CRESST POLICY BRIEF 9 • SPRING 2007
National Center for Research on Evaluation, Standards, and Student Testing

# The Practical Benefits of Growth Models for Accountability and the Limitations Under NCLB

**Pete Goldschmidt**
*National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*
UCLA/California State University, Northridge

**Kilchan Choi**
*National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*
UCLA

**Conclusions**

- Growth models can ameliorate problems in the NCLB system, including subgroup reporting problems and minimum group size.
- Each student should count equally towards a school's classification of meeting AYP or not.
- Student growth trajectories should be estimated along with confidence intervals to account for measurement and other sources of error.
- The advantages of growth models are lost if states must track individual student growth, as required under the Growth Model Pilot program.
- Pilot models ought to account for the effects of factors beyond school control.
- Multiple indicators of school quality should be used to measure more fairly and accurately school and student improvement.

The No Child Left Behind (NCLB) Act of 2001 (NCLB, 2002) requires states to monitor student and school performance based on "adequate yearly progress" (AYP), which essentially is a count of the number of students meeting a specified target. States, many of which have a growing number of schools and school districts entering NCLB's "needs improvement" status, have urged the U.S. Department of Education to consider alternate ways to measure and report student progress because AYP disproportionately identifies certain schools as failing (Choi, Goldschmidt, & Yamashiro, 2005). Some researchers have predicted that by the 2013-2014 school year, nearly all schools and school districts will not meet AYP requirements, even many of America's highest achieving schools in affluent areas (Goldschmidt, 2006; Linn, 2005).

In November 2005, U.S. Secretary of Education Margaret Spellings announced a Growth Model Pilot program (U.S. Department of Education, 2006) to which states may submit alternative accountability models to monitor schools. As of

February 2007, five states—Tennessee, North Carolina, Delaware, Arkansas,[1] and Florida[1]—have had their growth models approved for use in 2006-2007. Growth models are a different way to track student progress compared with current NCLB requirements that use AYP. This policy brief addresses the broader topic of education accountability models, or systems, describes both status and growth accountability models, and provides several policy recommendations.

## Purpose of an Accountability Model

First and foremost, policymakers must consider the purpose of an accountability model. Without knowing what policy intends to accomplish, policymakers or educators cannot choose from a myriad of model options or make valid inferences from model results.

Differing purposes of accountability models often depend on who uses the results. Parents are interested in information for the purpose of enrolling their children in "good" schools. The general public wants to know how well their local schools are doing. Education policymakers use accountability results to enforce state or federal achievement goals and often to monitor school performance in order to levy sanctions or provide rewards. Whatever the use, all audiences share the common assumption that accountability results are accurate and that valid inferences and good decisions can be made based on those results. Importantly, all accountability models will likely result in some intended consequences—higher test scores, for example. But they are also likely to produce unintended consequences, such as teaching to the test rather than to broader content standards. Each consequence must be weighed accordingly.

No model can guarantee a specific outcome. Even in combination with effective rewards and sanctions, the best model will not ensure higher achievement. What models can do, however, is provide accurate results that encourage improved learning, quality decision making, and confidence in the entire accountability system.

## Types of Accountability Models

There are two general approaches to monitoring school performance: status models and growth models. Status models use a single year's assessment results as an indicator of school performance and attach decision rules to those results. Growth models use 2 or more years of assessment results as an indicator of school performance and attach decision rules to changes in performance. Although the annual required target that schools must meet under NCLB is termed adequate yearly progress, this model is a status model because it simply counts the number of students meeting the target in that year.[2]

NCLB requires that 100% of students must demonstrate proficiency in reading and mathematics by 2013-2014 for those schools receiving Title I funding. Furthermore, schools must demonstrate adequate yearly progress towards the 100% proficiency target. A school that does not meet the annual target (set by each state) faces increasingly severe sanctions based on the number of contiguous years that the school misses its target. NCLB presumes that monitoring the percentage of students who are proficient in reading and mathematics is sufficient to

identify schools that are doing a good job and schools that need improvement.

Unfortunately, this assumption has several flaws. First, because schools are held accountable for performance by student subgroups, large, diverse schools are less likely to meet their targets simply because they have more subgroups and hence more opportunities to miss achieving their AYP goals (Novak & Fuller, 2003). Second, simply monitoring the percentage of students in a school who score at or above the proficient level in comparison with an annual target percentage places too much emphasis on student enrollment characteristics (a school that routinely receives a large influx of limited English proficient students each year will be at a disadvantage in comparison with a school that receives very few). Third, monitoring school performance based on a single year assumes that current student performance is a function of only the current year's instruction—ignoring past years. Fourth, reducing scores to a single cut-point (proficient or above vs. below proficient) loses a significant amount of information about student performance (Thum, 2003). In most cases, a school will not receive credit for moving students up within an achievement level, nor will it be sanctioned if students move down within a level.

## Status Models

Most policymakers and the public are familiar with a basic status accountability model. Policymakers using this model believe they are answering the question "On average, how are students performing this year?" For example, Figure 1, based on data from our research with a sample of urban schools, shows that half of the schools are meeting their target and half


**Figure 1.**

2002-2003 percent proficient in mathematics. The achievement goal (annual measurable objective) is the AYP target for 2004-2005.

are not. Policymakers and the public would conclude that School A, with 35% of its students proficient or above, is performing better than School U, with fewer than 20% of its students proficient or above. School K would be considered the best school and School L would be considered the worst.

However, this status model fails to answer important questions. To what extent is previous student performance influencing current performance? What student background factors are influencing achievement? How does current performance relate to achieving the 100% proficiency target? How accurate is this model in identifying schools in need of improvement? These questions have promoted interest in growth models as an alternative to status models.

Figure 2 illustrates current performance but also shows growth and decline during the 2003-2004 school year. The same schools met the target as in the previous year, but School K, the highest performing school in 2002-2003, dropped 9% between the 2002-2003 school year and the 2003-2004 school year. Should School K be sanctioned? Perhaps. Further investigation reveals that there was a significant change in the performance of students entering and exiting School K between the two years. The 2003-2004 second-grade students had about 15% fewer students performing at the proficient (or above) level compared with the 2002-2003 second graders. Also, the 2003-2004 fifth graders had 10% fewer students proficient than the 2002-2003 fifth graders (see Figure 3). In summary, the new School K students did not perform as well

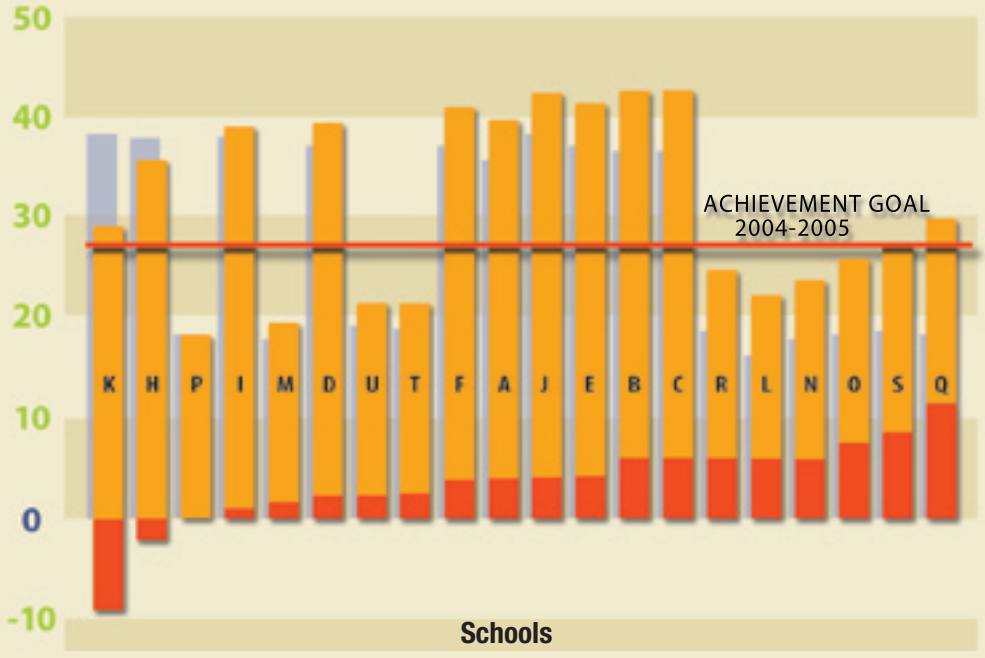
**Figure 2.**

2003-2004 percent proficient in mathematics

1-year change in percent proficient

2002-2003 percent proficient in mathematics

The achievement goal (annual measurable objective) is the AYP target for 2004-2005.



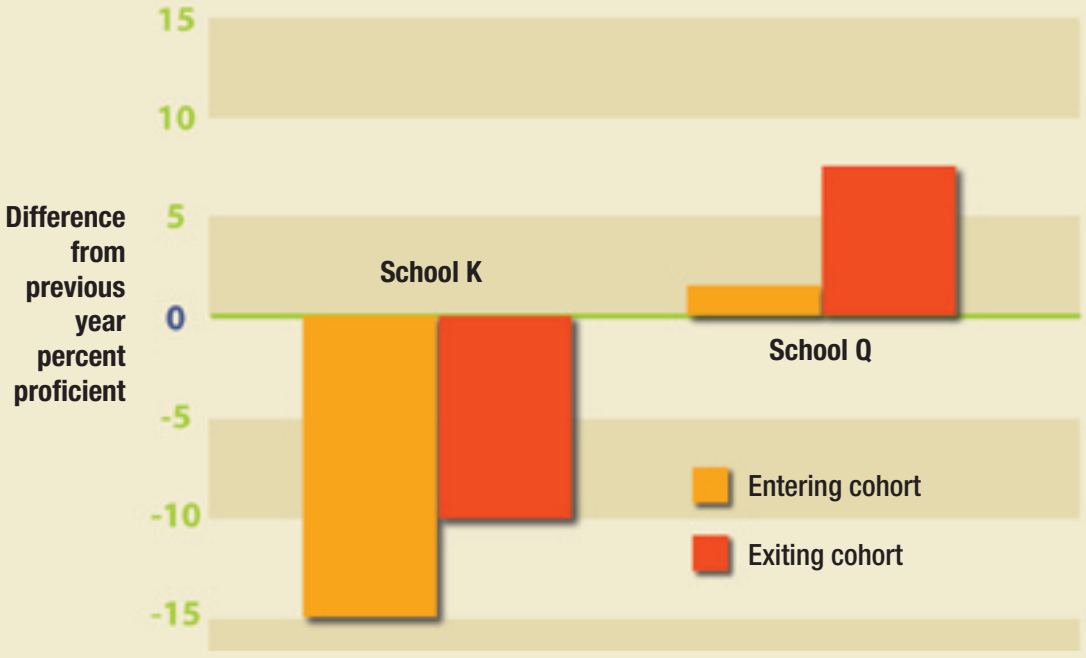



**Figure 3.**
Contribution of entering (second grade) and exiting (fifth grade) cohorts on school performance.

as the students whose place they took, while the scores of students tested in both years remained about the same. Such results are likely more a result of changes in student populations than changes in school quality. School quality seldom varies substantially in just a single year.

Figures 2 and 3 show that School Q, which is among the poorly performing schools in the sample, had the largest 1-year gain of all schools. Should School Q be rewarded? Perhaps, but just as was the case for School K, the dramatic 1-year change is more likely a result of demographic changes outside of the school's control than improved teaching and learning. Basic status models like these provide limited information for decision makers.

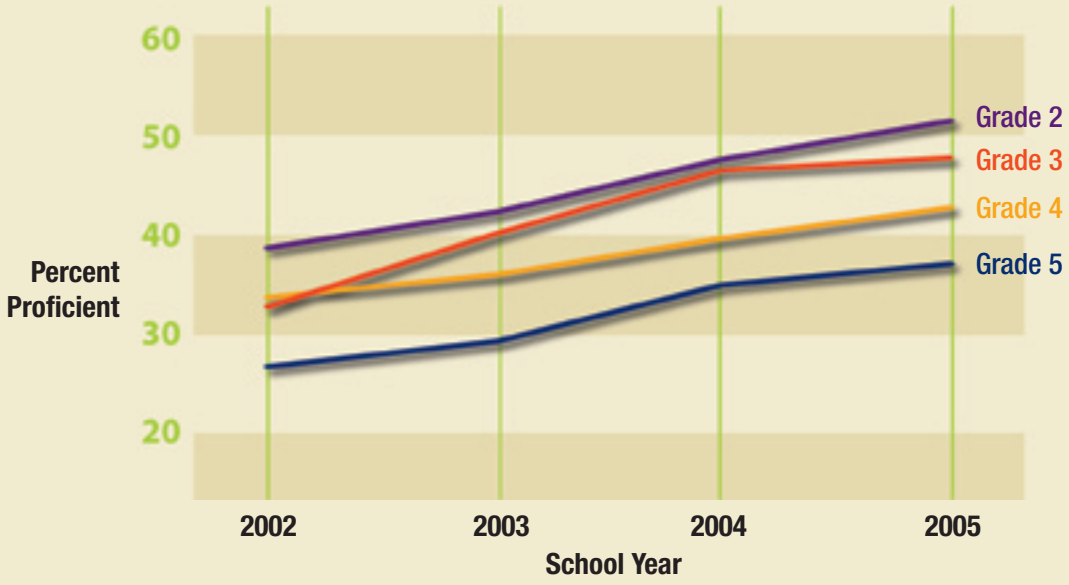
## School Improvement Models

School improvement accountability models have some benefits over simple status accountability models. Figure 4 demonstrates how schools (those presented in Figure 1) are improving based on the percentage of students proficient in each grade over time. For example, slightly less than 40% of second graders were proficient in mathematics in 2002 whereas a little more than 50% of second graders were proficient in 2005. One benefit of a school improvement model is that scaling issues are less important because the same grade is modeled over time.

Second, AYP is easily reported from a school improvement model. In fact, schools may be

**Figure 4.**

School improvement model, 20-school sample.




monitored based not on the current percent proficient or on the change in percent proficient, but rather on whether they are making sufficient progress towards proficiency targets at some future year. A third benefit of a school improvement model is that it incorporates both where a school is currently and whether its growth is enough to reach a specified target. Although the school improvement model in Figure 4 presents how group performance is changing over time, it does not provide much information about how *individual* student performance is changing over time. For that we turn to a growth model.

## Growth Models

As in Figure 4, Figure 5 shows that subsequent groups of students in the same grade are doing better. For example, in 2002 a little less than 40% of second graders are proficient, in 2003 a little more than 40% of second graders are proficient, and in 2005, about 51% of second graders are proficient.

Figure 5 also displays what happens to those same second-grade students as they move up through the grades. Based on the same students (panel data), this is considered a true growth model. The 2002 second-graders' performance is relatively flat over time. As third-grade students in 2003, about 40% of the same students are proficient. As fifth-grade students in 2005, the percent proficient drops slightly to about 37% proficient. The key difference between Figures 4 and 5 is that in Figure 4, sequential cohort performance is tracked over time, whereas in Figure 5, the same students are tracked over time.

Thus, Figures 4 and 5 illustrate that the type of model used can produce very different results.

The sequential cohort approach suggests significant improvement, whereas the growth model approach based on the same students shows either no growth or some decline. The correlation of school rankings based on these two types of longitudinal designs ranges from .25 to .75 depending on subject area and time frame. Even more complex accountability models have emerged that further highlight the differences between models (see Goldschmidt and Hara, 2005, and Choi, 2006a and 2006b, for additional details).[3]

Is one approach preferable to the other? Both methods can provide results from which valid inferences can be made about schools. But in the past 18 months interest has focused on the U.S. Department of Education's Growth Model Pilot program, which many states have attempted to pursue as an alternative to current AYP reporting methods. A key interest is improved validity for identifying schools in need of improvement.

## Growth Model Pilot Program

Proposed growth models under the U.S. Department of Education pilot program can be classified into three categories: (1) models that use a previously known state average growth, (2) models that use individual student growth to predict future growth, and (3) value tables.

The first type, *state average growth,* examines a student's current test scores and projects those scores 3 years into the future based on the current 3-year state average growth. Current-year AYP determinations are based on those projections. The second type of model, *individual student growth,* uses multiple years of *individual* student test scores and estimates that growth into the future. If the current estimated growth is sufficient to meet the proficiency target in a set time frame (e.g., 3 years in the future), then the school is given credit for that student being proficient in the current year. The third type of model, *value tables*, awards a specific

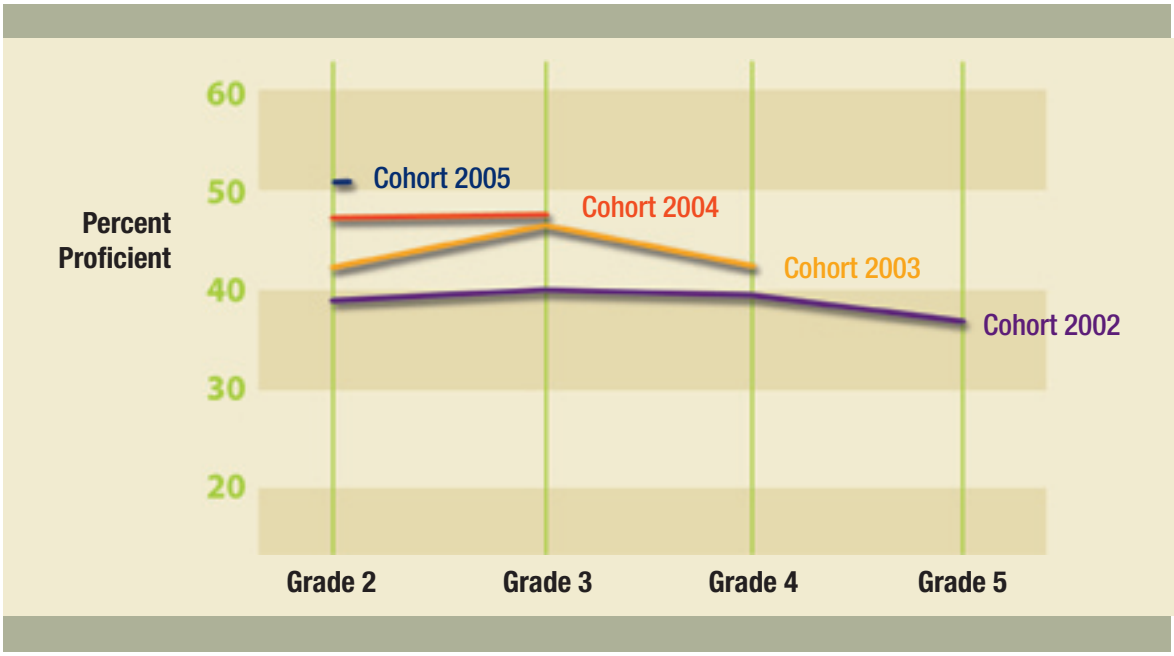



**Figure 5.**
Change in student performance over time.

value for student movement among proficiency levels from one year to the next. For example, a student would earn 100 points for a school if the student moved from *basic* in 2005 to *proficient* in 2006. A school's current-year status (i.e., meeting AYP targets) is determined by the *average point value* of its students. The values in a value table are arbitrarily set based on the values (determined by the state) placed on changes in proficiency categories.

As mentioned earlier, the U.S. Department of Education's Growth Model Pilot program (U.S. Department of Education, 2006) identified seven core principles that pilot programs must meet. The principle of tracking individual student growth combined with the prohibition of aggregating estimated or observed growth for determining AYP makes it difficult to use various types of growth models. Estimating individual student growth parameters may not be precise enough for practical purposes. Thus, using individual student growth parameters for "counting" significantly reduces the rationale for applying growth modeling techniques (Choi, 2006a). As a result, pilot growth models in two states show almost no change in the number of schools making or not making AYP compared with the existing status model. Both models adhered to the Growth Model Pilot program principles.

Another core Growth Model Pilot principle states that growth models must "set expectations for annual achievement based on meeting grade-level proficiency, not on student background or school characteristics" (U.S. Department of Education, 2006). Thus, achievement targets for the 2013-2014 school year must be fixed at the same level for all the students regardless of their characteristics or prior achievement levels. This places potentially unobtainable expectations for growth on initially poor performing students, as well as placing different growth goals for initially poor performing students compared with students initially performing well. These expectations mitigate the potential benefits of using growth models for accountability.

The U.S. Department of Education requires that states validate pilot growth models by comparing overlap in schools not meeting AYP using a state's current NCLB accountability system compared with its new proposed growth model. The inference is that models with the greatest overlap of schools identified as in need of improvement (or not) using both methods are better than models with less overlap. Unfortunately, this test, combined with the other requirements, makes the growth models essentially the same as the current status systems, and thus eliminates virtually all of the benefits from investing in a growth model (for more details, see Choi, 2006a).

## A Better Approach

England uses value added models[4] to monitor school performance. Despite its data limitations (assessments are not administered in contiguous grades and are not vertically scaled), the English education ministry produces school results that are widely accepted by stakeholders. As in the U.S., the English models also use individual student data but focus on mean school growth estimates that incorporate student background characteristics as well. England emphasizes monitoring and reducing achievement gaps

rather than lowering expectations for schools, especially those with large numbers of at-risk students (Ray, 2006).

Although we can argue the merits and caveats of England's value added models, the over-arching difference between the two countries is that the U.S. is attempting to monitor schools based on a single school quality indicator, test scores, whereas England uses several indicators. In the U.S., both status and growth models depend on test scores as the sole indicator of school quality. Placing high stakes onto a single indicator to evaluate school quality means that other important indicators, such as high school graduation or percentage of Advanced Placement courses passed by students, are left out. Furthermore, monitoring schools by just test scores has unintended consequences (Stecher, 2006). Some of the negative consequences are that subjects not tested receive less emphasis and instructional time and that curriculum narrows to the real or expected content of the test (Stecher, 2006).

## Recommendations

Our analyses of both status and growth models during the past few years lead us to the following conclusions.

First, although growth models are more expensive to develop and maintain, and more difficult for the public to understand, the promise that they may help level the playing field for schools, districts, and states is certainly worth pursuing. Growth models ameliorate several major problems in the current NCLB system, including subgroup reporting problems and minimum group size. Currently, scores from students who fall into multiple subgroups can count against a school several times. Also, different minimum group sizes across states pose additional accuracy and fairness issues. Growth models can eliminate these problems if individual student growth is used to generate overall school growth estimates.

Second, within the spirit of No Child Left Behind, we feel that schools should be held accountable for every student's achievement. To that end, each student should count equally towards a school's classification of meeting AYP or not meeting AYP.

Third, monitoring growth requires estimates of student growth trajectories—and because trajectories are estimates, confidence intervals should be used to account for measurement error or other potential sources of error. This is especially important for projections beyond the span of the data.

Fourth, we do not advocate estimating individual growth trajectories and counting the number of students in a school that meet a certain criterion, as in value tables. This method is approved under the current growth model principles, but some of the growth model advantages are lost because states must count individual students meeting certain growth expectations. The growth model thus becomes a status model that at best delays not making AYP for a few years.

Fifth, ultimately meeting the 100% proficiency goal is a function of school processes and not the accountability model. The accountability model, along with decision rules placed on the model, merely monitors progress towards that end. If growth towards proficiency is the key

element to monitor along with performance gaps, then pilot models ought to be able to account for the effects of factors beyond school control—such as a student's school readiness (e.g., initial performance status, or other student background characteristics associated with school readiness such as low socioeconomic status). States should have the option to include some of these factors in order to examine how schools are helping students who begin school at a disadvantage.

Finally, as stated in the CRESST Standards for Educational Accountability Systems (Baker, Linn, Herman, & Koretz, 2002), we encourage the use of multiple indicators of school quality to more fairly and accurately measure school and student improvement.

## Notes

[1] As of this writing, Arkansas' and Florida's growth models are approved but cannot be implemented until their assessment systems receive departmental approval.

[2] With some complications and exceptions, such as subgroups, minimum *n,* and the safe harbor provision.

[3] The studies by Goldschmidt and Hara (2005) and Choi (2006a, 2006b) used data from districts whose underlying organizational structures are significantly different. The models also used somewhat different approaches, but still led to relatively similar conclusions.

[4] Value added models are similar to growth models except that the emphasis is on random effects as opposed to fixed effects for growth models (for more detailed discussion of value added models see Choi et al., 2005, and Goldschmidt et al., 2005).

## References


Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002, Winter). *Standards for educational accountability systems* (CRESST Policy Brief 5). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Choi, K. (2006a). *Growth-based school accountability systems: Key issues and suggestions* (Invited paper prepared for the U.S. Department of Education). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Choi, K. (2006b, April). *A new value-added model using longitudinal multiple-cohorts data.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Choi, K., Goldschmidt, P., & Yamashiro, K. (2005). Exploring models of school performance: From theory to practice. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement* (NSSE Yearbook, Vol. 104, Part 2, pp. 119-146). Chicago: National Society for the Study of Education. Distributed by Blackwell Publishing.

Goldschmidt, P. (2006, April). *Practical considerations for choosing an accountability model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Goldschmidt, P., & Hara, M. (2005, April). *Are there really good schools: The role of changing demographics within schools.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., et al. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Washington, DC: Council of Chief State School Officers.

Linn, R. L. (2005). *Test-based educational accountability in the era of No Child Left Behind* (CSE Rep. No. 651). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Novak, J., & Fuller, B. (2003). *Penalizing diverse schools? Similar test scores, but different students, bring federal sanction* (PACE Policy Brief). Berkeley, CA: Policy Analysis for California Education.

Ray, A. (2006, October). *Value added implementation in England.* Paper presented at the First International Meeting on Value Added and School Accountability, Santiago, Chile.

Stecher, B. (2006, September 13). *"No Child" leaves too much behind.* Washingtonpost.com. Retrieved December 12, 2006, from http://www.washingtonpost.com/wp-dyn/content/article/2006/09/11/AR2006091100581_pf.html

Thum, Y. M. (2003). Measuring progress toward a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods and Research, 32,* 153-207.

U.S. Department of Education. (2006). *No Child Left Behind. Growth models: Ensuring grade-level proficiency for all students by 2014.* Retrieved March 21, 2007, from www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf

## Secretary's Core Principles for the Growth Model Pilot Project

### U.S. Department of Education

*(www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf)*

1. Ensure that all students are proficient by 2014 and set annual goals to ensure that the achievement gap is closing for all groups of students;
2. Set expectations for annual achievement based on meeting grade-level proficiency, not on student background or school characteristics;
3. Hold schools accountable for student achievement in reading/language arts and mathematics;
4. Ensure that all students in tested grades are included in the assessment and accountability system, hold schools and districts accountable for the performance of each student subgroup, and include all schools and districts;
5. Include assessments in each of grades 3–8 and in high school for both reading/language arts and mathematics, and ensure that they have been operational for more than one year and receive approval through the NCLB peer review process for the 2005-06 school year. The assessment system must also produce comparable results from grade to grade and year to year;
6. Track student progress as part of the state data system; and
7. Include student participation rates and student achievement on a separate academic indicator in the state accountability system.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in the publication do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.



UCLA
GRADUATE SCHOOL
OF EDUCATION AND
INFORMATION STUDIES

UCLA
Center for the Study of Evaluation

CSE/CRESST
GSE&IS Building, Mailbox 951522
Los Angeles, California 90095-1522

ADDRESS SERVICE REQUESTED

Presorted Standard
U.S. Postage
**PAID**
Permit No. 400
Laguna Beach, CA