Mississippi Department of Education

Development of a Growth Model for the 2009 Statewide Accountability System

Report #2

This report completely replaces Report #1. All of the pilot growth models were re-run deleting the ten high schools that did not have data for calculating a growth composite in 2009. The new model runs used the final QDI values that reflected any data changes resulting from the request for review process and the AYP appeals process.

Report #2 describes the development and evaluation of sixteen separate pilot growth models. It presents statistical information for each pilot prediction equation and summary statistics for the growth statuses and QDI ranges that are used to determine final accountability labels within the approved accountability system. There were only subtle differences in the final results under the different pilot models, but there were larger differences among the models in terms of simplicity, prediction accuracy at the student level, and interpretation of the growth composite values. The report identifies the model that produces robust predictions and is the easiest to explain and interpret.

Prepared for the Office of Research and Statistics

by
Stephen W. Hebbler

September 21, 2009

Prediction Equations

The final student level matched datasets containing the 2008 and 2009 MCT2 results and the 2008/2009 Algebra I and Biology I results were used to develop prediction equations. Multiple regression analyses were set up to examine two dependent variables for the MCT2 predictions, two types of predictors for the MCT2, and two methods for standardizing the residuals. Separate analyses were run at the school level and the district level. A total of 16 pilot growth models were examined as shown in Table 1.

Table 1

Model ID[1]	Dependent Variable for MCT2	MCT2 Predictors	Standardization Method
Model 001 School Model 001 District	2009 Scale Score	2008 Language SS, 2008 Mathematics SS	Student Residuals Standardized
Model 002 School Model 002 District	SS Change 2008 to 2009	ITP, IRM[2]	Student Residuals Standardized
Model 003 School Model 003 District	2009 Scale Score	ITP, IRM	Student Residuals Standardized
Model 004 School Model 004 District	SS Change 2008 to 2009	2008 Language SS, 2008 Mathematics SS	Student Residuals Standardized
Model 005 School Model 005 District	2009 Scale Score	2008 Language SS, 2008 Mathematics SS	School & District Mean Residuals Standardized
Model 006 School Model 006 District	SS Change 2008 to 2009	ITP, IRM	School & District Mean Residuals Standardized
Model 007 School Model 007 District	2009 Scale Score	ITP, IRM	School & District Mean Residuals Standardized
Model 008 School Model 008 District	SS Change 2008 to 2009	2008 Language SS, 2008 Mathematics SS	School & District Mean Residuals Standardized

[1]For all pilot models, the SATP prediction equations were identical:
2009 Grade 8 Algebra I scale score predicted from 2008 Grade 7 MCT2 Language SS and Math SS,
2009 Grade 9 Algebra I scale score predicted from 2008 Grade 8 MCT2 Language SS and Math SS, &
2009 Grade 9 Biology I scale score predicted from 2008 Grade 8 MCT2 Language SS and Math SS.

[2]These variables are the same as those used for predicting MCT scale score change in the 2003 through 2007 school level accountability system. A description of the variables and their use appeared in the manual developed for the earlier accountability model. It is presented below, adapted for the MCT2.

Given each student's initial achievement on each section of a test and the average performance for all students in the state, a regression model can be developed to predict each student's gain (or scale score) taking into account the student's overall initial achievement and regression to the mean. On the MCT2, the two factors above are operationalized as follows.

1. The overall achievement level of the student at the beginning of the school year is the combined scale scores on the language arts and mathematics tests minus the state mean for the combined scale scores for those tests. The resulting value is a measure of the degree to which the student's overall achievement on the MCT2 compares to the typical achievement for all students in the state. This variable is called the Index of True Proficiency (ITP).
2. The overall value used to correct for regression to the mean is the student's scale score at the beginning of the year on one of the MCT2 tests (language arts or mathematics) minus the state mean scale score for that test. The resulting value is a measure of the degree to which the student's score on that test compares to the typical achievement for all students in the state. This variable is called the Index of Regression to the Mean (IRM).

Prediction Accuracy

The regression equations in the pilot growth models predict performance at the student level. Although the predictions are not accurate enough for use at the student level, the positive and negative prediction errors tend to cancel each other, so average residual values for groups of students within a school or district are much more accurate. Tables 2 through 4 present R-Squared (R^2) values for each prediction equation. R^2 indicates the proportion of variance in the dependent variable accounted for by the prediction equation (e.g., 0.58 means 58%). Generally, a higher R^2 value indicates better predicting ability. The formula for R^2 is shown below.

$R^2 = SS_{Model} / SS_{Total}$ where, $SS_{Total} = SS_{Model} + SS_{Error}$

The R^2 values for the equations that predict MCT2 scale score change were all low. That may be due to a lack of variance since the SS changes from grade to grade on the MCT2 were very small (some slightly negative). R^2 values for the MCT SS change prediction equations in the growth model used from 2003 through 2007 were higher and were similar to the new equations for predicting MCT2 and SATP scale scores.

To ensure the most accurate predictions, students included in the regression analyses had to

- meet full academic year (FAY) at the district level in 2007/2008 and 2008/2009,
- have MCT2 scores from 2007/2008, and
- have 2008/2009 score(s) from MCT2, Grade 8 Algebra, Grade 9 Algebra, or Grade 9 Biology.

Table 2

R^2 for MCT2 Language Arts Predictions (Scale Score or Change)						
Model(s)	D.V.	Grade 3>4	Grade 4>5	Grade 5>6	Grade 6>7	Grade 7>8
001, 005	SS	0.58	0.61	0.62	0.57	0.60
002, 006	Change	0.09	0.18	0.16	0.21	0.26
003, 007	SS	0.58	0.61	0.62	0.57	0.60
004, 008	Change	0.09	0.18	0.16	0.21	0.26

Table 3

R^2 for MCT2 Mathematics Predictions (Scale Score or Change)						
Model(s)	D.V.	Grade 3>4	Grade 4>5	Grade 5>6	Grade 6>7	Grade 7>8
001, 005	SS	0.57	0.60	0.62	0.64	0.56
002, 006	Change	0.27	0.16	0.19	0.23	0.17
003, 007	SS	0.57	0.60	0.62	0.64	0.56
004, 008	Change	0.27	0.16	0.19	0.23	0.17

Table 4

R^2 for SATP Predictions (Scale Score)			
Model(s)	Grade 8 Algebra I	Grade 9 Algebra I	Grade 9 Biology I
The same equation was used for all pilot models.	0.44	0.45	0.61

Standardization Procedures

Once the regression equations were developed, they were applied to the student level data to yield raw residuals (actual performance minus predicted performance). The student level residuals were standardized based on the statewide distribution of raw residuals. Additional computer programs averaged the student level raw residuals at the school and district levels. Those mean residual values were then standardized using the distributions of school and district level mean values. Characteristics of different standardization methods are shown below.

Table 5

Interpretation at the School Level		
Standardized Student Level Residuals Averaged by School	A value of zero (0) indicates that the typical student in the school "just met" his/her predicted performance.	Values above and below zero represent distances from the predicted value in student level SD units.
Raw Student Residuals Averaged by School then Standardized	A value of zero (0) indicates that the performance of the school is at the mean of the school level performance distribution. The state mean in that distribution can represent performance at, above, or below the predicted level of performance.	Values above and below zero represent distances from the statewide school level performance mean value in school level SD units.
Interpretation at the District Level		
Standardized Student Level Residuals Averaged by District	A value of zero (0) indicates that the typical student in the district "just met" his/her predicted performance.	Values above and below zero represent distances from the predicted value in student level SD units.
Raw Student Residuals Averaged by District then Standardized	A value of zero (0) indicates that the performance of the district is at the mean of the district level performance distribution. The state mean in that distribution can represent performance at, above, or below the predicted level of performance.	Values above and below zero represent distances from the statewide district level performance mean value in district level SD units.

Notes:

There is a separate and unique prediction equation for each test. Standardization of the residual values (at the student or school and district levels) is required in order to combine the results across tests for different grades and subjects.

It is assumed that the conversion of the student level scale scores for the predictors into the ITP and IRM variables was a way to standardize the predictors in the earlier growth model. The technique was developed by David Thissen of the L.L. Thurstone Psychometric Laboratory at the University of North Carolina. Presumably, if the scale score differences on different tests from year to year are similar, this standardization of the predictor variables should produce residuals that can be combined and averaged without further adjustment. Although the ITP and IRM variables were used in the earlier Mississippi growth model, the resulting residuals still had to be standardized because there were very large differences in the MCT scale score gains across grades and subjects.

Calculation of a Growth Composite

The standardization procedures described above produce comparable values for each separate test. Every school will have some combination of standardized values depending on the grade configuration and the SATP courses (if any) taught at the school. Most districts have standardized values for all of the tests. Since the standardized residual values are directly comparable, they can be weighted (for the number of students contributing data toward the standardized value) and summed to yield a single growth composite for the school or district. The growth composite is interpreted differently depending on whether the weighted values were derived by standardizing the residuals at the student level or at the school and district levels. The interpretation information shown in Table 5 is applicable to the growth composite. A separate growth composite was calculated for each school and district under each pilot growth model.

Combining Pilot Growth Model Results with the Achievement Model

Computer programs combined the results from each pilot growth model with the QDI already calculated for each school and district. If the growth composite value was zero (0) or greater, the school or district was assigned the growth status label, “Met.” If the growth composite value was less than zero (0), the school or district was assigned the growth status label, “Not Met.”

Those programs also applied the recently calculated high school completion information – High School Completion Index (HSCI) and the 5-year cohort graduation rate -- according to the approved accountability model. Finally, a final pilot accountability label was assigned to each school and district based on the combined results from the pilot growth model, the achievement model and (for schools and districts with a graduating class) the high school completion information.

A sizable negative correlation has generally been found between annual student achievement (the QDI value) and certain demographic characteristics at a school or district -- particularly, percentage of students eligible for free lunch and percentage of minority students. Since the pilot growth models predict future student performance based on each student’s previous test scores, an effective growth model should not exhibit significant correlations with those demographic variables.

An accountability model comparison form (on page 7) includes the following information:

- the number and percentage of schools/districts assigned each growth status (Met, Not Met, NA),
- the number and percentage of schools/districts falling into each 2009 QDI (Achievement) range,
- a cross-tabulation of N and % for each growth status/QDI range combination, and
- a correlation matrix showing the relationships among the following variables – QDI, pilot growth composite, HSCI, pilot accountability label, % free lunch eligible, % minority, N-count.

Although the correlation between the QDI and school demographics was large (r = -.77 for free lunch eligibility; r = -.66 for minority percentage), the correlation between the pilot growth composite values and school demographics was much smaller (r = -.28 for free lunch eligibility; r = -.28 for minority percentage).

As in the accountability model used from 2003 through 2007, higher performing schools met their growth expectations at a greater rate than lower performing schools. Thus, there is a moderate positive correlation between the QDI and the growth composite (r = +.54).

Since a school’s performance on the growth component has only a slight affect on its assigned accountability label, the correlation between the QDI and the pilot accountability label (which was given an ordinal scale numeric value for this purpose) was very high (r = +.94)

The correlation values at the district level between the QDI and growth composite values and demographics were larger (more negative). However there was still a pronounced difference between the QDI correlations (r = -.85 for free lunch eligibility; r = -.81 for minority percentage) and the growth composite correlations (r = -.46 for free lunch eligibility; r = -.48 for minority percentage).

Results on some of the models were identical. For example, the results using the more complex ITP/IRM variables vs. simpler scale score values on the MCT2 are identical because the prediction coefficients are based on the relationships between variables rather than the way the variables are scaled. Since the calculation of ITP and IRM involve subtracting a statewide constant from each student's scale score values (singly or combined), the relationship remains the same. Basically, the results for all models standardized the same way (student level or school/district level) are identical.

In terms of the number of schools and districts falling into each growth status/QDI range cell, differences in the pilot models under the two standardization methods were subtle. However, there are some differences that identify certain models as clearly superior to others.

- The prediction equation statistics presented earlier indicate that the pilot growth models that predict MCT2 SS change (Models 002, 004, 006, and 008) are not as robust as those that predict the scale score on the MCT2 (Models 001, 003, 005, and 007). Although the prediction errors generally cancel within large groups, the pilot growth models use a minimum cohort n-count of 10, so robust prediction equations are desirable.

- The prediction models that use simple MCT2 scale scores as predictors (Models 001, 004, 005, and 008) are simpler and easier to understand than those that use the ITP/IRM variables (Models 002, 003, and 006) and the results using either set of predictors are identical.

- The school and district level growth composites are more meaningful and easier to interpret if a value of zero (0) represents a typical student "exactly meeting prediction" (Models 001, 002, 003, and 004) rather than statewide average performance within the school or district level distribution of mean residual values (Models 005, 006, 007, 008).

Using the three criteria above, one can construct a table showing the degree to which each model satisfies those criteria. Model 001 is the only pilot growth model that satisfies all three criteria, so it is recommended as the model to be used in the statewide accountability system for 2009.

Table 6

Model ID	Robust Predictions	Simple Predictors	Meaningful Growth Composite
Model 001	**Yes**	**Yes**	**Yes**
Model 002	No	No	Yes
Model 003	Yes	No	Yes
Model 004	No	Yes	Yes
Model 005	Yes	Yes	No
Model 006	No	No	No
Model 007	Yes	No	No
Model 008	No	Yes	No

Pilot Accountability Labels Using Growth Model 001

After combining the growth statuses from pilot Model 001 with the 2009 QDI ranges and applying the high school completion information (where appropriate), each school and district was assigned a pilot accountability label based on the approved model. Figure 1 (on page 8) shows the school level pilot accountability labels and Figure 2 (on page 9) shows the district level pilot accountability labels.

Reporting Growth Model Results

The following growth model data should be included in each school and district accountability report.

- One line for each prediction cohort (e.g. MCT2 Grade 3 to Grade 4 Language Arts) containing the following columns
    - Cohort N-Count (number of students in the cohort)
    - Mean Raw Residual (average distance above/below prediction in scale score points)
    - Mean Standardized Residual (average distance above/below prediction in student SD points)
    - Weighted Mean Standardized Residual (the cohort's contribution to the growth composite)
- A summary line showing a sum for the following columns
    - Cohort N-Count (the sum is the total number of data points used for the growth composite)
    - Weighted Mean Standardized Residual (the sum is the Growth Composite)

Approval of a Growth Model and Future Development

Once a growth model has been selected for use in 2009, it must go through the state's Administrative Procedures Act (APA) public review process. Due to the technical nature of the growth model, the APA submittal should outline the basic structure of the selected model and present that structure in relatively simple terms.

Details regarding data preparation, selection criteria for inclusion in prediction cohorts, minimum number of students required for calculating growth composites, and possible additional statistical adjustments (e.g., application of confidence intervals) should be left open for further examination and development.

The APA submission should include statements indicating that additional "production mode" development of the growth model will be conducted. In addition to annual revision of the initial prediction equations, the development process will include:

For 2010

- Equations for predicting 2010 Grade 10 Algebra I scale score from 2008 Grade 8 MCT2
- Equations for predicting 2010 Grade 10 Biology I scale score from 2008 Grade 8 MCT2
- Equations for predicting 2010 English II (multiple choice) scale score from 2008 Grade 8 MCT2

For 2011:

- Equations for predicting 2011 U.S History (from 1877) scale score from 2008 Grade 8 MCT2

The approved overall accountability model indicates that the Grade 5/8 Science Test will be included beginning in 2010. The addition of that test data in the Achievement Model will simply require the addition of the Science distribution statistics to the QDI calculation formulas. Use of the Grade 5/8 science test in the Growth Model has not been determined.

2009 Accountability Model Development Comparison Form

Model ID	Total N and %		N and % in Each 2009 QDI Range										Correlation Matrix					
			<100		100-132		133-165		166-199		200-300			%FL	%Min	Tot-N	Growth	QDI
001 - 004 School Level	Met	402 50.4%	6	1%	55	7%	158	20%	144	18%	39	5%	QDI	-.77	-.66	+.21	+.54	---
	Not	396 49.6%	55	7%	156	20%	136	17%	45	6%	4	1%	Growth	-.28	-.28	.00	---	+.54
	Total	798	61	8%	211	26%	294	37%	189	24%	43	5%	Label	-.72	-.60	+.19	+.61	+.94
Standardization Method: Standardizes residuals at the student level (meaning zero – 0 represents exactly meeting expectation).																		
001 - 004 District Level	Met	66 43.7%	0	0%	10	7%	27	18%	27	18%	2	1%	QDI	-.85	-.81	+.25	+.68	---
	Not	85 56.3%	8	5%	45	30%	27	18%	5	3%	0	0%	Growth	-.46	-.48	+.13	---	+.68
	Total	151	8	5%	55	36%	54	36%	32	21%	2	1%	Label	-.76	-.73	+.26	+.70	+.94
Standardization Method: Standardizes residuals at the student level (meaning zero – 0 represents exactly meeting expectation).																		
005 - 008 School Level	Met	397 49.8%	6	1%	55	7%	156	20%	141	18%	39	5%	QDI	-.77	-.66	+.20	+.54	---
	Not	401 50.3%	55	7%	156	20%	138	17%	48	6%	4	1%	Growth	-.28	-.28	.00	---	+.54
	Total	798	61	8%	211	26%	294	37%	189	24%	43	5%	Label	-.71	-.60	+.19	+.61	+.94
Standardization Method: Standardizes residuals at the school level (0 represents average performance in school growth composite distribution).																		
005 - 008 District Level	Met	76 50.3%	1	1%	12	8%	31	21%	30	20%	2	1%	QDI	-.85	-.81	+.25	+.68	---
	Not	75 49.7%	7	5%	43	28%	23	15%	2	1%	0	0%	Growth	-.46	-.48	+.13	---	+.68
	Total	151	8	5%	55	36%	54	36%	32	21%	2	1%	Label	-.76	-.74	+.25	+.69	+.94
Standardization Method: Standardizes residuals at the district level (0 represents average performance in district growth composite distribution).																		

Figure 1
Pilot Growth Model 001 (School Level)

MCT2 SS Predicted by MCT2 Language SS & Math SS
SATP SS Predicted by MCT2 Language SS & Math SS

QDI Range (2009 Values)	Growth Status[1]			High School Completion Variables
	Not Met 396 (49.6%)	Met 402 (50.4%)		
200-300 43 (5.4%)	High Performing 4 (0.5%)	Star School[2] 31 (3.9%)		HSCI >= 230 or Grad Rate >= 80%
		High Performing 8 (1.0%)		HSCI < 230 and Grad Rate < 80%
166-199 189 (23.7%)	Successful 45 (5.6%)	High Performing[2] 131 (16.4%)		HSCI >= 200 or Grad Rate >= 75%
		Successful 13 (1.6%)		HSCI < 200 and Grad Rate < 75%
133-165 294 (36.8%)	Academic Watch 136 (17.0%)	Successful 158 (19.8%)		
100-132 211 (26.4%)	At Risk of Failing 156 (19.6%)	Academic Watch 55 (6.9%)		
Below 100 61 (7.6%)	Failing 55 (6.9%)	Low Performing 6 (0.8%)		

[2]Note: Schools without a graduating class are assigned this accountability label.

[1]Met indicates a growth composite of 0 or above; Not Met indicates a negative growth composite value. The value in each cell represents the number of schools assigned to that label by the pilot model.

Figure 2
Pilot Growth Model 001 (District Level)

MCT2 SS Predicted by MCT2 Language SS & Math SS
SATP SS Predicted by MCT2 Language SS & Math SS

QDI Range (2009 Values)	Growth Status[1]			High School Completion Variables
	Not Met 85 (56.3%)	Met 66 (43.7%)		
200-300 2 (1.3%)	High Performing 0 (0.0%)	Star District[2] 2 (1.3%)		HSCI >= 230 or Grad Rate >= 80%
		High Performing 0 (0.0%)		HSCI < 230 and Grad Rate < 80%
166-199 32 (21.2%)	Successful 5 (3.3%)	High Performing[2] 20 (13.3)		HSCI >= 200 or Grad Rate >= 75%
		Successful 7 (4.6%)		HSCI < 200 and Grad Rate < 75%
133-165 54 (35.8%)	Academic Watch 27 (17.9%)	Successful 27 (17.9%)		
100-132 55 (36.4%)	At Risk of Failing 45 (29.8%)	Academic Watch 10 (6.6%)		
Below 100 8 (5.3%)	Failing 8 (5.3%)	Low Performing 0 (0.0%)		

[2]Note: Districts without a graduating class are assigned this accountability label.

[1]Met indicates a growth composite of 0 or above; Not Met indicates a negative growth composite value. The value in each cell represents the number of districts assigned to that label by the pilot model.