State of Arizona
Department of Education

Tom Horne
Superintendent of
Public Instruction

July 2, 2007

Addendum to Arizona's Growth Proposal
Adjustments for Regression to Mean and Error in Gains Scores

This addendum describes how the Arizona growth model will address the issues regarding error raised in the peer review. In their comments the peers were concerned about regression to the mean and errors in the gains scores. After discussion with the peers, U.S. ED staff, and its accountability advisory committee, the Arizona Department of Education proposes to correct for these problems by using the predictions of a regression model. This decision is based on the superior predictive validity of the proposed method. The proposed method makes the correct prediction 80 percent of the time for most subjects and grades, compared to 50-70 percent for the other methods. The other two methods tend to err on the side of avoiding false positives leading to a high number of false negatives.

Proposed Method. Using the Predicted Score with Correction for Measurement Error.

This method corrects for error by using the prediction of a student's score resulting from a regression equation as well as the standard error of the estimate. These corrections are made to ensure that schools do not receive credit for spurious changes that may be a function of statistical error. Instead, our goal is to ensure that schools receive AYP credit for students who appear to be truly scoring above their expected targets.

ADE will regress current year scores on the previous year's scores using a regression model with school fixed effects. The estimated coefficients will be used to generate predicted scores for the current year for each student. The standard error of the prediction is then used to evaluate whether the predicted value is truly larger than the growth target or whether the difference is due only to measurement error.

Specifically, the following equation will be estimated by subject and grade:

$$Scalscor_{it} = \alpha_j + \beta Scalscor_{i,t-1} + \varepsilon_i . \qquad (1)$$

$Scalscor_{it}$ is the scale score of student *i* on the AIMS test for the current year.

$Scalscor_{i,t-1}$ is the scale score of student *i* on the AIMS for the previous year.

α_j is the fixed effect for school *j*.

ε_i is a normally distributed error term with mean zero and standard deviation σ.

The estimated coefficients are then used to generate predicted scores, $PRED_{it}$, for each student:

$$PRED_{it} = \hat{\alpha}_j + \hat{\beta} Scalscor_{i,t-1}.$$

Also estimated is the standard error of the prediction: $SEPRED_{it.}$ The lower bound value found using the 97.5th percentile of the t distribution is then calculated for the prediction:

$$Lower_{it} = PRED_{it} - t_{2.5}\, SEPRED_{it};$$

where $t_{2.5}$ is the 97.5th percentile of the t distribution.

If $Lower_{it} \geq$ Growth Target, then the student is deemed to have met the growth target. This estimate is used to evaluate whether the fitted value at time *t* is truly larger than the expected score at time *t*.

The standard error of the prediction is calculated using the following formula:

$$SEPRED_{it} = \sqrt{h_i s^2}$$

where $h_i = x_i \left(X'X\right)^{-1} x_i'$; X is the matrix of regressors, x_i is the *i*th row of X, and s^2 is the mean squared error.

The estimates of the parameters, the predicted values, and the standard errors of the predicted values are all generated using the SAS PROC MIXED procedure. Since we assume there are no random effects and the variance is constant across schools, the procedure is equivalent to ordinary least squares.

Examples

The following table shows the results for 4th grade math for a single school.

Student	3rd Grade Score	3rd Grade Performance Level	4th Grade Score	Predicted 4th Grade Score	Standard Error of Prediction	Lower Bound	Growth Target	Met Growth Target?
A	362	FFB	447	417	4.27	409	407	Y
B	409	A	456	455	4.27	447	438	Y
C	456	M	470	493	4.27	485	469	Y
D	521	E	579	546	4.27	537	513	Y

For this grade, subject, and school the estimated parameters are:

$$\hat{\alpha}_j = 123.26;$$
$$\hat{\beta} = 0.8111.$$

Sample calculations are given for student A:

$$PRED_A = 123.26 + 0.8111 \text{ X } 362 = 417$$

$$LOWER_A = 417 - 1.96 \text{ X } 4.27 = 409.$$

where $t_{2.5}$ with 62,600 degrees of freedom is approximately 1.96.

The following example illustrates how the expected score at time *t* is determined for student A:

$$\text{Annual growth target} = (496 - 362)/(6 - 3) = 45.$$

The student must score 362 + 45 = 407 on the 4th grade math test in order to meet the growth target.

Since 409 > 407 the student met the growth target.

Predictive validity of proposed method.

The table below provides evidence regarding the predictive validity of the proposed method. Following the suggestion of U.S. ED staff, the time horizon used to set the growth targets was shortened to two years (2005 to 2007) instead of three. The table compares if a student met the growth target in 2006 to the student's proficiency in 2007. For example, the first cell in the table says that of the students who took the 3rd grade math test in 2005, 16 percent did not meet their growth target in 2006 and were not proficient in 2007. The cell below it says that 11 percent met their growth target in 2006 but were not proficient in 2007. The model makes a correct prediction 80 percent of the time, with the percentage of false positives tending to be slightly higher than the percentage of false negatives.

Percent Accurate Predictions—Proposed Method				
			Proficient 2007?	
Subject	Grade 2005	Met Growth Target in 2006?	No	Yes
Math	3	No	16%	7%
		Yes	11	65
	4	No	19	7
		Yes	13	61
	5	No	23	19
		Yes	4	53
	6	No	19	4
		Yes	16	60

Reading	3	No	20	12
		Yes	7	61
	4	No	15	5
		Yes	14	66
	5	No	22	12
		Yes	7	59
	6	No	19	4
		Yes	14	63