Establishing Annual Measurable Achievement Objectives for Limited English Proficiency Students: Five Options

Barbara A. Badgett
University of Nevada, Las Vegas

Chad W. Buckendahl
Robert A. Spies
Buros Center for Testing
University of Nebraska, Lincoln

October 2006

Abstract

Title III of No Child Left Behind (NCLB) requires states to establish Annual Measurable Achievement Objectives (AMAO) as benchmarks against which the annual performance of students' with Limited English Proficiency (LEP) can be compared. This paper discusses five potential strategies for establishing AMAO for LEP students: 1) growth, 2) observed empirical, 3) status, 4) improvement status, and 5) judgmental. For each of these options, we describe the measurement strategy, offer illustrations from other states when applicable, and analyze the strengths and weaknesses based on educational and political goals. Specific challenges of accountability systems for subpopulations are also discussed.

Establishing Annual Measurable Achievement Objectives for Limited English Proficiency Students: Five Options

Introduction

Federal legislation requires state departments of education to choose standardized assessments to measure English literacy skills for their Limited English Proficient (LEP) student populations. According to No Child Left Behind's (NCLB) Title III requirements, proficiency targets, or Annual Measurable Achievement Objectives (AMAO), must be established as benchmarks against which students' annual English language proficiency, as determined by the standardized assessment of choice, can be compared. With enhanced visibility of special populations (e.g., students with disabilities, students identified as Limited English Proficiency) within states, there are a number of questions raised regarding appropriate strategies for accountability systems. Although many of these discussions started with the "regular" education population, they have expanded to the full student population. The additional challenge of a shift in the historical curricular/instructional focus to more academic content expectations and the heterogeneity of the distributed abilities within these sub-populations have broadened the conversations without consensus on how to respond to the legislated expectations.

The primary purpose of this paper is to describe potential strategies for establishing AMAO for LEP students. These strategies are discussed separately and characterized as being more compliant or less compliant with respect to the federal legislation. For each of these options, we describe the measurement strategy, offer illustrations from other states (when applicable), and analyze the strengths and weaknesses of each strategy. These options include: 1) growth, 2) observed empirical, 3) status, 4) improvement status, and 5) judgmental.

Title III specifies three requirements that states must include when establishing Annual Measurable Achievement Objectives. These requirements are listed here:

1. annual increases in the number or percentage of children making progress in learning English;
2. annual increases in the number or percentage of children attaining English proficiency by the end of each school year, as determined by a valid and reliable assessment of English proficiency; and
3. making adequate yearly progress for limited English proficient children as described in Title I.

Table 1 serves as an advanced organizer for discussing this topic in terms of both measurement and policy considerations. Thirteen considerations are listed in this table and provide comparative information for five potential measurement strategies for establishing AMAO for LEP students. Some policy questions are implicit in selecting an accountability model. These may include: 1) the capacity to select a more or less complex model of assessment and 2) the need to conform to current NCLB regulations.

Table 1. Technical and policy considerations for establishing AMAO within states

Considerations by accountability model	Growth	Observed Empirical	Judgmental	Status	Improvement Status
Underlying purpose	Rank/rate schools based on performance change	Rank/rate schools based on performance change	Rank/rate schools based on current performance	Rank/rate schools based on current performance	Rank/rate schools based on current performance
Major issues for consideration					
Ratings generally understood	Likely	Likely	Very Likely	Very Likely	Very Likely
Requires more than 1 year of data	Yes	Yes	No	No	No
Unique student ID required	Yes	Yes	No[a]	No[a]	No[a]
Potentially confounds student & school effects	Less Likely	Less Likely	Yes	Yes	Yes
Implementation time	Moderate/varies	Moderate/varies	Quick	Quick	Quick
Implementation Process	Moderate/varies	Simple	Simple	Simple	Simple
Possible to measure within school inequities in performance	Possible	Possible	Limited	Limited	Limited
Costs	Moderate	Low	Low	Low	Low
Measures change for individual students	Yes	Yes	No	No	No
Intended consequences	Reward growth	-Reward growth at student level -Growth targets based on observed data	Reward high performing school	Reward high performing school	Reward high performing school
Unintended Policy consequences	-May ignore high achieving schools -May ignore within school inequalities	-May ignore high achieving schools -May ignore within school inequalities	-Fosters status quo -Ignores within school inequalities -Rewards schools with "favorable" enrollment	-Fosters status quo -Ignores within school inequalities -Rewards schools with "favorable" enrollment	-Fosters status quo -Ignores within school inequalities -Rewards schools with "favorable" enrollment
Unintended Educational consequences	-Perceived different standards for different subgroups -Focus on students of all abilities	-Perceived different standards for different subgroups -Focus on students of all abilities	-Does not reward student achievement growth (school improvement) -Reduce incentives for high-quality teachers to teach -Focus on students just below proficiency	-Does not reward student achievement growth (school improvement) -Reduce incentives for high-quality teachers to teach -Focus on students just below proficiency	-Does not reward student achievement growth (school improvement) -Reduce incentives for high-quality teachers to teach -Focus on students just below proficiency
Currently approved by EducationDepartment for NCLB	Under consideration	No	Yes	Yes	Yes

Note. Adapted from "Policymakers' guide to growth models for school accountability: How do accountability models differ?" by P. Goldschmidt, P. Roschewski, K. Choi, W. Auty, S. Hebbler, R. Blank, & A. Williams, (2005). Retrieved February 1, 2006, from http://www.ccsso.org/Whats_New/Press_Releases/7461.cfm

[a]Future disaggregation of data would require unique student ID.

The first section of this paper describes options that state departments of education have considered in establishing their AMAO policy that may be less compliant with NCLB. Some of these have also been discussed for the regular assessment programs.

Growth models

An increasingly popular model that is being used or considered by states for the regular education population is a growth model. Growth models offer stakeholders an opportunity to track proficiency level or achievement score progress from one year to the next at individual and group levels. In theory, this allows the comparison of students' or student cohorts' current year performance to their own previous year's performance (Blank & Cavell, 2005). For example, in this approach students' fourth-grade scores are compared to their own third-grade scores instead of another fourth-grade cohort's previous year's scores.

Performance improvement is reflected in student or cohort growth scores. In basic growth models, a growth score is the difference between a student's or cohort's previous year's score and current score. A school's growth score is an average of student or cohort growth scores (Goldschmidt, Roschewski, Choi, Auty, Hebbler, & Williams, 2005; Blank & Cavell, 2005). More complicated growth models designed to address additional factors and answer more complex questions require expertise in sophisticated psychometric and statistical methods. For example, with value added growth models states can statistically control for confounding variables (e.g., background, prior achievement) and more accurately identify the effects of teachers, schools, programs, and districts.

With data for three or more time points, growth models allow states to begin to identify growth trajectories. That is, they are able to establish whether students are on track to be proficient (Gong, 2006). Through regression analyses, states adopting growth models can identify the expected probability that individual students or cohorts of students will reach proficiency.

Categories of growth models

Growth models can be established according to measures of either relative/normative or absolute/criterion-referenced growth. Relative/normative growth measures establish whether a student's growth is greater, lesser, or typical as compared to the school's population or other representative norm group's sample (Smith & Yen, 2005). It is possible that a relative/normative growth model would provide valuable information to inform instruction or policy change within schools or districts.

Absolute/criterion-referenced growth measures, on the other hand, involve the evaluation of student growth with regards to content and performance standards set by the state. With these measures, growth is expected to result in students reaching established achievement scores by a certain date. Absolute growth measures offer indications of progress towards criteria that are more consistent with NCLB requirements (Goldschmidt et al., 2005).

There are several requirements for the effective implementation of growth models. First, the assessment system of choice must offer a vertically linked scale across grade levels and subjects (Goldschmidt et al., 2005). Vertically linked scales report scores from multiple tests on one scale. Each test covers different portions of the scale,

which allows reported scores for successive grades or ability levels to clearly illustrate growth or development along a continuum (Lerner, 1986). The use of vertical scales is controversial because it requires satisfying strong, hard-to-test, assumptions regarding dimensionality and linearity (Haertel, 2005). For example, in terms of science achievement, the use of a vertical scale requires the assumption that a student's performance on a 7th-grade earth science assessment is directly comparable to his/her performance on an 11th-grade physics assessment. It is possible those achievement relationships are not linear and that different dimensions of science are being tested.

Second, periodic (e.g., annual) administrations of statewide assessments in successive grades are necessary. The results of those assessments must be interpretable based upon state-established performance standards for defining achievement levels (Blank & Cavell, 2005).

Third, statewide, individual student record data systems that allow for the tracking of students across time and schools are necessary. These systems must be capable of intra-district and intrastate tracking of students (CCSSO, 2006). Intrastate tracking of students is only possible if there is a common state testing system.

Finally, states must have at least two-years of data for students for whom they track progress (Gong, 2006). Intuitively, it makes sense that to estimate any change or growth, more than one measure is needed. If states have more than two years of data for individual students, growth trajectories can be examined (Goldschmidt et al., 2005).

Advantages of growth models

Growth models have several advantages: 1) These models result in data that can be aggregated to answer questions addressed through either status or improvement model data; 2) Student learning can be assessed at both individual and group levels; 3) Rates of progress for different subgroups, classrooms, schools, or districts can be assessed and compared when data meet minimum statistical assumptions; 4) Levels of growth—and whether they are above, below, or at average—can inform policy decisions and instruction on an individual or cohort level (van der Ploeg & Thum, 2004); 5) Student growth trajectories can be established, whereby future proficiency levels can be predicted and necessary interventions can be made to target students who are not expected to reach proficiency; and 6) Students' varying initial proficiency levels are acknowledged so that progress at all levels is recognized.

Disadvantages of growth models

Although the benefits of implementing a growth model may make it an attractive option for stakeholders, important resource demands should be considered. Additional costs would be incurred through: 1) developing an appropriate infrastructure; 2) creating a standard setting process for growth; 3) acquiring psychometric expertise to analyze, interpret and explain results; 4) establishing supporting data systems; 5) training relevant staff and creating materials; and 6) addressing the perceived incongruence between NCLB and growth models (Betebenner, 2005).

The first additional cost – developing appropriate infrastructure – may be a deterrent to states with limited resources for technical demands. The cost of the second resource concern – creating a standard setting process for growth – can be minimized by integrating the process into existing standard setting processes (Goldschmidt et al., 2005)

that have already been budgeted. On the other hand, if status standards have already been established for Limited English Proficiency students, growth standards must be set in their own process at a cost comparable to that of previous LEP standard setting processes.

The third additional cost – acquiring psychometric expertise to handle technical issues associated with these models – would also be a demand on resources. The number of staff members and time required of them will depend largely upon the complexity of the questions the state requests to be answered through the data. As questions become more specific or complex, the statistical methods and technical requirements become more sophisticated, time consuming, and labor intensive. The fourth expense—establishing supporting data systems—should not be substantial for states equipped with data systems that identify individual students who can be tracked across time and location. The interconnectivity of this system is critical to the successful implementation of any growth model and serves as a basis for growth model analyses. On the other hand, if a state decides to investigate more sophisticated growth questions, data systems might require upgrades. The fifth expense—training staff and creating materials—cannot be avoided. In order to understand and effectively utilize the results of growth measurement analyses, teachers, staff, administrators, and legislators must receive adequate training. In addition, the media, parents, and public should be informed about the new system, how it works, and how to interpret the results obtained through it. Current training and validation materials will have to be adapted. (Goldschmidt et al., 2005)

According to NCLB guidelines, growth models are not currently accepted as an indicator of AMAO achievement. However, this might soon change. On November 18, 2005, U.S. Secretary of Education Spelling announced an initiative to pilot growth models in 10 states. Then, on May 17, 2006, the U.S. Department of Education reported Secretary Spelling's approval for Tennessee and North Carolina to use growth models to show AYP for the 2005-2006 school year (USDE, 2006). Six other states will soon be considered for approval of growth model approaches to AYP for the 2006-2007 school years. Additionally, in a Report for Congress, Riddle (2005) explained that two bills that would allow growth models as additional annual yearly progress (AYP) indicators are currently before the 109th Congress. If passed, these bills would also allow improvement tracked through growth models to serve as an additional indicator to justify reducing the number of schools marked as failing to make AYP. Although Congress must pass these bills in order for growth models to be used in reporting AMAO achievement to the US Department of Education, states' use of growth models is not contingent upon approval from Congress. States can still use growth models and take advantage of the aforementioned benefits.

Observed empirical models

Tracking individual student growth annually allows for a variation on a growth-model approach that may establish targets that are more reasonable. Growth targets established through an observed empirical model would be based on observed individual student annual growth from time one to time two. The observed averages of growth for cohorts 3-5, 6-8, and 9-12 in each subject would then be used in decision-making about reasonable targets for the respective cohort/subject combinations.

The advantage of this model over the growth model example illustrated in the previous section is that, with the observed empirical model, target annual growth is not

determined based on legislatively imposed regulations. Instead, real data is taken into account to make informed projections about the types of growth that can be reasonably expected. Proficiency attainment target setting would still be possible with this approach, but those expectations may be more rational and empirically based.

The disadvantage to an observed empirical approach is that it would not be compliant with federal legislations. It is likely that reasonable projected growth targets based on observed empirical data would not result in expecting all students to attain proficiency by 2014. On the other hand, it may be more likely that these established growth targets would be achieved. Growth-model targets that only take into account the growth needed in order to reach 100% proficiency by 2014 do not take into account the amount of growth that is actually possible for students. Thus, it is less likely that those targets will be met.

The following sections describe options that may be more compliant with the current federal regulations.

Status models

One strategy that states have chosen to provide estimates of the annual percentages of students meeting English Language Proficiency may be characterized as status models. These methods involve determining a proficiency level and then comparing that proficiency level to a target (i.e., AMAO; Goldschmidt et al., 2005). The development of a status model begins with the determination of initial minimum starting points for expected increases in the percentage of students reaching proficiency each year. Initial expected percentage increases (AMAO) for each subject/grade level are set to be equal to whichever of the following yields a greater value: 1) the fraction of students from the lowest performing subgroup who are categorized as proficient or higher during the baseline year, or 2) the fraction of students from the 20% lowest-achieving statewide schools who have reached proficient or higher achievement levels during the baseline year (Riddle, 2005).

Once a baseline for expected increases is established and threshold levels of achievement are determined, students would be classified according their standardized test scores (using a state's achievement classification descriptions) into one of the following performance categories: below basic, basic, proficient, or advanced. According to baseline increase estimates and the percentage of students falling into each of the above categories, AMAO can then be determined. AMAO are the targets or status bars against which the percentage of students reaching proficiency will be measured. AMAO should be established according to a schedule that will ensure that all students, regardless of their threshold proficiency levels, are on track to reach English language proficiency by 2014.

Categories of status models

Two categories of status models can be used to establish AMAO: conditional and unconditional. With an unconditional status model, unadjusted percentages of proficient students (i.e., mean school performance) are measured against targets. In contrast, conditional status models take into account additional factors beyond a school's influence that might explain some variability in student performance. The primary difference between these two categories of status models is that conditional models assume that

students bring with them background experiences that may affect their progress, whereas unconditional models do not (Goldschmidt et al., 2005). For the Limited English Proficient (LEP) student population, one of the background variables of interest might be students' prior experience with formal schooling.

An illustration of a conditional model is documented in Kentucky. The Kentucky Department of Education (KDE) has taken the prior experience with formal schooling background variable into account in the development of AMAO for its LEP student population. Students were divided into subgroups based on whether they had received formal schooling in their home countries and cohorts based on the number of years educated in the United States. Two groups were first devised based on the former factor: whether students had been formally schooled in their countries of origin. Each of those groups was then subdivided into cohorts based upon their years of US schooling. The group with little or no previous formal education (group one) was divided into seven cohorts: 1) less than two years, 2) two to three years, 3) three to four years, 4) four to five years, 5) five to six years, 6) six to seven years, and 7) seven years or more years of US school instruction. The group with previous schooling (group two) was divided in five cohorts: cohorts 1-4 are the same as the other group and cohort five includes students who had received five or more years of US schooling. Because of the different starting points, there were different intermediate timing expectations for students depending on the subgroup (condition) to which they apply. Specifically, students with less previous experience are given more time to meet the expectations. Group one students are expected to reach English Language proficiency within seven years; group two is expected to reach the same target in five years based on these differential background conditions.

The KDE assigned each of group two's cohorts different percentage increase targets (AMAO) for student progression to successive achievement levels and proficiency attainment. All, half, or none of the students are to move one proficiency level each year so that by 2014 all students will have attained proficiency or better. The annual percentage increases for all group-two students were set at 50% because the majority of the group's students were determined to be performing at the beginning level of achievement (Kentucky Department of Education, n.d.).

Conditional status models, such as that created by Kentucky policy makers, do not hold students with different characteristics to unequal standards. Students are held to the same standards regardless of academic background, but their progress is expected to follow a different timetable. These models do not enable schools to hide poor school performance; instead, they bring it to light and allow districts or states to make between-school comparisons of students with similar characteristics (Goldschmidt et al., 2005). However, Linn (2003) has argued that they are delaying the inevitable given the same expectation of "ultimate" performance (i.e., proficiency attainment).

Status model variations on target performance

According to NCLB, AMAO can be set according to linear increases in intermediate goals, or using a combination of linear and stepwise increases. Linear increases are assessed annually, whereas stepwise increases are measured after a set number of years. Stepwise achievement level increase schedules are allowed according to

NCLB guidelines, with two caveats: 1) the bar must be raised at least once every three years, and 2) during the initial period the bar must be raised after no more than two years.

For example, Arizona has chosen to use a combination of linear and stepwise increases to set their AMAO. Policy makers set their AMAO to extend over three-year increments from 2001 to 2010 and one-year increments from 2010 to 2014 (Bell, 2005). Stepwise increases in intermediate goals can be useful in a number of ways: 1) to allow districts time to fine tune and implement programs targeted at the lowest-achieving students, 2) to provide sufficient time for districts to align their curricula and materials with standards, 3) to allow schools the opportunity to appropriately prepare students for new types of assessments (see Bell, 2005).

Advantages of status models

There are five important advantages to status models. First, status models present information about how schools or subgroups are performing and how much progress must be made in order to meet targets (Linn, 2005). Second, they hold all students and schools to the same standards, although conditional models hold may hold cohorts to different timetables according to relevant characteristics. Third status models are consistent with NCLB requirements. According to NCLB, current status must be reported in terms of the percentage of students at proficiency levels. Fourth, ratings revealed by status models may be more easily understood and may not necessitate additional training for stakeholders. Fifth, status models require no more than one year's data and data system and analytical requirements are relatively low and thus less costly than other models (e.g., growth models).

Disadvantages of status models

The disadvantages of status models should also be noted. First, status models may put schools comprised of initially low-performing populations at a disadvantage compared to schools serving initially high-achieving students (Linn, 2005). That is, status models hold schools responsible for ability characteristics of their student populations over which they have no control. Second, status models do not account for prior performance and require targets to be met regardless of staring points. Third, achievement gains are only recognized in terms of AMAO accomplishment. Status models do not recognize achievement gains above proficiency levels or gains that do not bring students to AMAO targets. Fourth, pressure to achieve AMAO might result in teachers' concentrated efforts being directed toward students performing just below proficiency levels, thereby neglecting those at proficiency and those well above or below proficiency (Riddle, 2005).

Improvement Status Models

In contrast to status models, improvement models offer measures of change between different cohorts. With this model, cohorts' annual progress toward proficiency is compared to the progress of other cohorts (Goldschmidt et al., 2005). With improvement status models, states are able to address whether students' current levels of performance are superior to the performance levels of the previous students in that grade. Students' initial statuses would first be established for each subject/grade combination. AMAO would then be calculated for each subject grade combination based upon the

difference between the baseline proficiency percentage and the 100% proficiency target divided by the number of years between baseline data collection (2004) and 2014 ([100-baseline]/# of years). Beginning at the baseline percentage, each year the percentage of students reaching English proficiency for subject/grade combinations should increase by the percentage result from the above operation. Assuming 2004 baseline data indicates that 20% of ELL students have reached 4th-grade English reading proficiency, then that percentage of students should increase by 8% each year ([100-20]/[2014-2004]). The key is to set the percentage increase rates at a level that puts all students on schedule to reaching proficiency by 2014.

For example, let's assume that 40% of third-grade students reached or exceeded proficiency in English writing. To calculate reasonable AMAO for third-grade writing, the current percentage of students deemed proficient or better (40%) would be subtracted from the target percentage (100%) and the residual would be divided by the number of years between 2005 and 2014 (9). The equation would look like this: (100-40)/9=6.67. This result suggests that approximately 7% of third-grade students should reach English writing proficiency each year.

Advantages

There are two major advantages to this approach. First, if status models for AMAO are not met in terms of aggregate student achievement, states can still meet federal standards under safe harbor if the number of students not meeting proficiency standards decreases by 10% from the previous year. Second, the data management and analyses requirements are the least demanding of the models presented in this white paper. That is, states would merely obtain and group annual data by subject (i.e., speaking, listening, reading, and writing) and by grade level and then make year-to-year comparisons.

Disadvantages

Because of their simplicity, improvement status models may present an attractive option for establishing AMAO, but they present important disadvantages. First, improvement status models result in an evaluation of schools, school districts, or the state—not an evaluation of student progress (van der Ploeg & Thum, 2004). Thus, the results of these accountability assessments are only able to reveal whether schools/districts/states are meeting the AMAO, not which teachers/schools/districts' practices are or are not effective. Thus, although improvement status models may offer valuable information in terms of NCLB accountability requirements, they offer little information in terms of student assessment that can inform practice. Teachers' knowledge of their school failing to meet AMAO is not going to help them understand how to improve their pedagogy. Second, improvement status models involve comparing the proficiency percentage of students at a particular grade level to those of students who were at that grade level the previous year. Students' current proficiency statuses are not compared to their own previous proficiency statuses. Comparing different groups of students in order to establish how much proficiency progress is being made might present a serious confound concerning qualitative differences between the groups being compared. Although for large samples of fairly stable populations these concerns may not be relevant, this year's fourth-grade cohort may be comprised of very different students

than next year's fourth-grade cohort. Changes in proficiency levels might reflect learning, or they might simply reflect qualitative differences inherent in the two groups (Goldschmidt et al., 2005). The use of rolling-averages or uniform averages of multiple years of data to establish improvement targets may better control for misrepresentation of student learning due to a single year's aberrant data.

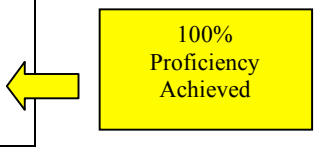
Judgmental models

There may be more reasonable alternatives for setting AMAO that still honor the intent of the federal legislation. The Judgmental model, which is a modified version of status models, provides a more customized plan that takes into account subject/grade cohorts' initial proficiency levels. This might be a more educationally and politically defensible method of establishing AMAO. The underlying structure of a judgmental model would be the same as a status model. The difference lies in how the goal structure is devised. Initial proficiency levels would first be established. A goal structure would then be devised according to a schedule that expects all students to reach proficiency by 2014.

This goal structure could be set according to a linear or linear/stepwise combination scale. In table 2 we offer an example of a goal structure. AMAO are set at 9% progress from one level to the next based on the observed distribution of students' performance. Note that these goal targets are uniform but could be adjusted between the Proficient and Advanced levels to reflect a greater priority placed on the transition from Progressing to Proficient.

Table 2.
Third-grade reading AMAO Goal Structure: 9% Progress goal

Year	Proficiency Level			
	Below Basic	Basic	Proficient	Advanced
2005	61.3	10.5	19.8	8.4
2006	52.3	10.5	19.8	17.4
2007	43.3	10.5	19.8	26.4
2008	34.3	10.5	19.8	35.4
2009	25.3	10.5	19.8	44.4
2010	16.3	10.5	19.8	53.4
2011	7.3	10.5	19.8	62.4
2012	0	8.8	19.8	71.4
2013	0	0	19.6	80.4
2014	0	0	10.6	89.4



Advantages of status models

Judgmental models would likely afford states many of the benefits of status models: 1) they would present information about how schools or subgroups are performing and how much progress must be made in order to meet targets (Linn, 2005); 2) students and schools would be held to the same ultimate standards, but according to customized timetables; 3) judgmental models, while not completely consistent with NCLB requirements, recognize the intent of federal legislation; 4) ratings revealed by

judgmental models are likely to be more easily understood (as compared to growth models) and may not necessitate additional training for stakeholders; and 5) this approach has the same data and analytical requirements as status models and thus are less costly than other models (e.g., growth models). Over and above the advantages shared with status models, judgmental models afford states the ability to establish targets according to current performance and what type of progress is necessary to reach the eventual goal (i.e., 100% proficiency). Therefore, judgmental models, as previously mentioned, might be more politically and educationally defensible. In addition, the previously discussed modifications of status/improvement status models would likely help reconcile some disadvantages of those models: 1) status models do not account for prior performance and require targets to be met regardless of staring points; and 2) improvement status models involve comparing the proficiency percentage of students at a particular grade level to those of students who were at that grade level the previous year. While judgmental models would also involve comparing group performance, projected rates of progress toward proficiency may be more reasonably accomplished.

Disadvantages of judgmental models

Judgmental models also share in common some of the disadvantages of status and status improvement models. First, they may put schools comprised of initially low-performing populations at a disadvantage compared to schools serving initially high-achieving students (Linn, 2005) because they have more distance to travel on the ability continuum. Second, achievement gains are only recognized in terms of AMAO accomplishment. Thereby the models fail to recognize achievement gains above proficiency levels or gains that do not bring students to AMAO targets. Third, pressure to achieve AMAO might influence teachers' concentrated efforts potentially resulting in those students at or well below proficiency to be neglected. In addition, as noted above, the design of judgmental models is not in complete compliance with NCLB.

Summary

In this paper, we have discussed possible strategies for establishing AMAO for LEP student population, including 1) growth, 2) observed empirical, 3) status, 4) improvement status, and 5) judgmental models. Status and improvement status models are the most compliant with federal legislation but may lead to unreasonable targets that do not consider students' prior or existing achievement levels. Judgmental models are similar to status models. Target-setting procedures are similar to those of status models, but with judgmental models students' initial achievement levels are taken into account. Although not specifically endorsed by federal regulation, judgmental models do recognize the basic intent of previous federal legislation. Both growth models and observed empirical models have a different assessment focus than the previously described models and are not endorsed by federal regulations. However, these models allow for analysis of annual, individualized student data—the results of which may be used for accountability and assessment purposes. Observed empirical models, like growth models, involve tracking individual students' annual growth, but growth-target decisions are based on observed empirical data of past student growth levels. Therefore, these models may result in more reasonable targets than growth models and are easier to implement than the traditional growth models.

Ultimately, the decision of state departments of education must reflect and be aligned with the current educational assessment policies of their respective states. Assessment models, including judgmental, status, and improvement status more adequately reflect federal perspectives and may be viewed as focusing on summative approaches that are designed for policy makers who may wish to compare growth within and between school districts.

References

Antal, J., Calliope, M., Wen-Hung, C., Winter, P., & Ferrara, S. (2005). *English Language Development Assessment (ELDA) technical report: 2005 Operational and field test administrations*. Washington, DC: American Institutes for Research.

Bell, R. (July 30, 2005). *Arizona's school accountability system Technical manual Volume III: Title III accountability* (Arizona Department of Education). Retrieved February 1, 2006, from http://www.ade.az.gov/asd/lep/

Betebenner, D. (2005*). The Importance of Performance standards in measures of student growth.* Presentation at the Council of Chief State School Officers 2005 National Conference on Large-Scale Assessment, San Antonio, TX.

Blank, R. & Cavell, L. (2005, March). *Memorandum to Chief State School Officers: Growth Model Primer.* Retrieved February 1, 2006, from http://www.ccsso.org/content/pdfs/Growthmemo.pdf.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Retrieved February 1, 2006, from http://www.ccsso.org/Whats_New/Press_Releases/7461.cfm

Gong, B. (2006, January). *Considering student growth for school accountability decisions: In the context of the NCLB growth models pilot*. Presentation at the Council of Chief State School Officers Growth Models Seminar, San Diego, CA.

Council of Chief State School Officers (CCSSO) (2006, January). Growth models: CCSSO summary of questions and answers. Retrieved February 10, 2006, from http://www.ccsso.org/projects/Accountability_Systems/Growth_Models/7724.cfm

Haertel, E. H. (2005, August). *Using a longitudinal student tracking system to improve the design for public school accountability in California*. Retrieved March 3, 2006, from Stanford University, School of Education Web site: http://ed.stanford.edu/suse/faculty/displayRecord.php?suid=haertel

Kentucky Department of Education. (n.d.) *Title III English language proficiency annual measurable achievement objectives (AMAO) 2003-2004 state report*. Retrieved January 30, 2006, from http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/District+Support/Meetings+and+Trainings/Fall+Assessment+Meeting,+2005.htm

Lerner, R. M. (1986). *Human development* (2nd ed.). New York: Random House.

Linn, R. (2005, April). *Issues in the design of accountability systems* (Center for the Study of Evaluation No. 650). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32, 3-13.*

Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions. A paper in the series: Implementing the state accountability system requirements under the No Child Left Behind Act of 2001.* Retrieved February 2, 2006, from http://www.ccsso.org/Publications/Download.cfm?Filename=AYPpaper.pdf.

Riddle, W. (2005, August). Report for Congress. Adequate yearly progress (AYP): Might growth models be allowed under the No Child Left Behind Act? Retrieved February 15, 2006, from http://www.opencrs.com/rpts/RL33032_20050815.pdf.

Smith, R. & Yen, W.M. (2005). *Models for evaluating grade-to-grade growth*. Presentation at Louisiana Middle School Association, 2005. Retrieved Feb 5, 2005, from www.ccsso.org/content/pdfs/Smith_ModeslforEvaluatingGrade.pdf

United States Department of Education (2006, May). *Press release: Secretary Spelling approves Tennessee and North Carolina growth model pilots for 2005-2006*. Retrieved September 15, 2006, from http://www.ed.gov/news/pressreleases/2006/05/ 05172006a.html

van der Ploeg, A. & Thum, Y. M. (2004, December*). Finding additional value in new accountability systems*. Retrieved February 3, 2006, from http://www.ncrel.org/policy/pubs/pdfs/value.pdf