hypothesis suggests that a set of universal semantic primitives and the rules for their combinations into lexical items underly all languages. These semantic components do not represent external physical properties of objects, but rather the categories or principles according

to which objects and situations are classified. In other words, the semantic primitives represent the psychological conditions according to which the physical and social environment is processed. These semantic features are identical in all languages; they are part of the capacity for language common to human beings.

d. The semantic feature acquisition hypothesis (Clark, 1973)

This hypothesis states that when children start to use the first words, they do not know the full meaning; they only have partial entries for them in their lexicon. These partial entries correspond to partial aspects of the external objects. The development of semantic knowledge consists of adding more features of meaning to the lexical entry of the word. When children are able to combine all the critical features of the word in their lexical entry, the meaning for that particular word will be equivalent to the meaning for adults. Children do not develop adult meaning for all words at the same time, but rather gradually and separately for each word. More general features are acquired first, specific features are added later.

e. The referring - reference theory

(Macnamara, 1982).

The act of referring is the basic concept in this theory. As explained by Macnamara "referring is a three-place predicate: a person uses a word to refer to an object" (p. 227). In order to learn new words, the child has to connect the word with the object by interpreting the speaker's intention of referring. Because the child can not get it all right the first time, this process is facilitated whenever a child hears an adult utter a name in the presence of a salient object.

Reference, on the other hand, is a property of certain symbols that are acquired from the acts of referring. Reference is a relation between words and objects. The child needs the ability to detect which words are referring and to attribute to them the permanent capacity to refer. This permanent capacity is reference.

The mental ability that capacitates the child to connect the reference with the object is, according to Macnamara, a primitive of cognitive psychology which is an innate ability, as are also some conceptual categories, principles, and evaluative devices that facilitate the process of learning a language. Macnamara says: "The child does not learn what referring is, nor does he learn to refer, ... a child naturally interprets certain events as acts of referring" (p. 228) by imposing an explanation on an observable event. As a consequence the child learns the meaning of words.

f. The lexical contrast theory

(Clark, 1983)

A new theory has been advanced recently by Clark (1983) that should be included in the review of the theories of semantic development. It has been called the Lexical Contrast Theory. This theory of meaning acquisition, that is still in the process of being fully developed, relies on lexical contrasts of meaning rather than on semantic features. Two basic principles, namely "contrast" and "conventionality" govern the acquisition of meaning and account for the direction children as well as adults follow in the process of acquiring word meanings. The <u>Principle of Contrast</u> states that "the conventional meanings of every pair of words (or word-formation devices) contrast" (p. 820). The <u>Principle of Conventionality</u> states that "for certain meanings, there is a conventional word or word-formation device that should be used in the language community" (p.820). In the process of acquiring communication skills, children use the conventional words for objects, situations, and states on the assumption that these words have consistent meanings from one occasion to the next. When children and adults feel the need to make further distinctions or to convey meanings for which they do not have words in their lexicon, a lexical gap exists. In order to fill these gaps in their vocabularies, children and adults will try to look for new words. They assume that these newly acquired word meanings contrast with those meanings already known to them.

Evaluation of the theories and selection of word

characteristics

Three word characteristics surface from these theories of semantic

development which possibly should be considered as good measures of word

difficulty: the semantic features, the universal semantic primitives,

and the concrete-abstract dimension. The semantic features, however, and

the universal primitives do not appear to be objective and quantifiable

measures. The linguists have not provided specific guideliness to objec-

tively identify what these features or universal primitives might be.

As Clark (1983) indicates

> There is considerable evidence against the semantic feature
> hypothesis. The most critical concerns semantic features
> themselves: the criteria for their identification are vague,
> their status vis-a-vis features postulated in accounts of
> adult meanings is unclear, and only certain domains of the
> lexicon allow word meanings to be decomposed into features or
> components of meanings. (p. 819)

It seems that the features could be different in each particular

individual and consequently they are in the domain of individual differ-

ences rather than in the domain of objective measures. The semantic

primitives have been defined as .categories or principles to classify

objects and situations. These semantic primitives seem to be epistemolo-

gical principles, in a Kantian or Piagetian sense, rather than word

characteristics. They reside in the mind not in the words; consequently,

this study that looks for word characteristics as measures of diffi-

culty, cannot consider the semantic primitives for measuring purposes.

It is true that words relying directly on these primitives could be eas-

ier than words further removed from them. However, linguists have not

indicated clearly which words are close or far from these semantic prim-
itives. The dimension concrete-abstract is objective and measurable in a
dichotomous way. This dimension is accounted for by the fact that the
procedure proposed in this study considers first the content or meanings
of each word. This procedure implies that the list of words in both lan-
guages will have equal numbers of concrete and abstract words.

Macnamara's theory offers clear support for the frequency of use
selection as a measure of word difficulty. The theory indicates that
children learn the meaning of words by interpreting the act of refer-
ring. It also indicates that children may not get the meaning all right
the first time they hear a person using a word to refer to an object.
Consequently, the more frequently a word is heard to refer to an object,
the better its meaning will be learned.

The Lexical Contrast Theory, as it stands today, appears to have more
explanatory power than previous theories. It offers two basic principles
that account for meaning acquisition in children as well as in adults,
and a theoretical explanation for some phenomena observed in children's
language not fully explained by other theories, such as overextensions,
use of general-purpose words, coinages, etc. The theory does not offer,
however, any basis for identifying word characteristics that could serve
as objective measures of word difficulty.

The review of other partial and less comprehensive theories of mean-
ing acquisition or semantic development points out some other word char-
acteristics that could be considered as measures of word difficulty.
Clark (1983), in discussing how words are acquired, mentions three

important groups of words: basic, superordinate, and subordinate. Basic words refer to terms that children acquire first because they are of the greatest utility to them (Brown, 1958). Superordinate and subordinate words are labels for more general or more specific categories respectively. According to these word characteristics only three levels of difficulty could be established, this is obviously not enough to cover the whole range of vocabulary. In addition, although some attempts have been made to identify basic levels of words (Rosch et al., 1976) these levels are not the same from language to language (Clark, 1983). The boundaries that separate category membership are highly diverse and unclear, particularly for superordinate and subordinate words. This fact would constitute a serious difficulty in classifying words as belonging to one or another group. These three word categories as well as others similar to them (count and mass nouns, object, situation, and state words, simple and complex meanings), are, to some extent, indirectly accounted for in the procedure followed in this study because the content or meaning of the English and non-English words is taken into consideration.

Finally, another important linguistic aspect is that some words have several meanings. Number of meanings could be considered as a measure of word difficulty because it has been hypothesized that words with multiple meanings have multiple entries in the lexical memory. This assumption was tested by Rubenstein, Garfield, and Millikan (1970) and they concluded that homographs -words with several meanings- have separate entries in the lexical memory for each distinctive meaning. The multiple entry hypothesis was supported also by Jastrzembski and Stanners (1975)

and by Craik and Tulving (1975). Consequently, when the individual is requested to verbalize the meaning of a word presented auditorily, the recall of any meaning will be easier for words with more entries in the dictionary than for words with fewer entries. Based on these research findings, number of meanings could be considered an objective measure of word difficulty. However, from a practical point of view, this measure presents some problems. For example, it would be difficult to determine what a specific "meaning' is in contrast to "other different meaning" for a particular word. Although some entries in the dictionary for a particular word are listed as separate, however, they are so close in meaning that they could be considered as having one meaning rather than two, three, or more, and consequently they should be grouped. In which instances this grouping of meanings should take place could be a matter of discussion and could introduce some subjective judgments into a method that tries to be as objective as possible. The review of these theories has not presented any other semantic aspects which could be more objective and quantifiable than meaning and frequency of use. Consequently, these two word characteristics were selected for the present study.

Summary

The literature related to the issues involved in this study has been reviewed. The concept of bias and the results of empirical studies on bias support the idea that mental tests are not biased against minori-

ties. The problems of translating vocabulary tests and the procedures of translating them into Spanish have been reviewed and discarded as unsatisfactory. Some theoretical aspects of linguistic and semantic development were also examined with the intention of pointing out the relationship between frequency of use and other important linguistic aspects of words. Finally, the theories of semantic development have been analyzed in order to determine which linguistic variables can be used to make English and non English lists of words equal. Of all word characteristics that surfaced from the review of the theories of semantic development and the consideration of the theoretical aspects of linguistic and semantic development, only two, meaning and frequency of use, have been considered in the present study. They are clear, objective, and in the case of frequency of use, quantifiable and easy to analyze statistically. The methodology presented in chapter III is based on these two word characteristics.

CHAPTER III

METHODOLOGY

As pointed out previously, the purpose of this study was to investigate linguistic variables which could account for word difficulty in translation or construction of vocabulary tests to be used in two languages. It was assumed that any languages could be used. English was chosen for the obvious reason that it is the language of instruction in American schools. Spanish was the comparison language chosen because most of the bilingual children in this country are Spanish speaking. The investigator is also Spanish speaking and familiar with the problems of translation and administration of tests to Spanish speaking children.

This study consisted of two parts. In the first part a lexical analysis of a set of English words and non-English words was conducted with the purpose of compiling two lists, one in English and one in Spanish, which are similar in content and difficulty as measured by the frequency of use. The Spanish words comprise the experimental list to be used in part two of this study. A second non-English (Spanish) list, comparable to the first in meaning, was compiled using word by word translation matching the English words and without controlling the word characteristic frequency of use which was incorporated into the experimental list. This second set of Spanish words served as a control list.

In the second part of the study, and for the purpose of experimental validation of the translation procedures, the English list was administered to an English-speaking sample and the Spanish experimental and control lists were administered to a Spanish-speaking sample similar to the English sample. Subjects were required to verbalize the meanings of the words. These responses were scored following the general scoring principles given by Wechsler (1974, p.161-162. See Appendix Four) in the WISC-R manual. These principles have been commonly accepted as adequate for psychological testing purposes.

It was hypothesized that: (a) the English sample and the Spanish sample would obtain results not significantly different on the experimental list, and (b) the results of the Spanish sample on the control list would be significantly different from the results obtained by the English speaking sample on the same list.

Part One. The development of the English and Spanish lists

Twenty four English and twenty four Spanish words which matched the meaning and frequency of use of the English words were selected from Eaton's (1967) word frequency dictionary. In this dictionary the words, English and Spanish, are grouped according to their frequency of use into seven groups of diminishing frequency. The words are followed by a number and the letters "a" or "b." The number after a word indicates the frequency by thousand of use to which the word belongs. The letters "a" or "b" indicate respectively the first or the second five hundred in that thousand. Thus, Distance 1b and Distancia 1b occur in the second five hundred of the first thousand, indicating that they have approximately the same frequency of use, and therefore in this study were considered equal in difficulty according to the selected criterion of difficulty. Trecho 4a, however, that has the same meaning as Distance as well, occurs in the first five hundred of the fourth thousand, indicating that its frequency of use is different. Therefore it was considered in this study to be more difficult than Distance and Distancia. For the purposes of statistical analysis the letters "a" and "b" were assigned the numerical values of "0.0" and "0.5" respectively. The English and the Spanish experimental words selected for this study were followed in Eaton's Dictionary by the same number and letter, with the exception of the words in the 6th and 7th groups. These were followed only by the numbers, because at these levels the English words are not followed by the letters "a" or "b" in the Eaton's dictionary.

There were two words in each group from 1a to seven. There was an equal number of nouns and adjectives in both lists. The number of verbs was not controlled because English words are frequently nouns and verbs, which is not so in Spanish. For the purpose of statistical analysis the words were arbitrarily classified into three difficulty levels: (a) easy, words from groups 1a to 2b in Eaton's dictionary, (b) medium, from 3a to 4b, and (c) difficult, from 5a to seven. Thus there were eight words at each difficulty level. The selected words met the two following criteria for equivalency:

(a) The content equivalency was controlled by choosing an English word and its corresponding Spanish counterpart in Eaton's dictionary. Both words were considered equivalent in content because Eaton grouped the English and non English words by their similar thought content or general meaning. In this dictionary only the most general and important meaning of each word is considered when the highest frequency is specified, the so called primary meaning, which might be neither the original nor the etymological meaning of the word, but the meaning which has the most frequent usage (Zipf,1949b). The Index of Equivalency is 100%, indicating that all selected Spanish words are equivalent in thought content to their corresponding English words.

(b) The second criterion of equivalency, difficulty as measured by frequency of use, was controlled by Eaton's dictionary as well. Eaton has paired English and non English words according to frequency of use; for the purpose of this study, that is assumed to be a good measure of word difficulty. This Index of Equivalency is

also 100%, indicating that all selected Spanish experimental words
are equally difficult as their corresponding English words.

The Spanish list so compiled was the experimental list. In order to com-
pile the control list, 24 Spanish words were selected from the Eaton's
list, words that match the meaning of the English words but not the fre-
quency of use. For reason of statistical comparisons, these control
words were also arbitrarily incorporated into the three difficulty lev-
els, not according to their corresponding frequency of use, but accord-
ing to the frequency of use of the English and Spanish experimental
lists. The lists of selected words are presented in Appendix One.

Part Two. Experimental validation

Following the construction of the English and the Spanish experimental and control lists, the English list was administered to English speaking subjects and both Spanish lists were administered to Spanish speaking samples, one Mexican and one Puerto Rican. The following null hypotheses were tested:

Ho 1: There will be no significant performance differences between the mean of the total English sample and the mean of the total Spanish speaking sample on the experimental list.

Ho 2: There will be no significant performance differences on the experimental list, first between the mean of an English sample matched to a Mexican sample and secondly between the mean of an English sample matched to a Puerto Rican sample and the means of the Mexican and Puerto Rican samples.

Ho 3: There will be no significant performance differences among the means of the total English sample and the means of the total Spanish sample across levels of word difficulty (easy, medium, and difficult) on the experimental list.

Ho 4: There will be no significant performance differences on the experimental list, first among the means of an English sample matched to a Mexican sample and secondly among the means of an English sample matched to a Puerto Rican sample and the means of the

Mexican and Puerto Rican samples across levels of word difficulty (easy, medium, and difficult).

Ho 5: There will be no significant performance differences between the mean of the total English sample and the mean of the total Spanish speaking sample on the control list.

Ho 6: There will be no significant performance differences on the control list, first between the mean of an English sample matched to a Mexican sample and secondly between the mean of an English sample matched to a Puerto Rican sample and the means of the Mexican and Puerto Rican samples.

Ho 7: There will be no significant performance differences among the means of the total English sample and the means of the total Spanish sample across levels of word difficulty (easy, medium, and difficult) on the control list.

Ho 8: There will be no significant performance differences on the control list, first among the means of an English sample matched to a Mexican sample and secondly among the means of an English sample matched to a Puerto Rican sample and the means of the Mexican amd Puerto Rican samples across levels of word difficulty (easy, medium, and difficult).

It was expected that null hypotheses numbers one, two, three, and four would not be rejected, but null hypotheses numbers five, six, seven, and eight would be rejected.

Subjects.

One hundred and sixty (160) eighth graders from Chicago schools were selected for this study. Pairs of children, one English and one Spanish speaking (Mexican and Puerto Rican) were selected from the same regular classroom, children who had never been referred for psychological evaluation. They were matched on age, sex, and academic achievement as estimated by their teachers. There is some recent evidence that demonstrates that teachers are capable of making accurate judgements of the achievement levels of their pupils (Hoge and Butcher, 1984). Students could not be matched on academic achievement by their results on the Iowa Test of Basic Skills (ITBS) because some Hispanic subjects were selected from bilingual programs and frequently these students do not have ITBS results in their files. The Mexican and Puerto Rican subjects were not matched among themselves because they were taken from different schools. The total sample was distributed as follows: 80 English speaking and 80 Spanish speaking, 40 Mexicans and 40 Puerto Ricans. There was an equal number (40) of boys and girls in each group, English and Spanish. A numerical description of the sample is presented in the following table.

TABLE 1

Number of Subjects by Sex and Race.

	English	Hispanics		Total
		Mexican	Puerto Rican	
Boys	40	20	20	80
Girls	40	20	20	80
	80	40	40	160

The Hispanic groups were Spanish speaking. To be considered Mexican or Puerto Rican both parents had to be born in Mexico or Puerto Rico respectively. The children could have been born either in their parents' country or in the United States. The English speaking children were all white, born in the United States, and they spoke no language other than English. The Spanish speaking children, although they spoke some English, all spoke Spanish fluently and without any accent. They were tested by talking in Spanish to a bilingual teacher in the school. If these children could be called bilinguals, their bilingualism was of the kind characterized as "bilingual parallelism" by Anastasi and Cordova (1953) which implies that the child is able "to express himself in all types of situations in at least one language..., the second language provides a parallel means of expression in some or all situations" (p. 3).

Procedure

The 24 words were given individually to each subject by experienced school psychologists, saying: "This is a vocabulary test. I want to know how many words you know. I am going to say a word and you tell me what the word means". The English speaking sample was administered the English list, while the Spanish speaking samples were administered both, the experimental and control lists, in two sessions no less than two weeks apart. Half of the Hispanic children were administered the experimental list first, and half the control list first. The same subject samples controlled for individual differences.

In order to control for the "I don't know" effect when the difficulty of the words increased, the difficulty was counterbalanced by alternating the easy, medium, and difficult words as defined in this study. It has been observed, especially with children, that after three or four consecutive failures when the level of difficulty of test items increases, they tend to say "I don't know" and stop making an effort to provide the right answers. Most intelligence tests have discontinuance rules after a specific number of consecutive failures. Since this study required that all words be administered to all children, the counterbalanced difficulty of words prevented the "I don't know" effect from taking place. The list of words in the order given to the subjects is presented in Appendix Two.

The responses were scored two, one, or zero according to the scoring criteria provided by Wechsler in the WISC-R manual (Wechsler, 1974,

p.161-162. See Appendix Three).

Design and Statistical Analysis

The design consisted of three lists of 24 words each. There was an English list and two Spanish lists -Experimental and Control. The lists were arbitrarily divided into three levels of difficulty -easy, medium, and difficult. The lists were administered to English and Spanish-speaking subjects matched on age, sex, and academic achievement. The subjects were half boys and half girls. The Spanish subjects were half Mexicans and half Puerto Ricans. The independent variable was the frequency of use of the words, the dependent variable was the score obtained by the subjects. This experimental paradigm permitted multiple comparisons between the English and the two Spanish lists across levels and across words for the total Spanish sample, for Mexicans and Puerto Ricans, and for boys and girls separately. The significance of differences between means were calculated with a MANOVA for lists and levels, and with univariate analyses for words. The SPSS computer program was used for all statistical analyses.

Summary

This chapter has presented the procedures that were followed to compile the English and the two Spanish lists, experimental and control. Twenty four English words were selected from different levels of frequency of use. The 24 Spanish words in the experimental list matched the corresponding English words in meaning and frequency of use. The Spanish control words, on the other hand, matched the meaning but not the frequency of use. In the second part, the experimental validation, eight null hypotheses were formulated and the lists of words were administered to English and Spanish speaking samples to verify whether or not the selected variables, particularly frequency of use, made both lists, English and Spanish experimental, equal in difficulty. It was assumed in the present study that obtaining similar results with the English and Spanish experimental lists and different results with the English and Spanish control lists would confirm the validity of frequency of use as an adequate measure of word difficulty.

CHAPTER IV

RESULTS OF THE EXPERIMENTAL VALIDATION

After the administration of the lists to the corresponding samples,

the verbal responses were scored according to the criteria explained in

the Procedure section of the previous chapter. Means and standard devi-

ations were calculated for the three lists, English, Spanish experimen-

tal, and Spanish control, for the three levels of word difficulty, Easy,

Medium, and Difficult, and for single words. Means and standard devia-

tions were also calculated for boys and girls, and for Mexicans and

Puerto Ricans separately, for the whole word list and for the three lev-

els of word difficulty. These results are shown in Tables 2, 3, and 4.

Tests of significance of differences between the mean of the total Eng-

lish sample and the total Spanish samples were conducted for the whole

list, for the three levels of word difficulty, and for the individual

words, comparing the English list with the Spanish experimental and with

the Spanish control lists. The means of the whole list were also com-

pared for boys and girls separately, for English and Mexicans, and for

English and Puerto Ricans. Only matched children were compared among

themselves in any test of significance of differences. Consequently,

there were 80 subjects in each group for the total sample comparisons

and only 40 in each group of boys and girls and Mexicans and Puerto

Ricans. When these two ethnic groups were compared with the English

group, the English subjects with their matched Mexican or Puerto Rican

counterparts were the population of comparison.

TABLE 2

Means and Standard Deviations of Performance Scores for the
Three Lists by List, Level of Word Difficulty, and Sex.

	English		SpanExp(1)		SpanCont(2)	
	M	SD	M	SD	M	SD
List	1.033	.882	1.050	.879	.884	.877
Level I	1.435	.722	1.421	.659	.634	.835
Level II	1.017	.918	1.190	.879	.740	.899
Level III	.646	.812	.537	.832	1.278	.753
Boys	1.061	.883	1.095	.874	.939	.873
Girls	1.005	.880	1.004	.883	.829	.878

(1) Spanish Experimental List.

(2) Spanish Control List.

TABLE <u>3</u>

Means and Standard Deviations of Performance Scores for the Two
Spanish Groups by List, Level of Word Difficulty, and Sex
in the Experimental List and by List in the Control List.

	<u>SpanExpMex</u>(<u>1</u>)		<u>SpanExpPR</u>(<u>2</u>)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
List	1.008	.872	1.091	.885
Level I	1.353	.688	1.490	.623
Level II	1.112	.892	1.268	.861
Level III	.559	.828	.515	.837
Boys	1.060	.875	1.131	.872
Girls	.956	.867	1.052	.897

	<u>SpanContMex</u>(<u>3</u>)		<u>SpanContPR</u>(<u>4</u>)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
List	.872	.881	.895	.874

----

(1) Mexican Group in the Spanish Experimental List.

(2) Puerto Rican Group in the Spanish Experimental List.

(3) Mexican Group in the Spanish Control List.

(4) Puerto Rican Group in the Spanish Control List.

TABLE 4

Means and Standard Deviations of Performance Scores
by Words in the Three Lists.

English	M	SD	Span.Exp.(L)	M	SD	Span.Con.(2)	M	SD
1. Fire	1.900	.301	Fuego	1.900	.301	Lumbre	1.462	.778
2. Deep	1.487	.655	Profundo	1.625	.700	Hondo	1.662	.501
3. Distance	1.512	.551	Distancia	1.275	.449	Trecho	.087	.284
4. Famous	1.725	.527	Famoso	1.600	.492	Célebre	.425	.725
5. Slave	1.400	.586	Esclavo	1.450	.548	Siervo	.837	.848
6. Bitter	1.112	.856	Amargo	1.187	.713	Acerbo	.037	.191
7. Mercy	.850	.872	Misericordia	.937	.847	Clemencia	.300	.603
8. Calm	1.500	.711	Calmar	1.400	.586	Sosegar	.262	.545
9. Display	1.050	.912	Desplegar	.775	.899	Exhibir	.812	.812
10. Perish	.412	.774	Perecer	.687	.922	Sucumbir	.050	.219
11. Peasant	.700	.862	Campesino	1.587	.774	Labriego	.112	.420
12. Fortunate	1.437	.793	Afortunado	1.312	.865	Venturoso	.225	.594
13. Mansion	1.837	.488	Mansión	1.625	.769	Vivienda	1.712	.599
14. Contend	.400	.739	Competir	.987	.584	Contender	.050	.219
15. Nourish	.962	.877	Nutrir	1.137	.896	Alimentar	1.812	.505
16. Artificial	1.337	.794	Artificial	1.412	.790	Postizo	1.150	.843
17. Convent	.487	.811	Monasterio	.637	.917	Convento	1.375	.891
18. Treacherous	.237	.428	Pérfido	.100	.341	Traidor	1.112	.503
19. Exquisite	.525	.655	Primoroso	.375	.643	Exquisito	1.187	.730
20. Slap	1.712	.455	Bofetada	1.750	.515	Bofetón	1.462	.745
21. Caution	1.075	.823	Cautela	.912	.943	Precaución	1.425	.707
22. Incredible	.950	.809	Inverosímil	.437	.793	Increible	1.562	.613
23. Deplore	.037	.191	Deplorar	.050	.219	Lamentar	1.212	.774
24. Scaffold	.150	.479	Cadalso	.037	.191	Horca	.887	.795

(1) Spanish experimental list.
(2) Spanish control list.

The results of the comparisons for the English and Spanish experimental lists are shown in Table 5. The difference between the means of the two lists was not significant ($F_{(3838, 1)}$ = .343, $p > .05$), neither was the difference for boys ($F_{(1918, 1)}$ = .733, $p > .05$) nor for girls ($F_{(1918, 1)}$ = .0006, $p > .05$. Consequently null hypothesis one was not rejected. The results of the comparisons between means of the English sample matched to the Mexican and Puerto Rican groups and the means of these two ethnic groups, Mexican and Puerto Rican, are shown in Table 6. For the experimental list the difference was not significant either for Mexicans ($F_{(1918, 1)}$ = .7005, $p > .05$) or for Puerto Ricans ($F_{(1918, 1)}$ = 2.7002, $p > .05$). Consequently, null hypothesis two was not rejected.

TABLE 5

Means and Significance of Differences between Means
for English and Spanish Experimental Words.

English	M	Spanish	M	F	Sig.of F
1. Fire	1.900	1. Fuego	1.900	0.0	1.000
2. Deep	1.487	2. Profundo	1.625	1.642	.202
3. Distance	1.512	3. Distancia	1.275	8.926	.003**
4. Famous	1.725	4. Famoso	1.600	2.399	.123
5. Slave	1.400	5. Esclavo	1.450	.309	.579
6. Bitter	1.112	6. Amargo	1.187	.362	.548
7. Mercy	.850	7. Misericordia	.937	.413	.521
8. Calm	1.500	8. Calmar	1.400	.940	.334
Level I	1.435		1.421	.132	.716
9. Display	1.050	9. Desplegar	.775	3.684	.057
10. Perish	.412	10. Perecer	.687	4.171	.043*
11. Peasant	.700	11. Campesino	1.587	46.879	0.000**
12. Fortunate	1.437	12. Afortunado	1.312	.907	.342
13. Mansion	1.837	13. Mansion	1.625	4.347	.039*
14. Contend	.400	14. Competir	.987	31.079	0.000**
15. Nourish	.962	15. Nutrir	1.137	1.556	.214
16. Artificial	1.337	16. Artificial	1.412	.358	.550
Level II	1.017		1.190	11.902	.001**
17. Convent	.487	17. Monasterio	.637	1.200	.275
18. Treacherous	.237	18. Perfido	.100	5.044	.026*
19. Exquisite	.525	19. Primoroso	.375	2.131	.146
20. Slap	1.712	20. Bofetada	1.750	.237	.627
21. Caution	1.075	21. Cautela	.912	1.346	.248
22. Incredible	.950	22. Inverosimil	.437	16.356	.000**
23. Deplore	.037	23. Deplorar	.050	.147	.701
24. Scaffold	.150	24. Cadalso·	.037	3.793	.053
Level III	.646		.537	5.653	.018*
List	1.033		1.050	.343	.558
List-Boys	1.061		1.095	.733	.392
List-Girls	1.005		1.004	.0006	.979

* Significant at .05 level.
** Significant at .01 level.

TABLE 6

Means and Significance of Differences between Means among
  Ethnic Groups in the Experimental and Control Lists.

	English $\underline{M}$	SpaExpMex $\underline{M}(\underline{1})$	$\underline{F}$	Sig. of $\underline{F}$
List	1.0416	1.0083	.7005	.403
Level I	1.4218	1.3531	1.5607	.212
Level II	1.0343	1.1125	1.1877	.276
Level III	.6687	.5593	2.8415	.092

	English $\underline{M}$	SpaExpPR $\underline{M}(\underline{2})$	$\underline{F}$	Sig. of $\underline{F}$
List	1.0250	1.0916	2.7002	.100
Level I	1.4500	1.4906	.5631	.453
Level II	1.0000	1.2687	14.6030	.000**
Level III	.6250	.5156	2.8078	.094

	English $\underline{M}$	SpaConMex $\underline{M}(\underline{3})$	$\underline{F}$	Sig. of $\underline{F}$
List	1.0416	.8729	17.7760	.000**
Level I	1.4218	.6531	153.6819	0.000**
Level II	1.0343	.7125	20.1763	0.000**
Level III	.6687	1.2531	87.7229	0.000**

	English $\underline{M}$	SpaConPR $\underline{M}(\underline{4})$	$\underline{F}$	Sig. of $\underline{F}$
List	1.0250	.8958	10.2606	.001**
Level I	1.4500	.6156	183.6707	0.000**
Level II	1.0000	.7687	10.2825	.001**
Level III	.6250	1.3031	121.4232	0.000**

* Significant at .05 level.
** Significant at .01 level.
(1) Mexican Group in the Spanish Experimental List.
(2) Puerto Rican Group in the Spanish Experimental List.
(3) Mexican Group in the Spanish Control List.
(4) Puerto Rican group in the Spanish Control List.

At the three levels of word difficulty, the difference between means for the total sample (See Table 5) was not significant at level one, Easy words ($F(1278, 1) = .132$, $p>.05$). The difference was significant, however, at level two, Medium words, ($F(1278, 1) = 11.902$, $p<.01$) and at level three, Difficult words ($F(1278, 1) = 5.653$, $p<.05$). At level two the Spanish mean ($M = 1.190$) was greater than the English mean ($M = 1.017$), whereas at level three the English mean ($M = .646$) was greater than the Spanish mean ($M = .537$). Null Hypothesis three was consequently rejected at the Medium and Difficult levels, but was not rejected at the Easy level.

At the three levels of word difficulty the comparisons between the English sample matched to the Mexican group and the Mexican group were not significant at any level of word difficulty (See Table 6): Level I, Easy words ($F(638, 1) = 1.5607$, $p>.05$), Level II, Medium words ($F(638, 1) = 1.1877$, $p>.05$), and Level III, Difficult words ($F(638, 1) = 2.8415$, $p>.05$). For the English - Puerto Rican comparisons, only the difference at Level II was significant ($F(638, 1) = 14.6030$, $p<.001$). The difference was not significant either at Level I ($F(638, 1) = .5631$, $p>.05$) or at Level III ($F(638, 1) = 2.8078$, $p>.05$). Null hypothesis four was not rejected for English - Mexican comparisons at the three levels of word difficulty, nor rejected at level one and three for English and Puerto Rican comparisons. Null hypothesis four was rejected, however, at level two for the Puerto Rican group.

When the means for individual words were compared with 158 and 1 degrees of freedom, seven pairs of words were significantly different, one at level one, four at level two, and two at level three (See Table

5). Four English words (Distance, Mansion, Treacherous, and Incredible) yielded greater means than their corresponding Spanish experimental words (Distancia, Mansion, Perfido, and Inverosimil). Three Spanish words (Perecer, Campesino, and Competir) yielded greater means than their English counterparts (Perish, Peasant, and Contend).

When the number of individuals who knew the meanings of the words regardless of one or two point responses was considered, similar results were obtained. Chi-square comparisons were conducted to determine the significance of differences. As shown in Table 7 only seven pairs of words showed significant differences. Five of them (Peasant - Campesino, Mansion - Mansion, Contend - Competir, Treacherous - Perfido, and Incredible - Inverosimil) were the same pairs that showed significant differences when the means were compared. Distance - Distancia and Perish - Perecer were significantly different when the means were compared but not when the number of individuals who knew their meanings was considered for comparisons. The opposite was true for Bitter - Amargo and Exquisite - Primoroso that only showed significant differences when compared according to the number of individuals who knew their meanings, but not when compared according to their corresponding means. These findings support the conclusions reached with the previous findings.

TABLE 7

Number of Individuals who knew the Meaning of Words and
Chi-Square for Significance of Differences.
English and Spanish Experimental Lists.

	English	N		Span.Exp.	N	Chi-Square
1.	Fire	80	1.	Fuego	80	0.000
2.	Deep	73	2.	Profundo	70	.260
3.	Distance	78	3.	Distancia	80	.506
4.	Famous	78	4.	Famoso	80	.506
5.	Slave	76	5.	Esclavo	78	.172
6.	Bitter	55	6.	Amargo	67	4.968*
7.	Mercy	43	7.	Misericordia	49	.918
8.	Calm	70	8.	Calmar	76	1.954
9.	Display	49	9.	Desplegar	37	3.618
10.	Perish	19	10.	Perecer	30	2.666
11.	Peasant	35	11.	Campesino	66	30.016**
12.	Fortunate	65	12.	Afortunado	59	1.290
13.	Mansion	76	13.	Masion	66	5.070*
14.	Contend	20	14.	Competir	66	53.198**
15.	Nourish	48	15.	nutrir	52	.426
16.	Artificial	64	16.	Artificial	65	.038
17.	Convent	23	17.	Monasterio	28	.718
18.	Treacherous	20	18.	Perfido	6	8.269*
19.	Exquisite	35	19.	Primoroso	22	4.604*
20.	Slap	80	20.	Bofetada	77	1.358
21.	Caution	56	21.	Cautela	42	3.160
22.	Incredible	52	22.	Inverosimil	21	25.742**
23.	Deplore	3	23.	Deplorar	4	0.000
24.	Scaffold	8	24.	Cadalso	3	1.560

* Significant at .05 level.
** Significant at .01 level.

The same statistical procedures were conducted with the English and the Spanish control lists. The results of these comparisons are shown in Table 8. The difference between the mean of both lists was significant for the whole list with the total sample ($F(3838, 1) = 27.506$, $p < .001$). It was also significant for boys ($F(1918, 1) = 9.2280$, $p < .01$), for girls ($F(1918, 1) = 19.239$, $p < .001$), and for the three levels of word difficulty, Easy ($F(1278, 1) = 336.947$, $p < .001$), Medium ($F(1278, 1) = 29.628$, $p < .001$), and Difficult ($F(1278, 1) = 207.676$, $p < .001$) with the total sample. As shown in Table 6, the difference between the English sample matched to Mexican and Puerto Rican groups and the Mexican and Puerto Rican groups was significant for the whole list for Mexicans ($F(1918, 1) = 17.776$, $p < .001$) and for Puerto Ricans ($F(1918, 1) = 10.260$, $p. < 01$) as it was for both ethnic groups at the three levels of word difficulty (Mexican, Easy $F(638, 1) = 153.6819$, $p < .001$, Medium $F(638, 1) = 20.1763$, $p < .001$, Difficult $F(638, 1) = 87.7229$, $p < .001$, Puerto Ricans, Easy $F(638, 1) = 183.6707$, $p < .001$, Medium $F(638, 1) = 10.2825$, $p < .01$, Difficult $F(638, 1) = 121.4232$, $p < .01$. Consequently, null hypotheses five, six, seven, and eight were rejected.

TABLE 8

Means of Performance Scores and Significance of Differences
between Means for English and Spanish Control Words.

	English	M		Spa.Cont	M	F	Sig. of F
1.	Fire	1.900	1.	Lumbre	1.462	21.959	0.000**
2.	Deep	1.487	2.	Hondo	1.662	3.592	.060
3.	Distance	1.512	3.	Trecho	.087	422.503	0.000**
4.	Famous	1.725	4.	Celebre	.425	168.201	0.000**
5.	Slave	1.400	5.	Siervo	.837	23.781	0.000**
6.	Bitter	1.112	6.	Acerbo	.037	119.976	0.000**
7.	Mercy	.850	7.	Clemencia	.700	21.480	0.000**
8.	Calm	1.500	8.	Sosegar	.262	152.447	0.000**
	Level I	1.435			.634	336.947	0.000**
9.	Display	1.050	9.	Exhibir	.812	3.021	.084
10.	Perish	.412	10.	Sucumbir	.050	16.224	.000**
11.	Peasant	.700	11.	Labriego	.112	29.969	0.000**
12.	Fortunate	1.437	12.	Venturoso	.225	119.676	0.000**
13.	Mansion	1.837	13.	Vivienda	1.712	2.088	.150
14.	Contend	.400	14.	Contender	.050	16.472	.000**
15.	Nourish	.962	15.	Alimentar	1.812	56.320	0.000**
16.	Artificial	.337	16.	Postizo	1.150	2.094	.150
	Level II	1.017			.740	29.628	0.000**
17.	Convent	.487	17.	Convento	1.375	43.385	0.000**
18.	Treacherous	.237	18.	Traidor	1.112	140.355	0.000**
19.	Exquisite	.525	19.	Exquisito	1.187	36.432	0.000**
20.	Slap	1.712	20.	Bofeton	1.462	6.553	.011*
21.	Caution	1.075	21.	Precaucion	1.425	8.315	.004**
22.	Incredible	.950	22.	Increible	1.562	29.096	0.000**
23.	Deplore	.037	23.	Lamentar	1.212	173.556	0.000**
24.	Scaffold	.150	24.	Horca .	.8875	50.412	0.000**
	Level III	.646			1.278	207.676	0.000**
	List	1.033			.884	27.506	0.000**
	List-Boys	1.061			.939	9.228	.002**
	List-Girls	1.005			.829	19.239	.000**

* Significant at .05 level.
** Significant at .01 level.

TABLE 9

Number of Individuals who knew the Meaning of words
and Chi-Square for Significance of Differences..
English and Spanish Control.

English	N	Span.Cont.	N	Chi-square
1. Fire	80	1. Lumbre	80	0.0
2. Deep	73	2. Hondo	79	3.280
3. Distance	78	3. Trecho	7	122.316**
4. Famous	78	4. Celebre	23	80.660**
5. Slave	76	5. Siervo	30	59.146**
6. Bitter	55	6. Acerbo	3	73.128**
7. Mercy	43	7. Clemencia	18	16.556**
8. Calm	70	8. Sosegar	17	70.764**
9. Display	49	9. Exhibir	44	.640
10. Perish	19	10. Sucumbir	5	9.606**
11. Peasant	35	11. Labriego	6	27.578**
12. Fortunate	65	12. Venturoso	10	75.920**
13. Mansion	76	13. Vivienda	74	.106
14. Contend	20	14. Contender	4	12.548**
15. Nourish	48	15. Alimentar	76	28.098**
16. Artificial	64	16. Postizo	58	1.240
17. Convent	23	17. Convento	58	25.662**
18. Treacherous	20	18. Traidor	75	78.380**
19. Exquisite	35	19. Exquisito	71	36.226**
20. Slap	80	20. Bofeton	68	10.898**
21. Caution	56	21. Precaucion	70	7.318**
22. Incredible	53	22. Increible	76	21.164**
23. Deplore	3	23. Lamentar	63	92.840**
24. Scaffold	8	24. Horca	50	47.706**

* Significant at the .05 level.
** Significant at the .01 level.

When the means for individual words were compared (See Table 8), twenty pairs of words showed significant differences between the English and the Spanish samples, whereas only four showed no significant different results. The pairs of words with no significant results were Deep - Hondo, Display - Exhibir, Mansion - Vivienda, and Artificial - Postizo.

Similar results were obtained when the number of individuals who knew the meaning of the words rather than the means were compared. As shown in Table 9, nineteen pairs of words showed significant differences, whereas five pairs were not significantly different. Four of these five pairs were the same pairs that showed no differences in the comparisons of their corresponding means. The pair Fire - Lumbre was not significantly different in the number of individuals who knew the meaning of the word. All of the subjects in both samples knew in some way the meaning of the words.

Since, in the case of the English and Spanish control lists, the frequencies of use were different, it was very important to examine the results for single words to see the relationship between frequency of use and score. As shown in Table 10 there were fifteen English words whose frequency of use was greater than that of the corresponding Spanish control words and nine Spanish control words with higher frequency of use than that of the corresponding English words. With the exception of only three pairs of words (Deep - Hondo, Mansion - Vivienda, and Artificial - Postizo) all the words showed a direct relationship between their frequency of use and the corresponding mean, i. e., the more frequently a word is used in the language (the lower its Frequency of Use Index) the higher the obtained mean. The difference between the means

of these three pairs, however, was not significant.

Finally, Spearman Rank Order correlations were conducted to determine to what extent the rank order of the means of the English words was correlated with the rank order of the means of the Spanish words in both the experimental and control lists. As expected the correlation for the English and Spanish experimental lists was significant (Rho = .7867 > .485 the value required for significance at .01 level) and for the English and Spanish control lists was not significant (Rho = .1604 < .343 the value required for significance at .05 level). Although the obtained rank order of the words according to their corresponding means was not perfect when compared with the rank order of the words according to their frequency of use, the correlation between the obtained (means) and the expected (frequency of use) orders was significant at the .01 level both in the English (Rho = .6096 > .485) and the Spanish (Rho = .6290 > .485) lists.

TABLE 10

Frequency of Use Indices and Mean Performance Scores
for English and Spanish Control Lists.

	English List	Frequency of use	M		Spanish List	Frequency of use	M
1.	Fire	1.0	1.900	1.	Lumbre	2.5	1.462
2.	Deep	1.0	1.487	2.	Hondo	2.0	1.662*
3.	Distance	1.5	1.512	3.	Trecho	4.0	.087
4.	Famous	1.5	1.725	4.	Celebre	2.0	.425
5.	Slave	2.0	1.400	5.	Siervo	5.0	.837
6.	Bitter	2.0	1.112	6.	Acerbo	6.0	.037
7.	Mercy	2.5	.850	7.	Clemencia	5.5	.300
8.	Calm	2.5	1.500	8.	Sosegar	3.5	.262
9.	Display	3.0	1.050	9.	Exhibir	5.0	.812
10.	Perish	3.0	.412	10.	Sucumbir	4.5	.050
11.	Peasant	3.5	.700	11.	Labriego	4.0	.112
12.	Fortunate	3.5	1.437	12.	Venturoso	4.0	.225
13.	Mansion	4.0	1.837	13.	Vivienda	3.5	1.712*
14.	Contend	4.0	.400	14.	Contender	6.5	.050
15.	Nourish	4.5	.962	15	Alimentar	2.5	1.812
16.	Artificial	4.5	.337	16.	Postizo	6.5	1.150*
17.	Convent	5.0	.487	17.	Convento	3.0	1.375
18.	Treacherous	5.0	.237	18.	Traidor	2.5	1.112
19.	Exquisite	5.5	.525	19.	Exquisito	2.5	1.187
20.	Slap	5.5	1.712	20.	Bofeton	6.0	1.462
21.	Caution	6.0	1.075	21.	Precaucion	4.0	1.425
22.	Incredible	6.0	.950	22.	Increible	5.0	1.562
23.	Deplore	7.0	.037	23.	Lamentar	2.5	1.212
24.	Scaffold	7.0	.150	24.	Horca	6.0	.887

* Pairs for which word with higher Frequency of Use Index
shows a lower mean.

Summary

The results of the present investigation clearly indicate that the pre-

dictions made were accurate. It was predicted that most of the Spanish

experimental words would show no significant different results and most

control words would show significant different results when compared

with their corresponding English words. Seventeen experimental words
showed no significant differences with regard to their corresponding
English words whereas twenty control words showed significant differ-
ences. Similar results were obtained when the number of individuals who
knew the meaning of the words were compared rather than the scores. At
the three levels of word difficulty (easy, medium, and difficult) only
the easy level showed no significant differences in the experimental
list. The results were significantly different, however, at the medium
and difficult levels. In the control list the difference was signifi-
cant at the three levels of word difficulty. When the two ethnic groups
were compared separately, only the Puerto Rican group showed a signifi-
cant difference at the medium level in the experimental list. Both,
Mexican and Puerto Rican ethnic groups, however, showed significant dif-
ferences at the three levels of word difficulty in the control list.
Frequency of use and score showed a direct relationship for twenty one
words; that is, the words used more frequently in the language obtained
higher scores than the words of lower frequency of use. Finally, sig-
nificant rank order correlations were obtained between the order accord-
ing to frequency of use and the order according to the scores obtained
with both the English and the Spanish experimental lists.

CHAPTER V

DISCUSSION

This investigation was conducted with the purpose of finding linguis-
tic variables which could account for word difficulty in vocabulary test
construction or translation. If such variables are found, they would
allow test constructors or translators to compile bilingual vocabulary
lists of words with the assurance that both versions--English and Span-
ish in this case--will be equally difficult. Such test versions would
provide a very useful tool for psychologists testing linguistic minori-
ties because, if both versions are similar in content and difficulty,
they will provide comparable results and eliminate bias in vocabulary
testing.

The results of the present investigation are discussed in terms of
the assumptions formulated and the implications of the findings. Sug-
gestions for further investigation will also be made.

Assumptions

The results of the present investigation present empirical evidence in favor of the assumptions made in this study.

Assumption one: <u>The</u> <u>frequency</u> <u>of</u> <u>use</u> <u>is</u> <u>one</u> <u>valid</u> <u>and</u> <u>practical</u> <u>measure</u> <u>of</u> <u>word</u> <u>difficulty</u>.

As shown in the previous chapter, the comparisons between the English and the Spanish experimental lists, in which the frequency of use was equal for each pair of words in both lists, provided not significantly different results for the total sample, for boys and girls, or for Mexican and Puerto Rican groups. The comparisons, however, for English and Spanish control words, in which the frequency of use was different in each pair, provided significantly different results for the total sample, for boys and girls, and for the Mexican and Puerto Rican groups. Also, no significantly different results were obtained at Level I of word difficulty, easy words. When the words of Level II and Level III were compared, the results were somewhat conflictive because the difference between the English and the Spanish experimental group of words was significant at both levels. A closer look at the individual words, however, clarifies in part these conflictive results. At Level II, four pairs of words were significantly different and two pairs at Level III. The relationship between the characteristics of the sample and the meaning of words can explain four of these obtained differences. The fact

that the individuals in the English sample were children born and raised in metropolitan areas whereas most of the individuals in the Spanish sample came from or were exposed to rural environments, could explain the difference in favor of Mansion for English speaking children and in favor of Campesino for Spanish speaking children. Mansion is a concept common in urban environments, whereas Campesino is a rather common concept in rural areas. Since the Spanish verb Competir has its corresponding noun Competicion, which is constantly used to refer to sports competitions, it is not surprising that the Spanish sample scored higher in this word than the English sample on Contend which does not have an equivalent noun used as frequently as Competicion in Spanish. It might be possible that the reason for the higher score of the English word Incredible could be found in the association of this word with the popular TV program "The Incredible Hulk." The analysis of the English speaking children's responses suggested this interpretation. No plausible explanation can be found for the differences of the other three pairs of words. Consequently, only three pairs of words, one at each level of word difficulty, remain without explanation.

Similar results were obtained when the number of individuals who knew the meaning of words, rather than the means, were compared. These results support the first assumption and also indicate that the procedure proposed in the present study is valid at least for Mexican and Puerto Rican groups, and for both boys and girls.

Assumption Two: <u>Words</u> <u>found</u> <u>in</u> <u>vocabulary</u> <u>tests</u> <u>usually</u> <u>represent</u>
<u>and</u> <u>should</u> <u>represent</u> <u>concepts</u> <u>that</u> <u>are</u> <u>common</u> <u>to</u> <u>all</u> <u>cultural</u> <u>and</u> <u>lin</u>-
<u>guistic</u> <u>groups</u> <u>living</u> <u>in</u> <u>the</u> <u>United</u> <u>States</u>.

The fact that the results for some pairs of words were not as pre-
dicted, indicates that frequency of use, although valid and practical,
is not a perfect measure of word difficulty. Consequently, when fre-
quency of use is going to be considered as a measure of word difficulty
in vocabulary test construction, some precautions must be taken. The
selected words must be "neutral" in the sense that they should not be
words which favor one segment of the population more than another. It
is a clear finding of this study that the students from a rural environ-
ment scored higher than their counterparts on the word <u>Campesino</u> which
represents a rural concept, whereas urban students scored higher than
their counterparts on the word <u>Mansion</u> which is a word more common in
urban than in rural environments.

The "neutrality" of vocabulary words could be a problem in the United
States, particularly in urban areas, where a variety of ethnic groups
attend school together. Test constructors should be very sensitive to
cultural differences and they must be very well informed about the con-
notations of the words they select for testing in order to prevent the
introduction of bias in their tests. The words selected must represent
concepts that are common to all cultural and linguistic groups living in
the United States.

Assumption Three: Children's vocabulary is similar to adult vocabulary.

According to present results this seems to be the case. The high significance between the expected and the obtained rank order correlations supports this assumption. The expected order (according to frequency of use) was based on adult use of words, whereas the obtained order (according to mean values) was calculated with the children in the sample. The fact, however, that the students in this study scored very high on the pair Slap (M = 1.712) - Bofetada (M = 1.750), words of low frequency of use, could indicate that children's vocabulary has moved some words of low frequency into higher levels of use. The concept Slap - Bofetada appears to be part of the daily living experience of most children. The English sample scored high also on the word of low frequency of use Incredible (M = .950) because of its association with the TV program "The Incredible Hulk." This association could be the reason why the Spanish sample scored high on the Spanish control word of low frequency Increible (M = 1.562) which has the same root and meaning as Incredible.

In the construction of vocabulary tests for children it would be very important to carefully scrutinize the words and to consider whether or not there are some words that although of low frequency of use in the general population are currently frequently used in children's language.

Assumption Four: The frequency of use of most words remains constant

<u>for</u> <u>long</u> <u>periods</u> <u>of</u> <u>time</u>.

This assumption, similar to assumption number three, is supported by the fact of the significant rank order correlations between the expected and the obtained orders. The expected orders were based on the frequency of use of words compiled and ranked by Thorndike in 1932 for English words and by Buchanan in 1927 for Spanish words. The obtained rank, obviously was calculated recently. It should be noted, however, that this finding does not **rule** out the possibility of some alterations in the frequency of use structure of the lexicon. New words have been added to the lexicon, words that were not used in Thorndike's or Buchanan's time. These new words, however, do not appear to have caused significant alterations in the structure of the lexicon, but rather additions and probably small alterations. For test construction and translation purposes the available frequency of use dictionaries in English and in Spanish seem to be useful and reliable. As pointed out earlier, some precautions must be taken when neologisms are part of the vocabulary list of words.

<u>The</u> <u>problem</u> <u>of</u> <u>bias</u> <u>in</u> <u>mental</u> <u>testing</u>

Another important issue addressed in the present study was the problem of bias in mental testing. The results of this investigation present empirical evidence from another perspective in favor of the non-bias position with linguistic minorities. In this study the item difficulty in vocabulary testing was controlled in order to make the pairs of words

equally difficult in both languages, English and Spanish, in the experimental list and different in difficulty in the control list. Previous studies have only administered the English original and one translated version to different samples and then they have compared the obtained results. This procedure is not totally appropriate because there is no guarantee that the items in both versions were equally difficult. A tentative measure of word difficulty was introduced in this study making both words of each pair equally difficult in the experimental list. When the samples were compared, most pairs of words showed not significant differences in the experimental list, whereas most pairs showed significant differences in the control list in which the difficulty was different for each word of the pair. These results imply that when there is assurance that both English and Spanish versions are equally difficult, vocabulary tests are not biased, at least with Spanish speaking students. It would be very useful to replicate this investigation with other languages and linguistic minority groups.

It was pointed out in Chapter I that the excessive attention devoted to the suspected bias of measurement instruments has prevented a conscientious analysis of the social structures that might be the real sources and maintaining factors of bias in society. It was also pointed out that what the tests have done is to detect the results of inequality in society rather than to cause such disturbing inequalities. The findings from the National Assessment of Educational Progress and from the College Entrance Examination Board reported by Jones (1984) support this last assertion. Jones provides evidence that permits him to conclude that "the gap between White and Black average verbal and quantitative

achievement levels for elementary and secondary students has appreciably narrowed over the 1970s" (p.1209). These results are discussed in terms of the social changes, such as school desegregation, positive changes of attitude of white students toward black students, career opportunities for black citizens, improvement of incomes for many black families, and attitudinal or motivational changes which are the result of the broader participation of blacks in American society. Tests, rather than being biased, have been sensitive to these social changes and have reflected in their results these positive social changes. The same sensitivity seems to have been true in the past. The social inequalities in society rather than being the result of testing, were the cause of poor performance of black students.

The problem of uncontrolled translations

The significant differences obtained with the comparisons between the English and the Spanish control words clearly indicate that using uncontrolled translations of the words in vocabulary tests is a detrimental practice that should be discontinued by professionals testing linguistic minority students. These uncontrolled translations provide results that by no means are comparable with the test norms because in these test materials there is no assurance that English and Spanish words are equally difficult. As the results of the present investigation suggest, when there is a need to translate the words of a vocabulary test, it is advisable to select words that match the frequency of use of the original English words. Frequency of use, although not per-

fect, appears to be one good, reliable, and highly economical measure of word difficulty.

This measure should be used, however, within the scope and the limits of the present investigation. In this investigation a very specific type of vocabulary test has been discussed, the Wechsler-like vocabulary test. In this type of test the word is presented orally and the student is required to verbalize the meaning of the word. It is questionable that such a measure would be adequate with other types of vocabulary tests in which guessing and excluding alternatives is possible. In these tests, although there is control of the stimulus word if trans- lated according to frequency of use, there is no control of the possible alternatives which could be excluded and of the guessing processes. The translation of this type of test is much more complex than the transla- tion of the Wechsler-like vocabulary tests and consequently their trans- lation procedures should be a matter of further investigation.

Ancillary Results

A careful scrutiny of the Spanish control list reveals that the Span- ish words that match the meaning of the English words at the Easy and Medium level of difficulty have a lower frequency of use than the corre- sponding English words, whereas the opposite is true at the difficult level. As shown in Appendix One, the mean score for English words is at Level I, Easy M = 1.75 and at Level II, Medium M = 3.75, and for Spanish Level I M = 3.81 and at Level II M = 4.56. At Level I all eight Spanish

words have lower frequency of use than the English words. At Level II six Spanish words have lower frequency of use and two have higher frequency of use than their corresponding English words. The mean score at Level III, Difficult, is for English M = 5.875 and for Spanish M = 3.94. Only one Spanish word has a higher frequency of use than its English counterpart. These facts imply that the Spanish control words in this study are more difficult at Easy and Medium level and easier at the Difficult level than the corresponding English words. The results obtained with the samples parallel these differences in difficulty. As shown in Table 8, the English mean for Easy words, Level I is M = 1.435 and for Medium words, Level II M = 1.017, whereas the means for the Spanish sample are Level I M = .634 and Level II M = .740. These results are reversed at the Difficult level, English Level III M = .646, Spanish Level III M = 1.278. Table 10 shows that, with the exception of only three, all pairs of words show a direct relationship between word difficulty, as measured by frequency of use, and score.

These findings unlike truly experimental findings are biproducts of the procedures that were followed to select the experimental and control words. The experimental were words that matched the meaning and frequency of use of the English words, whereas the control were words that had a different frequency of use than the English words. This selection was done with the purpose of investigating the relationship between frequency of use and score while holding constant the meaning of the words. From the pool of words that could have been selected for the control list, the Spanish experimental words were excluded, consequently reducing the number of alternatives for translation. Since the English words

at Level I and III were the extremes in the list, a sort of statistical regression toward the mean could have happened in the process of selecting control words. This was particularly true with Level I (Easy) words, because it was very difficult, and in the case of words in the 1.0 category impossible, to select control words that were easier. At Level III, however, it was possible, although difficult, to select words of lower frequency, because the Spanish words in Eaton's dictionary go beyond the 7.0 level. Since the words in vocabulary tests are ordered according to difficulty, as measured by the passing percentage in most cases, it is possible that the same phenomenon, a sort of statistical regression toward the mean, happens when the words are translated into other languages. This possibility is reduced by the fact that no words are excluded from the pool of all possible alternatives for translation. It must be considered, however, that many English words do not have a counterpart in meaning that matches their frequency of use, and consequently variation in difficulty must occur. It is very important that this finding be fully investigated because of the implications for testing linguistic minority children. It might be possible that uncontrolled translations of vocabulary tests provide a pattern of difficulty in which the translated words are more difficult than the original words at the beginning of the vocabulary list and easier at the end of the list. Since vocabulary tests usually have discontinuance rules after a specific number of incorrect responses, most children referred for psychological evaluation are not generally administered the whole list because early in the test they meet the criterion for discontinuance. As a consequence they are administered an incomplete list of words that is more difficult than the original English version. It would be very

important to investigate whether or not this imbalance in difficulty in fact exists and, if it exists, whether it could explain, at least in part, the low vocabulary scores of many Spanish speaking children. This will be the focus of this researcher's ongoing work.

This investigation was initiated with a practical purpose in mind, with the intention of finding an objective and reliable method of construction and translation of bilingual vocabulary tests. The uncontrolled translations used frequently by professionals testing linguistic minority students in their native languages, do not offer the guarantee of being similar to the original version, and consequently the results obtained with these versions are not comparable to the norms obtained with the original version. When the same norms are used with the original and the translated versions, it is imperative that both versions be similar in content and difficulty. The traditional procedure to measure item difficulty has been the passing percentage of the individuals in the standardization samples. This is an impractical procedure in instances in which samples are not available or translations of many tests have to be made to many different languages. Consequently it would be very useful to find an objective and reliable procedure that avoids these problems.

The present investigation was conducted with the purpose of finding linguistic variables which can be utilized to construct or to translate

bilingual vocabulary tests, making both versions similar in content and difficulty. In this way bias can be reduced when testing linguistic minority students. The review of the theories of semantic development revealed several word characteristics which were considered and discarded as measures of word difficulty. Meaning was selected in order to make both lists equal in the most fundamental characteristic of words, their meaning. Frequency of use was selected because it is an objective, quantifiable, and statistically easy to analyze characteristic. In regard to frequency of use as a measure of word difficulty it was assumed that it is a valid and practical measure that can be utilized with different linguistic groups as well as with adults and children, and finally that frequency of use remains constant for long periods of time.

A secondary purpose of the present investigation was to provide additional support from another perspective to the claim that vocabulary tests are not biased against linguistic minorities. Most of the empirical evidence has been concerned with the statistical properties of the tests in different groups. These studies have administered the English original and the translated version to two different samples, and then they have compared the obtained results. It is felt that this procedure is not totally appropriate because there is no guarantee that the items in both versions are equally difficult. In the present investigation a tentative measure of word difficulty was introduced making both words of each pair --English and Spanish-- equally difficult.

In order to verify empirically whether or not frequency of use is a good measure of word difficulty, one English and two Spanish lists --ex-

perimental and control-- of 24 words each were compiled. The words were
arbitrarily divided into three levels of word difficulty --easy, medium,
and difficult-- according to their frequency of use. The experimental
Spanish words matched the meaning and frequency of use of the corre-
sponding English words, whereas the control words matched the meaning
but not the frequency of use. The English list was administered to 80
English-speaking eighth graders and the Spanish experimental and control
lists were administered to 80 Spanish-speaking students (40 Mexicans and
40 Puerto Ricans) matched to the English counterparts in grade, age,
sex, and academic achievement. There was a total of 40 boys and 40
girls in the sample. Their verbal responses were scored two, one, or
zero points according to the scoring criteria provided by Wechsler in
the WISC-R manual. Means and standard deviations were calculated for
the whole list, for the three levels of word difficulty, and for single
words across ethnic and sex groups. The means were compared for statis-
tical significance of differences with multivariate and univariate pro-
cedures. Most of the comparisons between the English and the Spanish
experimental, either with Mexicans or Puerto Ricans and with boys and
girls, were not significant, whereas most English - Spanish control com-
parisons were significant. The differences with the experimental list
were obtained mainly with words in the medium difficulty level. These
significant differences, however, were explained in terms of the rela-
tionships between the meaning of the words and the sample characteris-
tics. Words whose meaning had a rural connotation (Campesino - Peasant
) were known better by the Spanish sample which was mostly of rural ori-
gin, whereas words with urban connotation (Mansion - Mansion) were
known better by the English sample mostly of urban descent. Similar

results were obtained when the number of individuals who knew the mean-
ing of words, regardless of one or two point responses, was considered.
Chi-square comparisons were conducted to determine the significance of
the differences. Finally, Spearman rank order correlations were con-
ducted to determine to what extent the rank order of the means of the
English words were correlated with the rank order of the Spanish words
in both the experimental and control lists. The correlation between the
English and the Spanish experimental list was significant, but not the
correlation between the English and the Spanish control list. The rank
order correlation between the obtained (means) and the expected (fre-
quency of use) orders was also significant in both the English and the
Spanish lists.

The results of the present investigation present empirical evidence
in favor of the assumptions made. The fact that most pairs of words
matched in frequency of use showed no significant different results
whereas most of the pairs with different frequency of use showed differ-
ent results indicates that frequency of use is a valid and practical
measure of word difficulty at least with Mexican and Puerto Rican groups
and with boys and girls. Some words were sensitive to the cultural
background of the students (urban or rural), which indicates that the
words of bilingual vocabulary tests should be "neutral" in the sense
that they should not be words which favor one segment of the population
more than another. The significant rank order correlations between the
expected and the obtained orders support the assumption that children's
vocabulary is similar to adult's vocabulary. The expected order
(according to frequency of use) was based on adult use of words, whereas

the obtained order (according to mean values) was calculated with the children in the sample. The significant correlation supports also the assumption that the frequency of use of most words remains constant for long periods of time. The expected orders were based on lists compiled in 1932 by Thorndike for English words and in 1927 by Buchanan for Spanish words. The obtained rank obviously was calculated recently.

The results of the present investigation present empirical evidence from a different perspective than the traditional studies, in favor of the non-bias position of vocabulary test with linguistic minorities. A tentative measure of item difficulty was introduced in this study making the words of the experimental list equally difficult as their counterparts in the English list. The non significantly different results obtained imply that when there is assurance that both English and Spanish versions are equally difficult, vocabulary tests are not biased, at least with Spanish-speaking students.

Careful attention, because of its implications for testing, was paid to the fact that the Spanish control words have a lower frequency of use than their corresponding English words at the easy and medium level of difficulty whereas the opposite is true at the difficult level. This was not considered a truly experimental finding but rather a byproduct of the selection of words procedures. When selecting the control words, from the pool of words that could have been selected, the experimental words were excluded, consequently reducing the number of alternatives for translation. However, since the English words at the easy and difficult levels are the extremes in the list, a sort of statistical regression toward the mean could have happened in the process of select-

ing control words. Since the items of most tests are usually arranged in order of increasing difficulty, the same regression toward middle values can happen in any translation of test items, making the items of the translated version more difficult at the beginning and easier at the end. Since most tests have discontinuance rules after a specific number of incorrect responses, most children referred for psychological evaluation are not generally administered the whole list of items because early in the test they meet the criterion for discontinuance. As a consequence they are administered an incomplete list of items that might be more difficult than the items in the original version. It was suggested that this fact should be fully investigated in further studies.

In summary, the present investigation suggests that the use of uncontrolled vocabulary test translations to test linguistic minority students must be discontinued because this might not provide the same pattern of item difficulty as the original version. When a translation of vocabulary test has to be made or when a bilingual vocabulary test has to be constructed, frequency of use should be considered as a good measure of word difficulty. Frequency of use can be used for this purpose within the limitations and with the precautions discussed in the present study.

REFERENCES

Anastasi, A. (1976). Psychological testing. (4th ed.), New York: Mac-
    Millan.

Anastasi, A. and Cordova, F. (1953). Some effects of bilingualism upon
    the intelligence test performance of Puerto Rican children in New
    York. Journal of Educational Psychology, 44, 1-19.

Anglin, J. M. (1970). The growth of word meaning. Cambridge, Mass.:
    MIT Press.

Argulewicz, E. N., Bingenheimer, L. T., and Anderson, C.C. (1983).
    Concurrent validity of the PPVT-R for Anglo-American and Mexican-A-
    merican students. Journal of Psychoeducational Assessment, 1,
    163-167.

Argulewicz, E. N. and Abel, P. R. (1984). Internal evidence of bias in
    the PPVT-R for Anglo-American and Mexican-American children. Jour-
    nal of School Psychology, 22, 299-303.

Beck, J. L., Perfetti, C. A., and Mc Keown, M. G. (1982). Effects of
    long term vocabulary instruction on lexical access and reading com-
    prehension. Journal of Educational Psychology, 74, 506-521.

Bierwisch, M. (1970). Semantics. In J. Lyons (Ed.), New horizons in
    linguistics. Penguin Books.

Boder, D. P. (1925). La Escala Binet-Simon-Terman-Mexico (B-S-T-M) (2a
    ed.).

Borude, R. R. (1971). Immediate recall as a function of frequency of
    word usage (count): A replication. Psychologia: An International

Journal *of* *Psychology* *in* *the* *Orient*, 14, 184-186.

Brown, R. (1958). How shall a thing be called? *Psychological* *Review*, 65, 14-21.

Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired associate learning. *Journal* *of* *Experimental* *Psychology*, 63, 409-412.

Carroll, J. B. and Horn, J. L. (1981). On the scientific basis of ability testing. *American* *Psychologist*, 36, 1012-1020.

Cerda, E. (1960). Nuestra adaptacion espanola de las escalas de W-B II, WISC y WAIS. *Revista* *de* *Psicologia* *General* *y* *Aplicada*, 15 77-90.

Chomsky, N. (1970). Remarks on nominalization. In R. A. Jacobs and P. S. Rosembaum (Eds.), *Readings* *in* *English* *transformational* *grammar*. Walthan, Mass.: Ginn and Co.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive* *development* *and* *the* *acquisition* *of* *language*. New York: Academic Press.

Clark, E. V. (1977). First language acquisition. In J. Morton and J. C. Marshall (Eds.), *Psycholinguistics*. *Developmental* *and* *pathological*. Ithaca, N.Y.: Cornell University Press, p.1-72.

Clark, E. V. (1977b). Strategies and the mapping problem in first language acquisition. In J. Macnamara (Ed.), *Language* *learning* *and* *thought*. New York: Academic Press.

Clark, E. V. (1979). Building a vocabulary: Words for objects, actions, and relations. In P. Fletcher and M. Garman (Eds.), *Language* *acquisition*. Cambridge: University Press.

Clark. E. V. (1983). Meanings and concepts. In P. H. Mussen (Ed.), *Hand-*

book of child psychology. (Fourth Edition). V. III Cognitive development. J. H. Flavel and E. M. Markman, Editors. New York: John Wiley and Sons.

Cohen, J. (1957a). The factorial structure of the WAIS between early adulthood and old age. Journal of Consulting Psychology, 21, 283-290.

Cohen, J. (1957b). A factorial-analytically based rationale for the Wechsler Adult Intelligence Scale. Journal of Consulting Psychology, 21, 451-457.

Cohen, J. (1959). The factorial structure of the WISC at ages 7.6, 10.6, and 13.6. Journal of Consulting Psychology, 23, 285-299.

Condas, J. (1980). Personal reflections of the Larry P. trial and its aftermath. School Psychology Review, 9, 154-158.

Craik, F. I. M. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 104, 268-294.

Creelman, M. B. (1966). The experimental investigation of meaning: A review of the literature. New York: Springer.

Cronbach, L. J. (1960). Essentials of psychological testing, (2nd. ed.). New York: Harper.

Eaton, H. S. (1967). An English-French-German-Spanish word frequency dictionary. New York: Dover Pub.

Flaugher, R. (1978). The many definitions of test bias. American Psychologist, 33, 671-679.

Germain, J. (1970). La medida de la inteligencia. Madrid: Espasa-Calpe.

Glasser, A. J. and Zimmerman, I. L. (1967). Clinical interpretation of

the Wechsler Intelligence Scale for Children. New York: Grune and
    Stratton.

Glucksberg, S. and Danks, J. H. (1975). Experimental psycholinguistics.
    An introduction. Hillsdale, N.J.: Lawrence Erlbaum Ass.

Gordon, R. A. and Rudert, E. E. (1979). Bad news concerning IQ tests.
    Sociology of Education, 52, 174-190.

Guiraud, P. (1954). Les Characteres statistiques du vocabulaire.
    Paris: Presses Universitaires de France.

Hoge, R. D. and Butcher, R. (1984). Analysis of teacher judgements of
    pupil achievement levels. Journal of Educational Psychology, 76,
    777-781.

Howes, D. H. and Solomon, R. L. (1951). Visual duration threshold as a
    function of word probability. Journal of Experimental Psychology,
    41, 401-410.

Hunter, J. E., Schmidt, F. L., and Rauschenberger, J. (1984). Methodo-
    logical, statistical, and ethical issues in the study of bias in
    psychological tests. In C. R. Reynolds and R. T. Brown (Eds.),
    Perspectives on bias in mental testing. New York: Plenum Press.

Jackendoff, R. (1975). Morphological and semantic regularities in the
    lexicon. Language Journal of the Linguistic Society of America,
    51, 639-671.

Jastrzembski, J. and Stanners, R. F. (1975). Multiple word meanings and
    lexical search speed. Journal of Verbal Learning and Verbal Behav-
    ior, 14, 534-537.

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

Johnston, W. T. and Bolen, L. M. (1984). A comparison of the factor

structure of the WISC-R for blacks and whites. Psychology in the

Schools, 21, 42-44.

Jones, L. V. (1984). White - Black achievement differences. The nar-

rowing gap. American Psychologist, 39, 1207-1213.

Jones, R. (1978). Protection in evaluation procedures: Criteria and

recommendations. In L. Morra (Ed.), Developing criteria for evalua-

tion and the protection in evaluation procedures provision of Pub-

lic Law 94-142. Washington,D.C.: U.S. Office of Education, Bureau

of Education for the Handicapped.

Kaufman, A. L. (1975). Factor analysis of the WISC-R at 11 age levels

between 6.6 and 16.6 years. Journal of Consulting and Clinical

Psychology, 43, 137-147.

Kaufman, A. L. (1979). WISC-R research: Implications for interpreta-

tion. School Psychology Digest, 8, 5-27.

Macnamara, J. (1982). Names for things. A study of human learning.

Cambridge, Mass.: The Massachusetts Institute of Technology Press.

Macnamara, J., Cleirich, A., and Kellaghan, T. (1972). The structure of

the English lexicon: the simplest hypothesis. Language and Speech,

15, 141-148.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult

intelligence. (5th ed.). Baltimore: Williams and Wilkins.

McNeill, D. (1970). The acquisition of language. New York: Harper.

Mercer, J. (1978). Protection in evaluation procedures. In L. Morra

(Ed.), Developing criteria for the evaluation of protection in

evaluation procedures provision of Public Law 94-142. Washing-

ton,D.C.: U.S Office of Education, Bureau of Education for the

Handicapped.

Murray, A. M. and Mishra, S. P.   (1983).   Interactive effects of item content and ethnic group membership on performance on the McCarthy Scales. Journal of School Psychology, 21, 263-270.

Nelson, K. (1973).   Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, 38, (Serial Number 149).

Oakland, T. (1983). Concurrent and predictive validity estimates for the WISC-R IQs and ELPS by racial-ethnic and SES groups. School Psychology Review, 12, 57-61.

Oller, J.W. (1983).   Are we testing for intelligence or language? In S.S. Seidner (Ed.), Issues of language assessment. Vol.II Language assessment and curriculum planning. Illinois State Board of Education.

Olmedo, E.L. (1977).   Psychological testing and the Chicano: A reassessment. In J.L. Martinez Jr. (Ed.), Chicano psychology. New York: Academic Press.

Olmedo, E.L. (1981).   Testing linguistic minorities. American Psychologist, 36, 1078-1085.

Postal, P.M. (1966).   Review article: Andre Martinet, Elements of general linguistics. Foundations of Language, 2, 151-186.

Ramos Lopez, E. C. (1970). Adaptacion de los subtests de vocabulario de las escalas de inteligencia Stanford-Binet Form L-M y Wechsler para ninos. Revista de Psicologia General y Aplicada, 102, 52-68.

Reschly, D. J. (1979).   WISC-R factor structures among Anglos, Blacks, Chicanos, and native American Papagos. Journal of Consulting and Clinical Psychology, 46, 417-422.

Reschly, D. J. (1980). Non biased assessment. Iowa State University,

Department of Psychology, December (Unpublished manuscript).

Reynolds, C. R. (1982). The problem of bias in psychological assess-
ment. In C.R. Reynolds and T.B.Gutkin (Eds.), The Handbook of
school psychology. New York: John Wiley and Sons, p.178-208.

Reynolds, C. R. (1983). Test bias: In God we trust; All others must
have data. Journal of Special Education, 17, 241-260.

Reynolds, C. R. and Brown, R. T. (1984). Bias in mental testing. An
introduction to the issues. In C. R. Reynolds and R. T. Brown
(Eds.), Perspectives on bias in mental testing. New York: Plenum
Press.

Reynolds, C. R. and Piersel, W. C. (1983). Multiple aspects of bias on
the Boehm Test of Basic Concepts (Forms A nad B) for white and for
Mexican-American children. Journal of Psychoeducational Assess-
ment, 1, 135-142.

Rodriguez-Bou, I. (1950). A Study of the parallelism of the English and
Spanish vocabularies. Rio Piedras, Puerto Rico: University of
Puerto Rico.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., and Boyes-Braem, P.
(1976). Basic objects in natural categories. Cognitive Psychol-
ogy, 8, 382-439.

Ross, R. J. and Reschly, D. J. (1983). An investigation of item bias
on the WISC-R with four sociocultural groups. Journal of Consult-
ing and Clinical Psychology, 51, 144-146.

Rubenstein, H., Garfield, L. and Millikan, J. A. (1970). Homographic
entries in the internal lexicon. Journal of Verbal Learning and
Verbal Behavior, 9, 487-494.

Samuda, R. J. (1975). Psychological testing of American minorities:

    Issues and consequences. New York: Dodd, Mead.

Sanchez, G. I. (1934a).   Bilingualism and mental measures: A word of

    caution.   Journal of Applied Psychology, 18, 765-772.

Sanchez, G. I. (1934b).   The implications of a basal vocabulary to the

    measurement of the abilities of bilingual children.   The Journal of

    Social Psychology, 5, 395-402.

Sandoval, J.,   Zimmerman, I. L.,   and Woo-Sam, J. M.   (1983).   Cultural

    differences on WISC-R verbal items.   Journal of School Psychology,

    21, 49-55.

Silverstein, A. B. (1969).   An alternative factor analytic solution for

    Wechsler's Intelligence Scales.   Educational and Psychological

    Measurement, 29, 763-767.

Silverstein, A. B. (1977).   Alternative factor analytic solutions for

    the Wechsler Intelligence Scale for Children-Revised.   Educational

    and Psychological Measurement, 37, 121-124.

Solomon, R. L. and Howes, D. H.   (1951).   Word frequency, personal val-

    ues, and visual duration thresholds.   Psychological Review, 58,

    256-270.

Sternberg, R. J. and Powell, J. S. (1983).   Comprehending verbal compre-

    hension.   American Psychologist, 38, 878-893.

Terman, L. M. and Merrill, M. A.   (1973).   Stanford-Binet Intelligence

    Scale.   Boston: Houghton Mifflin.

Travers, R. M. W. (1982).   Essentials of learning.   (Fifth edition). New

    York: Macmillan Publishing.

Vance, H. B. and Wallbrown, F. H.   (1978). ' The structure of intelli-

    gence for black children: A hierarchical approach.   Psychological

    Record, 28, 31-39.

Van Hagen, J. and Kaufman, A. L. (1975). Factor analysis of the WISC-R for a group of mentally retarded children and adolescents. Journal of Consulting and Clinical Psychology, 43, 661-667.

Vernon, P. E. (1979). Intelligence: Heredity and environment. San Francisco: W. H. Freeman.

Wechsler, D. (1959). Escala de Inteligencia Wechsler para Ninos. (2a ed.), Manual del Departamento de Instruccion Publica. San Juan de Puerto Rico.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation.

Werner, H. and Kaplan, E. (1952). The acquisition of word meanings: A developmental study. Monographs of the Society for Research in Child Development, N. 51.

William, U. (1971). The construction of standardized test for Welsh-speaking children. Educational Research, 14, 29-34.

Winter, W. (1961). Impossibilities of translation. In W. Arrowsith and R. Shattuck (Eds.), The craft and context of translation. Austin, Texas: The University of Texas Press.

Zimmerman, I. L. and Woo-Sam, J. M. (1973). Clinical interpretation of the Wechsler Adult Intelligence Scale. New York: Grune and Stratton.

Zipf, G. K. (1949a). Human behavior and the principle of least effort. An introduction to human ecology. Cambridge, Mass.: Addison Wesley Press.

Zipf, G. K. (1949). The psychobiology of language. Houghton Mifflin.

APPENDIX A

LIST OF WORDS, THEIR FREQUENCY OF USE, MEANS AND SD BY

GROUPS AND BY GENERAL LIST.

English	fr.	Spanish Exp.	fr.	Spanish Con.	fr.
		Easy			
Fire	1.0	Fuego	1.0	Lumbre	2.5
Deep	1.0	Profundo	1.0	Hondo	2.0
Distance	1.5	Distancia	1.5	Trecho	4.0
Famous	1.5	Famoso	1.5	Celebre	2.0
Slave	2.0	Esclavo	2.0	Siervo	5.0
Bitter	2.0	Amargo	2.0	Acerbo	6.0
Mercy	2.5	Misericordia	2.5	Clemencia	5.5
Calm	2.5	Calmar	2.5	Sosegar	3.5

M = 1.75	M = 1.75	M = 3.81
SD = .60	SD = .60	SD = 1.58

English	fr.	Spanish Exp.	fr.	Spanish Con.	fr.
		Medium			
Display	3.0	Desplegar	3.0	Exhibir	5.0
Perish	3.0	Perecer	3.0	Sucumbir	4.5
Peasant	3.5	Campesino	3.5	Labriego	4.0
Fortunate	3.5	Afortunado	3.5	Venturoso	4.0
Mansion	4.0	Mansion	4.0	Vivienda	3.5
Contend	4.0	Competir	4.0	Contender	6.5
Nourish	4.5	Nutrir	4.5	Alimentar	2.5
Artificial	4.5	Artificial	4.5	Postizo	6.5

M = 3.75	M = 3.75	M = 4.56
SD = .60	SD = .60	SD = 1.40

English	fr.	Spanish Exp.	fr.	Spanish Con.	fr.
		Difficult			
Convent	5.0	Monasterio	5.0	Convento	3.0
Treacherous	5.0	Perfido	5.0	Traidor	2.5
Exquisite	5.5	Primoroso	5.5	Exquisito	2.5
Slap	5.5	Bofetada	5.5	Bofeton	6.0
Caution	6.0	Cautela	6.0	Precaucion	4.0
Incredible	6.0	Inverosimil	6.0	Increible	5.0
Deplore	7.0	Deplorar	7.0	Lamentar	2.5
Scaffold	7.0	Cadalso	7.0	Horca	6.0

M = 5.875	M = 5.875	M = 3.94
SD = .79	SD = .79	SD = 3.94

General List		
M = 3.79	M = 3.79	M = 4.06
SD = 1.83	SD = 1.83	SD = 1.42

APPENDIX B

LIST OF WORDS IN THE ORDER THEY WERE GIVEN TO THE SUBJECTS

English	Spanish Exp.	Spanish Con.
1. Fire	Fuego	Lumbre
2. Display	Desplegar	Exhibir
3. Convent	Monasterio	Convento
4. Deep	Profundo	Hondo
5. Perish	Perecer	Sucumbir
6. Treacherous	Perfido	Traidor
7. Distance	Distancia	Trecho
8. Peasant	Campesino	Labriego
9. Exquisite	Primoroso	Exquisito
10. Famous	Famoso	Celebre
11. Fortunate	Afortunado	Venturoso
12. Slap	Bofetada	Bofeton
13. Slave	Esclavo	Siervo
14. Mansion	Mansion	Vivienda
15. Caution	Cautela	Precaucion
16. Bitter	Amargo	Acerbo
17. Contend	Competir	Contender
18. Incredible	Inverosimil	Increible
19. Mercy	Misericordia	Clemencia
20. Nourish	Nutrir	Alimentar
21. Deplore	Deplorar	Lamentar
22. Calm	Calmar	Sosegar
23. Artificial	Artificial	Postizo
24. Scaffold	Cadalso	Horca

APPENDIX C

VOCABULARY - SCORING CRITERIA

General Scoring Principles

(From WISC-R Manual)

2 points

1. A good synonym ("A hat is a cap," "Join means unite," "Brave means courageous").

2. A major use ("A knife is for cutting," "An umbrella keeps the rain off you").

3. One or more definitive features or primary features of objects ("A clock has hands that move around a dial," "A diamond sparkles in the sun and is very valuable").

4. A general classification to which the word belongs ("A donkey is an animal," "A thief is a criminal").

5. A correct figurative use of the word ("Procrastination is the thief of time").

6. Several less-definitive but correct descriptive features which cumulatively indicate understanding of the word ("A bicycle has wheels and pedals," "A nail is thin, pointy at the end, and made of metal").

7. Verbs: A definitive example of action or a causal relation ("You clock a horse to see how fast he can run," "You can join pieces of

paper with glue").

1 point

In general, a response which is not incorrect but shows poverty of content.

1.A vague or less pertinent synonym ("A donkey is something like a horse," "A fable is a proverb," "Hazardous means poisonous").

2.A minor use, not elaborated ("A knife is to eat with," "An umbrella is to keep off the sun").

3.An attribute which is correct but not definitive or not a distinguishing feature ("A clock has hands," "A nail is for hammering," "A diamond goes on a ring").

4.An example using the word itself, not elaborated ("Join the army," "Gamble money").

5.A concrete instance of the word, not elaborated ("Brave means you fight a bear," "Nuisance is when your kid brother won't leave you alone").

6.A correct definition of a related form of the word (defining "gambler"instead of "gamble," "seclusion" instead of "seclude," "spy" instead   of "espionage").

0 points

1.Obvious wrong answers.

2.Verbalisms ("Alphabet soup," "A brave man"), when no real understanding is shown after inquiry.

3.Not totally incorrect responses, but ones which, even after questioning, are very vague or trivial or show great poverty of content ("A bicycle has a seat," "A belfry is real high").

APPROVAL SHEET

The dissertation submitted by Jose Miguel Tamayo
has been read and approved by the following committee:

  Dr. Anne M. Juhasz
  Professor of Educational Foundations, Loyola

  Dr. Ronald R. Morgan
  Associate Professor of Educational Foundations, Loyola

  Dr. Daniel C. O'Connell, S.J.
  Professor of Psychology, Loyola
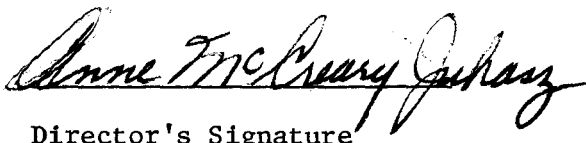
  Dr. Martha Wynne
  Assistant Professor of Educational Foundations, Loyola

The final copies have been examined by the director of the
dissertation and the signature which appears below verifies
the fact that any necessary changes have been incorporated
and that the dissertation is now given final approval by the
Committee with reference to content and form.

The dissertation is therefore accepted in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy.

April 15, 1985

Date   Director's Signature