# Why Measuring Growth is Especially Important in Evaluation of English Language Learners

Marie Miller-Whitehead

Paper presented at the Annual Meeting of
AMTESOL

Florence, AL
January, 2005

Address comments to the author at:

Marie Miller-Whitehead
P. O. Box 273
Florence, AL 35631
Mmillerwhitehead1@yahoo.com

# Why Measuring Growth is Especially Important in Evaluation of English Language Learners

Marie Miller-Whitehead

Teachers, by definition, believe that their subjects are important and that there are goals and objectives for learning that their students should achieve, based usually on some agreed upon criteria. The unit test, the mid-term exam, and the vocabulary quiz: these assessments are designed to measure the student's attainment of specific course objectives, and teachers usually expect that students will do very well on them. Native speakers of English, for example, begin studying language arts in kindergarten. Even very young children have a sense of English letters and words because the vast majority have been exposed to environmental print on television, cereal boxes, games, and the like. Most young children of average ability will begin school with a schema for learning English. Most will follow a natural progression of attaining skills in reading and writing. They have heard English since birth, and have spoken it since age two. By the time these children reach middle school or high school, they will have studied English, informally and formally, for anywhere from 13 to 15 years. Not all will be "A" students in English, however.

Now, imagine that you are a student who has recently arrived in the United States. You are placed in an English class where you are expected to read short stories, talk about them, and write about them. Your native language has a different alphabet, or perhaps no written alphabet at all. You have been learning English for 2 years. No matter how long and hard you study, you cannot possibly catch up with the rest of the class, all of whom are native speakers of English. You are doomed to failure, because at best you can only do about half of each test and even reading a short story of 10 to 12 pages takes you several hours of looking up words in the dictionary. You are now surrounded by English every day, but listening to a strange language all day is tiring, so every chance you get, you spend with your friends who speak the same language as you.

Almost everyone who has taught English Language Learners has faced the dilemma of knowing that a non-native student is performing well below grade level. You, as a trained observer, know that the student has average or perhaps even above average ability. How can you encourage the student to continue trying, even though his work is not as accurate or polished as that of the rest of his classmates, even those of lower ability? Not only do you wish to encourage this student, you also need some personal validation of your teaching skills. You know that the student is learning and making progress, but based on his speaking and writing skills he is still far below the class average. You have decided to maintain a portfolio of this student's work to show his parents and counselors how much progress has been made. Fortunately for you, the scores of your English Language Learners will not be reflected in your school's SAT10 scores. You teach several ESL students, and their test scores would bring down your school's average. Some states exempt ESL students from taking the statewide assessments unless they have been

continuously enrolled in a U.S. school for at least 3 years. Other states, such as Alabama, require that all enrolled students participate in the annual statewide assessments. Your ESL students are all from the same country, and all are making progress in their language skills at about the same rate, but you do not feel that it would be to their advantage or the school's for them to participate in the annual statewide assessment. However, under No Child Left Behind only one half of 1% of students in any specified subgroup may be tested using an alternative assessment, unless the group is too small to obtain statistically reliable results. Your ESL students can be tested, but with some agreed-upon accommodations.

You do not teach ESL full time because there are so few of these students at your school, so keeping up with ESL issues is an extra preparation for you. However, you have colleagues who are full time ESL teachers, and their students come from very different language backgrounds. Their students do not all progress at the same rate because their L1s are not the same. For some students, English is an L3 or even an L4. These students have developed sophisticated strategies for language learning already and have a strong knowledge of cognates and language structure. These students will exit the ESL program very quickly; others will take much longer. For most of the latter, English is really an L2; they have little in the way of language scaffolding to build upon. According to Rong and Preissle (1998), only 10% of Vietnamese children who entered the United States in the early 1990s reported oral English proficiency; however, approximately 75% of these immigrant children reported that they spoke English very well after 10 to 15 years in the United States. Of Hispanic children who entered the United States during the same time period, approximately 97% spoke a language other than English at home; only 3% reported that English was the only language spoken in the home. Also, 12% of Hispanic children ages 6 to 16 were not enrolled in school in 1990, four times higher than the percentage of Asian children.

Is there a methodology that will place these students on an equal footing with their native English-speaking classmates, if not in actual achievement, at least in terms of their improvement? There is. You might wish to develop a method to measure how much knowledge students have gained in your class. This methodology is variously known as a student gain score, growth score, or value-added score. Why might student growth or gain be valid for use in ESL? Most experienced teachers know that parent education, socioeconomic status, student ability, and motivation all have an effect on student academic achievement. These factors also apply to ESL students, with the additional confounding effect of interference from their L1 (native language). Research has shown that it may take several times longer for students of the same ability level to learn a second language, depending upon its commonality to their L1. Thus, an equitable system would provide some means of showing improvement based on the student's past achievement, rather than compared to that of students of different ability levels and different socioeconomic status and different native languages.

The federal No Child Left Behind Act (NCLB) has recognized the effect these factors have on student achievement, especially pertinent to the education of special populations of students. However, the NCLB also holds school accountable for teaching all students, even those who do not speak English. As a result, in addition to criteria for reaching proficiency in basic skills, the legislation mandates that schools and districts demonstrate Adequate Yearly Progress

(AYP) in reaching their goals. For example, for subgroups that are below proficient (such as LEP students in English), the 10% rule says that a district has achieved Adequate Yearly Progress if the number of the group below proficient is reduced 10% from the previous year, and progress is made on other objectives, or if the group meets or exceeds statewide annual objectives. Additionally, achievement objectives are to be set for subgroups such that there is a baseline established and equal incremental increases in achievement. Under current provisions of NCLB, LEP students who have been in the United States for three or more consecutive years must be assessed in English and language arts in English.

A requisite for the use of a student growth score is that the student must be in a teacher's group of students for a certain number of days during the school year. In Tennessee, a student must be present for 150 days of the school year for a value-added score to be computed. The rationale for the use of the measure of student growth is that it is independent of the student's ability, ethnicity, socioeconomic status, or the geographic location of the school (McLean & Sanders, 1984). Thus, it is possible for students in a remedial class in a poor inner city school to attain gain or growth scores equivalent to or higher than those of gifted students in an accelerated class in a wealthy neighborhood. Another requisite is that the assessment instrument should be capable of measuring a wide range of student achievement, from several grade levels below to several grade levels above the class grade level. The assessment should also be valid and reliable. Generally speaking, this means that the assessment should be norm-referenced. In Tennessee, where I spent most of my teaching career, value-added gain scores are computed from a student's yearly scores on the California Test of Basic Skills or its newer version, the TerraNova. McGraw-Hill has developed special versions of these assessments for use in the state of Tennessee. The TVAAS researchers have recommended that holistic teaching and teaching integrated subject matter are more consistent with good test scores than teaching isolated facts and skills that have been tested for in the past (Sanders & Horn, 1993).

However, perhaps your students do not take a statewide assessment such as the CTB5, ITBS, or SAT10. What measures might you use? ESL students are often expected to obtain a passing score on a test of English knowledge such as the TOEFL. However, even though schools and programs establish cut scores, students may take TOEFL practice tests or the actual test several times before they qualify to exit the ESL program. The TOEFL is norm-referenced and measures a wide range of student achievement, so it or a test like it would be an appropriate measure.

Whatever assessment is eventually selected, how might a classroom teacher compute a measure of student growth for his or her class? There are many models, including the very sophisticated statewide Tennessee Value-Added Assessment System (TVAAS), a model that computes a "teacher" and "school" effect for each school district in Tennessee, based on scale scores on the CTBS or TerraNova. I began working with the Tennessee data soon after the implementation of Tennessee's Education Improvement Act in 1992. The statistical model is complex and very sensitive, but the concepts may be applicable to the needs of the classroom teacher, particularly where there are wide variations in student achievement and ability levels. The test is administered each year to students in grades 3-8 in language arts, reading, math,

science, and social studies. Thus, each student receives a scale score in five subjects each year, and in writing assessments at specified grade levels.

Here is an example that shows grade level, standard score, percentile rank, and standard deviations for the ITBS language arts subtest, and demonstrates how a teacher might determine the amount of growth a student should demonstrate to maintain or exceed his or her baseline percentile rank. In this example, a baseline scale score of 174 ranks at the 50th percentile for grade 3. At grade 4, a scale score of 174 ranks at the 26th percentile. Thus, a student who received the same scale score in grades 3 and 4 would reduce his percentile rank from 50th to 26th. In effect, the student would have a negative growth of .64 standard deviations. A student's test score would need to increase by .89 standard deviations to maintain standing at the 50th percentile (Table 1).

Table 1
Standard Scores and Percentile Ranks for Grade 3 and Grade 4 on ITBS Language Arts

| | Percentile Rank | |
|---|---|---|
| Standard Score | Grade 3 | Grade 4 |
| 170 | 42 | 21 |
| 171 | 44 | 22 |
| 172 | 46 | 23 |
| 173 | 48 | 25 |
| 174 | 50 | 26 |
| 175 | 52 | 27 |
| 176 | 54 | 29 |
| 177 | 57 | 30 |
| 178 | 58 | 31 |
| 179 | 61 | 33 |
| 180 | 62 | 34 |
| 181 | 64 | 36 |
| 182 | 66 | 37 |
| 183 | 69 | 39 |
| 184 | 71 | 40 |
| 185 | 73 | 41 |
| 186 | 74 | 42 |
| 187 | 76 | 44 |
| 188 | 77 | 45 |
| 189 | 78 | 47 |
| 190 | 79 | 48 |
| 191 | 81 | 50 |
| Standard Deviation | 19.05 | 24.25 |

Now, suppose that your first year ESL student received a baseline standard scale score in grade 4 of 170. While this score places the student in the 21st percentile (below proficient), suppose that in grade 5 the student has improved and receives a score that places him in the 40th percentile. While this is still a lower score than you would like, the student's growth or gain has far exceeded that of most of his classmates.

Other examples are provided in Table 2-4 (Bratton, Horn, & Wright, 1996). Table 2 shows a classroom example, given a national norm gain in math of 25 scale score points. Notice that TVAAS uses three years of student data to compute a growth score, or value-added score because the TerraNova is administered only once each year, in April. The average gain for the Class #1 shown in Table 2 is 31 points, compared to the national norm gain of 25 points for the same three years. What are the average gains for the next two classes in Table 2? What are some sources of error in computing gain scores in this way?

Table 3 shows typical scale scores for the CTB math subtest across grades 2 through 8. U.S. norms are given on the first row, and scale scores for four years from 1992 to 1995 in the remaining rows. To follow a cohort of students, you look on the diagonal for each succeeding year and subtract from the previous year's score. While you can compute these yourself, the TVAAS annual report to schools provides information in the format of Table 4 (Bratton et al.).

Table 4 shows an actual TVAAS score report printout that is sent to schools and school districts each year. The "G" designation signifies that the gain is significantly above the national norm gain for the same period of three years, while "R" signifies a gain that is significantly below national norm gain, and "Y" indicates that the gain is not large enough to be significant after accounting for measurement error, even if the number is higher than the national norm.

What are some reasons that these gain scores are almost always computed from student performance on multiple-choice, standardized assessments? How difficult would it be to compute student growth on assessments such as the DIBELS or Student Oral Language Observation Matrix (SOLOM)? What would be some sources of error?

Table 2

Computing gains at the classroom level

Grade 5 Mathematics National norm gain = 25

| **Student** | **SS for 1994** | **SS for 1995** | **Average SS** | **1995 - 1994** |
|---|---|---|---|---|
| Aaron | 783 | 782 | 783 | - 1 |
| Aileen | 734 | 774 | 754 | + 40 |
| Adam | 715 | 770 | 743 | + 55 |
| Amanda | 716 | 761 | 739 | + 45 |
| Alan | 721 | 743 | 732 | + 22 |
| Amy | 717 | 743 | 730 | + 26 |
| Arnold | 714 | 741 | 728 | + 27 |

Average gain Class #1 = 31

Table 2 (cont.)

| | | | | |
|---|---|---|---|---|
| Barbara | 711 | 727 | 719 | + 16 |
| Barry | 699 | 736 | 718 | + 37 |
| Betsy | 719 | 713 | 716 | - 6 |
| Benjamin | 687 | 730 | 709 | + 43 |
| Brenda | 706 | 709 | 708 | + 3 |

Average gain Class #2 = +19

| | | | | |
|---|---|---|---|---|
| Carl | 674 | 731 | 703 | + 57 |
| Caroline | 678 | 724 | 701 | + 46 |
| Charles | 676 | 722 | 699 | + 46 |
| Chloe | 672 | 711 | 692 | + 39 |
| Christopher | 658 | 704 | 681 | + 46 |
| Colleen | 668 | 679 | 674 | + 11 |

Average gain Class #3 = ????

Table 3

Math -- Estimated Mean Scale Scores

| Grade | 2 | 3 | 4 | 5 | 6 | 7 | 8 | (% of Norm) |
|---|---|---|---|---|---|---|---|---|
| USA Norm | 615.0 | 675.0 | **701.0** | 726.0 | 745.0 | 760.0 | 778.0 | |
| 1992 | 638.0 | 682.9 | 717.4 | 741.9 | 750.2 | 768.9 | 784.0 | |
| 1993 | 634.6 | 684.8 | 713.7 | 739.6 | 751.5 | 768.0 | 782.0 | |
| 1994 | 634.6 | 689.5 | 715.5 | 740.2 | 756.1 | 770.7 | 784.8 | |
| 1995 | 639.7 | 686.3 | 714.7 | 740.7 | 758.8 | 772.7 | 784.0 | |

Table 4

Math -- Estimated Mean Gains and (in parentheses) their Standard Errors

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| USA Norm | 60 | 26 | **H 25.0** | 19 | 15 | 18 | |
| 1993 Mean Gain | 46.9 (0.7) | 30.8 (0.6) | 22.2 (0.5) | 9.6 (0.5) | 17.7 (0.5) | 13.2 (0.5) | 86.1 (0.9) |
| 1994 Mean Gain | 55.0 (0.7) | 30.6 (0.6) | 26.5 (0.5) | 16.5 (0.5) | 19.1 (0.5) | 16.8 (0.5) | 101.0 (0.9) |
| 1995 Mean Gain | 51.8 (0.7) | **C 25.2 (0.6) D** | 25.3 (0.5) | 18.6 (0.5) | 16.6 (0.5) | 13.3 (0.5) | 92.4 (0.9) |
| **L** | | | | | | | |
| 1995 3-Year-Avg Gain: | 51.2 (0.4) R* | 28.9 (0.3) G | 24.7 (0.3) R | 14.9 (0.3) R* | 17.8 (0.3) G | 14.4 (0.3) R* | 93.2 (0.4) |
| 1994 3-Yr-Avg Gain: | 49.4 (0.4) R* | 34.2 (0.4) G | 27.0 (0.3) G | 12.9 (0.3) R* | 19.4 (0.3) G | 16.7 (0.3) R* | 97.9 |
| 1995 Mean Gain | | | | | | | |
| 1994 3-Yr-Avg Gain: | 2.4 S | | | 5.6 S | | -3.4 NS | |

There are many ways to evaluate student progress, and the statistical methods for computing gain or difference can become complex. Depending upon the needs of your students

and your own teaching style, you may use an entirely different approach. Some programs have their own evaluation models that specify how student progress is to be assessed. Perhaps the previous examples are methods that you already use in your class to assess your own teaching and learning. However, these examples may give you some ideas of how the process works in other states. Perhaps you will try some of these and find that your ESL students actually obtain gain or growth scores that are well above that of state or national norms!

NOTES

References

Bratton, S. E, .Jr., Horn, S. P., & Wright, P. (1996). Using and interpreting Tennessee's Value-Added Assessment System: A primer for teachers and principals.

McLean, R. A., & Sanders, W. L. (1984). Objective component of teacher evaluation—A feasibility study. (Working paper No. 199). Knoxville: University of Tennessee College of Business Administration.

Rong, X. A., & Preissle, J. (1998). Educating immigrant children: What we need to know to meet the challenges. Thousand Oaks, CA: Corwin Press.

Sanders, W. L., & Horn, S. (1993). An overview of the Tennessee Value-Added Assessment System (TVAAS). Knoxville: Tennessee Value-Added Assessment Research Center.