A Comparative Study Of The Language Assessment Battery, The Woodcock-Munoz Language Survey And The Portfolio Assessment Results Of Language Learners

Louise Langlois Chaker

Follow this and additional works at: https://scholarship.shu.edu/dissertations

Part of the Bilingual, Multilingual, and Multicultural Education Commons, Educational Assessment, Evaluation, and Research Commons, and the Elementary Education and Teaching Commons

Recommended Citation

A COMPARATIVE STUDY OF THE LANGUAGE ASSESSMENT BATTERY, THE WOODCOCK-MUNOZ LANGUAGE SURVEY AND THE PORTFOLIO ASSESSMENT RESULTS OF LANGUAGE LEARNERS

BY

LOUISE LANGLOIS CHAKER

Dissertation Committee

Dr. Juan Cobarrubias, Ph. D., Mentor
Dr. Anthony J. Collella, Ph. D.
Dr. Judith Lawrence, Ed. D.
Dr. Mel Shay, Ed. D.

SETON HALL UNIVERSITY
UNIVERSITY LIBRARIES
SO. ORANGE, NJ 07079-2871

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Education

Seton Hall University
1999
ACKNOWLEDGMENT

It is with great pleasure that I take this opportunity to thank Dr. Juan Cobarrubias, my mentor, for all his help in the development and completion of this dissertation. My heartfelt gratitude is also extended to the other members of the committee, Dr. Anthony Colella, Dr. Judith Lawrence and Dr. Mel Shay whose time, guidance and encouragement was greatly appreciated.

I would also like to acknowledge the other members of the faculty of Seton Hall University who have provided throughout these years the enriching environment that has helped me in this work.

A very special expression of gratitude goes to Mrs. Imogene Johnson. Her support, encouragement, and confidence in the completion of this dissertation were truly motivating.

Ms. Carmen Delgado must be thanked for kindly providing knowledge and expertise essential to the task.

I would like to express my gratitude to the superintendent of the school district where the study took place.

Finally, my most appreciative thanks go to the teachers who gave me their assistance in providing the information necessary for the completion of this dissertation.
DEDICATION

to

my parents

Fernand Langlois

and

Stella Fournier

whose example

of ingenuity and hard work

have guided me throughout my life.

To my

husband, Peter

and my children

Anne Marie, Martin and Carl

who have encouraged me

and shared every step of the way.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... ii
DEDICATION .......................................................................................................................... iii
LIST OF TABLES .................................................................................................................... vii
LIST OF FIGURES ................................................................................................................ ix

I. INTRODUCTION .............................................................................................................. 1

The Problem ....................................................................................................................... 1
Purpose of the Study ........................................................................................................... 2
Research Question ............................................................................................................ 3
Research Hypotheses ......................................................................................................... 3
Definition of Terms ........................................................................................................... 5
Limitations of the Study ..................................................................................................... 13
Significance of the Study ................................................................................................. 14
Organization of the Study ............................................................................................... 15

II. REVIEW OF RELATED LITERATURE ........................................................................ 16

Assessment: A look at the Last Twenty Years ............................................................. 17
The Home Language Survey ......................................................................................... 22
Teacher Judgment .......................................................................................................... 29
Standardized Achievement Test ................................................................................... 32
Language Proficiency Test ............................................................................................ 35
A New Look at Assessment: Authentic Assessment ................................................... 38
   Portfolios: the True Value of the Student’s Ability .................................................. 40
   Summative and Formative Assessment ........................................................................ 42
   Elements Appropriate for Inclusion in the Portfolio ................................................... 43
   Baseline Assessment .................................................................................................... 47
   Student’s Attitude: an Aspect of the Portfolio ............................................................ 47
   The Portfolios of LM/LEP Students ............................................................................. 48
   Oral Language Assessment ........................................................................................ 52
   Reading Assessment .................................................................................................... 54
   Writing Assessment ..................................................................................................... 57
   Collection and Analysis of Data .................................................................................. 60
   Time Demands and Other Concerns .......................................................................... 62
Summary ............................................................................................................................ 63
III. RESEARCH METHODOLOGY .................................................. 64
  Instrumentation and Procedures ......................................... 64
  Participants ....................................................................... 66
  Procedures ........................................................................ 67
    Data Collection ................................................................ 67
    Data Analysis .................................................................. 68

IV. ANALYSIS OF DATA .............................................................. 70
  Participant Data .................................................................. 70
  Reliability Analyses ........................................................... 72
  Research Hypotheses .......................................................... 72
  Supplemental Analyses ........................................................ 84

V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS ............. 86
  Summary ........................................................................... 86
    Purpose of the Study ....................................................... 86
    Design of the Study ........................................................ 86
    Results ........................................................................... 87
  Conclusions ....................................................................... 87
    Findings and Discussions ................................................ 87
    Conclusion ....................................................................... 100
    Recommendations .......................................................... 100

REFERENCES ......................................................................... 102

APPENDIXES

A2. Kindergarten Reading Rubric ............................................. 111
A3. Kindergarten Writing Rubric .............................................. 113
A4. Oral Proficiency Rubric .................................................... 115
A5. First and Second Grade Reading Rubric .............................. 117
A6. First and Second Grade Writing Rubric .............................. 119
A7. Intermediate Reading Rubric ............................................. 121
A8. Intermediate Writing Rubric .............................................. 123
B. Letter of Permission for Instrument Utilization ................... 125
C1. School Administrator Consent .......................................... 127
C2. Consent Granted ............................................................. 129

v
C3. Teachers Consent............................................. 131
D1. Parental Consent........................................... 133
D2. Parental Consent French Version....................... 135
D3. Parental Consent Spanish Version..................... 137
E1. Student Assent............................................ 139
E2. Student Assent French Version......................... 141
E3. Student Assent Spanish Version......................... 143
F. Teachers Worksheet......................................... 145
G. Teachers Interview Questions........................... 147
LIST OF TABLES

1. Selected State Entry/Exit Methods ......................................................... 23
2. Types of Data Used by School Districts to Assess LEP ............................ 27
3. Reading/Writing Portfolio Sample Contents ........................................... 46
4. Academic Language Functions .............................................................. 53
5. Frequency Distribution on Grade Level .................................................. 71
6. Length of Stay in Program ..................................................................... 71
7. Interrater Reliability Coefficients ........................................................... 72
8. Chi$^2$ Analysis of Pass Rates for LAB and W-MLS Total Scores .......... 74
9. Chi$^2$ Analysis of Pass Rates for LAB and W-MLS Oral Subtests ........ 74
10. Chi$^2$ Analysis of Pass Rates for LAB and W-MLS R/W Subtests ....... 74
11. Chi$^2$ Analysis of Pass Rates for LAB and LPA Total Scores .............. 76
12. Chi$^2$ Analysis of Pass Rates for LAB and LPA Oral Subtests .......... 77
13. Chi$^2$ Analysis of Pass Rates for LAB and R/W Subtests ................. 77
15. Chi$^2$ Analysis of Pass Rates for W-MLS and LPA Oral Subtests ....... 79
16. Chi$^2$ Analysis of Pass Rates for W-MLS and LPA R/W Subtests ....... 79
17. Chi$^2$ Analysis of Exit Rates for grades K-3 and for grades 4-8 on LPA .. 81
18. Chi$^2$ Analysis of Exit Rates for grades K-3 and for grades 4-8 on LAB .. 82
19. Chi² Analysis of Exit Rates for grades K-3 and for grades 4-8 on W-MLS..... 82
20. Chi² Analysis of Exit Rates for Spanish and Creole Students on LPA.......... 83
21. Chi² Analysis of Exit Rates for Spanish and Creole Students on LAB.......... 83
22. Chi² Analysis of Exit Rates for Spanish and Creole Students on W-MLS....... 83
23. Chi² Analysis of Exit Rates for LAB and W-MLS Total at 3.5 CALP........... 85
24. Chi² Analysis of Exit Rates for LPA and W-MLS Total at 3.5 CALP........... 85
25. Agreement for the W-MLS and the LAB........................................... 95
26. Agreement for the W-MLS and the LPA........................................... 96
LIST OF FIGURES

1. Five Stages of Portfolio Program Development................................. 50
2. Literacy Development Checklist....................................................... 55
3. Learning Strategies Self-Assessment.................................................. 58
4. Self-Assessment of Reading Ability..................................................... 58
5. Sample Portfolio Analysis Form.......................................................... 61
Chapter I

INTRODUCTION

The Problem

The process of assessment of Language Minority/Limited English Proficient (LM/LEP) students for the purpose of placement has varied minimally since specialized language classes have been offered to these students. This is not necessarily evidence that this process is effective or efficient. The population served by second language programs is extremely varied. Wide variances in the areas of ability and academic background exist with these students, as they do to a lesser extent in the mainstream classes. We also see a wide variety of language backgrounds in the students serviced by language programs. These variances are met daily in individual language classes. The assessment of the diverse population of the language classes is still mainly based on the evaluation of discrete linguistic points.

Second Language instruction is also confronted with newly emerging problems that augment the complexity of language assessment. We are now seeing an increase in the number of illiterate students among LM/LEPs. The special needs of these students should be identified at assessment and specialized language services should be provided. Questionable referrals of LM/LEPs to Special Education departments remain a problematic area (Posner, B., 1989; Gonzalez V., Brusca-Vega, R. & Yawkey, T., 1997).
In such cases, language assessment instruments remain solely what they are intended for, to give an indication of the student's knowledge of the language being assessed.

Being able to identify accurately those students who genuinely require language learning assistance is more than ever becoming a necessity. Educational resources are limited and should be applied to those specific needs to which they are assigned. Retaining in a program students for whom the services are not profitable, dilutes the quality and the effectiveness of instruction for those in need of language instruction. However, releasing into the mainstream a student who is not ready to cope with a school curriculum centered on the development of higher order thinking skills and a literature based language program, will create problems for both the bilingual student and his new class. Accurate and appropriate assessment will satisfy the need to correctly include in language programs only those students who will benefit from them.

**Purpose of the Study**

The purpose of this study is to analyze the placement results of two Language proficiency tests and the results of the assessment of the language portfolios of LM/LEP students at the elementary level. More specifically, similarities and differences amongst and between the results of the Language Assessment Battery (LAB), the Woodcock-Munoz Language Survey (W-MLS) and the Language Portfolio Assessment (LPA) of participating students will be examined. The study will compare the rates of exit for the participating students relying on recognized cut-off scores for each test. The comparisons will highlight if there is agreement between the tests on the selection of exit candidates for participation in the mainstream program. The study will seek to determine reasons for the similarities and differences not only in the final results of each test, but also between
the two sub-tests, Oral (Listening/Speaking) and Reading/Writing for the tests examined in the study. Teachers of the participating students will be surveyed to determine the perceived reasons for the lack or presence of correlation amongst and between the assessment forms. Aside from comparing the exit rates of the two Language assessment tests and the portfolios of the subjects, the exit rates will be examined when grouped in primary and intermediate grade levels. The exit rates will also be examined when grouped in the two native languages (Spanish and Haitian-Creole) represented in the participants.

Besides providing a description of how the various assessment modalities discussed in the study compare to each other, it is hoped that an image of how particular conditions relative to characteristics of the population or the assessment format can influence results.

The Research Question

Do the LAB, W-MLS, and LPA consistently identify LM/LEP students in need of bilingual instruction?

Research Hypotheses

H1: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the W-MLS total scores are used.

H1A: There is a significant difference in the pass rates of students when assessed with the Oral sub-tests of the LAB from when the W-MLS Oral sub-tests are used.

H1B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing sub-tests of the LAB from when the W-MLS Reading/Writing sub-tests are used.
H2: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the LPA total scores are used.

H2A: There is a significant difference in the pass rates of students when assessed with the Oral sub-tests of the LAB from when the Oral rubric of the LPA is used.

H2B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing sub-tests of the LAB from when the Reading and Writing rubrics of the LPA are used.

H3: There is no significant difference in the exit rates of the students when assessed with the W-MLS total scores from when the LPA total scores are used.

H3A: There is no significant difference in the pass rates of students when assessed with the Oral sub-tests of the W-MLS from when the Oral rubric of the LPA is used.

H3B: There is no significant difference in the pass rates of students when assessed with the Reading/Writing sub-tests of the W-MLS from when the Reading and Writing rubrics of LPA are used.

Sub-Hypotheses

H1: There is a significant difference in the exit rates of the LAB, W-MLS, and LPA when grade level clusters (K - 3 and 4 - 8) are considered.

H2: There is no significant difference in the exit rates of the LAB, W-MLS and LPA when native language (L1) of the students is considered.

Definition of Terms

Language Minority / Limited English Proficient student (LM/LEP student): Federal legislation defines these students as those who:

1. Were not born in the United States or whose native language is not English.
2. Come from households where a language other than English is dominant.

3. Are American Indian or Alaska natives and come from households where a language other than English has an impact on their language proficiency.

4. For any of the above reasons have sufficient difficulty in the English language to keep them from being successful in classrooms where English is the language of instruction, or prevent them from participating in our society. (Bilingual Education Act, 1994)

**Language Assessment Battery (LAB)** (New York City Board of Education, 1982): LAB is a language assessment instrument developed by the New York City Board of Education in 1976; the most recent version was published in 1982.

The purpose of the English version of the LAB is to assess the language proficiency (reading, writing, listening and speaking) of all non-native speakers of English whose English language proficiency may be limited (New York City Board of Education, 1982).

The test consists of 4 levels with 2 parallel forms at each level. The test has an English and a Spanish version. The format of the LAB is:

- **Level I** - grades K, 1, 2
- **Level II** - grades 3, 4, 5
- **Level III** - grades 6, 7, 8
- **Level IV** - grades 9, 10, 11, 12

All levels of the LAB test assess listening, speaking, reading and writing ability individually or in a combined way. The tests are in the multiple-choice format except for
the interview portion of the Speaking sub-test. The average administration time is estimated at one hour.

At level I, the first part of the Listening/Speaking test is an interview that contains a set of 22 oral questions or simple commands directed mainly at vocabulary assessment. The 14 remaining questions require grammatical responses. The student is marked for both, grammatical correctness and relevancy. The Reading/Writing test has a separate 28 item section for kindergarten and for beginning of first grade. In this test, students are asked to select a grapheme and a picture after an oral prompt of phonetic significance. Students are also asked to read single words and two word sentences and select an appropriate picture. For first and second grade, the 28 item test requires selection of a single word from an oral prompt, selection of an appropriate picture from a single word or a simple sentence prompt. The test also includes sentence selection from a picture prompt. The final exercise requires sentence completion from a written prompt referred to as Cloze procedure.

At level II, in the first 16 items of the listening test, students are asked to select a picture from a sentence prompt. For the next 14 items, students are asked to select an appropriate response for an oral prompt. The speaking test is an interview which contains 26 items. Questions are either centered on pictorial prompt or directed at the student’s daily activities. Answers are marked for both grammatical correctness and relevancy. The reading test is a 36 item Cloze exercise organized in progressive order of difficulty. The writing test is a 20 item exercise of sentence completion with targeted grammatical or semantic points.
At level III, the listening test includes 20 items where the students select the appropriate response to an oral prompt. The next 18 items, is a Cloze exercise, where the text is read to the students, and students select the appropriate answer. The speaking test is a 26 item test in interview format. Questions are either centered on a pictorial prompt or student's daily activities. Answers are marked for both grammatical correctness and relevancy. The reading test is made up of four Cloze passages containing a total of 55 items placed in increasing order if difficulty. The writing test is a 20 item sentence completion or sentence selection exercise aimed at discrete grammatical or semantic points.

Level IV of the LAB is not relevant to this study. The cut-off marks used for this study are issued from a Norming Study of the Language Assessment Battery and Maculaitis Assessment Program (New Jersey State Department of Education, 1989). Woodcock - Munoz Language Survey (W-MLS) (Woodcock and Munoz-Sandoval, 1993). The English Survey and the Spanish Survey are sets of individually administered tests which primarily provide measures of cognitive academic language proficiency (CALP) (Cummins, 1984).

Cummins (1982, 1983, 1984) observed two types of language proficiency: basic interpersonal communication skills (BICS), and cognitive academic language proficiency (CALP). BICS can be defined as everyday communicative skills demonstrated in the pronunciation, vocabulary and grammar that take place within daily conversations. BICS relates to proximity and immediacy of needs, goals, and environment. On the other hand, CALP relates to proficiency in the academic context, the sort of language that students encounter in class. The skills displayed will involve semantics and abstract context.
Students can usually master communicative skills in about two years; however, proficiency in the academic language can take up to seven years.

There has been some discussion on the topic of the distinction between BICS and CALP. While CALP is seen as a determinant of student’s ability to perform in school settings, Wald (1984) recognizes the importance of BICS as a factor that influences academic achievement. He points out that the “match of language form and function in everyday situations to the system required in the classroom is very much involved in the outcome of academic achievement” (p. 59). He adds that the quality of the classroom communication will influence the interaction of teacher and student as well as the level of teacher’s expectation of the student’s output.

CALP is not an easily defined entity. Troike (1984) links it to Oller’s construct of general language proficiency, which is closely related to IQ measures. Wald (1984) on the other hand envisions CALP as literacy skills rather than as some less clearly understood form of cognition. Edelsky, C., Huddelson, S., Flores, B., Barkin, F., Altwerger, B., and Jilbert, K. (1983) go further by calling it “test-wiseness, an ability that incorporates a desire to do well on artificial (as opposed to real-world) test” (p. 6).

Cummins and Swain (1983) realize the danger of misinterpretation in the use of acronyms to represent concepts such as the ones discussed here. These concepts may end up taking a dimension different from the one conceived by the author. Cummins (1984a, 1994) no longer sees the BICS and CALP distinction as a focal point in his theory of language proficiency. He now refers to dimensions of contextual support and cognitive complexity. The evolution of this concept does not take away from the research in
language development the BICS and CALP distinction generated, it also served to
dissipate some of the criticism directed towards bilingual education Baral (1987).

The Woodcock-Munoz Language Survey is designed to provide cut-off points for
five levels of CALP for individuals of 2 years of age and up. Within each survey is found
4 tests.

Test 1: Picture Vocabulary
Test 2: Verbal Analogies
Test 3: Letter - Word Identification
Test 4: Dictation

The four tests are grouped to form an Oral Language Cluster (Tests 1 and 2), a
Reading/Writing Cluster (Tests 3 and 4) and a Broad English Ability Cluster (an average
of both clusters).

The Picture Vocabulary test measures the ability to name a pictured object. The
items are organized in progressive order of difficulty.

The Verbal Analogy test assesses the ability to comprehend and complete a word
relationship. The vocabulary remains simple throughout, but the relationships become
increasingly complex. In essence, the test measures the ability to reason with the
language.

The Letter-Word Identification test measures orthographic learning. Students are
asked to identify successively, rebus, letters, and words in increasing order of difficulty.
Students do not have to know the meaning of the words, they only have to demonstrate
the ability to read them.
The Dictation measures prewriting ability, letter production, and later, knowledge of punctuation, capitalization, spelling, and usage. This portion is administered as a traditional dictation.

Tests 1, 2, and 3 are given individually, while Test 4 may be given to a whole group. Basal and ceiling rules are to be followed, in order to minimize testing time. Adjustments may be made to ensure accurate starting and ending point.

**Language Portfolio Assessment (LPA):** The Language portfolio is a collection of individual student's work containing evidence of Written Language, Reading and Oral Language. The portfolio supplies sufficient evidence to document the ease at which the student independently and creatively uses language in these three areas. Published tests, for example, are not to be included in the portfolio assessment for the purpose of this study.

Self-assessment opportunities, usually conducted through teacher-student conferences, are to be given. The level to which self-assessment is performed by the student is, in itself, a demonstrator of the student's ability to use the language. Therefore, if the student's level of proficiency with the language warrants it, the teacher should take advantage of this opportunity to include in the rating of the portfolio, the student's ability to self-assess as evidence of oral language ability.

The authentic portfolio assessment process will illustrate a broad repertoire of classroom-based language activities. Moya and O'Malley (1994) describe the character of the portfolio contents: "Focussing on authentic language proficiency across sociolinguistic contexts and naturally occurring language tasks acknowledges the holistic and integrative nature of language development and focuses on communicative and
functional language abilities rather than attainment of discrete fragmented skills" (p.6).

The rating is achieved through the use of scoring rubrics (see Appendixes A1 to A8) which give a numerical value to the teacher's assessment of the student's work. Rubrics are scales that describe levels of performance for various descriptors qualifying a specific skill. These descriptors are selected to represent the breadth of performance required to describe a skill. The holistic nature of the rubrics ensures that Oral language assessment for example, does not bear mainly on grammatical accuracy, but that fluency and expression are also taken into account. Each level of performance is attributed a numerical grade. These are added for the various descriptors within the rubric in order to yield a score. Minimal levels of performance necessary for adequate performance in the mainstream class are selected within the rubric to establish a cut-off score.

Elements included in the portfolio are descriptive of the level of academic language controlled by the students (Pierce-Valdez and O'Malley, 1992). The items selected for inclusion serve to clarify the categories described in the three language rubrics.

Written Language is displayed through multiple samples of creative writing products including but not restricted to:

1. Process writing samples which display evidence of developmental stages (brainstorming, drafts, final copy).

2. Student selected dialogue journal entries.

3. Comprehensive independent class-work that demonstrates the student's level of writing ability.
The attribution of a numerical grade to the written part of the portfolio is done through a writing rubric which uses the Holistic Criteria developed by Sharon Moya (Moya and O’Malley, 1994) of the Evaluation Assistance Center at Georgetown University, Washington D.C. Permission was granted by the author to use her Holistic Criteria for this study.

Reading ability is evidenced through an anecdotal record of the student’s performance as maintained by the teacher. Notes reflect rate, intonation and use of punctuation. The categories listed in the Reading Rubric are the areas of documentation that the teacher maintains. The teacher’s comments are descriptive of the student’s ability to decode and recognize vocabulary, make meaning of the selection, and his or her capacity to apply knowledge gained from the reading. From the accumulated notation, the teacher can assign an accurate rating for each category of the rubric. The categories within each of the three levels of Reading Rubrics vary to reflect developmental changes between the grade levels. These categories are representative of the skills stressed in The Treasury of Literature, (Farr and Strickland, 1993) the reading series used in the district where this study is conducted. This researcher developed the Reading rubrics using, as a basis, the skills featured in the district’s reading series guides. The rubrics were presented to a jury of experts comprised of administrators and teachers of mainstream classes at the grade levels the study is concerned with, as well as to the Bilingual-ESL Supervisor.

Oral Language is assessed through “story retelling” activities that are to take place during the reading lesson. Teachers keep an anecdotal record of the activities. Here again, assessment is based on the categories listed in the Oral Proficiency Rubric which is the assessment measure for the Bilingual-ESL Department of the participating district. In
this study, the purpose of restricting oral assessment to “story retelling” activities is threefold. First, it confines oral language assessment to an academic purpose of language (cognitive academic language proficiency, CALP) and away from basic communication (basic interpersonal communication skills, BICS) (Cummins, 1984). Next, it gives homogeneity to the data collected in the various classes and grades included in the study. Finally, it links two language related activities: reading and text discussion, for a common purpose: the assessment of reading and that of oral expression in the academic context. $L_1$ will be identified as the native language of the students. In this study it will either be Spanish or Haitian Creole depending on which linguistic group is addressed. $L_2$ will be identified as English, the target language of this population.

Limitations of the Study

The conclusions of this study will be limited to:

1. The data collected during the 1996-97 academic school year.

2. Those students in the population targeted whose parents consented to their participation.


4. New Jersey primary schools listed under the New Jersey State District Factor Grouping A.

5. Interrater analysis for LPA verified for 47% of the cases.
Significance of the Study

This study will look at the levels and areas of differences for the LAB test, the W-MLS and the LPA of bilingual students at the elementary school level. Will the outcome of the tests be maintained when the approach of the three tests is different. The study will examine if the tests can reach a common decision for a very important question in the life of the LEP student, that of placement in or out of a bilingual program.

The current homogeneous application of testing procedures that yield a single numerical value to students' knowledge is still the rule. If educational changes that focus on a literature based curriculum, and higher order thinking skills are to be reflected in assessment, actual student's work should be the base for this assessment. The study will help in fostering increased reliance on the teachers' assessment portion of multi-criteria assessment. However, discrepancies between the language assessment tests examined in this study and teacher's assessment may occur. The study will look at the areas in assessment where divergence occurred, and seek explanations through teachers' feedback.

Since the Woodcock-Munoz Language Survey is a recently published language assessment test, this researcher has not been able to locate in the literature any research associated with this instrument. This work will therefore be an initial effort at comparing an instrument meant to assess academic language to teachers' assessment of students' work and to the results of the LAB which is a comprehensive measure of student's language.
placement of language students in and out of language programs in a typical urban district.

Organization of the Study

Chapter I presented limitations of language assessment in bilingual classrooms today. Reliance on language assessment tests focusing on grammatical structure such as the LAB offers but a limited view of students' capability. Looking at their level of proficiency with academic language through a test such as the W-MLS offers a new vantage point, but it is still through that same limited window, that of a test format. Examining students' production through an assemblage of their work in a portfolio, gives a deeper understanding of their potential, but this information may be suggestive. The problems that pertain to this study were discussed. Concerns over possible divergence in assessment results of the LAB, the W-MLS and the LPA results were expressed. These differences can result in the misplacement of bilingual students. The chapter concludes by clarifying areas the study hopes to enlighten. Chapter II delves more deeply into the literature. It examines the evolution of language assessment and looks specifically at various forms of language assessment. Chapter III describes the methodology of the study. Chapter IV analyzes the data collected for the study, states findings and examines these findings in relation to the hypotheses. Chapter V summarizes the study, discusses the findings, draws some conclusions and presents recommendations for future research.
Chapter II

REVIEW OF RELATED LITERATURE

Ever since our educational system has been concerned with the specialized instruction of the students whose native language is not English, much thought has been given to the formulation of various types of programs for the particular needs of a varied population. Somehow, in putting our efforts in program offerings, we have neglected to refine the process of placing these students in and out of Language programs. Assessment and placement is then regarded as a byproduct of the program. The result has been that many children end up in programs designed for LM/LEP (Language Minority/Limited English Proficient) students when they should not, using up the limited allocations directed towards children who really need the programs. The reverse is also true when students are pushed out of bilingual programs prematurely without having reached the level of language mastery required to assure their success in the mainstream. Second Language assessment needs to go beyond the process of sorting students for exit or retention in bilingual programs and look within the Language proficiency tests to determine the students' success and failure. Teachers must also be allowed to make decisions on the placement of their students. These decisions should not be made arbitrarily, but they should be based on work that approximates what the student will be asked to perform in the mainstream class. A review of the various assessment modalities
offered educators helps us understand the options that can best assist teachers and administrators in the placement of LM/LEP students.

Assessment: A Look at the Last Twenty Years

The identification of students who are to participate in bilingual programs is often seen as problematic and divergent views exist on the topic. Some research (Gonzalez et al., 1997; Oller, 1992; Cummins, 1984) asserts that students who have reached a passing communicative level of proficiency are often pushed out of bilingual programs too early. These students often have not reached the threshold of mastery with the academic language they will require in order to be successful in the monolingual class. These students end up failing not because of their limitations, but because schools send them unprepared into the mainstream. Rossell and Baker (1988) on the other hand point that it has been common practice to err on the side of over inclusion rather than release into the mainstream candidates that require language assistance. This has led to the inclusion of students who were not in need of the specialized teaching therefore depriving LM/LEP students of essential language instruction. The pioneer programs were especially guilty of this kind of practice. Rossell and Baker (1988) have reviewed the most important court decisions that have influenced the direction that specialized instruction has taken to this day. They point out that one of the earliest court decisions that was concerned with the bilingual student’s needs is United States v. Texas (1971/1972, as cited in Rossell and Baker, 1988). The court based its decision to have districts provide bilingual education to “all economically disadvantaged” students from age 3 to grade 6 without realizing that monolingual English speaking children of Mexican ancestry might be
included. Serrano v. Portales (1974, as cited in Rossell and Baker, 1988) demonstrated the same type of decision when the court ruled that everyone gets special language instruction until junior high school where the instruction may then continue until graduation if English language achievement is still low. In effect the court ordered that the entire population of one school, which was 85% Hispanic, follow a bilingual program. Clearly again, the level of English proficiency of students is ignored as well as the likely inclusion of monolingual students in a language program. Keyes v. School District No. 1 (1983, as cited in Rossell and Baker, 1988) introduced the use of a home language survey to identify any non-English speaking person in the student’s home. However, the courts created the possibility of including a student who speaks no or very little Spanish in a transitional bilingual class if he scores at or below the 29th percentile in a standardized assessment test.

The most important decision in our field remains Lau v. Nichols (1973/1974, as cited in Rossell and Baker, 1988). Its text talks about “national-origin, minority-group children” who display an “inability to speak and understand the English language” (Rossell and Baker, p. 593) as recipient of special services. The authors point out that the Supreme Court relied on the Department of Health, Education, and Welfare’s 1970 interpretative guidelines for compliance with the 1964 Civil Rights Act. However, it offers no insight as to how these recipients will be determined.

The first court decision to specify the level of English required by a student to participate in the regular instructional program was stated in Aspina of New York v. Board of Education of New York (1973/1974, as cited in Rossell and Baker, 1988). There, all
Spanish surnamed students were to take the Language Assessment Battery (LAB) in both English and Spanish. Students scoring below an arbitrary 20th percentile in the English test and having a higher test score in Spanish than in English were assigned to the Bilingual Education Program.

In 1981, the courts ordered major changes in bilingual education remedy for Texas schools. However, these have been revised on appeals. The sensible decision in United States v. Texas (1981, as cited in Rossell and Baker, 1988) takes into consideration the standard error of measurement of 4% in standardized tests and makes it part of the cut-off.

...a student who takes the test and scores at the 40th percentile would be just as likely to score as high as forty-four percent or as low as thirty-six percent, if he took the test again. In establishing its cut off score, the state of California compensated for this standard error by reducing its minimum proficiency four points from forty to thirty-six percent. That approach is clearly superior from the standpoint of psychometrics to fixing a single cut-off score which ignores the standard error of measurement...Students scoring below that level shall be classified as LEP. Students scoring from thirty-six through forty-three shall be classified by LPAC on the basis of its informed discretion. Students scoring at the 44 level or higher shall be classified as proficient in English, unless they fail to demonstrate reasonable proficiency on an oral English language test. (Rossell and Baker, p.595)
The purpose of the LPAC (Language Proficiency Assessment Committee) was to decide on the placement of the students who fit the thirty-six through forty-three window. Though the idea of taking into account the standard error in defining the cut-off seems a good idea, it remains that the cut is still an arbitrary point. The use of standardized tests as a placement tool is also under scrutiny (De Avila, 1990). The most valid part of this decision is the establishment of an assessment committee, comprised of the professionals who are most familiar with the student's capability and progress, and who are most likely to make the best decisions on his chances of success in the monolingual class.

Finally, in 1987, the state court of California examined for the first time in Jimenez v. Honig (1987, as cited in Rossell and Baker, 1988) the reliability of standardized test scores as determinant for participation in or exit from bilingual programs. Up until then, the thirty-sixth percentile was the sole determinant for entry and exit. Review by the school appraisal team, or by the school or district language team for those students scoring between the thirty-first and thirty-fifth percentile became available. In 1985, an additional component allowed a student who had been enrolled in a bilingual program for three years, to be evaluated by a language appraisal team for reclassification, regardless of the student's score on his standardized test. This was in response to complaints by educational personnel that numerical cut-offs were unreliable indicators of a student's language performance:

Norm reference tests are based on comparisons to an identified norm. The normal distribution of scores which are curved, consigns examinees to every point of the curve from the lowest to the highest and always relegates some examinees to
positions below a certain percentile. Because the scores of LEP students on norm referenced tests are based on the performance of nonminority students, their scores tend to be lower than those of their English-proficient counterparts. But lack of English proficiency is not necessarily the cause of poor test performance. Some students whose primary language is English score, according to the normal distribution, below a certain percentile for reasons other than language deficiency. Were these very same English-speaking students placed in a bilingual classroom where exit eligibility is determined by their test scores alone, they would be ineligible for reclassification (Rossell and Baker, p. 597).

However, the court did not change the thirty-sixth percentile as entry criterion, and added some contingencies for its use at exit.

Here in New Jersey, there have been three exit procedure changes since the mid-eighties. The state initiated a multi-exit criteria procedure in 1991. This procedure reinstates criteria requirement that had been abandoned in the spring of 1988. The reason for the initial change to single exit criteria, which was the language assessment test, was that the bilingual student was held to a higher standard than the monolingual student (NJ Dept. of Education, 1987). With multi-exit criteria, the students now have to satisfy an array of requirements such as: minimum passing grade on a recognized language assessment test such as the LAB, reach the recognized passing percentile on a recognized standardized achievement test, be working at the appropriate academic grade level, and receive teacher's recommendation.
The review of legal decisions concerning the placement of LM/LEP students shows some progress in the process of defining the appropriate educational programs for these students. What started as the assignment of students to bilingual education programs on the basis of Spanish sounding surnames, now recognizes Language Proficiency Assessment Committees as capable to decide of a student's ability to function in the monolingual class. The assessment tools and criteria presently used in the placement and assessment of LM/LEP students should reflect the objectives of our curricula and the philosophy of our instruction (Newman, C., Smollen, L. and Lee, J., 1995). They should give a reliable and valid measurement of the students' language(s) if we want our programs to serve the population that requires our expertise. A cursory look at the tabulation of entry/exit methods by state (DeGeorge, pp. 53-55) in Table I, points to a definite reliance on common types of instrument. De Avila (1990) adds that the screening procedures do not vary between the states that have formalized processes. This is confirmed in Table 2 which describes the percentages of districts using various assessment methods (Fleischman and Hopstock, p.17). The use of the Home Language Survey is common as an initial identifier of students likely to need second language instruction.

The Home Language Survey

The Home Language Survey gives a first glance at whether a student is in contact with a language other than English in his home environment. Obviously, that is all that this instrument is meant to be: an indicator. It cannot absolutely determine participation in the program. De Avila (1990) explained that the information released by parents can, at
<table>
<thead>
<tr>
<th>State</th>
<th>Identification Methods</th>
<th>Exit Methods</th>
<th>Instruments Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>Parent questionnaire</td>
<td>Holistic assessment</td>
<td>JTBS Comprehension</td>
</tr>
<tr>
<td></td>
<td>Language observation questionnaire</td>
<td>Standardized achievement test score</td>
<td>Teacher Fluency Survey</td>
</tr>
<tr>
<td></td>
<td>Language assessment instrument</td>
<td></td>
<td>Battle Culture-Free Self Concept Inventory</td>
</tr>
<tr>
<td>Arizona</td>
<td>Home language survey</td>
<td>Reassessment at least every two years</td>
<td>BSM I &amp; II</td>
</tr>
<tr>
<td></td>
<td>Language assessment</td>
<td>Teacher evaluation</td>
<td>LAS I &amp; II</td>
</tr>
<tr>
<td></td>
<td>Teacher observation/opinion</td>
<td>Student performing at grade level</td>
<td>IPT I &amp; II</td>
</tr>
<tr>
<td></td>
<td>Parental statement</td>
<td>Parental opinion and consultation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Objective assessments of English oral language</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Objective assessment of reading and writing skills.</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>Home language survey</td>
<td>Mastery of English language curriculum</td>
<td>BINL</td>
</tr>
<tr>
<td></td>
<td>English oral/aural proficiency test</td>
<td>Oral/aural proficiency testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Literacy testing</td>
<td>Parental evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teacher evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Above 36th percentile on standardized criterion referenced test; some discretion allowed</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>Parent/teacher checklist</td>
<td>Oral language achievement test score</td>
<td>Standardized achievement test</td>
</tr>
<tr>
<td></td>
<td>Oral language test</td>
<td>Standardized reading and math pre/post test scores</td>
<td>Oral interview</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Selected self-concept scale</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teacher evaluation and assessment</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Spanish/English pre-test</td>
<td>Score at or above 50% on achievement tests</td>
<td>Standardized achievement test</td>
</tr>
<tr>
<td></td>
<td>Language proficiency tests</td>
<td>Attainment of average academic grades</td>
<td>Oral interview</td>
</tr>
<tr>
<td></td>
<td>Standardized achievement test in reading, math, language</td>
<td>Teacher evaluation and assessment</td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>Home information survey</td>
<td>English language proficiency test score</td>
<td>LAS</td>
</tr>
<tr>
<td></td>
<td>Language assessment test</td>
<td>Standardized achievement test</td>
<td>BINL</td>
</tr>
<tr>
<td>Idaho</td>
<td>Home language survey</td>
<td>Teacher observation</td>
<td>BINL</td>
</tr>
<tr>
<td></td>
<td>Language assessment test</td>
<td>Close reading test</td>
<td>Brigance-C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standardized test score</td>
<td>LAB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LAS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IPT</td>
</tr>
<tr>
<td>Illinois</td>
<td>Home language survey</td>
<td>Above average English proficiency for native English speakers at grade level in district</td>
<td>LAS</td>
</tr>
<tr>
<td></td>
<td>Student language</td>
<td>Assessment of listening, understanding, speaking, reading and writing</td>
<td>BSM</td>
</tr>
<tr>
<td></td>
<td>Assessment of average English proficiency for native English speakers at grade level in district</td>
<td>Below average English proficiency for native English speakers at grade level in district</td>
<td>IDEA</td>
</tr>
<tr>
<td></td>
<td>Academic history</td>
<td>Evaluation of some variables used in identification procedure</td>
<td>BINL</td>
</tr>
<tr>
<td></td>
<td>Additional factors as determined by SEA and district</td>
<td></td>
<td>BOLT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Others with approval of SEA</td>
</tr>
<tr>
<td>State</td>
<td>Identification Methods</td>
<td>Exit Methods</td>
<td>Instruments Used</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>Indiana</td>
<td>Teacher observation and referral Cumulative grades and records Speech test Parent information Informal assessment School consultation team Achievement test Criterion referenced test Language proficiency test</td>
<td>Student grades Teacher evaluation Achievement test scores</td>
<td>ITBS LAS SAT MAT PPVT CAT Gates-McGintie Language test Articulation test Gunn Reading Test</td>
</tr>
<tr>
<td>Kentucky</td>
<td>Teacher observation Parental information Kentucky Essential Skills Test Tutor observation Course grades Oral language proficiency Criterion-referenced test</td>
<td>Standardized test scores Classroom performance Teacher recommendations Oral proficiency tests Writing test</td>
<td>Kentucky Essential Skills Test Teacher observation Woodcock Language Proficiency Battery Davis Diagnostic Test for ESL Students ITBS SAT</td>
</tr>
<tr>
<td>Louisiana</td>
<td>Parental information Language proficiency assessment Standardized achievement test</td>
<td>Standardized achievement test scores Teacher observation</td>
<td>CAT SRA CTBS SAT MAT Criterion-referenced test</td>
</tr>
<tr>
<td>Maine</td>
<td>Home language survey Oral Language proficiency assessment Informal oral observation</td>
<td>Standardized achievement test scores Oral language proficiency test scores State achievement test score Teacher observation Course grade</td>
<td>LAB BSM CELT BINL IPT MAP</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Home language survey Teacher referral Oral interview Language proficiency test</td>
<td>Language proficiency test score Standardized achievement test scores Language continuum instrument Course grades Teacher recommendation Parental input</td>
<td>BSM CTBS Cloze reading test ESL test Metropolitan Reading Survey</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Teacher referral Parental information Informal assessment Language proficiency test Comprehensive student record Standardized achievement test scores Tutor observation Speech test Course grade Criterion-referenced test</td>
<td>Teacher judgment Standardized achievement test scores Language proficiency test scores</td>
<td>Teacher-made language proficiency instrument Standardized achievement test Criterion-referenced test</td>
</tr>
<tr>
<td>State</td>
<td>Teacher referral</td>
<td>Informal assessment</td>
<td>Parent information</td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Mississippi</td>
<td>Teacher referral</td>
<td>Informal assessment</td>
<td>Parent information</td>
</tr>
<tr>
<td>New York</td>
<td>Home language identification</td>
<td>Score below 23rd percentile on an English language assessment instrument approved by the Commissioner of Education</td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>Teacher referral</td>
<td>Informal assessment</td>
<td>Parental information</td>
</tr>
<tr>
<td>Ohio</td>
<td>Home language survey</td>
<td>Oral/aural proficiency testing</td>
<td>Literacy testing</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Teacher observation and referral</td>
<td>Speech test</td>
<td>Parent information</td>
</tr>
<tr>
<td>Texas</td>
<td>Home language survey</td>
<td>Oral language proficiency tests (English and/or Spanish)</td>
<td>Informal assessment (teacher/parent interview, student interview, teacher survey)</td>
</tr>
<tr>
<td>State</td>
<td>Screening Instrument</td>
<td>Test scores/Teacher judgment</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Teacher/counselor referral</td>
<td>Standardized achievement test</td>
<td>CTBS Teacher-made test</td>
</tr>
<tr>
<td></td>
<td>Parent information</td>
<td>scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal testing</td>
<td>Teacher judgment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehensive student records</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Speech test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standardized achievement test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language proficiency test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutor observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Criterion-referenced test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>Home language survey</td>
<td>Oral English language assessment score</td>
<td>ITP LAS</td>
</tr>
<tr>
<td></td>
<td>Oral English language assessment score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standardized achievement test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher referral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>One year or more deficiency in grade level in language</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key To Tests Listed**

BINL: Basic Inventory of Natural Language  
BOLT: Bilingual Oral Language Tests  
Brigance-C: Brigance Comprehensive Inventory of Basic Skills – English and Spanish  
Brigance-D: Brigance Diagnostic Assessment of Basic Skills – Spanish  
CAT: California Achievement Test  
CELT: Comprehensive English Language Test  
CTBS: Comprehensive Test of Basic Skills  
FLA: Functional Language Assessment  
ITP: Idea Oral Language Proficiency Test  
ITBS: Iowa Test of Basic Skills  
LAB: Language Assessment Battery  
LAS: Language Assessment Survey  
MAP: Macularis Assessment Program  
MAT: Metropolitan Achievement Test  
MRT: Metropolitan Readiness Test  
PIAT: Peabody Individual Achievement Test  
PPVT: Peabody Picture Vocabulary Test  
QSE: Quick Start in English  
SAT: Stanford Achievement Test  
SRA: Science Research Associates, Inc.  
TAP: Total Academic Proficiency  
WRAT: Wide Range Achievement Test

Table 2

Types of Data Used by School Districts to Assess LEP

<table>
<thead>
<tr>
<th>Data Used to Determine LEP Status</th>
<th>% of Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Proficiency tests in English</td>
<td>83.1</td>
</tr>
<tr>
<td>Home language survey</td>
<td>76.6</td>
</tr>
<tr>
<td>Teacher judgment</td>
<td>69.4</td>
</tr>
<tr>
<td>Achievement tests in English</td>
<td>52.3</td>
</tr>
<tr>
<td>Teacher ratings of oral proficiency</td>
<td>49.2</td>
</tr>
<tr>
<td>Writing Samples in English</td>
<td>45.5</td>
</tr>
<tr>
<td>Oral proficiency tests in native language</td>
<td>34.4</td>
</tr>
<tr>
<td>Literacy tests in English</td>
<td>33.9</td>
</tr>
<tr>
<td>Achievement tests in native language</td>
<td>11.6</td>
</tr>
<tr>
<td>Other</td>
<td>7.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Used to Reclassify Students</th>
<th>% of Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher judgment</td>
<td>78.6</td>
</tr>
<tr>
<td>Achievement in English</td>
<td>75.0</td>
</tr>
<tr>
<td>Oral proficiency tests in English</td>
<td>74.4</td>
</tr>
<tr>
<td>Teacher ratings of English proficiency</td>
<td>57.0</td>
</tr>
<tr>
<td>Writing samples in English</td>
<td>52.8</td>
</tr>
<tr>
<td>Literacy tests in English</td>
<td>29.8</td>
</tr>
<tr>
<td>Other</td>
<td>10.3</td>
</tr>
</tbody>
</table>


times, be misleading:

One is left wondering about the validity of asking parents, who are unclear about their own linguistic habits, to make retrospective judgments about their children’s linguistic patterns. While parents are certainly in the best position to know, the use of technically unevaluated questions could well be problematic. Some parents may simply not know how well the child speaks a language that the parents do not speak
Some parents may well be loath to admit the use of a language other than English. Some may be fearful about placing their child in a special program and are confused over the purpose of the survey or fear reprisal from immigration authorities and so on. (De Avila, p.204)

Canales (1992) also added that some parents will not respond truthfully on the Home Language Survey for fear that their children may be placed in a program that is not conducive to learning English. The survey may also point to a student who lives in a home where a language other than English is spoken, but that student may be unable to score high enough on the language assessment test or the achievement test for reasons other than language. Had the same student come from a monolingual home, he might also have scored below the cut-off. The Home Language Survey disregards the differences of language usage for the language(s) spoken in the immigrant home and it fails to tell us how much of the home environment influences the child’s language. It does not tell us about language dominance in the student, neither does it tell us how much influence the language outside of the home has had on the student.

Rossell and Baker (1988) pointed out that California has made an effort to correct this problem. The home language survey used in California includes questions such as:

1. Which language did your son or daughter use when he or she began to talk?
2. What language does your son or daughter most frequently use at home?
3. What language do you use most frequently to speak to your son or daughter?
4. Name the languages in order most often spoken by the adults at home. (p.606)
These are questions that should narrow down the student population to only those who would benefit from a Bilingual/ESL program, but even the best home language survey can steer a student in the wrong program. School personnel awareness of the limitations of entry procedures and the language specialist’s vigilance with student placement are essential to avoid errors.

Teacher Judgment

Teacher and staff judgment should be a key element of student placement in language programs. Young, M., Hopstock, P., Rudes, B., Fleischman, H., Zheler, A., Shaycoft, M., Goldsamt, M., Bauman, J., Burkheimer, G., and Rattner, M. (1984, as cited in Rossel and Baker, 1988) report that two-thirds of LM/LEP students are screened through procedures which include some form of staff judgment. Other studies (Mace-Matluck, 1982; Cardoza, 1984) have also found teachers recommendation having major influence in the exit of students from bilingual programs. This however is a variance from most mandated regulations.

Most laws, regulations and many court decisions attempt to eliminate teacher judgment. Thus a superficial consideration of these data suggests widespread noncompliance with the legal restrictions on using teacher judgment. Schools have been wise to allow teacher judgment to temper the methods that are legally prescribed (Rossell and Baker, 1988).

Some states recognize the value of teacher judgment in placement decisions; they therefore include teacher’s recommendation as one of the multi-exit criteria that the student must satisfy to leave the program. This can be seen in Table 1 (De George, 1988).
Gandara and Merino (1993) noted that California features a “five item multi-criteria process” featuring among others: writing sample, teacher judgment, and parental consent. However, a passing score on the Language Proficiency Test and success on the Standardized Achievement Test are enough to effect a change in the language proficiency status of the student that would reclassify him as Fully English Proficient. In such cases, the recommendation, being one of many criteria, may be seen as just a stamp of approval, and is not a deciding factor. The emphasis invariably remains on the Language Assessment Test and a passing grade on the Achievement Test.

Whether teacher’s input is given credence or not, the research points to it as a reliable source for decision. Rossell and Baker (1988) mention two studies that validate teacher’s judgment. Ullibarri, D., Spencer, M., and Rivas, G., (1981) examined how teacher rating, achievement test data, and oral language proficiency, correlated in a sample of 900 elementary students in five different districts. A variety of achievement and language tests were used along with rating in the areas of English language proficiency, math and reading competency, and chances for achievement in English monolingual classes. The study found that across the grades, teachers’ ratings were the best predictors of academic success. The Southwest Educational Development Laboratory in a 7-year study of 800 students from kindergarten to grade four reached similar conclusions. The study looked at correlations between the LAS, which is a language proficiency test, teacher’s ratings, and taped samples of students’ speech in various contexts. Teacher’s ratings and the speech samples were highly correlated, but the LAS correlated poorly with either of the two measures. These studies found that teachers’ ratings were the best
predictors of academic success. Teachers are aware of test limitations and are confident in their ability to make correct decision in their placement.

They [teachers] expressed little confidence in the tests. Generally, users felt that teacher's judgment was more likely to be a valid measure of language proficiency and capability of succeeding in an all-English-medium classroom, than any test that they had been using. However, project staff had continued to employ the test in the entry/exit process in order to satisfy state of federal regulations or to give the appearance of objectivity in project decision making. (Southwest Reg. Lab., 1979, as cited in Rossell and Baker, p.621)

These researchers have pointed out that even if legislation strives to exclude teacher's judgment from placement decisions there has been a wise, natural reliance on teacher's input. Teachers' judgment is based on daily experience with the student in the context of the interactive classroom, and this alone is a fair means of assessment. However, problems of reliability, validity and possible conflicts of interest related to employment status may be a concern. Therefore, these decisions are best not made in isolation (De Avila, 1990). Relying on the teacher's judgment based on concrete evidence with the support of department head and school administrator along with other teachers working with the student in the context of the LPAC will expose every possibility in order to bring about the best decision (Pelavin and Baker, 1987, as cited in Rossell and Baker, 1988). The language assessment team is therefore an ideal format for expression of staff judgment. Can assessment tools available to the professionals making placement
decision, accurately, reliably and fairly describe the LM/LEP student? Cummins (1982, 1983, 1984) tells us that the LM/LEP student must be able to go beyond conversational skills in L₂ and must be able to control academic language and higher order thinking skills in his second language before he can compete on equal footing with his monolingual peers. Assessment, to be effective and accurate, must therefore take into consideration the academic output of the student. At exit, we must establish if the bilingual student can manage the work that is being done in the “classroom down the hall”. The two most scrutinized forms of assessment are the Standardized Achievement Test and the Language Assessment Test. Are these tests determinant of likelihood of success in the monolingual class?

Standardized Achievement Test

The problem with using a Standardized Achievement Test to evaluate the language learner’s achievement, lies in the fact that the test is used for a purpose other than what it is intended for. However, these tests have been used consistently by most states as language proficiency assessment measures for LM/LEP students (Canales, 1992; Newman et al., 1995; Moya and O’Malley, 1994). The achievement test is a measure of the English-monolingual population’s knowledge of its native language. It is designed in such a way that 30 to 40% of the English native language population, depending on the cut-off, will be below that cut-off point. Native English speakers could therefore be classifiable as LM/LEP. The bilingual students come mainly from disadvantaged socio-economic background. Disadvantaged students score lower than the average population. It is therefore safe to say that a higher proportion of bilingual students than the average
population will fall below the cut-off by the fact that these students are socio-economically disadvantaged and that they are tested with an instrument that is not language sensitive. The appropriate testing instrument would be able to discriminate if the student has acquired enough English to function with ease in society as well as with the academic language that will make it possible for him to be successful in the monolingual class. Cummins (1982, 1983, 1984) discussion of BICS and CALP has demonstrated that we have been successful at facilitating the acquisition of communicative language, however, through this process, we may have been exiting students who were not necessarily equipped to meet the academic challenge of the monolingual class. A lack of reading/writing tests designed for LEP assessment has lead districts to rely on Standardized Assessment Tests for the placement of LEP students. Duncan and De Avila (1988) see four problematic areas with this use of norm-referenced tests.

1. The tests assess across a wide variety of academic knowledge. Math computation, for instance, may be included in the test.

2. Not all the items on the tests are of equal probability across all levels of ability. The tests provide a wide range of items in terms of difficulty.

3. The tests do not assess oral proficiency.

4. The students may not understand the tests' instructions if their oral skills are below a certain level.

Rossell and Baker (1988) noted that states have set up their cut-offs at various
arbitrary levels according to what they deem is a determinant of success. Walton (1989, as cited in De Avila, 1990) noted that the cut-off marks of various states may be as low as the twenty-third percentile. He added that there was no indication that the cut-offs were empirically selected and that their selection was founded on Chapter I criteria and availability of resources. The test, however, does not clarify the reasons for the student’s difficulties when he fails to reach the cut-off (Rossell and Baker, 1988). The actual purpose of these minimum passing grade cut-offs is to determine who in the monolingual population will be receiving compensatory education. As noted earlier, while this mark determines the monolingual population’s measure of success in mastery of academic skills, it is not meant to be a measure of language acquisition for the bilingual population. Many times students with low level of English end up selecting answers at random on multiple-choice tests yielding a flawed measure that can affect classification. Neither does the achievement test give us a valid measure of the LM/LEP student’s knowledge of the language nor does it give us an idea of academic capability. Only an achievement test in the student’s native language can give us a measure of his academic knowledge. A score of 20 could have been the assessment of a bright student as well as that of a weak one. We should be able to differentiate these students in order to tailor the academic program we are to offer students of different abilities, as the weak student may have already reached his full potential while the other is just at the threshold of learning (Rossell and Baker, 1988).

Newman et al. (1995) noted that the assessment of isolated skills does not correspond to the current literature based curriculum and ignores the student’s motivation
and prior knowledge. They added that the use of traditional assessment may force teachers to abandon modern teaching practices to revert to skill-based format in order to prepare their students for the tests. Hamp-Lyons (1992) added that these tests may be related to proficient writing, as the student demonstrates he can select accurately the appropriate discrete elements, however, they do not demonstrate the production of proficient writing.

We have seen how the use of standardized tests for the purpose of classification can misdirect students into programs where they do not belong. Standardized tests should never be used on students with minimal English. However, these can become a diagnostic tool with the more advanced students (Pelavin and Baker, 1987, as cited in Rossell and Baker, 1988). Still, only teacher judgment and observation data can validate its potential effectiveness in individual cases (Canales, 1992).

Language Proficiency Test

Placement of LM/LEP students into language programs is based on the results of Language Proficiency Tests. Often, the language test is used at exit as well as entry. At entry, the test may be the only criterion, at exit, it may be used in conjunction with a standardized assessment test, or it may figure as a component of a multi-exit criteria process. The importance of the language proficiency test in the placement process cannot be denied. Rossell and Baker (1988) point out that both Texas and California have developed commissions to study and identify valid language assessment tests that could be used in their respective states. They found that there were no psychometrically acceptable tests of language competency in existence. These researchers have produced
an extensive list of similar studies. They have also demonstrated the lack of validity and
reliability of available instruments. Although specifically designed to measure second
language development, language proficiency tests share the flaws noted in achievement
tests. Cut-offs are determined arbitrarily. Where mandates are such, that the language
test is the sole or main criterion, a single point or even a few can determine placement in
or out of the program. De Avila (1990) saw this procedure as dangerous and recognized
Kerlinger’s (1973, as cited in DeAvila, 1990) approach of determining cut-offs through
the use of a bandwidth to accommodate variations.

Just as was the case for achievement tests, a language test cannot determine
if a student failed because of language or because he is doing poorly in class. If language
proficiency tests were valid, they would yield common results regardless of which test is
used. However, Ulibarri et. al (1981) found that the results of three oral proficiency tests,
the Language Assessment Survey (LAS), Bilingual Syntax Measure (BSM) and Basic
Inventory of Natural Language (BINL) failed to show commonality in test results.
Russell and Baker (1988) report a study by Cervantes (1979) that presents similar results.
De Avila (1990) points out that the difference in the level of difficulty (P-Values)
between the different language assessment tests contribute to the variances in results,
"thus, students could be kept out of programs through the use of easy tests to identify
eligibility, or kept in through the use of a very difficult test" (p.30). Research conducted
by Pelavin and Baker (1987) attempting to find agreement in the results of various tests
of language proficiency demonstrated that there existed more agreement among the tests
in the entry sample than in the exit sample. They hypothesized that the disagreement
between tests was occurring for those students who knew some English. The tests were more in agreement for the students who knew no or very limited English. When students begin to possess a demonstrable amount of L₂, the tests lose their accuracy. The intermingling of language and non-language factors, which are difficult to identify give results that cannot tell us if the student is classified for reasons of language or because of factors not related to language.

Ramirez, Yen and Ramey (1986, as cited in Rossell and Baker, 1988) have also demonstrated lack of consistency in results. They analyzed the test results of the IDEA Proficiency Test (ITP), a test approved by the state of California for the purpose of evaluating language proficiency. Even with minimal exposure to English, there should have been some growth, but the researchers found that in many cases there was none. They observed loss for more than half of the students. On a sampling of 573 kindergartners, 40% of students had made no progress in English and 17% knew less English two years later. Results were similar with 1st graders: 50% showed no progress and 13% demonstrated less English. The 3rd grade sample indicated 48% had made no progress and 7% knew less English two years later. Considering the lack of consistency in the assessment outcome of LEP students, we may ponder with De Avila (1990): “If the assessment of language proficiency fails to produce a linear relationship, the entire system can be questioned” (p. 30).

There are obvious limits to language proficiency tests. De George (1988) noted that a sole measure of language assessment may not be sufficient to point out the
language needs of LEP students, and that these tests are often not, on their own, suitable to predict student expertise in the completion of academic work.

A New Look at Assessment: Authentic Assessment

The previously discussed tests (language proficiency tests and standardized achievement tests) are required sources of information for current assessment of the LM/LEP. However, they only give us a narrow description of the student at a single point of his experience. These tests are a constrained simulation of the student’s possible output. Traditional assessment evokes knowledge as discrete, hierarchically arranged units in which increased expertise is represented along a single dimension. This is not an actual reflection of the teaching that goes on in class nor is it a rendition of the students’ potential. Oller (1992) and Canales (1992) suggest that language competence is a more unitary concept that is best assessed through more integrated means. Evaluation of students’ progress should integrate the multiple processes and activities that take place in learning.

Assessment should actually evoke the ability of the student to engage in complex performances of the processes and strategies used by students engaged in such performances, or of the mental representations that students bring to bear in solving complex problems. (Camp, 1993, p.184)

A typical problem associated with the use of traditional assessment methods, especially the use of the achievement test, is that the curriculum tends to focus on the test because of the high stakes attached to the test result. In our zeal to have students succeed, we have come to equate teaching to being successful in the test. Therefore,
instead of the student learning vocabulary from the work they read and applying it in
their writing they end up memorizing lists that are likely to be used in the test (Gitomer,
1993). If teaching "to the test" is a common though not recognized strategy in
education, perhaps we have reached a point where we are developing assessment that
can demonstrate that teaching to the test is a valid use of instructional time.

Systematically valid assessment instruments are those that foster the type of
learning and performance that are deemed critical to the educational mission.
Assessment activities that are systematically valid are worthwhile learning tasks in and
of themselves (Gitomer, 1993). The following characteristics are descriptive of
activities that are usable for authentic assessment.

Characteristics of Activities

1. Related to real world skills.
2. Integrate knowledge and skills.
3. Extend flexible time frame.
4. Display evidence of collaborative work.
5. Complex and challenging.
6. Allow opportunity for support.
7. Consistent with goals for learning.

Authentic assessment by its definition includes performance assessment, where
the student is asked to complete a task that will demonstrate his mastery of a skill or
combination of skills (Camp, 1993).
Portfolios: The True Value of the Student’s Ability

Among the various possible formats for performance assessment: open-ended questions and problems, essays, hands-on activities, extended projects, computer simulations, portfolios, the latter, appears to be the most immediately accessible and least technical means to integrate complex performances into assessment and to bring assessment procedures into the everyday life of the classroom. Teachers have always kept folders of students’ work; however, these may have been the repositories of various tests and quizzes. The task now focuses on collecting data that mirrors the Reading/Writing curriculum, and chronicles individual student development. Wolf (1989) defines a portfolio as “a sequenced collection of work that gives insight on the evolution of artistic thinking” (p. 24) or it can be seen as a collection of student works that feature his effort, progress and achievements (Paulson, Paulson & Meyer, 1991). It evidences performances based on authenticity, is on-going, multidimensional, and provides collaboration between teacher and student (Valencia, 1990). Clearly there exists in the literature a notion that portfolios should produce a picture of both educational processes and outcomes. The portfolio enables the viewer (assessor) to see what the producer (student) can do and how he thinks, works, and develops.

As the focus is on the student, authentic assessment becomes a personalized form of assessment where a representative sample of the student’s work is selected as a means to evaluate growth. This allows us to eliminate variations in performance because of lack of familiarity with the topic, lack of linguistic or contextual knowledge because of age or capability, or having to deal with a topic that triggers emotional response. When
used appropriately, portfolio processes are more reliable than psychometric assessment (Yawkey, T. D., Gonzalez, V. and Juan, Y., 1994). Baker (1991) finds that “the most useful studies on alternative assessment in the ERIC educational database are those conducted in the writing assessment area” (p.3). These have focused the evaluation on:
1. Qualities central to writing performance.
2. Use of exemplars of writing and rubrics to inform scorers’ judgment insuring reliability by having scorers internalize the criteria on which their judgments are to be based (Frederiksen and Collins, 1989).

For assessment to be meaningful, it must underlie clear goals through adequate methodology. Stiggins (1991) decries the lack of assessment literacy among American educators. He defines “assessment literates” as those who can recognize that assessment targets are unclear. He points out that assessment methods are missing their target, that samples of performance are inadequate, that there are specific extraneous factors creeping into assessment data and that assessment results are unclear. The training of educators in assessment methods is essential to the creation of assessment forms that are more valid, reliable and appropriate. Wiggins (1989) points to the lack of target identification in assessment. He sees these targets as standards of what we value as excellence in education. Our current forms of assessment lack these benchmarks or exemplars that excellent teaching and excellent work can be set to. He also adds that at exit level, the work should take several varied forms to which a common set of standards (criteria) should be applied. Assessment focuses on a series of successive and progressive approximation to the exemplar. What we now need is “clear specification of exit level
results against which student work can be continuously compared... [assessment that is] criterion referenced rather than normative, longitudinal rather than periodic, output rather than input driven” (French, p. 257-8).

**Summative and Formative Assessment**

The trend to develop portfolios in the classroom is increasing. Sheppard (1994) noted that elementary school teachers are more likely to use portfolio assessment than middle and high school teachers and that in over half the programs using content-ESL, portfolio assessment and/or progress check lists were used. Johns and VanLeirsburg (1991) also demonstrated evidence of this trend where over a year; an increase of 25% was noticed in the number of teachers who use portfolios. The study indicated that teachers appreciated the portfolios to substantiate a referral, but found difficulties in basing a summative assessment on them. “Teachers have relied on numerical cut-offs for placement and now find it difficult to evaluate and pass judgment on a much broader output” (Johns and VanLeirsburg, p.8). At the present time, portfolios are mainly used as formative instructional tools. Teachers may find divergence in the idea of combining instruction and evaluation within the portfolio, but assessment should be aligned with instruction, and results should guide the direction of instruction. Task accomplishment often undergoes the influence of monitoring from the teacher where she takes advantage of the learning context to clarify, adjust her teaching or assist a student in need of help. In this case, the task is not suitable for summative assessment where validity and reliability of scores must be maintained. In a summative context, the student must be allowed to develop and complete the task without assistance in order to be able to demonstrate mastery of a standard or level of progress toward mastery. After
completion, it is possible to adjust instruction, remediate and offer opportunity to complete similar tasks in order to give evidence of learning.

To satisfy concerns for maintenance of standards in a summative evaluation, ideal circumstances require that personnel be experienced with student writing and that the students' work be assessed anonymously under controlled conditions. Standards are set at the beginning of the scoring session and reinforced throughout. The rater's application of the standards is monitored during the session and through statistical analysis afterward. Examples of the writing samples at each point of the scoring scale are published to make explicit the criteria and the standards applied in the evaluation, to encourage discussion, and make scores interpretable in non-technical terms (Camp, 1993). In the formative mode, the focus lies more on progress in performance. It may seem that portfolios tend to blur the line between summative and formative assessment. It remains the responsibility of the teacher to maintain the integrity of the summative function of the portfolio.

**Elements Appropriate for Inclusion in the Portfolio.**

Portfolio designers need to identify tasks and instruments to be used to measure whether goals have been attained. Standardized tests are often required for district accountability need. However, their result should be weighed along with those of anecdotal records, rating scales, teacher's observation checklist, and work samples. A combination of traditional and performance assessment measures provide a range of indicators of the student's ability level. The total information is more descriptive than the results of standardized assessment alone. Furthermore, having multiple indicators of student performance enables teachers to crosscheck one type of information against another. The variety and number of tasks found within the portfolio is likely to be
determined by the subject area, processes involved in the assessment, and characteristics of the students.

The ability to use the computer as a research and production tool can be displayed in the literacy portfolio. While at the threshold of a new century, students should be able to demonstrate their ability to use and communicate through the computer. The computer can also come to the help of the teacher in storing and retrieving print, voice, photographs, videotapes. Such a system is currently used at Linda Vista Elementary School in San Diego, California, one of the RJR Nabisco Foundation Next Century Schools (French, 1992).

Managing portfolios through an efficient and comfortable system is necessary for their continued maintenance. Newman et al. (1995) saw that if the task becomes too burdensome teachers just will not use portfolios, or they will use them as containers that serve no practical assessment purpose.

While guidelines may be given in construction and assessment of the literacy portfolio, certain features are recognized as required elements at all levels. Camp (1990) indicates that writing entries should evidence the whole process and not just the finished product, and that cooperative learning activities should be featured in the portfolio. Portfolios should give evidence of the students' awareness of the process and strategies and of what they value in their writing. The regularity in the presence of these elements assures us that we are sharing a common understanding of what the student's portfolio should include as well as other features having to do with the design, construction, and evaluation of portfolios (Camp, 1990, 1993; Wolf, 1989).
An intrinsic part of the completion of the portfolio is the student's participation in its individual development. Therefore, portfolios should provide the students with opportunities to engage in self-reflection and involvement in selecting the pieces for the portfolio. One should derive a sense of the student's activity and intentions in generating it. The student is given opportunities to display the materials that he is willing to share and that illustrate growth. In return, the student should derive some measure of control in the creation process, an understanding of the metacognitive aspect of his work, and a feeling of cooperation between himself and his teacher (Gonzalez et al., 1997). The process of conferencing should encourage self-editing and self-evaluation (Baker, 1991; Newman et al., 1995). The degree to which the student takes advantage of the opportunities provided by the portfolio depends for a large part on the student himself. He shares responsibility for the climate of the classroom, and the direction and motivation for the development of the portfolio activities. In programs where these elements are allowed to develop, the portfolio becomes a catalyst for enriched experiences alongside instruction. Test scores and other cumulative folder information can be included if they are clarified through the portfolio. Students should have the occasion to view models of portfolios and of the processes others have used to create, and reflect on them. Portfolios may serve many needs, but they should not conflict (Paulson et al., 1991).

Chamot and O'Malley (1992) present the elements that four districts selected to be included in their portfolio programs (see Table 3). One can see the priorities expressed by each one of them. Still many common elements transpire. Though it may be helpful
Table 3

Reading/Writing Portfolios: Sample Contents

**Arlington County Public Schools, Virginia, Elementary ESOL/IHILT Program**

Reading
- Teacher observation log
- Examples of what student can read
- Books / materials read
- Audiotape of students reading
- Test results, formal and informal
- Conferencing form
- Examples of skills mastered

Writing
- First piece of writing each year
- Learning log, dialog journal
- January and May writing samples
- Drafts and final products from different genres
- (personal narratives, exposition, letters, poems, essays, reports)
- Graphics (illustrations, diagrams)

**Stratham Memorial Elementary School, New Hampshire, Reading / Writing Portfolio**

Reading
- Favorite books / authors list
- Genre graph, indicating type of literature preferred
- Journal entries
- List of completed books

Writing
- Writing sample and cover sheet
- List of completed pieces
- Evaluation
- Goals and/or self-evaluation
- Annual narrative summary by student

**Orange County Public Schools Florida, Literacy Portfolio Components**

**Core Elements**
- Reading Development Checklist
- Three writing samples
- List of books read independently

**Optional Elements**
- Student self-assessment
- Reading journals
- Audio-tapes of student reading
- "Things I Can Do" List
- Test results, formal and informal
- Reading comprehension tests
- Reading records (miscue analysis) and anecdotal records

**Fairfax County Public Schools, Virginia, ESL Program**

**Core Elements**
- Two writing samples
- Informal reading assessment
- List of books read
- Reading/Writing checklist

Results of Degree of Reading Power test
(Grades 7-12)

**Optional Elements**
- Student self-assessment
- Audio/video tapes
- Teacher observations
- Dialogue journal excerpts
- Student selected work
to look at the decisions taken by a few districts, the choices made for each individual program will reflect its needs and intents. Other issues will influence the design of a portfolio program, such as the time span the portfolio should cover and the types of baseline entries that should be selected. Programs that develop portfolios of LM/LEP students should also clarify their exit standards and use raters who speak the native language of the student to judge the entries, even those in English, so that the characteristics of the product attributable to native language can be identified.

**Baseline Assessment**

The foundation of the literacy portfolio lies in the collection of production samples on which comparison for establishing level of growth or proficiency will be based. Some sampling will be done at entry in the program or at the beginning of the school year or marking period. The baseline sampling should attempt to emulate all the areas we mean to evaluate. The inclusion of work samples that range from inclusion into the program up to the present will clearly show growth, its rate and its extent.

It is of the utmost importance to include native language sampling of the LM/LEP student’s work to evaluate academic knowledge and native language literacy, since these may be the only possible way to estimate strength and later on, progress. Any academic record from the country of origin may also be included in the portfolio.

**Student's Attitude: An Aspect of the Portfolio**

Since the portfolio will give a unique chronological perspective of student growth, it is important to measure changes in perceptions and feelings as well as academic
outcomes. Walters (1992) sees portfolios as unique opportunities to assess products as well as perception and reflection. The opportunity for the student to express his attitudes, feelings, insights toward his achievement must be clearly offered. It will not necessarily come forth as a by-product of the student’s academic production. However, the ability of the student to pass a critical judgment on his work is founded in his participation in its individual development (Paulson, et al., 1991). French (1992) describes this as “thinking skills” which are not to be confused with those commonly recognized as the academic variety (i.e., higher order thinking skills). They display the ability of the student to self-assess and evaluate strength and weaknesses. They demonstrate the intrapersonal intelligence described by Gardner (1983) and confirm personal and intellectual maturity.

The Portfolios of LM/LEP Students

Portfolio development offers particular advantages to the LM/LEP student and his teacher that no other form of assessment can yield. Because the progress of the student is apparent not only in acquisition of knowledge, but also in language development, the assessment featured through the student’s portfolio will focus on both areas, thus offering multiple perspectives on the student. The portfolio should feature a wide variety of assessment formats such as direct observations, anecdotal records, check lists, and rating scales in order to get a complete and accurate description of his ability. The portfolio also offers a longitudinal record of the student’s progress (Gonzalez, V., et al., 1997). It offers an ideal tool to confer with the student or the parents and plan for instructional strategies.
O’Malley (1992) points out that the design of alternative assessment should begin with an analysis of the curriculum of the content area rather than with an analysis of the language syllabus, this will allow a better focus on the language that needs to be assessed. Multiple forms of assessment are required to give a varied perspective on the student’s academic growth. He sees alternative assessment “as a form of domain assessment. It must be valid for concepts, skills and language used, it must also be continuous in order to reflect student’s understanding and use of curriculum content over the year” (O’Malley, p.179). The author sees a need for new kinds of performance assessment instruments that will assess authentic academic tasks where higher order thinking skills and language functions are required from the student.

Since the design of each portfolio program will vary according to various factors, and that the influence of the environment will create fluctuations in how portfolios are administered at the classroom level, it is important that the foundation is laid clearly and that guidelines ensue. O’Malley (1992) recognizes 5 different stages of portfolio program development process that are schematized in Figure 1. This process is dependent on staff participation, therefore, it is important to offer school personnel informative in-services and secure widespread participation in the elaboration of the portfolio program.

The most problematic area of portfolio development lies in interpreting the data collected. We should be able to equitably demonstrate success or need for growth in particular areas or skills and at specific time or periods of the year for individual students. This assessment has to be made with a maximum of consistency (reliability) and with
Figure 1. Five Stages of Portfolio Program Development

the assurance that what is measured corresponds to what is meant to be measured (validity). To ascertain that the performance based assessment program will maintain its validity, it should include features such as:

clarifying the purpose of the assessment, identifying the consequences or specifying uses to be made with the results, and defining in explicit, observable terms the tasks and performance criteria to be considered in the assessment...

"authentic" performance assessment of student's performance on instructional tasks must: be accurate and viable, include the fundamental constructs of measurement, and demonstrate how they will contribute to the improvement of instruction and learning, especially for LEP students (Navarrete, p. 184).

Yawkey et al. (1994) present the following evaluation criteria as guidelines to help assess the Language portfolio of a LEP student:

1. Thoroughness of children's work samples.

2. Variety and type of products shown in the portfolio.

3. Children's generations of ideas and interpretations required by various examples.

4. Types of understandings, skills and attitudes...exemplified in the work samples.

5. Examples of...transfer between native and English language...levels.

6. Reliability of assessment of portfolio items with other related work samples not included in the portfolio. (p. 139)

The Language portfolio production for LM/LEP students is centered on the three major areas of literacy: reading, oral, and written English. The portfolio will also display
developing language experiences in areas of the LM/LEP student’s curriculum other than English, such as in Science and Social Studies.

**Oral Language Assessment**

Oral language assessment reflects an evaluation of the language the student uses in the classroom to complete daily tasks. This can take various forms such as story retelling, interviews, story completion, direct dialogues, response to picture cues, student self-evaluation. As time constraint is an important factor in the assessment of oral language, the testing of pairs or small groups of students is a viable alternative. The assessment of academic language should be performed through the means listed above. Academic language is the cognitively demanding and contextually reduced language of content area instruction and is critical for success in mainstream classrooms (Cummins, 1982, 1983, 1984). Academic language functions are the essential communication tasks that students must be able to perform in different content areas. They determine if the task will be simple or complex (Chamot and O’Malley, 1987). The classroom context will define the academic language that will need to be evaluated. Academic language may evolve using language functions such as identifying and describing content information, explaining a process, analyzing and synthesizing concepts, justifying opinions or evaluating knowledge (see Table 4). The language may be unidirectional: teachers and textbooks impart the information and students demonstrate their comprehension by answering oral and written questions. On the other hand, language may be interactive: teachers and students get involved in discussions, share ideas, and argue about values. Academic language functions are required in all content areas. The use of lower and higher order thinking skills is required to perform
Table 4

Academic Language Functions

<table>
<thead>
<tr>
<th>Language Function</th>
<th>Students Use Language to:</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seek Information</td>
<td>Explore the environment or acquire information</td>
<td>Use who, what, where, and how to collect information.</td>
</tr>
<tr>
<td>Inform</td>
<td>report, explain or describe information or procedures</td>
<td>Retell story or content-related information in own words, tell main ideas, summarize.</td>
</tr>
<tr>
<td>Analyze</td>
<td>Separation of whole into parts</td>
<td>Tell part or features of object or idea.</td>
</tr>
<tr>
<td>Compare</td>
<td>Analyzes similarities and differences in objects or ideas</td>
<td>Indicate similarities and differences in important parts or features of objects or ideas, outline diagram web, indicate how A compares / contrasts with B.</td>
</tr>
<tr>
<td>Classify</td>
<td>sort objects or ideas into groups and give reasons</td>
<td>Show how A is an example of B, how A is related to B, or how A and B go together but not C and D.</td>
</tr>
<tr>
<td>Predict</td>
<td>Predicts implications</td>
<td>Predict implications from actions or from started text.</td>
</tr>
<tr>
<td>Hypothesize</td>
<td>Hypothesize consequences</td>
<td>Generates hypotheses to suggest consequences from antecedents.</td>
</tr>
<tr>
<td>Justify</td>
<td>give reasons for an action, a decision, or a point of view</td>
<td>Tell why A is important, why you selected A, why you believe A.</td>
</tr>
<tr>
<td>Persuade</td>
<td>Convince another person of a point of view</td>
<td>Show at least two pieces of evidence or arguments in support of a position.</td>
</tr>
<tr>
<td>Solve problems</td>
<td>Determine solution</td>
<td>Given stated problem, reach solution.</td>
</tr>
<tr>
<td>Synthesize</td>
<td>Combine</td>
<td>Put A together with B to make C, predict or infer C from A and B, suggest a solution for a problem.</td>
</tr>
<tr>
<td>Evaluate</td>
<td></td>
<td>Select or name criteria to evaluate, prioritize a list and explain, evaluate an object or proposition, indicate reason for agreeing or disagreeing.</td>
</tr>
</tbody>
</table>


the functions. Lower order thinking skills and less cognitively complex tasks include recalling information, identify vocabulary, and give definitions. Higher order thinking skills involve using language to analyze, synthesize, and evaluate. There is clearly a relationship between the higher order thinking skills and the higher level of difficulty of academic language tasks. If society values the acquisition of higher order thinking skills
and higher level of academic language, our assessment should also reflect that we value their acquisition. We should also value the acquisition of strategies that facilitate the use of academic language and the mastering of the learning process. Scoring of oral language should be holistic with the focus on communication ability. Holistic scoring will ascertain that features such as grammatical accuracy and pronunciation do not dominate assessment. Depending on the objectives of instruction, if linguistic features need to be evaluated, they can be assigned a subscore lesser than the weight assigned to communicative ability.

Reading Assessment

Standardized tests focus on isolated reading skills, whereas performance assessment of reading skills focuses on reading comprehension. The literature lists several approaches: miscue analysis (Goodman and Burke, 1972), Individual Reading Inventory (Cunningham, P. M., Moore, S. A., Cunningham, J. W., and Moore, D. W., 1983), Cloze tests (Cohen, 1980; Oller, 1979) and anecdotal record (Goodman, Goodman and Hood, 1989). Pierce-Valdez and O’ Malley (1992) suggest an additional approach, the use of rating scales or checklists. The scales are advantageous, as they are systematic, require little teacher time and do not disrupt instructional time. The rating scales list performance tasks and the level at which each performance is demonstrated as shown in Pierce-Valdez and O’Malley (see Figure 2). Rating scales for reading comprehension focus on four main areas: reading skills, interest, applications, and reading strategies. Entries in each area can be filled in bi-weekly, monthly, or at the marking period.

Reading skills reflect the student’s ability to read a given text. Teachers may select appropriate objectives aligned with the curriculum. Decoding can be assessed through
Figure 2. Literacy Development Checklist

<table>
<thead>
<tr>
<th>Student</th>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>Academic Year</td>
</tr>
</tbody>
</table>

Mark: Effective X  Sometimes Effective /  Needs Work --

<table>
<thead>
<tr>
<th>Reading Processes</th>
<th>Quarter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. READING SKILLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluent decoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehends oral stories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literal comprehension in reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferential comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>II. INTEREST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiates own reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shows pleasure in reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selects books independently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samples a variety of materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>III. APPLICATIONS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participates in language experience story development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participates in reading discussion groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writes appropriate dialogue journal entries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chooses books of appropriate difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uses reading in written communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IV. READING STRATEGIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitors attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notices miscues that interfere with meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infers meaning based on:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word clues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarizes main ideas or key events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Links details to main idea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remembers sequence of events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicts conclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requests help if needed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Figure 2 is adapted from materials developed by the National Council of Teachers of English and by The Writing Lab of the University of New Hampshire
oral reading, and comprehension through questions on the text. Specific objectives such as inferential comprehension can be included. The level of interest in reading can be assessed by the behaviors displayed in regards to situations involving reading. The area concerned with applications describes how the student integrates reading with speaking and writing. Applications can include art production with written description and oral presentation along with a self-assessment on reading ability. Information can be collected to assess specific reading skills. Questions geared towards each skill can be asked to individuals or small groups, after silent reading of a passage. The level of accomplishment in each reading skill needs not be dependent on oral or written proficiency. For example, expression of the main idea can be stated in the student’s own words or it can be selected from a choice of possible answers. The evaluated skill can be assessed through prioritizing a given list, through the formulation of such a list or through the writing of a statement that expresses the student’s assessment of an idea.

Students who use reading strategies display better text comprehension than students who do not (Dole, Duffy, Roehler and Pearson, 1991). LM/LEP students also benefit from these strategies (Chamot & O’Malley, 1987; Padron and Waxman, 1988). Reading strategies assessment can shed light on what actions the student takes to assist his understanding and his memory of the materials he reads. Reading strategies evaluation can become part of the self-assessment of the portfolio through the development of a log. This can take the form of a list of strategies that the student identifies as desirable or recognizes as part of his learning behavior (see Figure 3, Pierce-Valdez and O'Malley, 1992). This checklist completed by the student will give the teacher valuable
information as well offer the student an opportunity to evaluate and plan behaviors for future activities.

Research on self-assessment of reading proficiency is rather limited, and most of it has been conducted at the university level. It reveals that self-assessment is reliable when it is associated with student's reasons for using the language (Oscarson, 1989). An example of self-assessment of reading ability is given in Figure 4 (Pierce-Valdez and O'Malley, 1992). It should be completed after a reading assignment is done.

**Writing Assessment**

Evidence of the quality of the student's written expression and of the process applied in the production will be found in the portfolio. This will be demonstrated through writing samples, examples of process writing, and journals. The topics should be developmentally appropriate, interesting and familiar to the students. It should encourage a writing style that is appropriate to the grade level. Students could be given a choice of topics thus increasing chances that he will be able to express himself on one of the topics. Topics should not be culturally biased. Offering neutral themes prevents references to areas of knowledge unfamiliar to students. Students could be encouraged to rely on the scope of their experience, the wealth of their culture, the depth of their sensitivity, to write on topics that are of interest to them. They should be encouraged to write on school subjects or discuss problems in science or mathematics that they find particularly interesting. Some LM/LEP students often perform better in the assignments that incorporate more technical language as they have not acquired the higher level of
Figure 3. Learning Strategies Self-Assessment

<table>
<thead>
<tr>
<th>Statement</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>I paid attention to the teacher.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I took notes when I listened.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I looked at the questions before I read.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I looked at my notes after class.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I repeated new words aloud.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I used new words in sentences.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Self-Assessment of Reading Ability

In reading a passage, I can:

<table>
<thead>
<tr>
<th>Reading Task</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Understand the main idea</td>
<td>All the time</td>
</tr>
<tr>
<td>2. Understand the details</td>
<td></td>
</tr>
<tr>
<td>3. Understand the vocabulary</td>
<td></td>
</tr>
<tr>
<td>4. Read quickly and still understand most of it</td>
<td></td>
</tr>
<tr>
<td>5. Find out the meaning of new words from the story</td>
<td></td>
</tr>
</tbody>
</table>

Note: Figures 3 and 4 were adapted from Performance and portfolio assessment for language minority students. Pierce-Valdez and O'Malley (1992).

comfort in L2 required in creative writing tasks. Directions, such as the length of time allowed writing the assignment, the resources that are permitted, and the scoring method should be explained to the students. Attention to these details will prevent flaws in assessment.

Students should be exposed to a broad range of quality exemplars of the type that
should be produced. Exemplars are the model on which students will learn mastery, the choice of exemplars is therefore very important. They can be issued from peer’s work and be the source of group pride. Some should also come from the monolingual class, for students to be able to compare their work to that of native speaker’s at their level and evaluate what needs to be accomplished. Other exemplars, depending on the student’s level of ability, can be envisioned: an author’s published work, or a relevant article may become a tool the student can use to improve his work. However, care must be evident in selection of exemplars. They must display attainable performance. Work that is too complex is likely to discourage or confuse the students. In the discussion of exemplars, distinction between form and style must be clarified. Opportunities for application of both must be elicited. Publication of class work gives purpose to the writing and demonstrates in its authenticity the full extent of the writing process. Publishing student’s writings also serves to inform staff and parents of the progress of the students.

Students should be aware of what is valued in quality writing. The criteria should be discussed with examples of how they are applied. Opportunities for cooperative learning should be given; peer editing and discussions will ensue. Scoring of writing samples should feature levels of proficiency within a grade level, such as beginning, intermediate, and advanced. Grading by more than one teacher is preferable. The information yielded from the grading will be worthy as long as the student’s work is representative of the student’s effort, and the quality of the scoring procedure is uniform within the students and between the raters.
Collection and analysis of data.

Some information contained in portfolios should be required and other optional. The required work should be consistent through the portfolios, this will insure that the portfolios contains the information required to make fair instructional decisions. The optional items will encourage the student to participate actively in selecting his best work or the pieces that appeal to him the most. Guidelines should be formulated to determine the frequency of collection, removal or upgrading of the contents. Palmer-Wolf, D., LeMahieu, P. G. and Eresh, J., (1992) describe a format which presents increasingly comprehensive information on the student: the portfolio is reviewed monthly, from which is built a portfolio for the marking period, and a year-end portfolio is drawn from these.

Aligning the student’s work with curricular objectives will guarantee the reliability of the assessment. Moya and O’Malley (1994) have developed a Portfolio Analysis Form (see Figure 5) to demonstrate the cohesion of goal and work. A form is advantageous, as it will indicate students’ deficiencies and possible instructional changes in view of their needs. A comprehensive and descriptive form also acts as a filter of the work that may not be aligned with the goals or objectives and shows the relationship of the work with the instructional goals. The authors see a variety of ways the information contained in the portfolio and described in the Portfolio Analysis Form can be used, such as: diagnosis and placement, monitoring of student’s progress, feedback on effectiveness of instruction, communication with other teacher, feedback to students, and communication with parents.
Figure 5. Sample Portfolio Analysis Form

<table>
<thead>
<tr>
<th>Performance Task</th>
<th>Contents Illustrating Student Progress</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary Comments:

Time Demands and Other Concerns

Kolls (1992) and Tierney, R.J., Carter, M.A. and Desai, L.E. (1991) comment on teachers’ concern with the time commitment required by portfolio assessment. This is mainly expressed by teachers who are at the implementation stage of portfolio assessment. This concern lessens as teachers develop a classroom environment that encourages students to become increasingly independent and responsible for their own portfolios. Time spent on portfolio building can be reduced by:

1. Making students responsible for the collection of information that goes in their own portfolio.

2. Identifying and listing the information that is required in the portfolio.

3. At the initial stage, developing portfolios for only the few students who need intensive monitoring.

4. Sharing responsibility of collection and interpretation of data with staff to free teachers for other instructional duties.

Much of the assessment to be included in the portfolio will take place during the performance of class activities and will be recorded through the use of instruments such as rating scales, therefore requiring a minimal amount of time. The area of portfolio requiring the most time will be conferencing with the student about various aspects of the portfolio, and this is an area where time should not be sacrificed. Portions of conferencing can also become a tool for oral assessment and assessment of academic language. Naturally, the student has to be made aware that elements of conferencing become summative whenever the case arises. Oral assessment through conferencing should never cause the student to feel threatened, it should only be attempted when the
student feels reasonably confident in the language that the conversation between teacher and student will feel comfortable, and will focus on the topic instead of the language itself. If oral assessment at conferencing becomes a source of stress, conferencing should keep its formative nature, and oral assessment should be performed in another context.

Authentic assessment presented in the form of a literacy portfolio gives an accurate and complete representation of the student’s accomplishment over time. The portfolio is not only an assessment tool, it is also motivational as the student participates actively in its development and its evaluation. Portfolio development does not go without effort, but with proper planning, it will reward both students and teachers.

Summary

Quality programs are required because of heightened demands on the students and teachers. Students are capable of surpassing themselves and are eager to demonstrate the caliber of their effort. We now have to empower them with the platform from which they can demonstrate their knowledge and creativity.

Our well meant attempts at describing the second language learner have served only to give a partial and sometimes flawed picture of his true worth. The result has been that many times errors in placement have resulted. We may err in basing placement decisions on standardized testing formats, and we may also err in surrendering placement decision solely to teacher’s judgment, which may be well meant but may sometimes be misguided. Can relying on assessment that is based on a representative sampling of the student work serve to develop the best possible educational plan for him? This research will attempt to see what various assessment measures tell us about the LM/LEP student.
Chapter III

RESEARCH METHODOLOGY

The literature featured various studies concerned with the comparison of language proficiency tests (Rossell and Baker, 1988; Pelavin and Baker, 1987). These tests are mainly concerned with the basic interpersonal communication skills (BICS) described by Cummins (1982, 1983, 1984). With today’s rich, integrated curriculum and the intellectual demands required for the acquisition of higher order thinking skills, there is need for a language assessment instrument that is concerned with the academic language demands the student faces (Newman et al., 1995). The results of an instrument such as the W-MLS, that can estimate the student’s level of familiarity with language irregularities, and the teacher’s assessment of the student’s classwork through the Language portfolio, could be concurring indicators of his likelihood to succeed in the mainstream class.

Instrumentation and Procedures

The English section of the Language Assessment Battery (LAB) is given to all the LEP students of the district at the end of each school year. These scores, along with the multi-exit criteria selected by the district are used to determine which students exit from the program.

The English section of the Woodcock-Munoz Language Survey (W-MLS) was given to the students who, and whose parents, had elected to participate in the study.
Both the LAB and the W-MLS were given at the end of the academic year, within three weeks of each other.

Written permission was requested from the district superintendent, the principals of the three schools where the study was to take place as well as from the Bilingual-ESL Department teachers (see Appendix C). The entire staff agreed to participate. Teachers were in-serviced on their volunteer participation in the completion of the rubrics and on the features and administration of the W-MLS. Although portfolios are developed in the classes and specific pieces are required for inclusion, such as unit tests and formal monthly writing samples, there is no district-wide recommended formula to score the combined work. For the purpose of reliability, parameters were set on what would be included in the assessment of students’ portfolios. For the study, teachers were asked to complete their assessment for the last trimester of the school year. They were also asked to exclude from their assessment any unit, district or standardized test that were part of the student’s collected work.

Three levels of rubrics (see Appendixes A1 to A8) were developed according to grade level or cluster of grade levels in order to establish a quantitative value to the teacher’s assessment. The levels are: Kindergarten, First and Second Grade, Intermediate (Grades Three to Eight). The rubrics were selected or developed to parallel the sub-tests of the W-MLS and the LAB. Therefore, for each of the three levels of rubrics, there is an Oral, a Reading, and a Writing Rubric.

The Oral Rubric used for grades one to eight is currently used in the district. Permission was granted by S. Moya, to use her Holistic Criteria as elements of the
Writing Rubric (see Appendix B). The Reading Rubrics were developed by this writer based on the reading series Treasury of Literature, (Farr and Strickland, 1995) which is presently used in the district where this study took place. The various categories featured in the Reading Rubrics are centered on the skills listed in the Treasury of Literature series’ guides. The rubrics were piloted on a group of 50 LM/LEP students the previous year. For developmental reasons, a set of rubrics for the three areas of language had to be developed specifically for the kindergarten level. These rubrics are based on the district’s kindergarten curriculum. All rubrics satisfied a jury of experts in the field. The jury of experts consisted of mainstream teachers, Language Art Supervisor, English as a Second Language Supervisor, and School Principal.

For 47% of the students in the study, the teacher is responsible for both ESL and Bilingual instruction. Students in these classes have only one language teacher. Therefore, for these students, it was not possible to compare the assessment of two teachers for one student, and establish interrater reliability.

Participants

All 120 students participating in the study are at the elementary level in an urban district in the state of New Jersey. They participate in self-contained bilingual classes. They receive 45 minutes of English as a Second Language instruction as part of their daily schedule. Spanish is the native language of 51% of the students, the remaining 49% speak Haitian Creole. The students included in the study are divided between three different schools. There are two magnet schools designated by language for grades K to 6. Students in grades 7 and 8 are at the Middle School.
A Parental Consent (see Appendixes D1 to D3) form in the native language was required of each potential participating student to be taken home. In turn, students who were granted parental permission signed a Student Assent (see Appendixes E1 to E3) to be included in the study. Of the 130 Parental Consent forms sent, 120 were returned assuring that 93% of the students would participate. Most of the students have been enrolled in the program three years or less. Many come with weak native language skills. Of those students who have been in the program one year or more, 23% are entitled to Basic Skills instruction because of their failure to reach the minimal proficiency level on the California Achievement Test (CAT 5). Students who have been enrolled in a language program for less than a year are not entitled to Basic Skills instruction, as it is a remedial program in nature.

Procedures

Data Collection

After having secured the various signed permission letters, the tests, rubrics and worksheets were distributed to teachers. The language assessment tests were administered and scored by ESL teachers, and ESL and Bilingual Education teachers completed rubrics. Each teacher was asked to enter the information on a schedule-like worksheet (see Appendix F). The worksheet included the following information for each student: name, grade, date of birth, gender, the number of years in the program, and the scores for the various language assessment tests and their sub-tests. The tests, rubrics, and worksheets were then returned to the writer for compilation.
Data Analysis

The analysis helped answer the following question:

Do the LAB, W-MLS, and LPA consistently identify LM/LEP students in need of bilingual instruction?

The analysis also examined the following hypotheses:

H1: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the W-MLS total scores are used.

H1A: There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the W-MLS Oral subtests are used.

H1B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the W-MLS Reading/Writing subtests are used.

H2: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the LPA total scores are used.

H2A: There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the Oral Rubric of the LPA is used.

H2B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the Reading and Writing Rubrics of the LPA are used.

H3: There is no significant difference in the exit rates of the students when assessed with the W-MLS total scores from when the LPA total scores are used.
H3A: There is no significant difference in the pass rates of the students when assessed with the Oral subtest of the W-MLS from when the Oral Rubric of the LPA is used.

H3B: There is no significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the W-MLS from when the Reading and Writing Rubrics of LPA is used.

Finally, the following sub-hypotheses were considered:

H1: There is a significant difference in the exit rates of the LAB, W-MLS and the LPA results of LEP students when grade level clusters (K-3 and 4-8) are considered.

H2: There is no significant difference in the exit rates of the LAB, W-MLS and LPA when the native language (L1) of the students is considered.

The data collected for the study was entered in SPSS (Statistical Program for the Social Sciences). Descriptive and inferential statistics will be used to provide a profile of the students and their test results. Frequency distributions as well as percentage charts will be presented. Chi² analyses will be used to determine if the exit rates or pass rates in each hypothesis are different from what would be found if chance alone were operating.

Along with the statistical analyses, the study will provide a qualitative discussion of the results through teachers' interviews. Individually, teachers will be provided with the results of the statistical analyses. They will be asked for their perceptions on the outcome of the analyses performed on the tests results. Each interview is expected to last approximately 30 minutes. The results will be presented in the fifth chapter of the study.
Chapter IV

ANALYSIS OF DATA

The purpose of this chapter is to present the results of the statistical analyses generated on the data collected in this study. The chapter begins with a presentation of basic descriptive statistics used to describe the students that participated in this study. This is followed by the results of hypothesis testing, where the statistics used to analyze each hypothesis are presented.

Participant Data

One hundred twenty students participated in this study. In the participant pool, (58, 48.3%) were males and (62, 51.7%) were females. The participant pool was also nearly equally split between Spanish (61, 50.8%) and Creole (59, 49.2%) students. The participant grade levels, which ranges from kindergarten to eight grade are presented in a frequency distribution in Table 5.

Table 6 describes the length of stay in the Bilingual Program for the participating students (see Table 6). It is to be noted that the students are distributed normally, with the largest number in nearly equal numbers in the 1 to 2 year and the 2 to 3 year segments.
Table 5

Frequency Distribution on Grade Level

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>18</td>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>21.7</td>
<td>36.7</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>9.2</td>
<td>45.8</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>10.0</td>
<td>55.8</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>8.3</td>
<td>64.2</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>10.0</td>
<td>74.2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>8.3</td>
<td>90.8</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>9.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 6

Distribution on Length of Stay (years) in Bilingual Program

<table>
<thead>
<tr>
<th></th>
<th>Less than 1 year</th>
<th>From 1 to 2 years</th>
<th>From 2 to 3 years</th>
<th>3 years</th>
<th>More than 3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C*</td>
<td>S*</td>
<td>C</td>
<td>S</td>
<td>C</td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>9</td>
<td>13</td>
<td>18</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>37</td>
<td>33</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

*C: Haitian Creole
S: Spanish
Reliability Analyses

Interrater reliability analyses were conducted for the teachers’ ratings of the Student’s Portfolios. A sample of 56 ratings was used for the reliability analyses, with subgroups that represented each level. The number of subjects in each grade level subgroup ranged from 3 to 12. However, interrater reliability analysis could not be conducted for the 4th, 7th, and 8th grades because of the insufficient number of cases with 2 ratings available. The interrater reliabilities are presented in Table 7. With 2 raters providing the evaluation of each subject in each of the grade level subgroupings, interrater reliabilities ranged from .73 to .99, indicating that the reliability of the teachers’ ratings for the LPA was sufficient.

Table 7

Interrater Reliability Coefficients for LPA

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>12</td>
<td>.99</td>
</tr>
<tr>
<td>K</td>
<td>6</td>
<td>.73</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>.96</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>.89</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>.94</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>.88</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>.98</td>
</tr>
</tbody>
</table>

Research Hypotheses

Chi² analysis was the primary statistical technique used to analyze the data collected for hypothesis testing since the data were categorical. In instances where the
Chi$^2$ analysis resulted in cells with expected values of less than 5, the Fisher Exact Test was used (Downie and Heath, 1974). The results of hypothesis testing are listed below:

H1: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the W-MLS total scores are used.

H1A: There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the W-MLS Oral subtests are used.

H1B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the W-MLS Reading/Writing subtests are used.

Hypotheses 1, 1A, and 1B regarding the exit/pass rates of students on the LAB and the W-MLS tests are presented in Tables 8 through 10. The Fisher Exact test (Downie and Heath, 1974) was used to analyze Hypothesis 1 regarding differences in the exit rates for participants when LAB and W-MLS total scores were considered. The results, presented in Table 8, indicate that the exit rates were significantly different than if chance alone were operating [Note: Fisher Exact p=.00001]. The frequencies indicate that more than expected students were identified by both tests as pass and fail. Therefore, we can conclude that both tests essentially exited and failed the same students when total scores were used.
### Table 8

**Chi$^2$ Analysis of Pass Rates for LAB and W-MLS Total Scores**

<table>
<thead>
<tr>
<th></th>
<th>W-MLS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LAB</td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td></td>
<td>11 (2.8)</td>
<td>11 (19.3)</td>
</tr>
<tr>
<td>Fail</td>
<td>4 (12.3)</td>
<td>94 (85.8)</td>
</tr>
</tbody>
</table>

Note: Fisher Exact $p = .00001$

### Table 9

**Chi$^2$ Analysis of Pass Rates for LAB and W-MLS Oral Subtest**

<table>
<thead>
<tr>
<th></th>
<th>W-MLS-Oral</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LAB Oral</td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td></td>
<td>7 (1.8)</td>
<td>24 (29.2)</td>
</tr>
<tr>
<td>Fail</td>
<td>0 (5.2)</td>
<td>89 (83.8)</td>
</tr>
</tbody>
</table>

Note: Fisher Exact $p = .00004$

### Table 10

**Chi$^2$ Analysis of Pass Rates for LAB and W-MLS R/W Subtest**

<table>
<thead>
<tr>
<th></th>
<th>W-MLS - R/W</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LAB R/W</td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td></td>
<td>15 (5.6)</td>
<td>6 (15.4)</td>
</tr>
<tr>
<td>Fail</td>
<td>7 (26.4)</td>
<td>82 (72.6)</td>
</tr>
</tbody>
</table>

Note: $X^2 = 26.08$, $df = 1$, $p = .00001$
The Fisher Exact test was used to analyze Hypothesis 1A regarding differences in the pass rates for participants when LAB and W-MLS oral subtest scores were considered. The results, presented in Table 9, indicate that the pass rates were significantly different than if chance alone were operating [Note: Fisher Exact p=.00004]. The frequencies in Table 9 indicate that more than expected students were identified by both tests as pass and fail. As a result, we can conclude that no significant differences in pass rates were experienced, that is both tests essentially passed and failed the same students when Oral subtest scores were used. The Chi^2 test was used to analyze Hypothesis 1B regarding differences in the pass rates for participants when LAB and W-MLS Reading and Writing subtests scores were considered. The results, presented in Table 10, indicate that the pass rates were significantly different than if chance alone were operating \([X^2=26.08, \text{df}=1, p=.00001]\). The frequencies in Table 10 indicate that more than expected students were identified by both tests as pass and fail. As a result, we can conclude that both tests essentially passed and failed the same students when the Reading and Writing subtest scores were used.

**H2:** There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the LPA total scores are used.

**H2A:** There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the Oral Rubric of the LPA is used.
H2B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the Reading and Writing Rubrics of the LPA are used.

The results of the analyses for Hypotheses 2, 2A, and 2B are presented in Tables 11, 12 and 13. The Chi² test was used to analyze Hypothesis 2 regarding differences in the exit rates for participants when LAB and LPA total scores were considered. The results, presented in Table 11 indicate that the pass rates were significantly different than if chance alone were operating \([X^2=43.33, \text{df}=1, p=.00001]\). The frequencies in Table 11 indicate that more than expected students were identified by both tests as pass and fail (see Table 11). As a result, we can conclude that both tests essentially exited the same students when the total scores were used. Also noteworthy is the fact the LPA identified 24 participants as exiting that the LAB failed to exit.

Table 11

<table>
<thead>
<tr>
<th></th>
<th>LAB</th>
<th>LPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td>Pass</td>
<td>22(8.4)</td>
<td>0(13.6)</td>
</tr>
<tr>
<td>Fail</td>
<td>24(37.6)</td>
<td>74(60.4)</td>
</tr>
</tbody>
</table>

Note: \(X^2=43.33, \text{df}=1, p=.00001\)
The Chi² test was used to analyze Hypothesis 2A regarding differences in the pass rates for participants when LPA and LAB Oral subtests scores were considered. The results, presented in Table 12 indicate that the pass rates were significantly different than if chance alone were operating \[X^2 = 38.33, df = 1, p = .00001\]. The frequencies in Table 12 indicate that more than expected students were identified by both tests as pass and fail. As a result, we can conclude that both tests essentially passed and failed the same students when the reading and writing subtest scores were used. Also noteworthy is the fact the
LPA passed 26 participants that the LAB failed.

The Chi^2 test was used to analyze Hypothesis 2B regarding differences in the pass rates for participants when LAB and LPA Reading and Writing subtests scores were considered. The results, presented in Table 13 indicate that the pass rates were significantly different than if chance alone were operating \([X^2=39.00, df=1, p=.00001]\). The frequencies in Table 13 indicate that more than expected students were identified by both tests as pass and fail. As a result, we can conclude that both tests essentially passed and failed the same students when the Reading and Writing subtest scores were used. Also note that the LPA passed 20 participants that the LAB failed.

H3: There is no significant difference in the exit rates of the students when assessed with the W-MLS total scores from when the LPA total scores are used.

H3A: There is no significant difference in the pass rates of students when assessed with the Oral subtests of the W-MLS from when the Oral Rubric of the LPA is used.

H3B: There is no significant difference in the pass rates of students when assessed with the Reading and Writing subtests of the W-MLS from when the Reading and Writing Rubrics of the LPA are used.

The results of the analyses for Hypotheses 3, 3A, and 3B are presented in Tables 14, 15 and 16.
Table 14

Chi$^2$ Analysis of Pass Rates for W-MLS and LPA Total Scores

<table>
<thead>
<tr>
<th>W-MLS</th>
<th>LPA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td>Pass</td>
<td>12(5.8)</td>
<td>3(9.3)</td>
</tr>
<tr>
<td>Fail</td>
<td>34(40.3)</td>
<td>71(64.8)</td>
</tr>
</tbody>
</table>

Note: $X^2=12.59, df=1, p=.0003$

Table 15

Chi$^2$ Analysis of Pass Rates for W-MLS and LPA Oral Subtest

<table>
<thead>
<tr>
<th>W-MLS - Oral</th>
<th>LPA-Oral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>Pass</td>
<td>7(3.2)</td>
</tr>
<tr>
<td>Fail</td>
<td>48(51.8)</td>
</tr>
</tbody>
</table>

Note: Fisher's Exact $p=.003$

Table 16

Chi$^2$ Analysis of Pass Rates for W-MLS and LPA R/W Subtest

| LPA-R/W |
|---------|---------|
| W-MLS R/W | Pass    | Fail |
| Pass      | 22(10.4)| 10(21.6)|
| Fail      | 17(28.6)| 71(59.4)|

Note: $X^2=26.13, df=1, p=.00001$
The Chi² test was used to analyze Hypothesis 3 regarding differences in the exit rates for participants when W-MLS and LPA total scores were considered. The results, presented in Table 14 indicate that the exit rates were significantly different than if chance alone were operating \( [X^2=12.59, df=1, p=.0003] \). The frequencies in Table 14 indicate that more than expected students were identified by both tests as exiting or failing (see Table 14). Therefore, we can conclude that both tests essentially exited and failed the same students when the W-MLS and LAB total scores were used. Also noteworthy is the fact the LPA identified 34 participants as exiting that the W-MLS failed to exit.

The Fisher Exact test was used to analyze Hypothesis 3A regarding differences in the pass rates for participants when W-MLS and LPA oral subtests scores were considered. The results, presented in Table 15 indicate that the pass rates were significantly different than if chance alone were operating \( [\text{Fisher Exact } p=.003] \). The frequencies in Table 15 indicate that more than expected students were identified by both tests as pass and fail. As a result, we can conclude that both tests essentially passed and failed the same students when the oral subtest scores were used. Also noteworthy is the fact the LPA assessment passed 48 participants that the W-MLS failed.

The Chi² test was used to analyze Hypothesis 3B regarding differences in the pass rates for participants when W-MLS and LPA Reading and Writing subtests scores were considered. The results, presented in Table 16 indicate that the pass rates were significantly different than if chance alone were operating \( [X^2=26.13, df=1, p=.00001] \). The frequencies in Table 16 indicate that more than expected students were identified by both tests as pass and fail (see Table 16). As a result, we can conclude that both tests
essentially passed and failed the same students when the Reading and Writing subtest scores were used.

Sub - Hypothesis 1: There are significant differences in the exit rates of the LAB, W-MLS and LPA when grade level clusters K – 3 and 4 – 8 are considered.

Chi² analyses were used to examine this sub-hypothesis where the grade clusters consisted of students in clusters K-3 and 4-8 and their exit rates were compared when each of the three assessment instruments were used. These results are presented in Tables 17, 18, and 19.

These results indicate that significant differences were found between the grade clusters when the LPA [Χ²=7.65, df=1, p=.005] and the LAB [Χ²=5.02, df=1, p=.02] were used. The frequencies in Tables 17 and 18 indicate that more K-3 students and less 4-8 students than would be expected by chance alone exited when the LPA and LAB were used (see Table 17 and 18). No significant differences in exit rates were found when the W-MLS was used [Χ²=2.12, df=1, p=.14].

Table 17
Chi² Analysis of Exit Rates for Grades K – 3 and 4 – 8 Students on LPA

<table>
<thead>
<tr>
<th>GRADE</th>
<th>LPA</th>
<th>Grade 3 (25.7)</th>
<th>Grade 4 (20.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>34</td>
<td>40 (32.7)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Χ²=7.65, df=1, p=.005
Table 18

Chi² Analysis of Exit Rates for Grades K – 3 and 4 – 8 Students on LAB

<table>
<thead>
<tr>
<th>GRADE</th>
<th>LAB</th>
<th>4-8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K-3</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>17(12.3)</td>
<td>5(9.7)</td>
</tr>
<tr>
<td>Fail</td>
<td>50(54.7)</td>
<td>48(43.3)</td>
</tr>
</tbody>
</table>

Note: X²=5.02, df=1, p=.02

Table 19

Chi² Analysis of Exit Rates for Grades K – 3 and 4 – 8 Students on W-MLS

<table>
<thead>
<tr>
<th>GRADE</th>
<th>W-MLS</th>
<th>4-8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K-3</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>11(8.4)</td>
<td>4(6.6)</td>
</tr>
<tr>
<td>Fail</td>
<td>56(58.6)</td>
<td>49(46.4)</td>
</tr>
</tbody>
</table>

Note: X²=2.12, df=1, p=.14

Sub-Hypothesis 2: There are no significant differences in the exit rates of the LAB, W-MLS and Language portfolio when native language of the students is considered.
Table 20

**Chi^2 Analysis of Exit Rates for Spanish and Creole Students on LPA**

<table>
<thead>
<tr>
<th></th>
<th>LPA</th>
<th>SPANISH</th>
<th>CREOLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>32(23.4)</td>
<td>14(22.6)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>29(37.6)</td>
<td>45(36.4)</td>
<td></td>
</tr>
</tbody>
</table>

Note: $X^2 = 10.47$, df=1, $p=.001$

Table 21

**Chi^2 Analysis of Exit Rates for Spanish and Creole Students on LAB**

<table>
<thead>
<tr>
<th></th>
<th>LAB</th>
<th>SPANISH</th>
<th>CREOLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>15(11.2)</td>
<td>7(10.8)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>46(49.8)</td>
<td>52(48.2)</td>
<td></td>
</tr>
</tbody>
</table>

Note: $X^2 = 3.24$, df=1, $p=.07$

Table 22

**Chi^2 Analysis of Exit Rates for Spanish and Creole Students on W-MLS**

<table>
<thead>
<tr>
<th></th>
<th>W-MLS</th>
<th>SPANISH</th>
<th>CREOLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>8(7.6)</td>
<td>7(7.4)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>53(53.4)</td>
<td>52(51.6)</td>
<td></td>
</tr>
</tbody>
</table>

Note: $X^2 = .04$, df=1, $p=.83$
Chi² analyses were used to examine this sub-hypothesis where the Spanish and Creole participants were compared on their exit rates when each of the three assessment instruments were used. These results are presented in Tables 20, 21, and 22.

The results indicate significant differences in exit rates when the LPA was used to determine proficiency \([X^2=10.47, \ df=1, \ p=.001]\). The frequencies in Table 20 indicate that more Spanish students and fewer Creole students than would be expected by chance alone exited when the LPA was used (see Table 20). No significant differences were found between the exit rates when the LAB \([X^2=3.24, \ df=1, \ p=.07]\) and the W-MLS \([X^2=.04, \ df=1, \ p=.83]\) assessments were used.

Supplemental Analyses

Additional analyses were conducted to determine the impact on the extent of agreement between the assessments when the CALP cutoffs were changed to 3.5 instead of 4. These results are presented in Tables 23 and 24.

Tables 23 and 24 present the results of the W-MLS at a CALP level of 3.5 with total LAB exit rates and those with LPA (see Tables 23 and 24). In both cases, the results were highly significant indicating that the rate of agreement between the assessments is significantly greater than if chance alone were operating. In Table 23 where CALP was set at 3.5, agreement with the LAB assessment existed for 92 cases (see Table 23). Also note that when CALP was set at 4.0, agreement was found in 105 cases.
Table 23

Chi² Analysis of Exit Rates for LAB and W-MLS Total Scores-3.5 CALP

<table>
<thead>
<tr>
<th>W-MLS Calp3.5</th>
<th>LAB</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>16(7.0)</td>
<td>6(15.0)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>22(31.0)</td>
<td>76(67.0)</td>
<td></td>
</tr>
</tbody>
</table>

Note: X²=12.98, df=1, p=.0001

Table 24 presents the results of the CALP 3.5 with total LPA exit rates (see Table 24). Like the prior assessment with LAB assessments, the results were highly significant indicating that the rate of agreement between the assessments is greater than if chance alone were operating. When CALP was set at 3.5, agreement with the LPA existed for 94 cases. Note that when CALP was set at 4.0, agreement was found in 83 cases.

Table 24

Chi² Analysis of Exit Rates for LPA and W-MLS Total Scores-3.5

<table>
<thead>
<tr>
<th>W-MLS Calp3.5</th>
<th>LPA</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>29(14.6)</td>
<td>17(31.4)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>9(23.4)</td>
<td>65(50.6)</td>
<td></td>
</tr>
</tbody>
</table>

Note: X²=33.93, df=1, p=.0001
Chapter V

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

Summary

Purpose of the Study

The purpose of this study was to examine whether different language assessment tests would yield the same results. If the ultimate purpose of the test is to determine the language competency sufficient to work in an English monolingual environment, do the results of a test based on knowledge of grammatical structure equate those based on assessment of academic language? Furthermore, are results constant when academic language is assessed through a language assessment test, as well as when it is based on the student's actual school work?

Design of the Study

In Chapter 1, the problem is presented along with the purpose of the study, the various hypotheses, definition of terms, and the limitations, significance and organization of the study.

Chapter 2 is a review of the literature on language assessment. The chapter was divided in two main parts where in the first one, a historical overview and the various language assessment alternatives are presented. The second part of the chapter is an in-depth presentation on portfolios for the purpose of language assessment.

Chapter 3 describes the instruments used in the study, the subjects, and the procedure for data collection and analysis.
Chapter 4 presents the analysis of the data collected for the study. Tables are provided to present the results of the analyses and brief narratives follow the tables for the purpose of explaining the results. Interrater reliability analyses are conducted for the purpose of validating the results.

Chapter 5 summarizes the study, presents results as well as recommendations for future research and study.

Results

The results of this study indicate that all three tests: the LAB, W-MLS and LPA, selected the same students for retention into the bilingual program or for release into the mainstream academic program. The Oral and Reading/Writing subtests of the LAB, W-MLS and LPA also support these results. The study points to differences in the exit rates between the primary and the intermediate grade levels when the results of the LPA and those of the LAB are examined. No such differences were noted when the grade level clusters were examined with the W-MLS. Finally, the analyses found that differences in the exit rates of the LPA existed between the two linguistic groups: Spanish speaking students exited at a faster rate than the Haitian Creole students. These differences were not present in the results of the LAB and the W-MLS.

Conclusion

Findings and Discussions

It would seem at first glance, that varying the approach of a language assessment test would be likely to change the outcome of the test, but the analysis of the collected data failed to show divergence. The following will present possible explanations for the outcomes. The section will also feature the teachers' input on the results of the study.
Comparing the W-MLS to the LAB.

Hypothesis 1: There is a significant difference in the exit rates of students when assessed with the LAB total scores from when the W-MLS total scores are used.

The results of the two first tests, the LAB and the W-MLS were examined through the first hypothesis. The result shows that there is no significant difference in the outcome of the tests. Although, as the hypothesis states, it was expected that a multiple choice test founded in large part on the evaluation of the student’s knowledge of grammatical structure such as the LAB would yield results that are different from those of a test where the student has to produce responses that qualify in most cases as examples of irregularities of English as for the W-MLS.

From an initial glance at the results displayed in Table 8, one might expect that the W-MLS and the LAB would fail to agree on who should exit or remain in the bilingual program. As can be seen, 15 (n=11+4) students were selected by one test and not the other as competent speakers of the language. While this number does not show that there is a significant difference in the outcome of the two tests, we should note that while not statistically significant, it matters which test is used for 15 students or 12.5% of the student population. The repercussions of this decision will be very real for these 15 students. The results of that decision will be felt during the following year, and possibly in the ones that follow. However, a point that supports the agreement of the tests, is that more students passed or failed both tests than were expected and fewer passed one and failed the other than were expected. Although the tests were different in their conceptual design, academic language does not exist in a vacuum, and basic interpersonal communication skills are necessary for the development of academic language. (Baral,
1987; Halle, G., Stansfield, C., and Rock, D., 1988; Cummins, 1983) We also recognize that correct grammatical usage is essential for academic language development. It is therefore possible to say that not only students who passed the LAB also passed the W-MLS, but they passed the W-MLS because they could also pass the LAB.

Sub-Hypothesis 1A: There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the W-MLS Oral subtests are used.

Sub-Hypothesis 1B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the W-MLS Reading/Writing subtests are used.

The analyses of the Oral and the Reading/Writing subtests of the LAB and W-MLS in hypotheses 1A and 1B are consistent with the result of Hypothesis 1, which analyzed the exit rates of both tests. One fact worth mentioning is the high number of students passing the LAB but failing the W-MLS (n=24) in the analysis of the oral subtests (Hypothesis 1A). This supports the premise that the LAB is effective at assessing BICS if we consider that the Oral sub-test is an interview where points are given for a correct answer even if it is not grammatically accurate. The W-MLS analyses oral language through precise vocabulary identification and analogies. Again, this numerical difference although substantial, is not sufficient to produce statistically significant differences between the oral portions of the two subtests.

Comparing the LAB to portfolio assessment.

Hypothesis 2: There is a significant difference in the exit rates of the students when assessed with the LAB total scores from when the LPA total scores are used.
The next group of hypotheses examines the results of the LAB and those of students' portfolios (LPA). Here again, the two assessment formats identified the same students as exiting or remaining in the bilingual program. It had been expected that because the LAB assesses mainly discrete grammatical points, and that the portfolios look at contextual application of language, the LAB would have been less restrictive than the portfolio assessment as far as exit is concerned. The statistical analyses showed otherwise.

The most striking aspect of the analyses is the high number of students who failed the LAB, but were successful with the portfolio assessment. The same outcome occurred with the analysis of Hypothesis 3, where the results of the W-MLS are compared to the results of portfolio assessment. Although a noticeably high number of students failed the LAB but had successful portfolio assessment, it did not keep the analysis from showing high correlation between the two tests. Even if the tests point to the same students for failure and success, it is worth investigating the teachers' perceptions of the reasons for which lack of agreement existed for some students between the Language Portfolio Assessment results and those of the LAB and the W-MLS. The next section will therefore present the teacher's perceptions of the results. Teachers are in an ideal position to discuss test results since they have daily interaction with the students in the classroom context.

Hypothesis 2A: There is a significant difference in the pass rates of students when assessed with the Oral subtests of the LAB from when the Oral rubric of the LPA is used.
Hypothesis 2B: There is a significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the LAB from when the Reading/Writing rubrics of the LPA are used.

The analyses for the Oral and Reading/Writing subtests supported the results of the comparisons for the total scores of the LAB to the total scores of the LPA.

Teachers' perception of the differences in outcomes.

The bilingual teachers were presented with the results yielded by the LAB and the W-MLS and those of the portfolio assessment. The teachers were of the opinion that the LAB and the W-MLS were overly restrictive and that students who had failed language assessment marginally might do very well in mainstream classes. They felt that in many cases, one more year in the bilingual program would not make much difference and may even slow these students' progress. In addition, they felt that allowing a marginal student to access the mainstream would be likely to challenge and motivate him or her. The teachers' perception of the academic level of the mainstream students in the district helped support this idea. The teachers felt that the average mainstream student would approximate the academic level of many marginal bilingual student. A consensus of the bilingual teachers was that if either the LAB or the W-MLS was given to the students of the district, the vast majority would not pass. They are justified in thinking so if one considers that the state norms were based on state-wide monolingual population and that the district where the study took place is ranked as one of 30 special need districts in the state.

Teachers discussed other reasons given to support the higher exit rate of portfolio assessment. The Kindergarten teachers noted that, as the program was a half-day session
for that grade level. The Reading/Writing sub-test should not bear as heavily as
Listening/Speaking for these students. The teachers felt that the time constraint forces
them to dwell more on matters of social skills, self-control, maturity, therefore
developing verbal aspects over Reading/Writing.

Five of the seven bilingual teachers participating in the study had students whom,
they felt, had problems other than language that prevented them from passing the
language assessment. These ranged from problems with sound perception, academic
difficulties from a perceptual nature, to lack of motivation. At times, these problems are
not severe enough that they should warrant a referral to the Child Study Team for
assessment, but they contribute to the student’s difficulties in his academic life. Students
with these difficulties often remain in bilingual programs longer than the average student
does which for this district is 2 to 3 years (see Table 6). Language assessment tests
cannot identify that failure is due to a reason other than language. Often, a sign that a
problem may exist with misplacement is the presence of a wide gap in scores between
subtests.

In one case, a teacher commented that a student had passed portfolio assessment
and failed the other two tests because she knew this student has difficulty controlling his
level of anxiety during most testing situations. In another individual case, a student who
had been in the bilingual program for a little over a year was able to pass the W-MLS but
was not successful with portfolio assessment. She had not mastered the language at a
level that would justify her exit out of the program. However, she had been able to raise
her W-MLS score to a passing level, mainly through her higher scores on the reading
section of the test, which is strictly a reading exercise centered on decoding a vocabulary
list. Obviously, such cases are rare, and teacher's vigilance should prevent such occurrence.

A comment by the teachers of the two oldest groups of students (grades 5-6 and grades 7-8) was that some students had willfully held back correct responses on the tests in order to remain in the bilingual class. While this may be extreme, teachers commented that it is traumatizing for many students to crossover to the mainstream when they have had the support of small group instruction from the same teacher over several years.

**The W-MLS compared to language portfolios assessment.**

Hypothesis 3: There is no significant difference in the exit rates of students when assessed with the W-MLS total scores from when the LPA total scores are used.

The next group of hypotheses compares the results of the W-MLS to those of the LPA. The third hypothesis is founded on the premise that as portfolio scores are based on the assessment of students' academic language, the scores of the W-MLS, which are also based on knowledge of academic language should yield the same results. The data analyses support the hypothesis: the same students passed or failed for both tests. As in the second hypothesis, a high number of students were successful with the portfolio assessment but failed the W-MLS (n=34). It would then appear that the W-MLS is more restrictive at exiting students than the LAB. However, as we will see further, additional students are likely to exit the bilingual program through the W-MLS. Teacher's opinion is then used to select which of the students that occupy a marginal bandwidth, are likely to be successful in the mainstream class.
Hypothesis 3A: There is no significant difference in the pass rates of students when assessed with the Oral subtests of the W-MLS from when the Oral rubric of the LPA is used.

Hypothesis 3B: There is no significant difference in the pass rates of students when assessed with the Reading/Writing subtests of the W-MLS from when the Reading and Writing rubrics of the LPA are used.

The additional hypotheses 3A and 3B shed more light on the nature of the W-MLS. Where the Oral subtest of the LAB and the Oral portion of the LPA correlated highly (29 cases passed both and 63 cases failed both), there was lack of agreement for 28 cases. The comparison of the Oral subtests for the W-MLS and the LPA, though still statistically significant, points to a lesser level of agreement. Here, there is lack of agreement for 48 cases (see Table 15). A parallel comparison can be drawn for the Reading/Writing portion of the W-MLS and that of the LPA. Table 16 points to a high level of agreement for the Reading/Writing subtest of the W-MLS and the LPA. There was agreement for 93 cases and lack of agreement in test outcome for only 27 cases. A lower level of agreement might be expected for the LAB/LPA Reading/Writing subtests comparison (see Table 13), but it is maintained, with agreement in outcomes for 98 cases and lack of agreement for 22 cases. The LAB, it should be noted, includes a section featuring the Cloze procedure in its Reading/Writing section. The Cloze procedure is a blank filling exercise, which is regarded as indicative of academic language proficiency. (Halle et al., 1988; Cummins, 1983). Though in this case, the exercise is in the multiple-choice format, it helps to raise the correlation of the Reading-Writing subtest of the LAB to the level of the corresponding subtest for the W-MLS and the LPA.
There is agreement on which students should exit or remain in the program. However, there is a noticed sensitivity with the LAB at selecting competent students in the Oral sub-test (see Table 12). The W-MLS, perhaps because of its nature (academic language assessment) and its format (single word answers) does not agree as highly as the LAB and the LPA on student oral competency. Once again, while statistical significance is important in the point of view of the researcher, lack of agreement between two tests will mean that according to which test is selected for the abiding decision, a student’s future will take a very different turn.

Change in CALP level.

The W-MLS determines exit by placing students’ scores on a graph. The placement is then identified as a location on one of five bandwidths representing CALP levels. A location on a bandwidth at the 4th or 5th level is a sign of language proficiency. In the case of marginal placement on a bandwidth of 3.5, the teacher’s professional opinion is required for exit out of the bilingual program. For this purpose, an additional analysis of the data was conducted to examine if the correlations were maintained or improved once the requirements for exit were changed.

Table 25

<table>
<thead>
<tr>
<th>Calp Level</th>
<th>Passed Both</th>
<th>Failed Both</th>
<th>Cases Agreeing</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11</td>
<td>94</td>
<td>105</td>
</tr>
<tr>
<td>3.5</td>
<td>16</td>
<td>76</td>
<td>92</td>
</tr>
</tbody>
</table>

Although it would seem that the most beneficial CALP level would be the 4th as it has the highest level of agreement, at the 3.5 level, Table 25 indicates that 5 more students, for a
total of 16, pass both tests while maintaining a high number of cases agreeing (see Table 25).

Table 26

<table>
<thead>
<tr>
<th>Clap Level</th>
<th>Passed Both</th>
<th>Failed Both</th>
<th>Cases Agreeing</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>12</td>
<td>71</td>
<td>83</td>
</tr>
<tr>
<td>3.5</td>
<td>29</td>
<td>65</td>
<td>94</td>
</tr>
</tbody>
</table>

The analyses of the 3.5 CALP level with the portfolio as indicated in Table 26, is of great significance, as the results of the Language Portfolio Assessment would determine whether a student should remain in or exit the bilingual program (see Table 26). In this case, teachers identified 17 students \((n=29 - 12)\) as, though borderline, sufficiently proficient to participate in mainstream classes. The number of cases in agreement for the 3.5 CALP level of the W-MLS/Language Portfolio Assessment comparison \((n=94)\) is very close to that of the 3.5 CALP level of the W-MLS/LAB \((n=92)\) therefore validating the consistency of the results for the purpose of using teacher’s recommendation with a 3.5 CALP level for placement of bilingual students.

**Grade level and rate of exit.**

Sub-Hypothesis 1: There is a significant difference in the exit rates of the LAB, W-MLS and LPA results of L.M/LEP students when grade level clusters (K - 3 and 4 - 8) are considered.

Another question raised in this research was whether the different tests examined reacted differently for different age groups. The teachers in this district have pointed out that students at the primary level (Grades K to 3) consistently exit at a higher rate than
the students at the intermediate level. This year, according to the LAB which is the instrument used in the district for placement of bilingual students, the rate was 17/67 or a little over 25% for the primary grades students, and 5/43 or nearly 12% for the intermediate grades.

Jim Cummins' (1979) concept of developmental interdependence describes how "the level of L₂ competence which a bilingual child attains is partially a function of the type of competence the child has developed in L₁ at the time when intensive exposure to L₂ begins." Therefore, a student who has a good knowledge of the structure of his native language has a distinct advantage at second language learning. This knowledge acts as an underpinning where transfer occurs and supports or contrasts the new structures. Though previous knowledge acquisition should give the intermediate students an advantage over their younger counterparts, teachers feel that the primary students learn their second language more naturally and more easily just as one would learn his or her mother tongue. They also feel that the younger students do not feel the burden of having to cope with the academic avalanche of the American curriculum that the intermediate grade students have to face. The older students are often deficient in their native language skills and poor in other academic areas.

In this study, we can see from the analyses of subhypothesis 1, that the LAB and the LPA clearly differentiate the two grade level groupings. The W-MLS however, exits and retains the expected number of students. Here again, the design of the tests can help shed some light on the reasons for these differences. The LAB and the LPA require respectively, knowledge and production of language structure. For the reasons quoted earlier by teachers, the primary students would be advantaged as far as exit is concerned.
On the other hand, the W-MLS focuses on vocabulary knowledge, decoding, intellectual manipulations of the language (analogies), nothing related to language structure recognition or production and consequently, nothing that structural knowledge of the native language could help. The W-MLS sees no difference between the two groups. However, without production of language, one can never really know the extent of the student competence in either his second or first language.

**Native language as a determinant of success.**

Sub-Hypothesis 2: There is no significant difference in the exit rates of the LAB, W-MLS and LPA when the native language ($L_1$) of the students is considered.

The analyses conducted for the second sub-hypothesis were meant to examine if there exist differences in the exit rates of the students when they are divided into two native language groups (Spanish and Haitian Creole). This hypothesis served as a means to verify the teachers’ assessment when a variable that should have no influence on the outcome was introduced, given that the two groups are under conditions that are seemingly constant: same curriculum, equivalent facilities, equivalent socio-economic conditions. In other words, all tests should yield the same outcome, regardless of the native language of the student.

The results indicated that both the LAB and the W-MLS showed no difference in outcome, as many Spanish speaking students as Creole students remained or exited the bilingual program as were expected. However, when the two linguistic groups were examined through the focus of their teacher’s assessment of the language portfolio, we found a rate of exit higher for the Spanish speaking students than for the Creole speaking students. What could explain that a linguistic group does better than another when no
differences were expected. Why is one group released into the mainstream at a higher rate than the other one is?

One reason that may come to mind is the suggestive nature of portfolio assessment. Although grading was guided by rubrics, and success was determined by cut-offs on each as well as on combinations of these rubrics, it is without a doubt that portfolio scoring is a suggestive matter. Assessment may vary from one individual to the other, even if coaching occurs and the availability of exemplars is present. For this study, the Interrater Reliability Analyses presented in Chapter 4 help demonstrate that the teachers' assessment is representative of the students' output. However, teachers pointed out to reasons that may be instrumental to the higher retention rate of the Creole students. Differences in staffing were brought out. For logistical reasons, there are multiple grade classrooms of three and four grade levels at the Haitian magnet school. To compensate, teachers have full-time assistants, but it remains that direct interaction with the teacher is reduced. Another difference pointed out by teachers is the initial literacy level of the students from Haiti. While there exist deficiencies for the Spanish speaking students, who come mainly from Central America, these are more severe for the Haitian students. According to UNESCO statistics, the rate of illiteracy in Haiti for 1995 would stand at 55% making it the highest for Latin America and the Caribbean (Education Information Service, 1995). For the same year, the gross enrollment in school for the primary and secondary grades was at the 40% level. Teachers see a reflection of these statistics in the classroom. Haitian newcomers are often not on grade level, or more so than their Spanish speaking counterparts. This situation is worse for the older students. Teachers have to hold back students to offer basics, which for motivated students can occur rapidly.
Students then have to progressively catch up to grade level. It is always hoped that for the older students, discouragement will not set in before the task comes to fruition.

Conclusion

The statistical analyses presented and discussed in this study point to convergence of outcome for the tests examined. It is hoped that the study helps to illustrate how the results of one test help complete or support the results of another.

The LAB and the W-MLS feature different approaches to language assessment. The common outcome of the tests is an indicator of their reliability. However, since placement decisions are made on an individual basis, the information yielded by student portfolios is valuable.

For both the LAB and the W-MLS, portfolio assessment is an ideal complement. For the LAB, portfolios help to demonstrate in a concrete manner, what it can only present in a multiple-choice format. The portfolio supports the outcome of the LAB. Though the LAB is not designed to use teacher's opinion for marginal cases, it may clearly support the teacher's opinion that a student, though minimally deficient, could be considered as an exit candidate on the strength of the work found in the portfolio.

For the W-MLS, the portfolio is a necessary component, not only because it supports the teacher's opinion for the placement of the student at the 3.5 CALP level, but the design of the W-MLS (single word responses) should require that the language portfolio be available to clarify the results.

Recommendations

Classroom time allocated to assessment is clearly taken away from instructional time except in the case of portfolio assessment, where the teacher must demonstrate
organizational skills. It is hoped that the following recommendations for future research and study will guide educators to devote more time and resources in the essential area of language assessment.

1. Continued research with the W-MLS is necessary as teachers and language program administrators need to know more about this new and revolutionary test.

2. Longitudinal research may be devoted to the outcome of placement in the mainstream of marginal students tested with the LAB when it is supported by portfolio assessment.

3. Language portfolio assessment is often qualified as subjective assessment. Research on how varying the content of the Language portfolio affects the results of assessment is valuable.

4. As portfolio assessment is an emerging form of language assessment, research that investigates creative ways of maintaining teachers’ efficacy and accuracy in maintaining and grading the portfolios is always welcomed.

   Instruction and assessment in the mainstream are driven by mastery of curriculum standards and acquisition of higher order thinking skills. These must also be a concern for Second Language instruction. We have reached a point where our assessment of bilingual students must reflect if their mastery of these standards and skills has reached a level that will make them competitive in the mainstream class. Language assessment needs to be thorough, accurate and impartial. Teachers need to be active participants in the decisions that affect the students they help. These decisions cannot be made arbitrarily. School districts need to focus on the professional development of Bilingual teachers. Assistance needs to be provided in developing and scoring language portfolios that will yield decisions that are educationally sound.
References


Camp, R. (1990). Thinking together about portfolios. The Quarterly of the National Writing Project and the Center for the Study of Writing, 12 (2) 8-14, 27.


Mace-Matluck, B. J. (1982). Teaching reading in the bilingual program with emphasis on transferability of Spanish reading skills to English reading. Paper presented at San Diego University BESC Multi-District Preservice Workshop, San Diego, CA.


Appendix A1

Kindergarten Oral Rubric
# Kindergarten Oral Rubric

**Student's Name:**

**Grade:**

**Total Score:**

**Date:**

**Signature of Rater:**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Oral Comprehension</th>
<th>Vocabulary</th>
<th>Pronunciation</th>
<th>Expression</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Consistently understands normal speech</td>
<td>Approximates that of a native Kindergartener</td>
<td>Clear and close to native speech</td>
<td>Adequate for a kindergartener</td>
<td>Follows directions accurately</td>
</tr>
<tr>
<td>4</td>
<td>Habitually understands normal speech</td>
<td>Occasional limits on some vocabulary</td>
<td>A few errors that do not detract from communication</td>
<td>Occasional errors that do not cause difficulty in comprehension</td>
<td>Can usually follow directions</td>
</tr>
<tr>
<td>3</td>
<td>Understands most of what is said when repeated</td>
<td>Vocabulary is simple</td>
<td>Problems in enunciation that interfere with communication</td>
<td>Noticeable errors, but ideas are obvious</td>
<td>Follows directions only partly, due to lack of comprehension</td>
</tr>
<tr>
<td>2</td>
<td>Understanding is limited to simple statements</td>
<td>Frequent errors or limits on vocabulary</td>
<td>Individual words are hard to understand</td>
<td>Makes mistakes that interfere with meaning</td>
<td>Directions are mostly mimicked</td>
</tr>
<tr>
<td>1</td>
<td>Has difficulty understanding even simple conversation</td>
<td>Limitations on vocabulary make communication very difficult</td>
<td>Speech is practically unintelligible</td>
<td>Speech is difficult to understand</td>
<td>Seldom follows directions</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate rating. Then add up for Total Score.

**Comments:**
Appendix A2

Kindergarten Reading Rubric
## Kindergarten Reading Rubric

<table>
<thead>
<tr>
<th>Rating</th>
<th>Name Writing/Recognition</th>
<th>Vocabulary Recognition</th>
<th>Tracking</th>
<th>Story Recall</th>
<th>Application of Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Recognizes and writes his/her name and that of others</td>
<td>Recognizes simple vocabulary once presented</td>
<td>Accurately uses tracking while following a story</td>
<td>Recalls accurately facts and details of a story</td>
<td>Recognizes sequence of events and cause/effect relationship</td>
</tr>
<tr>
<td>4</td>
<td>Recognizes his/her name and that of others. Writes his/her own name</td>
<td>Recognizes most of the vocabulary related to a theme</td>
<td>Tracking is accurate and sustained most of the time</td>
<td>Can recall many parts of a story</td>
<td>Can usually order the sequence of events and give cause/effect</td>
</tr>
<tr>
<td>3</td>
<td>Recognizes and writes his/her own name only</td>
<td>Recognizes some vocabulary</td>
<td>Tracking is fairly accurate</td>
<td>Recalls a few ideas in a story</td>
<td>Sometimes misses elements in sequencing and may give wrong cause/effect</td>
</tr>
<tr>
<td>2</td>
<td>Recognizes own name, but does not write it correctly as of yet</td>
<td>Occasionally recognizes some vocabulary</td>
<td>Use of tracking is starting to be apparent</td>
<td>Recalls only incomplete details</td>
<td>Makes mistakes in sequencing and cause/effect relationship</td>
</tr>
<tr>
<td>1</td>
<td>Does not recognize his/her own name</td>
<td>Rarely recognizes vocabulary</td>
<td>Tracking is hardly apparent</td>
<td>Single words found in the story are recalled</td>
<td>Minimal application of sequencing and cause/effect relationship</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate rating, then add up for Total Score.

**Comments:**
Appendix A3

Kindergarten Writing Rubric
### Kindergarten Writing Rubric

<table>
<thead>
<tr>
<th>Rating</th>
<th>Vocabulary</th>
<th>Story Development</th>
<th>Phonics</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Has strong command of vocabulary</td>
<td>Often uses conventional spelling</td>
<td>Has mastered sound-letter correspondence</td>
<td>Student can develop a story or complete one</td>
</tr>
<tr>
<td>4</td>
<td>Vocabulary is adequate</td>
<td>Uses invented spelling</td>
<td>Sound-letter correspondence is almost achieved</td>
<td>Main elements of story are present</td>
</tr>
<tr>
<td>3</td>
<td>Vocabulary is simple</td>
<td>Can illustrate responses</td>
<td>Significant deficits are noticed in phonics acquisition</td>
<td>Partial elements of story are present</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary is limited</td>
<td>Illustration does not correspond to expectation</td>
<td>Only a few correspondences are acquired</td>
<td>Only key words are present</td>
</tr>
<tr>
<td>1</td>
<td>Vocabulary is minimal</td>
<td>Mostly fails to respond</td>
<td>Minimal knowledge</td>
<td>No appreciable evidence of story</td>
</tr>
</tbody>
</table>

Directions: For each category, please circle appropriate rating, then add up for Total Score.

Comments:
Appendix A4

Oral Proficiency Rubric
### Oral Proficiency Rubric

**Student's Name:**

**Grade:**

**Total Score:**

**Date:**

**Signature of Rater:**

<table>
<thead>
<tr>
<th>Level</th>
<th>Comprehension</th>
<th>Fluency</th>
<th>Vocabulary</th>
<th>Pronunciation</th>
<th>Grammar</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Understands everyday conversation and class discussions without difficulties</td>
<td>Speech is fluent and effortless, approximating that of a native speaker</td>
<td>Use of vocabulary and idioms approximates that of a native speaker</td>
<td>Pronunciation and intonation approximate that of a native speaker</td>
<td>Grammatical usage and word order approximate that of a native speaker</td>
<td>Can convey meaning in complete, well detailed ideas</td>
</tr>
<tr>
<td>4</td>
<td>Understands nearly everything at normal speed, although occasional repetition may be necessary.</td>
<td>Speech is generally fluent with occasional lapses while student searches for correct manner of expression</td>
<td>Occasionally uses inappropriate terms or must rephrase ideas because of inadequate vocabulary</td>
<td>Always intelligible, though one is conscious of a definite accent and occasional inappropriate intonation pattern.</td>
<td>Occasionally makes grammatical or word order errors which do not obscure meaning</td>
<td>Ideas well put, although some details are missing</td>
</tr>
<tr>
<td>3</td>
<td>Understands most of what is said at slower-than-normal speed with repetitions</td>
<td>Speech is frequently disrupted by the student's search for correct manner of expression</td>
<td>Frequently uses the wrong words, conversation limited because of inadequate vocabulary</td>
<td>Pronunciation problems necessitate concentration of the listener and may lead to misunderstandings</td>
<td>Makes frequent errors of grammar and word order which occasionally obscure meaning</td>
<td>Expresses ideas on topic, but some clarity is lost</td>
</tr>
<tr>
<td>2</td>
<td>Has great difficulty following what is said. Can only comprehend social conversation spoken slowly with frequent repetitions</td>
<td>Usually hesitant, often forced into silence by language limitations</td>
<td>Misuse of words and very limited vocabulary make comprehension quite difficult</td>
<td>Very hard to understand because of pronunciation problems. Must frequently repeat to be understood</td>
<td>Grammar and word order errors make comprehension difficult. Must rephrase or restrict what is said to basic patterns</td>
<td>Can only express a few ideas about the context</td>
</tr>
<tr>
<td>1</td>
<td>Cannot understand even simple conversation</td>
<td>Speech is so halting and fragmentary as to make conversations almost impossible</td>
<td>Vocabulary limitations are so severe as to make conversation virtually unintelligible</td>
<td>Pronunciation problems so severe as to make speech virtually unintelligible</td>
<td>Errors in grammar and word order so severe as to make speech virtually unintelligible</td>
<td>Expression of ideas has little to do with the context that is discussed</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate level, then add up for Total Score.

**Comments:**
Appendix A5

First and Second Grade Reading Rubric
### Reading Rubric Grade 1 and 2

**Student's Name:**

**Grade:**

**Total Score:**

**Date:**

**Signature of Rater**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Decoding</th>
<th>Vocabulary</th>
<th>Meaning</th>
<th>Comprehension</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Effortless use of sound-letter relationship</td>
<td>Recognizes key words in context</td>
<td>Uses word parts (inflections, contractions) and contextual clues to clarify meaning</td>
<td>Finds main idea and sequence of events</td>
<td>Makes predictions on elements of text</td>
</tr>
<tr>
<td>4</td>
<td>Some occasional errors in use of sound-letter relationship</td>
<td>Usually recognizes key words</td>
<td>Often uses word parts and contextual clues to assist with meaning</td>
<td>Finds main idea and sequence of events most of the time</td>
<td>With most texts, is able to make predictions</td>
</tr>
<tr>
<td>3</td>
<td>Relies so heavily on decoding as it interferes with comprehension</td>
<td>Some key words are recognized</td>
<td>Attempts to use word parts and contextual clues to clarify meaning</td>
<td>Missing elements in statement of main idea and sequence of events</td>
<td>Occasionally makes predictions</td>
</tr>
<tr>
<td>2</td>
<td>Sound-letter correspondence is often erratic</td>
<td>Key word recognition is so difficult as to interfere with comprehension</td>
<td>Irregular use of word parts and context to clarify meaning</td>
<td>Often makes errors in stating main idea and sequence of events</td>
<td>Many errors are made in attempts at making predictions</td>
</tr>
<tr>
<td>1</td>
<td>Minimal attempts at decoding</td>
<td>Key word recognition is minimal</td>
<td>Minimal attempts at meaning clarification through text elements</td>
<td>Usually cannot state main idea and sequence of events</td>
<td>Usually cannot make predictions</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate rating, then add up for Total Score.

**Comments:**
Appendix A6

First and Second Grade Writing Rubric
# Writing Rubric Grade 1 and 2

Student's Name: ___________________________ Grade: _____________

Total Score: _______ Date: ___________ Signature of Rater: ___________

<table>
<thead>
<tr>
<th>Rating</th>
<th>Vocabulary</th>
<th>Organization</th>
<th>Transitions</th>
<th>Meaning</th>
<th>Mechanics</th>
<th>Story Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Vocabulary is varied and accurate</td>
<td>Stories are organized logically</td>
<td>Transition from one idea to another is apparent</td>
<td>Meaning is well expressed</td>
<td>Only a few mechanical errors are found</td>
<td>Shows a clear understanding of story development</td>
</tr>
<tr>
<td>4</td>
<td>Vocabulary is adequate for grade level</td>
<td>Some evidence of organization is seen in parts of the stories</td>
<td>Some transition of ideas is evident</td>
<td>Meaning is expressed but breaks down at times</td>
<td>Mechanical errors are present but do not disrupt communication</td>
<td>Shows a good understanding of story development</td>
</tr>
<tr>
<td>3</td>
<td>Vocabulary is simple</td>
<td>Organization is very simple</td>
<td>A few transitional markers are used</td>
<td>Meaning is frequently unclear</td>
<td>Mechanical errors affect communication</td>
<td>Shows a some understanding of story development</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary is limited and repetitious</td>
<td>Samples are comprised of only a few disjointed sentences</td>
<td>Sentences do not appear linked</td>
<td>Meaning only in a few sentences</td>
<td>Mechanical errors cause serious disruption in communication</td>
<td>Shows little evidence of story development</td>
</tr>
<tr>
<td>1</td>
<td>Responds with a few words</td>
<td>No complete sentences are written</td>
<td>No evidence of multiple ideas</td>
<td>No meaning is expressed</td>
<td>Errors are so many that they render most of the text unclear</td>
<td>No evidence of concepts of writing</td>
</tr>
</tbody>
</table>

Directions: For each category, please circle appropriate rating, then add up for Total Score.

Comments:
Appendix A7

Intermediate Reading Rubric
## Intermediate Reading Rubric

<table>
<thead>
<tr>
<th>Rating</th>
<th>Vocabulary</th>
<th>Meaning</th>
<th>Comprehension</th>
<th>Application</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Recognizes key words in context</td>
<td>Uses word parts (prefixes, suffixes, roots) and contextual clues to assist with meaning</td>
<td>Finds main idea, sequence of events and recognizes story elements</td>
<td>Can draw conclusion and summarize the text</td>
<td>Uses references to seek information</td>
</tr>
<tr>
<td>4</td>
<td>Usually recognizes key words</td>
<td>Often uses word parts and contextual clues to assist with meaning</td>
<td>Can usually find main idea, sequencing and elements of the story</td>
<td>Can draw conclusion and summarize for most texts</td>
<td>Often uses references</td>
</tr>
<tr>
<td>3</td>
<td>Some key words are recognized</td>
<td>Attempts to use word parts and contextual clues to clarify meaning</td>
<td>Some information missing in statement of main idea, sequencing and story elements</td>
<td>Summaries and conclusions are incomplete</td>
<td>Is beginning to use reference material</td>
</tr>
<tr>
<td>2</td>
<td>Key word recognition is so difficult as to interfere with comprehension</td>
<td>Irregular use of word parts and context to clarify meaning</td>
<td>Errors in statement of main idea, sequencing and story elements</td>
<td>Obvious errors in summaries and conclusions</td>
<td>Minimal use of references</td>
</tr>
<tr>
<td>1</td>
<td>Key word recognition is minimal</td>
<td>Minimal attempts at meaning clarification through text elements</td>
<td>Usually cannot state main idea, sequence of events and story elements</td>
<td>Usually cannot state conclusion nor summarize</td>
<td>Has difficulties using references</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate rating, then add up for Total Score.

**Comments:**
Appendix A8

Intermediate Writing Rubric
## Intermediate Writing Rubric

<table>
<thead>
<tr>
<th>Rating</th>
<th>Vocabulary</th>
<th>Organization</th>
<th>Transitions</th>
<th>Meaning</th>
<th>Mechanics</th>
<th>Topic Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Vocabulary is precise, varied and vivid</td>
<td>Organization is appropriate to writing assignment and contains introduction, development and conclusion</td>
<td>Transition from one idea to another is smooth and provides reader with clear understanding that topic is changing</td>
<td>Meaning is conveyed effectively</td>
<td>A few mechanical errors may be present but not to disrupt communication</td>
<td>Shows a clear understanding of writing and topic development</td>
</tr>
<tr>
<td>4</td>
<td>Vocabulary is adequate for grade level</td>
<td>Events are organized logically, but some part of the sample may not be fully developed</td>
<td>Some transition of ideas is evident</td>
<td>Meaning is conveyed but breaks down at times</td>
<td>Mechanical errors are present but do not disrupt communication</td>
<td>Shows a good understanding of writing and topic development</td>
</tr>
<tr>
<td>3</td>
<td>Vocabulary is simple</td>
<td>Organization is extremely simple or there may be evidence of disorder</td>
<td>There are few transitional markers or repetitive transitional markers</td>
<td>Meaning is frequently unclear</td>
<td>Mechanical errors affect communication</td>
<td>Shows some understanding of writing and topic development</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary is limited and repetitious</td>
<td>Sample is comprised of only a few disjointed sentences</td>
<td>No transitional markers</td>
<td>Meaning in only few sentences</td>
<td>Mechanical errors cause serious disruption in communication</td>
<td>Shows little evidence of discourse understanding</td>
</tr>
<tr>
<td>1</td>
<td>Responds with a few isolated words</td>
<td>No complete sentences are written</td>
<td>No evidence of multiple ideas</td>
<td>No meaning in production</td>
<td>Errors are so many that they render most of the text unclear</td>
<td>No evidence of concepts of writing</td>
</tr>
</tbody>
</table>

**Directions:** For each category, please circle appropriate rating, then add up for Total Score.

**Comments:**

Adapted from S. Moya
Appendix B

Letter of Permission for Instrument Utilization
Kőrmől Lépcső 2
H-1012 Budapest, HUNGARY
June 13, 1996

Louise Langlois Chaker
35 Cobb Road
Mountain Lakes, New Jersey 07046-1105
USA

Dear Ms. Chaker:

Please excuse the delay in responding to your request to use the Sample Holistic Criteria found in the Pierce-Valdez-O'Malley document in NCBE. I hope that the delay has not prevented you from going forward with your research plans.

There is no problem with your using these criteria to do your comparative study as long as referencing indicates their origin.

I would be interested to learn about your results once the study is completed because I continue to explore portfolios as a viable educational assessment alternative. I plan to be in Hungary for at least another year, so I can be contacted at this same address at least until July, 1997.

Good luck with your study.

Sincerely,

[Signature]

Sharon S. Moya, PhD
Appendix C1

School Administrator Consent
Dear Ms (principal),

The assessment of bilingual students is currently conducted in New Jersey through the use of a language assessment test as well as other selected criteria. A review of actual student work through the use of portfolio assessment may offer an alternative tool for the purpose of identifying exit candidates.

Your participation in allowing me to access the scores of your students' Language Assessment Battery, Woodcock-Munoz Language Test, and portfolio grades, will be greatly appreciated. Furthermore, I will require the assistance of your bilingual/ESL teachers in clarifying the results through their participation in an interview of approximately 10 minutes in length.

This study is being conducted as part of a doctoral dissertation at Seton Hall University, South Orange New Jersey, with Dr. Juan Cobarrubias of the College of Education and Human Services.

Your professional cooperation in carrying out this research is appreciated. Additionally, I want to assure you that neither you, the district, nor the students will be identified in the study. An abstract will be available to you upon request.

This project has been reviewed and approved by the Seton Hall University Institutional Review Board for Human Subjects Research. The IRB believes that the research procedures adequately safeguard the subject's privacy, welfare, civil liberties, and rights. The Chairperson of the IRB may be reached through the Office of Grants and Research Services. The telephone number of the Office is (201) 378-9809.

I have read the material above, and any questions I asked have been answered to my satisfaction. I agree to participate in this activity, realizing that I may withdraw without prejudice at anytime.

Subject or Authorized Representative                        Date
Appendix C2

Consent Granted
Dr. Superintendent
Board of Education
, NJ

Dear Dr.

Please find enclosed the information you requested. I hope you find the following to your satisfaction. Again, let me express to you my appreciation for the consideration you give my project.

Sincerely,

Louise L. Chaker

Approved
111196

cc.: Mr.
Appendix C3

Teachers Consent
Ms. (teacher)
(school)
(city), New Jersey

Dear Ms (teacher).

The assessment of bilingual students is currently conducted in New Jersey through the use of a language assessment test as well as other selected criteria. A review of actual student work through the use of portfolio assessment may offer an alternative tool for the purpose of identifying exit candidates.

Your participation in producing the various scores for your students’ Language Assessment Battery, Woodcock-Munoz Language Test, and portfolio grades, will be greatly appreciated. Furthermore, your assistance in clarifying the results through your participation in an interview of approximately 10 minutes in length will be most helpful.

This study is being conducted as part of a doctoral dissertation at Seton Hall University, South Orange New Jersey, with Dr. Juan Cobarrubias of the College of Education and Human Services.

Your professional cooperation in carrying out this research is appreciated. Additionally, I want to assure you that neither you, the district, nor the students will be identified in the study. An abstract will be available to you upon request.

This project has been reviewed and approved by the Seton Hall University Institutional Review Board for Human Subjects Research. The IRB believes that the research procedures adequately safeguard the subject’s privacy, welfare, civil liberties, and rights. The Chairperson of the IRB may be reached through the Office of Grants and Research Services. The telephone number of the Office is (201) 378-9809.

I have read the material above, and any questions I asked have been answered to my satisfaction. I agree to participate in this activity, realizing that I may withdraw without prejudice at anytime.

Subject or Authorized Representative

Date
Appendix D1

Parental Consent
Dear Parent:

I ask your permission to use your child’s scores for a language test, the:
- LAB - Language Assessment Battery (1/2 hour long) and
- his or her grades for Reading, Writing, and oral English.
Also, I ask your permission to give your child the
- W-MLS - Woodcock-Munoz Language Survey (1/2 hour long),
a language test that is not part of the District’s tests, and which will not count for a grade.

This study is part of my course at Seton Hall University. I want to look at how the language tests and the students’ grades match up. Your child’s name will not appear in the study, nor will his/her participation influence his/her grades.

A Seton Hall University committee that approves research involving children believes your child’s rights are safeguarded. You can telephone a representative of the University at (201) 378-9809.

I have read this letter. Questions I had about the study have been answered. I permit my child to participate. My child’s participation is voluntary but he/she has the right to withdraw his/her participation without any consequence to him/her.

Parent of Subject ___________ Date ___________
Appendix D2

Parental Consent, French Version
Cher parent:

Je vous demande la permission d'utiliser les résultats d'un test de langue de votre enfant:
- le LAB - Language Assessment Battery (durée de 1/2 heure) et
- ses notes de lecture, d'anglais écrit et oral.

Je vous demande aussi la permission de donner à votre enfant,
- le W-MLS - Woodcock-Munoz Language Survey (durée de 1/2 heure)

celui-ci est un test de langue qui ne fait pas partie des tests du District, et il ne comptera pas pour une note.

Cette étude fait partie de mes cours à l'Université de Seton Hall. Je veux voir si les tests de langue et les notes des étudiants sont semblables. Le nom de votre enfant ne sera pas dans cette étude, et sa participation n'influencera pas ses notes.

Un comité de l'Université de Seton Hall, qui approuve les recherches auxquelles les enfants participent, croit que les droits de votre enfant sont maintenus. Vous pouvez téléphoner à un représentant de l'université au (201) 378-9809.

J'ai lu cette lettre. On a répondu à mes questions sur cette étude. Je donne ma permission à ce que mon enfant participe. Sa participation est volontaire et il/elle a le droit de la retirer sans qu'il n'y ait de conséquence.

__________________________  ________________________
Parent du sujet                     Date
Appendix D3

Parental Consent, Spanish Version
Estimado Padre:

Estoy pidiendo su permiso para usar los resultados de un examen de lenguaje, el:

- LAB - Language Assessment Battery (1/2 hora de duración) y
- Las notas de lectura, escritura e inglés oral.

También, solicito su permiso para darle a su hijo el

- W-MLS - Woodcock-Muñoz Language Survey (1/2 hora de duración), un examen de lenguaje que no es parte de los exámenes del Distrito, y el cual no contara como una nota.

Este estudio es parte de mi curso en la Universidad de Seton Hall. Estoy interesada en comparar como los exámenes de lenguaje y las notas son similares. El nombre de su hijo no aparecera en el estudio, ni tampoco su participación influirá sus notas.

Un comité de la Universidad de Seton Hall que aprueba estudios en los cuales los niños participan, protege los derechos de su hijo. Ud. Puede comunicarse por teléfono con un representante de la universidad al (201) 378-9809.

Yo he leído esta carta. Preguntas que yo tenía acerca de este estudio han sido aclaradas. Yo doy permiso a que me hijo participe. La participación de mi hijo es voluntaria, pero el o ella tiene el derecho de retirar su participación sin ninguna consecuencia a el o ella.

Padre del Sujeto

Fecha
Appendix E1

Student Assent
Louise Langlois Chaker
Graduate Student
Seton Hall University
South Orange, NJ

Dear Student:

I study at Seton Hall University and I am looking at the scores of ESL (English as a Second Language) students like you, to see if language tests and school grades match up. I will need to use your scores for the

- LAB - Language Assessment Battery (1/2 hour long), and
- your grades for Reading, Writing and oral English.

Also I ask you to take the

- W-MLS - Woodcock-Munoz Language Survey (1/2 hour long), a language test that is not part of the District’s tests. The score for this test will not count for a grade.

Your name will not be in the study. The study will not change your grades in the report card. You are free to participate.

A Seton Hall University committee that approves studies that involve children believes that your rights are safe. Your parents know about this study, and you or they can telephone the University to ask about it. The number is (201) 378-9809.

I have read this letter. My questions have been answered. I agree to participate, but I may change my mind and this will not cause me any problems.

Subject (student) ___________________________ Date ___________________________
Appendix E2

Student Assent, French Version
Cher étudiant:

J'étudie à l'Université de Seton Hall et je veux voir si les résultats des tests de langue et les notes de classe des élèves ESL (anglais comme langue seconde) comme vous, sont semblables. J'aurai besoin de vos notes pour
• le LAB - Language Assessment Battery (durée de 1/2 heure), et
• vos notes de lecture, d'anglais écrit et oral.
Je vous demande aussi de prendre le test
• W-MLS - Woodcock-Munoz Language Survey (durée de 1/2 heure). celui-ci est un test de langue qui ne fait pas partie des tests du District, et il ne comptera pas pour une note.

Votre nom ne sera pas écrit dans cette étude. L'étude ne changera pas la note du carnet. Vous êtes libre de participer.

Un comité de l'Université de Seton Hall qui approuve les recherches auxquelles les enfants participent, croie que vos droits sont maintenus. Vos parents savent que cette étude est en progrès, et vous ou eux peuvent téléphoner à l'université au (201) 378-9809, pour vous informer.

J'ai lu cette lettre. On a répondu à mes questions. Je consent a participer, mais je peux changer d'idée, et cela ne me causera aucun problème.

Sujet (étudiant(e))

Date
Appendix E3

Student Assent, Spanish Version
Estimado Estudiante:

Yo estudio en la Universidad de Seton Hall y estoy comparando los resultados de ESL (Ingles como Segundo Lenguaje) de estudiantes como tu, para ver si los exámenes de lenguaje y las notas de la escuela son similares. Yo necesitaré usar tus resultados de el:

- LAB - Language Assessment Battery (1/2 hora de duración), y
- tus notas de lectura, escritura e inglés oral. También, solicito que tomes el
- W-MLS - Woodcock-Muñoz Language Survey (1/2 hora de duración), un exámen de lenguaje que no es parte de los exámenes del Distrito. El resultado de este examen no contará como una nota.

Tu nombre no será mencionado en este estudio. Este estudio no cambiará tus notas en las tarjetas de calificaciones. No estas obligado a participar.

Un comité de la Universidad de Seton Hall que aprueba estudios en los cuales niños participan protege sus derechos. Tus padres saben de este estudio, y tu o ellos pueden telefonear a la universidad para hacer preguntas. El numero es (201) 378-9809.

Yo he leído esta carta. Mis preguntas han sido aclaradas. Yo estoy de acuerdo en participar, pero puedo cambiar de idea y esto no me causará ningún problema.

Sujeto (estudiante)  Fecha
Appendix F

Assessment Project Worksheet
<table>
<thead>
<tr>
<th>Student Name</th>
<th>Grade</th>
<th>Gender</th>
<th># Years In Bilingual</th>
<th>LAB</th>
<th>WAMS</th>
<th>Teacher's Assessment</th>
<th>Combined</th>
<th>Total Combined</th>
<th>Reading</th>
<th>Writing</th>
<th>Oral</th>
<th>Oral (2)</th>
<th>Oral (3)</th>
<th>Oral (4)</th>
</tr>
</thead>
</table>
Appendix G

Teachers Interview Questions
TEACHER’S INTERVIEW QUESTIONS

After having reviewed the data for your class,

1. Can you determine individually the reasons for the presence or absence of correlations between the results of the LAB and those of the W-MLS?

2. Can you determine individually the reasons for the presence or absence of correlations between the LAB and the Language portfolio results?

3. Can you determine individually the reasons for the presence or absence of correlation between the W-MLS and the Language portfolio results?