

Notion Into Practice:  
A Systematic Testing Program for All LEP  
Students in a Large Urban District

Cecilia Oakeley and Hua Yang  
Dallas Public Schools

Introduction

With an unprecedented wave of immigrants sweeping the country, the number of students with limited English proficiency (LEP) in the public schools has been steadily increasing over the past two decades. Nationally, according to the data submitted by state education agencies to the U.S. Department of Education, from 1990-91 to 1994-95, the number of LEP students increased 44.8%, from 2,198,778 to 3,184,696 (Silcox, 1997). In the Dallas Public Schools, one of the largest school districts in the country, the number of eligible LEP students has almost tripled over the last 14 years. In 1995-96, there were 41,807 eligible LEP students in the district, accounting for nearly 30% of the total district population and representing a 15% increase over 1994-95 (Oakeley, July 1996).

The rapid increase in LEP students is a major cause in the soaring district budget for bilingual education (BE) and English as the Second Language (ESL) programs. With greater financial obligations, the demand for accountability from various education agencies at local, state, and national levels is increasing. Such demand was intensified with the current nationwide effort for setting high academic standards for all children, including LEP students. Under those conditions, the needs to assess the language development and academic progress of LEP students and to evaluate the effects of BE/ESL programs are assuming a top priority among local school administrators. Inclusion of LEP students in the large-scale testing programs to maximize the representativeness of students for accountability becomes an urgent task.

However, assessment of LEP students has always been a difficult and controversial issue. Traditionally, LEP students were often excluded from large-scale testing programs for various reasons (McDonnell, 1993). One concern originated from the historical implication that standardized tests may inadequately measure language minority students' cognitive ability (due to their low level of English proficiency) and, thus, categorized them into inappropriate classroom situations (Mclean, 1995). Other related concerns include the lack of test accommodations or adaptations for these students, which prompted school staff to believe these students could not participate meaningfully in the standardized tests. Nationally, prior to 1995, about half or more of

identified LEP students were excluded from the NAEP (National Assessment of Educational Progress) assessment based on specific criteria used with NAEP (NCES, July 1996).

Only recently, a growing consensus is emerging among educators and educational researchers of BE/ESL programs that exclusion of LEP students from large-scale testing programs due to their lack of English proficiency is not appropriate if schools and districts are held accountable for the performance of all of their students (Special Issues Analysis Center, 1994). As researchers have realized that:

“Evaluating these students without consideration for their special language needs is not an option; neither is removing them from all testing situation until they have learned enough English. Both scenarios neglect the need for information on the educational progress of the linguistically and culturally diverse segment of our school population and the programs that serve them.” (Anstrom, 1997)

### Objective of the Study

As the notion of including LEP students in large-scale assessment is becoming popular among researchers, the practical question of how to meaningfully include LEP students so that their needs and performances are recognized while ensuring that there is no misuse or misinterpretation of their scores still remains unanswered. More specific questions include the following:

1. What is the meaning of inclusion? Include LEP students into what kinds of testing programs? What are the purposes of inclusion?
2. What is the cut-off score that determines a student's exemption from normal testing in English? At what point are LEP students ready to be tested in English using the standardized test normally given to non-LEP students?
3. How can LEP student test scores be used to evaluate the effectiveness of BE/ESL programs?
4. Can Districts comprehensively include LEP student data in accountability systems?

In this paper, we present the experience in the Dallas Public Schools of finding a solution that provided reasonable answers to these questions during the last two years. A BE/ESL program evaluation was designed in 1994-95 with the help of a commercially developed test, the *Woodcock-Muñoz Language Survey (WMLS)*. In our practice, all enrolled district LEP students, with a language other than English, are first required to participate in the District BE/ESL test program administration of the *WMLS*. Then, students depending on the results, are assigned accordingly to the state-required criterion test, the *Texas Assessment of Academic Skills (TAAS)*, and/or the District testing program

(*ITBS/TAP*). As a result, in the school year of 1995-96, 96% of the LEP students participated in at least one of academic testing programs. In the following section, we first describe the process of setting up the guidelines and the initial implementation in the district, second, we outline the district's current policies and practices in assessing LEP students and their results, and then discuss why and how we think our practice reasonably answers the questions listed above.

### The Development of DPS's LEP Testing Programs

Prior to the 1994-95 school year, the District's and the State's testing programs for LEP students were using a very subjective process of classifying students for exclusion or inclusion in testing. Consequently a majority of the LEP students were left out of testing. This led to a weak BE/ESL program evaluation and the exclusion of one-third of the District's population from the accountability system.

During the fall of 1994, the District questioned whether a reliable method of documenting the progress of LEP students could be developed. The District was faced with several requirements to assess LEP students from different entities. First, a new Texas Administrative Code required that LEP students who did not take the *TAAS* be tested with an appropriate alternative assessment measure. Second, the District Accountability Task Force, concerned with the lack of information on LEP students' performance and the apparent lack of consistency in testing across schools, had encouraged the District to improve LEP testing procedures. Third, the director of the district's BE/ESL program, interested in baseline data on which to restructure the program, had requested comprehensive testing of all LEP students. And last, the Federal Court had also expressed an interest in a comprehensive evaluation of the BE/ESL program. All of these requirements could not be accomplished without some measure of progress of all students and a consistent method across schools to classify LEP students according to language proficiency. Therefore, an urgent need was developed to identify an assessment instrument for LEP students.

Identification of assessment instrument. The first step in identifying an assessment instrument was to review the literature. The review by Zuhler and et.al. (1994), provided information of the most frequently used tests by 83% of the districts. The review indicates that most tests elicit oral production for part or all of the responses. In addition, of the six most frequent used English language proficiency tests (not including the *WMLS* that came out in 1993), only two included items categorized as reading skill items. Saville-Troike (1991) indicated that older tests tended to focus on language and not language in relation to academic proficiency. Hence, tests vary in their

approach to assess language proficiency representing distinct approaches to the definition of language proficiency.

The selection of the test instrument in our District began by determining how language proficiency was defined by our District. Language proficiency was defined as the ability to be academically and cognitively proficient in English. With this in mind, the next step was to review the list of instruments approved by the Texas Educational Agency (TEA). Technical manuals as well as the actual test of each of the instruments were then reviewed for evidence of validity, reliability and errors of measurement, appropriate test development and revision procedures, appropriate scaling, norming, score comparability and equating procedures, and quality of the technical manual and user's guide. After this initial review, three tests were selected for further consideration. A comparative analysis of the three tests including the costs and testing times was completed. The District has been very conscious of the need for fiscal responsibility and for maximizing information while minimizing student testing time. A final selection was made after a thorough review, and the *WMLS* was recommend as the appropriate test.

Characteristics of *WMLS*. The *WMLS* was chosen for the following reasons: ease of administration, minimal instructional time for testing, ease of interpretation of scores, lower cost, and superior technical characteristics. The *WMLS* is primarily a measure of cognitive-academic language proficiency (CALP). CALP is defined as language proficiency in academic situations, or those aspects of language proficiency that emerge and become distinctive with formal schooling. CALP differs from basic interpersonal communication skill (BICS) which is acquired naturally and without formal schooling (Cummins, 1984). According to Cummins, measures of CALP are more relevant for an assessment of language proficiency in an academic setting than are measures of surface fluency, or BICS. The intent of the *WMLS* is to assess the student's proficiency with cognitively demanding and context-reduced language. Thus, the *WMLS* could be used for following purposes:

- To classify English or Spanish language proficiency from CALP level 1 (negligible English) to CALP level 5 (advanced English). And thus to help teachers understand a student's language abilities (the relative proficiency in each language and information concerning the age- or grade-levels at which a given individual will find certain oral language, reading, and writing tasks easy and difficult).
- To determine eligibility for bilingual services (doesn't use BICS or classification schemes based on a comparison to others).
- To assess a student's progress or readiness for English-only instruction.

- To provide information about program effectiveness.
- To describe the language characteristics of students in research studies.

The *WMLS* contains 4 tests that measure different aspects of language proficiency. The four tests are Picture Vocabulary, Verbal Analogies, Letter-Word Identification, and Dictation. Total testing lasts for 20 minutes and several types of scores are available including CALP levels (1-5), interpretive scores (grade equivalents, percentile ranks, and NCEs) and W-scores<sup>1</sup>.

Initial systemwide implementation in 1995. After *WMLS* was identified, the testing guidelines and procedures were drafted in spring 1995 for initial implementation. The basic concept was that each LEP student shall participate in the administration of the *WMLS*. Students should test with the *ITBS/TAP* (regular testing program for District accountability) and *TAAS* (State required testing program) as soon as it is appropriate. The initial plan was that the oral language proficiency measure of *WMLS* shall be administered prior to the administration of the *TAAS* and *ITBS/TAP* to determine which test is appropriate for a student.

Accordingly, the testing procedure required that each spring all LEP students would be tested with the oral test of the *WMLS*. For Kindergarten students there was no added testing required. First-grade students who participate in a BE program would also be administered the *SABE* (a Spanish norm-referenced test). Students in second-grade and above would be administered the 70 minute *ITBS/TAP* or the *SABE* according to the oral language proficiency score. Students would participate in the two-hour *TAAS* (Spanish for grades 3 and 4) at grades 3 and above as determined by the oral language proficiency score. Students who did not participate in the *ITBS/TAP* and *TAAS* would participate in the *WMLS* literacy examinations which include a 10-minute test of reading and writing and a 10-minute test of math computations. The final decisions regarding the testing taking of each LEP student were made by the Language Proficiency Assessment Committee (LPAC).

However, the initial plan was relaxed later in the year because some testing materials were delivered late and of some schools with large LEP populations were concerned about the logistics of full implementation of the proposed LEP testing program. As a result, instead of first testing with the oral *WMLS*, the LPAC would make a determination of the subsequent test (*ITBS/TAP*, *TAAS*, or *WMLS*-literacy part) without

---

<sup>1</sup> A *WMLS* W-score can be described as a rescaling of the original raw data. A W-score of 500 is equivalent to a 5th grade students achieving at the national norm. Intervals of W-scores make up the 5 different levels.

the oral portion of the *WMLS* being administered. In effect, the LPAC, ended up again determining the type of test to be administered based on subjective criteria.

Systemwide implementation in 1995-96. During the school year, 1995-96 it was determined that all LEP students would be required to take all four parts of the *WMLS* before any additional testing was to take place. The concept was to first determine the student's cognitive level in English to ascertain the readiness for the District's testing program. LEP students who scored *WMLS* broad ability levels 4 and 5 were to participate fully in the regular District or State testing programs. The remaining students had other options depending on their program or grade level. Elementary students in BE were to take the *SABE* (grades 1-6) and the Spanish version of the *TAAS* (grades 3 & 4) while other students were just to take the mathematics computation subtest of the *ITBS*. Secondary students with *WMLS* levels 1-3 were to take the course *Assessment of Course Performance (ACP)*-a District required secondary criterion test) and the *ITBS* mathematics computation subtest.

Systemwide implementation in 1996-97. The same testing procedures was applied in spring of 1997 with minor modifications. It was decided that 1) LEP students in prekindergarden and kindergarten would not have to test until the end of their first grade; 2) LEP students who had already attained a *WMLS* levels 4 or 5 would not have to retest with the *WMLS* since they were already at fluency levels (these students have not yet met State exit criteria); 3) more flexibility in policies was granted to schools to encourage students with *WMLS* level 3 to join the regular testing programs if schools consider it appropriate for the students.

Results of the last 2 years implementation. Table 1 summarizes the results of the last two-year implementation by showing the changes of the testing rates from 1993-94 through 1995-96. The 1993-94 school year, the year prior to the administration of the *WMLS*, was included for comparison purposes. The reader can note that the total number of students tested by at least one measurement has increased from 65% to 96% over the past three years. The percentage of students in the District and State testing plans decreased. These data suggest that students were not being appropriately assigned in the previous years as a result of the subjective classification process that existed. These results reflect our concept of inclusion. Inclusion does not necessarily mean that more students be tested, but instead, tested with an appropriate test to have meaningful results.

Table 1

Testing Rates of LEP Students on  
Different Testing Programs  
1993-94 to 1995-96

School year	Total LEPs served	BE/ESL Testing Program Percent	District Testing Program			State Testing Program		
		<i>WMLS</i> gr.Kn-12	<i>ITBS</i> gr.1-10	<i>SABE</i> gr.1-6	<i>ACP</i> gr.7-11	<i>TAAS-</i> Engl. gr.3-10	<i>TAAS-</i> Span gr.3-6	with at least one test
93-94	30,637	0%	37%	11%	85%	53%	0%	64%
94-95	34,467	35%	41%	22%	83%	56%	17%	84%
95-96	39,642	87%	26%	27%	65%	45%	12%	96%

Note. *ACP* was not given to grades 7-8 in the school years 93-93 or 94-95. The Spanish *TAAS* was not available for the school year 1993-94, only available for grades 3-4 in 1994-95, and for grades 3-6 in 1995-96.

The DPS model of including LEP students in the districtwide testing programs

The model of testing LEP students and including them in the districtwide testing programs that emerged after two years of piloting and modifying, is presented in the flow chart in Figure 1. This is the model currently used in the District. Features of the DPS model includes the following:

1. This systematically designed testing program serves multiple assessment purposes. The test results are used to measure not only individual LEP student's progress in English and academic development, but also the effects of the BE/ESL program, and accountability of schools and the district. The assessment purposes, corresponding to different testing programs, are summarized below in Table 2.
2. The model reflects the principle of fairness in assessment practice by providing equal and continuing opportunities to LEP students. Every student who entered the district with a language other than English was assessed by *WMLS* first and was then assigned to different learning and testing programs based on their *WMLS* scores. With the continuation of the testing program to take place annually, progress of LEP students is carefully monitored and they are promoted to regular testing program as soon as they are ready.

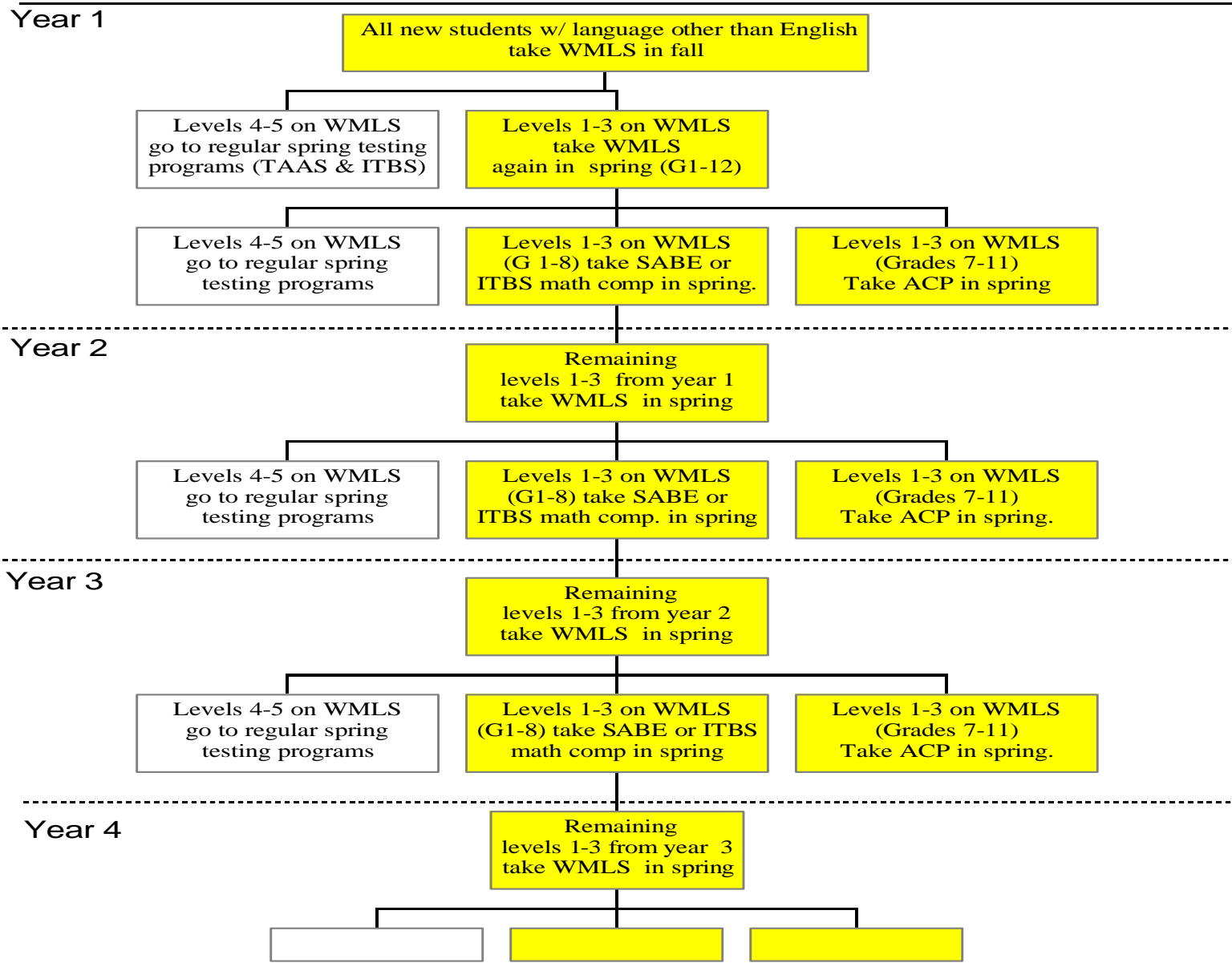
3. The model makes the District's accountability system more representative with the inclusion of LEP students' test data. The testing program for LEP students is an integrated part of the overall district assessment system. With test data now available on LEP students from at least one of the districtwide testing programs, one-third of the District's population was now added in the accountability system.

Table 2

Assessment Purposes of Different Testing Programs

Types of Testing Programs	Purposes of Testing
Initial <i>WMLS</i> in fall at the entry level for all students w/language other than Eng.	* Identification and placement
<i>WMLS</i> in spring for levels 1-3 LEP students on the initial <i>WMLS</i> .	* Screening for additional tests * Assessing annual growth in English * Evaluation of BE/ESL program
Regular spring testing programs ( <i>TAAS/ITBS</i> ) for Levels 4-5 LEP students on initial <i>WMLS</i> or on the spring <i>WMLS</i> .	* Assessing annual academic growth * State accountability ( <i>TAAS</i> ) * District accountability ( <i>ITBS</i> ) * Evaluating BE/ESL programs
<i>SABE</i> , or <i>ITBS</i> math comp, or <i>ACP</i> for levels 1-3 LEP on the spring <i>WMLS</i> .	* Assessing annual academic growth * Evaluating BE/ESL programs

# Flow Chart of Testing Programs for LEP Students in the DPS



## Validation

The LEP student testing program in the District provides our answers to the questions posed at the beginning of the paper. The following part of the paper discusses these questions by resending test results that have validated our testing program.

Meaning of inclusion. In our opinion, the meaning of inclusion is three-fold: 1) measuring academic growth of LEP students; 2) utilizing test data for BE/ESL program evaluation, and 3) including LEP students in accountability system. These three areas can be implemented only under the condition that LEP students are measured with a test appropriate to assess their language capabilities. And as stated before, inclusion does not necessarily mean that more students be tested, but instead, tested with an appropriate test to have meaningful results.

Table 3 shows the results of DPS's model for Spring 1996. The table details the different types of tests utilized for different purposes and the total percent tested by grade. From this table we conclude that 98.6% of LEP students were administered with at least one test.

Table 3

Percent of LEP Students Tested on Different  
Testing Programs by Selected Grades, 1995-96

Grade	Total LEP served	BE/ESL Testing Prog.	District Testing Program			State Testing Program		Total Tested	Percent with at least one test
		WMLS gr. Kn-12	ITBS/ TAP gr.1-10	SABE gr.1-6	ACP gr.7-12	TAAS- Engl. gr.3-10	TAAS- Span gr.3-6		
2nd	4,468	92%	22%	32%	n/a	n/a	n/a	4,342	97.2%
4th	3,773	90%	27%	13%	n/a	85%	11%	3,760	99.7%
6th	2721	87%	30%	7%	n/a	74%	8%	2,702	99.3%
8th	2046	83%	47%	n/a	39%	96%	n/a	2030	99.2%
10th	1098	79%	15%	n/a	88%	89%	n/a	1073	97.7%
Total	14,106	88%	28%	15%	13%	56%	5%	13,907	98.6%

Cut-off point that determines inclusion of LEP students in normal English testing.

Data in Tables 4 and 5 provide evidence which determines when LEP students are ready to participate in the standardized tests normally given to non-LEP students. Data indicate that students with *WMLS* levels of 4 or 5 are ready for the District and State testing programs.

Table 4 compares LEP students' *WMLS* levels with their *ITBS/TAP* reading percentiles. The data indicate 1) a pattern of increasing medians on the *ITBS* as the *WMLS* level increases and 2) a *WMLS* level of 4 translated to a median of equal or greater than the 40th percentile on the *ITBS*, which is the cutoff score for exiting the BE/ESL program in 3 out of the 5 grade comparisons (grades 2, 4, & 6). Table 5 compares LEP students' *WMLS* levels with their *TAAS* reading passing rates. Similar patterns were observed: 1) a pattern of increasing passing rates on the *TAAS* as the *WMLS* level increased and 2) more than 60% of the LEP students with levels 4 and 5 of the *WMLS* passed *TAAS* reading. In both tables, strong evidence exists to suggest that LEP students with *WMLS* levels 4 and 5 are ready to participate in the District and State testing programs.

Table 4

*ITBS/TAP* Reading Median Percentile Scores of LEP Students  
by *WMLS* Level and Grade - Spring 1996

<i>WMLS</i> Level	<u>2nd</u>		<u>4th</u>		<u>6th</u>		<u>8th</u>		<u>10th</u>	
	N	<i>Md</i>	N	<i>Md</i>	N	<i>Md</i>	N	<i>Md</i>	N	<i>Md</i>
1	6	8	6	12	2	7	16	8	0	n/a
2	18	15	35	18	44	20	197	12	38	11
3	249	28	288	28	314	30	465	21	77	17
4	575	45	548	47	267	40	97	27	12	33
5	30	67	21	58	6	56	4	44	2	16

Note. Testing policy indicated that only *WMLS* levels 1-3 were not required to take the *ITBS/TAP*.  
n/a=not applicable

Table 5

Percent of LEP Students Passing *TAAS*  
by *WMLS* Level and Grade - Spring 1996

<i>WMLS</i> Level	4th		6th		8th		10th	
	Total	%	Total	%	Total	%	Total	%
1	26	3.8	40	12.5	25	.0	134	7.5
2	168	8.9	157	12.1	226	11.9	259	19.7
3	725	30.3	581	39.9	464	39.7	284	48.2
4	491	67.2	245	73.1	83	62.7	42	66.7
5	17	88.2	5	100	2	100	4	25.0

Note. Testing policy indicated that only *WMLS* levels 1-3 were not required to take the *TAAS*.

If data have provided sufficient evidence that *WMLS* levels 4 and 5 can serve as the cut-off point for inclusion of LEP students to the District and State testing programs, what is the position of level 3, the midpoint between being included and being excluded. To answer this question, further analysis of the relationship between *WMLS* level 3 and the two tests was conducted. Data indicated a positive correlation of *WMLS* level 3 with *ITBS* reading percentiles for grades 1 through 9 but not for grades 10-12. Results of grades 1-9 were statistically significant at a p-Value of .001. The next question was whether a higher W-score within level 3 was predictive of a student's ability to meaningfully address the *ITBS* reading test. Results show that at each grade, the mean W-score generally coincided with an expected score at the 30th percentile on the *ITBS*. The mean was generally the midpoint of the 20 point W-score span that makes up level 3. These results indicate that students scoring at or above the midpoint at level 3 can be expected to score at or above the 30th percentile on the *ITBS* reading. The analysis of the relationship between *WMLS* level 3 students and their performance on the *TAAS* indicated that about one-third of those students will pass the test. The percentage increases for secondary students.

As a result of these analyses, LPACs were encouraged to consider testing high level 3 LEP students with the District's and State's testing programs.

Using LEP student test data for program evaluation. With the aforementioned testing plan in place, all LEP students now have one common measure, the *WMLS*. The District is now able to capture actual patterns of LEP students' academic status. As shown in Table 6, *WMLS* data reveal, that at the District level, the largest percentage (43.2%) of LEP students have a *WMLS* level of 3. The next largest group (28.6%) was level 2. A combined total of 13.4% of the LEP students had levels of 4 or 5. These districtwide trends repeated themselves when data were examined by grade with few exceptions. Among the students of 9th grade, 33% were in level 1 and 32% were in level 2. These high percentages can be explained by the fact that the majority of the incoming high school LEP are placed in grade 9 regardless of age due to the lack of information of academic credits from their homeland.

Table 6  
Distribution of *WMLS* Levels of LEP Students by Grade  
Spring 1996

Grade	Total Tested	Percent of <i>WMLS</i> Levels				
		1	2	3	4	5
PK	1627	18.2	49.6	24.6	6.9	0.7
Kn	4311	12.6	32.2	39.3	15.0	0.9
1st	4504	15.9	27.8	41.4	14.0	0.8
2nd	4105	10.4	25.4	47.9	15.6	0.8
3rd	3951	9.8	22.5	50.5	16.2	1.0
4th	3351	10.1	21.8	49.8	17.7	0.6
5th	2782	12.5	23.7	49.8	13.7	0.3
6th	2326	14.0	23.9	49.4	12.4	0.3
7th	2070	18.5	32.0	42.8	6.3	0.4
8th	1697	20.6	32.0	41.0	6.2	0.2
9th	1693	33.4	32.0	28.4	5.7	0.4
10th	858	21.4	36.7	35.8	5.5	0.6
11th	621	18.8	41.9	32.9	6.0	0.5
12th	<u>437</u>	<u>19.2</u>	<u>42.1</u>	<u>33.0</u>	<u>5.3</u>	<u>0.5</u>
Total	34333	14.7	28.6	43.2	12.7	0.7

To visualize the pattern, a graph was created using the *W*-scores of the *WMLS*. To get an average score for each grade, the mean *W*-score was calculated and studied. Figure 2 shows the mean *W*-score by grade. *WMLS* level 4 norm was placed in the graph to depict the growth curve and how the mean *W*-score increases with grade. Attaining at or above level 4 of the *WMLS* *W*-score interval indicates fluency with cognitive-academic language. The data in Figure 2 indicate that the mean *W*-scores of lower grades are lower than the level 4 interval but increase with the same slope or at the same rate as the growth curve. However, by grade 4, the mean *W*-scores of the District LEP students start to flatten out, falling further below the level 4 norm growth curve. One noticeable glitch in the line graph is the obvious dip at 9th grade. Again as explained previously, this may be the effect of the placement of newly arrived LEP students. These data suggest that the District should be concerned with the flattening out of the curve after grade 4.

A frequently asked question related to the effectiveness of the BE/ESL program evaluation is how long does it take for LEP students to become English proficient. Collier (1995), among others, concludes that it takes 4-7 years to become English proficient if the students are on grade level in their native language. In our BE/ESL program evaluation, we tried to answer this question by conducting two preliminary studies using *WMLS* data: 1) examining *WMLS* *W*-score data by the number of years students stay in the BE/ESL program; and 2) examining data on advancement of *WMLS* levels from one year to the next. The results are shown in Figures 3 and 4.

Figure 3 provides *WMLS* *W*-score results by years in the program of BE students in grades Kn - 6, while Figure 4 provides the same results of ESL students in grades Kn-12. Differences were found in the *WMLS* *W*-scores when the number of years in the program were included in the analysis; the more years in the program the higher the mean *W*-score was, regardless of the nature of the program. This pattern continued up to 5-6 years indicating a positive relationship between years in the program and *WMLS* *W*-score. As more longitudinal data is acquired, more insight regarding the length in the BE/ESL program will be available.

Figure 2

Figure 3

WMLS Mean W-Score of Kn-6 BE Students by Grade and Years in the Program

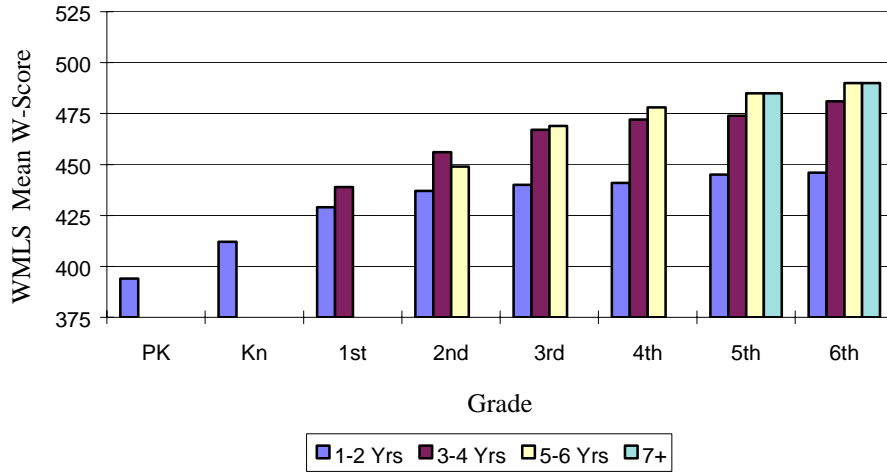
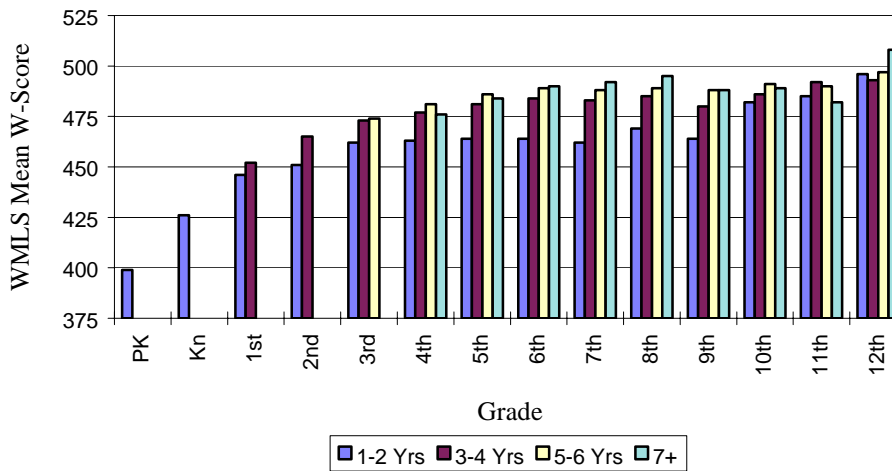


Figure 4

WMLS Mean W-Score of Kn-12 ESL Students by Grade and Years in the Program



The results of the second study, progression of levels, are shown in Table 7. The data are broken down by elementary and secondary students. Trends were slightly different between the two groups. For grades 1-6, the majority of the elementary students remained at the same level after one year of the program. At level 1, 40% advanced, at level 2, 35% advanced, and at level 3, 15% advanced one level. Smaller percentages advanced two levels. Secondary student data indicate that the majority also stayed at the same level but at higher percentages. Between 28% and 29% of level 1 or 2 students advanced to the next level. Those percentages are somewhat lower than those for the elementary students. Results from both studies suggest a lack of advancement of LEP students' cognitive proficiency. The District should be concerned of this lack of progression.

Table 7

Progression of *WMLS* Levels,  
Pre- and Post-Test Comparisons

Pre Test Level	Post Test Results									
	Remained at the same level		Advanced 1 level		Advanced 2 levels		Decreased 1 level		Decreased 2 levels	
	N	%	N	%	N	%	N	%	N	%
Grades Kn-6										
1	939	45.0	822	40.0	280	14.0	0	0.0	0	0.0
2	1648	52.0	1130	35.0	128	4.0	283	8.9	0	0.0
3	3138	71.0	676	15.0	11	0.2	549	12.0	57	1.3
4	644	60.0	29	2.7	0	0.0	366	34.0	21	2.0
5	18	36.0	0	0.0	0	0.0	20	40.0	9	18.0
Grades 7-12										
1	677	66.0	298	29.0	50	4.8	0	0.0	0	0.0
2	491	61.0	228	28.0	8	1.0	80	9.9	0	0.0
3	545	74.0	89	12.0	6	0.8	92	13.0	3	0.4
4	38	40.0	1	1.1	0	0.0	45	47.0	10	11.0
5	0	0.0	0	0.0	0	0.0	2	67.0	0	0.0

Note. Students decreasing one or more levels may be due to improper administration of the test. To correct this the District has put a plan in place to retrain test administrators if there are inaccuracies. Some students were retested. Data are not yet available from that study.

While the aforementioned *WMLS* results provided some insight to the program, other test results also provide meaningful information. *SABE* and Spanish *TAAS* results provide further insight on the effectiveness of the BE/ESL program; the *ITBS* and English *TAAS* results provide data on who are eligible to exit the program; course ACP's provide additional insight to the effectiveness of the courses by reviewing the passing rates.

In our evaluation, all of these outcome data were used in conjunction with data from classroom observations. The outcome and process data results were linked and, thus, making the evaluation complete. For instance, *WMLS* data indicate a lack of cognitive language in the upper grades as depicted by the flattening of the growth curve in Figure 2. These data were explained by the classroom observation data, which found classroom instruction to be weak in providing higher order thinking skills, especially at the upper grades. Hence, for the first time in our evaluation, we were able to interpret the outcome data with the process data. Accordingly, it was recommended in the 1995-96 program evaluation report that the curriculum be strengthened to include higher order thinking skills.

Can Districts comprehensively include LEP student data in accountability systems? With a testing policy that allows *all* LEP students to participate in the administration of some test, and the continuation of this testing program year after year, pre and post data will be available for a district's accountability system. In our District, all students have at least one appropriate measure and, therefore, will be included in the accountability system. LEP students are measured against each other on the same test. The purpose of the accountability system is to determine the amount of growth as compared to the District mean (Webster and et.al, 1997). This concept of assessing growth coincides with the opinions of an expert panel which made recommendations on students outcome variables for LEP students (Hopstock 1995). The focus group made three points on how outcome information should be used. One point was that "standards relating to change or growth should be used rather than standards or criterion scores developed for mainstream students."

### Conclusions

In this paper, we discussed our experience with the testing programs of LEP students and its results, our concept of inclusion of LEP students in the testing program, and the methods we used to determine the readiness of our LEP students in the District's regular testing program. We feel that conditional inclusion is the key to our success with these measures.

The data we explored provide answers to questions by examining the relationships of CALP measures of LEP students given by the *WMLS* and standardized academic achievement measures given by the *ITBS/TAP* and *TAAS* tests. Since these baseline results have so far proven to be successful, continued administration of the *WMLS* would provide additional information on the effectiveness of the BE/ESL programs. Other questions that can possibly be answered with LEP students' longitudinal are: Is the BE/ESL program more effective at certain grade levels? Can the number of years that students need to stay in BE/ESL programs be predicted? What areas of the curriculum need to be changed to improve performance? Searching for the answers to these questions are on our agenda.

### Reference

Anstrom (1997). Evaluation of Linguistically and Culturally Diverse Students and Programs that Serve Them. *BCBE Cross Currents*. Vol.1, No.2.

Collier, V. (1995). Acquiring a Second Language for School. *Directions in Language & Education, NCBE*. Vol., No.4, Fall 1995.

Cummings, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Austin, Texas: Pro-Ed.

Hopstock, P.J (1993). *Recommendations on Student Outcome variables for Limited English Proficient (LEP) Students*. Special Issue Analysis Center, *Task Order D170 Report*. Arlington, VA: Development Associates, Inc. (via internet, NCBE home page).

McDonnell, L.M. and Hill, P.T. (1993). *Newcomers in American School: Meeting the Educational Needs of Immigrant Youth*. RAND Publication, Santa Monica, CA.

Mclean, Z. Y. (1995). *History of Bilingual Assessment and its Impact on Best Practices Used Today*. New York State Association for Bilingual Education Journal. V10, pp. 6-12.

Oakeley, C. (1995). *Final Evaluation of the Bilingual Education and English-as-a-second language programs, 1994-95 (REIS95-440-2)*. Dallas, Texas: Dallas Independent School District, Department of School Evaluation.

Oakeley, C. (1996). *Final Evaluation of the Bilingual Education and English-as-a-second language programs, 1994-95 (REIS96-440-2)*. Dallas, Texas: Dallas Independent School District, Department of School Evaluation.

Saville-Troike, M. (1991). *Teaching and Testing for Academic Achievement: The Role of Language Development*. FOCUS: Occasional Papers in Bilingual Education, Number 4. Washington, DC: US Department of Education, Office of Bilingual Education and Minority Languages Affairs.

The National Center for Education Statistics (1996). Increasing the Inclusion of Students with Disabilities and Limited English Proficient Students in NAEP. *Focus on NAEP*, Vol. 2, No.1. July 1996.

Silcox, B. (1997). OBEMLA Releases SEA Report for 1994-95. *NCBE Cross Currents*. Vol.1, No.2., Winter 1997.

Webster, W. J., Mendro, R., Beardon, D. K., Bembry, K. L., and Jordan, H. (1997). *Rewarding Effective Schools - Theory and Practice in an Outstanding Schools Awards Program*. Paper presented at the Annual Meeting of the American Educational Research Association, March 1997, Chicago, Ill.

Zehler, A.M., Hopstock, P.J., Fleischman, H. L., and Greniuk, C. (1994). *An Examination of Assessment of Limited English Proficient Students*. Special Issue Analysis Center, *Task Order D070 Report*. Arlington, VA: Development Associates, Inc. (via internet, NCBE home page).  
Woodcock, R.W. and Muñoz-Sandoval, A.F., (1993). *Woodcock-Muñoz Language Survey Comprehensive Manual*. Chicago: Riverside.

Table X  
Assessment of LEP Students  
in the Dallas Public Schools (annual flow chart)

Time line	Assessment measure	Purpose(s)	Target population	Results/score	Treatment of the data
Fall (Aug., Sept., Oct.)	<i>WMLS</i> (all 4 parts, Eng. & Spanish)	identification Placement	* Newly enrolled LEP students	Broad ability score (Levels W1--W5)	LPAC decision based on data
Spring (Jan., Feb.)	<i>WMLS</i> (all 4 parts, English only)	*Assessment of annual growth * Screening for additional testing	Levels 1-3 students from Fall & last Spring tests	Broad ability score (Levels W1--W5)	LPAC decision based on spring-fall or spring-spring comparisons
Spring February	<i>TAAS</i> Writing (G4,8,10,11,12)	*Student academic annual growth	* Levels 4-5 students from fall & spring tests;	* <i>TAAS</i> objective mastering (student)	
April	<i>TAAS</i> math, reading (G3-8)	* ESL program evaluation	* Level 3 students optional;	* % of mastering on <i>TAAS</i> (group)	
April	Exit <i>TAAS</i> (G12)	* School/district accountability	* Levels 1-3 students enrolled for 3 yr.	* <i>ITBS</i> percentile	
April	<i>ITBS/TAP</i> (G1-9)				
May	<i>ACP</i> (G9-12)				
	<i>SABE</i> <i>TAAS</i> (Spanish) Exit <i>TAAS</i> (G12) <i>ITBS</i> math <i>ESL ACP</i>	*Student academic annual growth	* Levels 1-3 LEP students from spring test  * Students in BE program		