July 19, 2013

The Missouri Growth Model: Step by Step

The Missouri Growth Model provides a framework for identifying the contributions of Local Education Agencies (LEAs) *and* schools to student achievement. While the focus of this document is explaining growth as reported on an LEA's Annual Performance Report (APR), the word "school" may be substituted whenever an instance of "LEA" appears since the growth measures are also generated at the school-level.

While there are many ways to examine the relationship between school policies and student performance, measures of *student growth*—the change in student achievement across two or more points in time—take a uniquely nuanced perspective on the issue.

Overview

The basic premise of the Missouri Growth Model is that the scale scores on Missouri Assessment Program (MAP) grade-level exams are somewhat predictable. For example, knowing a student obtained a 600 scale score on the 4th grade MAP Mathematics assessment, the model assumes the student will probably score in the vicinity of a 620 scale score on the 5th grade Mathematics assessment. While both scores are below average for their respective grade levels, the model still predicts a learning gain.

It is impossible to know *exactly* how a student will score on an assessment based on the prior year's results—the model simply tries to come close, extrapolating from patterns observed over time across the entire statewide database of assessment scores. This is not to suggest the model predicts students to perform like the state as a whole. Rather, the full range of scores is accommodated so that predictions are about as accurate for low-scorers as they are for high-scorers. Predictions can be off because LEAs have the ability to set policies for themselves, and some policies may be more effective than others in promoting student achievement. The end goal of the Missouri Growth Model is to identify systematic differences in student growth among LEAs with comparable baseline performance.

Step 1: Compiling the Data

Currently, the Missouri Growth Model incorporates the four most recent years of MAP scores in the content areas of Mathematics and English/Language Arts. Only the grade-level assessments (i.e., grades 3-8) are accommodated by the model at this time. The data are arrayed to create "score pairs" whenever two consecutive assessment scores exist for the same student. Up to three score pairs may be included for a given student. For example, a student in 3rd grade in 2009 would be tracked through 2010, and his or her 3rd grade score from 2009 is paired with his or her 4th grade score from 2010 (pair one). Likewise, his or her 4th grade score from 2010 is paired with his or her 5th grade score from 2011 (pair two), and his or her 5th grade score from 2011 is paired with his or her 6th grade score from 2012 (pair three).

Rather than using the MAP scale scores per se, the scores are converted to "z-scores." Z-scores are a common transformation for data that exist on different scales. While a scale score of 640 on the 3rd grade English/Language Arts assessment is about average, the same score on the 4th grade

©2013 Missouri Department of Elementary and Secondary Education

English/Language Arts assessment would be well below average. Z-scores eliminate this problem: a z-score of zero always means "average," regardless of the specific assessment. The following example explains how a z-score is calculated:

Table 1: Calculating a Z-Score

STEP	EXPLANATION
1. Find the mean scaled score for the given assessment. Each combination of grade level, content area, and school year is treated as a different assessment in this context.	The mean is the sum of the scaled scores for all students with a valid score, divided by the number of students with a valid score.
2. Find the standard deviation of the scaled score for the same assessment.	Standard deviation measures the spread of the data. The greater the standard deviation, the further individual data points tend to be from the mean. For example, the numbers [1, 50, 99] will have a greater standard deviation than the numbers [49, 50, 51] even though the mean is the same.
3. Take the student's scaled score and subtract the mean. Then divide by the standard deviation. The result is the z-score.	If the mean is 640; the standard deviation is 38; and the student's actual scaled score is 700; then: Z=(700-640) / 38 Z= 1.5789 Z-scores can be either positive or negative.

The example below illustrates what a sample of scores pairs might look like for a single student:

Table 2: Preliminary Score Pairs for a Hypothetical "Student A"

Current Year	Grade Level (Current Year)	MAP Scale Score (Current Year)	MAP Scale Score Expressed as a Z-Score (Current Year)	Grade Level (Previous Year)	MAP Scale Score for Previous Year Expressed as a Z-Score
2012	8	760	1.30	7	1.10
2011	7	740	1.10	6	0.80
2010	6	720	0.80	5	0.60
2009	5	700	0.60		

This student has taken the MAP exam each year from 2009 through 2012. While his scale score for the content area was 700 in 2009 when he was a fifth grader, the equivalent z-score is .60, meaning the student's score was .60 standard deviations above the mean. This information from 2009 is carried over to 2010 (see blue arrow), and becomes paired with the student's 2010 score (.80, after standardizing). In this pair, the 2010 score is referred to as the *current year* score, and the 2009 score is referred to as the *previous year* score. The pairing process repeats itself twice more, with the student's 2010 and 2011 scores becoming paired with his 2011 and 2012 scores, respectively.

 ©2013 Missouri Department of Elementary and Secondary Education

To be clear, while all of the information presented in this example is required to prepare the data for analysis, the analysis itself uses only selected information from the score pairs in their preliminary state. Table 3 shows only the needed information.

Table 3: Preliminary Score Pairs for a Hypothetical "Student A" (Reduced)

Current Year	Grade Level (Current Year)	MAP Scale Score Expressed as a Z-Score (Current Year)	MAP Scale Score for Previous Year Expressed as a Z-Score
2012	8	1.30	1.10
2011	7	1.10	0.80
2010	6	0.80	0.60

Notice the fourth row, outcome year 2009, is eliminated because it is not paired with an exam score from the previous year—the model only uses data from the most recent *three* outcome years, i.e., years for which student scores are being predicted, at a time. *Outcome year* is synonymous with the term "current year" as it is used in Tables 2 and 3. The 2009 score still counts because it is one half of the 2010 score pair, however. In this way the model incorporates *four* years of assessment results.

Once the data from all students are arranged into score pairs, the pairs are grouped by LEA based on where the student was tested during the current year. Then, the LEA's average mobility and average prior-year exam score in both English/Language Arts and Mathematics are calculated. Average mobility is the percentage of students in the *current* year who were in-building less than a full academic year, and the average prior-year exam score is the average of the *previous year* test scores for all students taking the MAP test at the LEA in the *current year*.

The following example shows how these averages are calculated for a hypothetical LEA that tested three students in 2012, including the "Student A" described in the earlier example:

Table 4: Calculating Hypothetical "Anytown R-V" LEA Averages for Outcome Year 2012

Current Year	Student ID	Grade Level (Current Year)	In Building Less than Full Academic Year	MAP Scale Score Expressed as a Z-Score (Current Year)	MAP Scale Score for Previous Year Expressed as a Z-Score
2012	Student A	8	No	1.30	1.10
2012	Student B	4	No	1.05	0.80
2012	Student C	7	Yes	0.65	0.80
			33.3% Mobile		$= \frac{(1.10 + 0.80 + 0.80)}{3} = 0.90$

 ©2013 Missouri Department of Elementary and Secondary Education

These averages flow back to the score pairs with a matching "current year." For example, the Student A described earlier would have a complete score pair for outcome year 2012 that now includes the following:

Table 5: Complete Score Pair for a Hypothetical "Student A" for Outcome Year 2012

Current Year	Grade Level (Current Year)	MAP Scale Score Expressed as a Z-Score (Current Year)	MAP Scale Score for Previous Year Expressed as a Z-Score	In Building Less than Full Academic Year (Current Year)	LEA Percent Mobile	Average Prior-Year Exam Score
2012	8	1.30	1.10	No	33%	0.90

LEA averages are computed for each of the outcome years included in the model and flow back to each of the score pairs with a matching outcome year, completing the compilation of data.

Step 2: Doing the Math

At the core of the Missouri Growth Model is a statistical method called "two-stage regression." The key word here is "stage" because there are two very distinct operational phases in the analysis. In the first stage, the relationship between student test scores and a set of predictors is established. In this context, a *predictor* is a quantity or characteristic that has some ability to explain student test scores. The stage-one model uses the following predictors:

1. The student's prior-year assessment scores;
2. An indication of whether or not the student was in the building in which she took the test in the current year for a full academic year;
3. The average prior-year exam score for the LEA;
4. The LEA's average mobility

While not explicitly stated in the "Student A" and "Anytown R-V" examples, prior-year assessment scores at both the student and LEA levels are taken from English/Language Arts as well as from Mathematics, where possible. This is true even though a separate model is used for each content area. When predicting Mathematics scores, prior-year Mathematics scores are *always* required, and when predicting English/Language Arts scores, prior-year English/Language Arts scores are *always* required. When these *same-subject* predictor scores are missing, a score pair cannot be created. When only an *off-subject* score is missing, however, the model still creates a score pair. In this case, the average test score for the subject area, year, and grade in which the student should have been tested are set to the state average, and a special code is set indicating the score pair had a missing off-subject score.

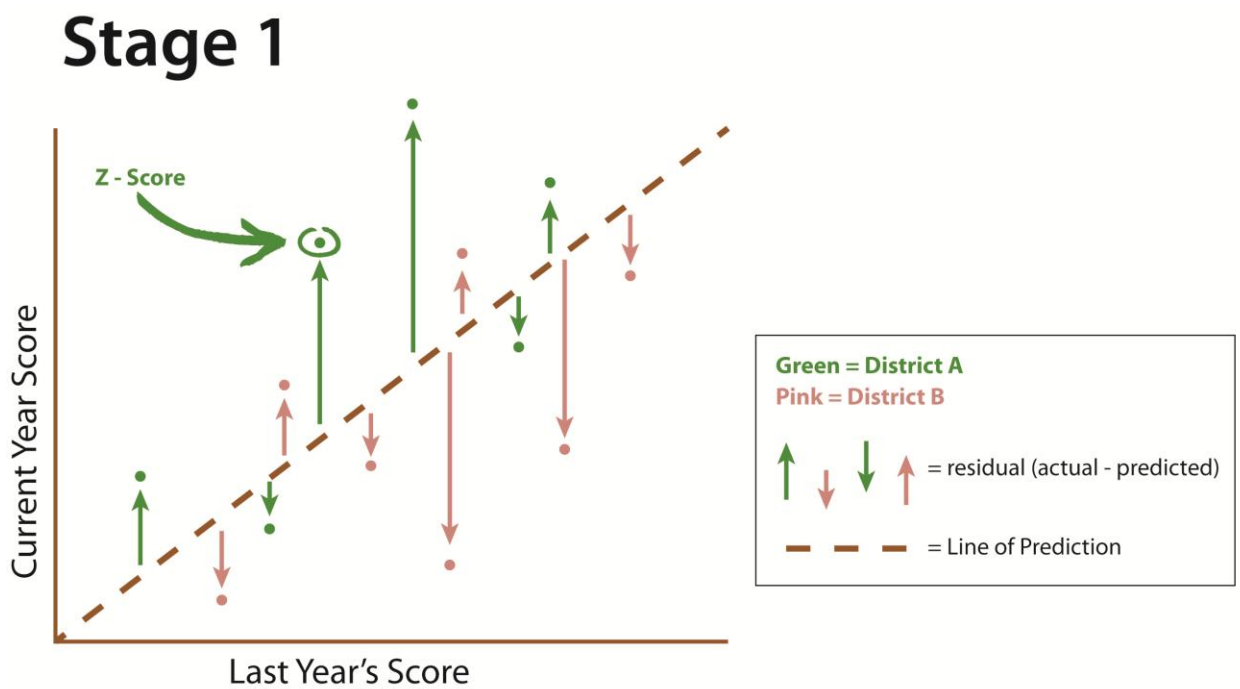
Since the presence of missing data may help explain student test scores, the model also includes the following predictors:

5. An indication of whether or not the student's off-subject, prior-year test score was missing;
6. The percentage of students with missing off-subject, prior-year scores

 ©2013 Missouri Department of Elementary and Secondary Education

A series of mathematical operations (automated by SAS, a statistical software package) uses information from these predictors to generate an equation relating the predictors to student test scores. The equation is of the form $y = x + z$, where y is the actual score and x and z are variables representing the predictors. For every score pair, the actual score (y) for the outcome year is predicted and the result is saved. Then, the predicted score for the outcome year is compared to the actual score to generate a **residual**.

Figure 1



The "residual" is the difference between the actual score and the predicted score. Residuals can be greater than zero. When they are greater than zero, they are an indication that the model underestimated the student's score—in other words, the actual score was higher than predicted. The opposite is true for scores less than zero—they indicate the student scored lower than the model predicted.

Step 3: Generating the Growth Measure for MSIP 5

Growth in MSIP 5 Standard 1

Once the residual is generated for each score pair, "stage one" is complete. At this point, it is possible to group the residuals any number of ways and begin to draw conclusions. In the context of MSIP 5 Standard 1, residuals are grouped by LEA. The score pairs assigned to a given LEA are determined on the basis of where the score was recorded for the outcome year. Note that if a student moves out of the LEA at any point during the three years included in the model, the student still contributes a score pair to the LEA as long as the student was tested there for any of the outcome years. More specifically, one score pair per outcome year tested at the LEA would be included in the LEA's growth calculation. If a student enrolls in the LEA mid-year and leaves the

 ©2013 Missouri Department of Elementary and Secondary Education

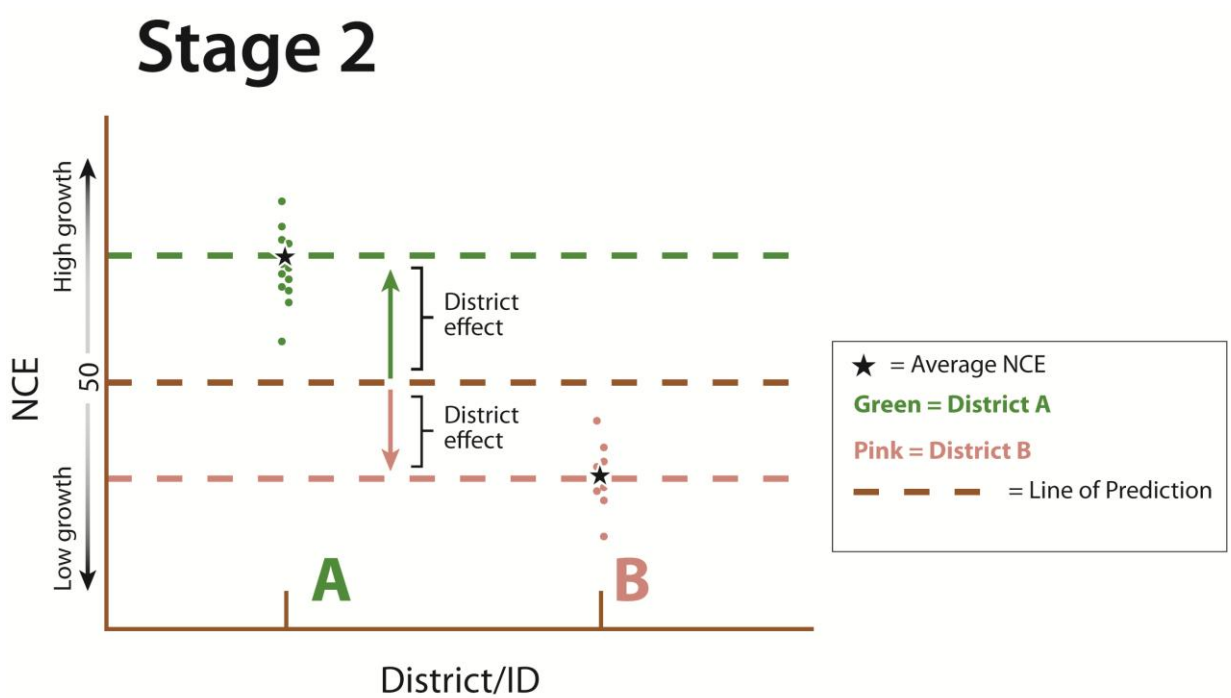
LEA prior to testing, the "sending" LEA will not be held accountable for the student's test score at the "receiving" LEA, but the receiving LEA will.

In "stage two" of the modeling procedure, the residuals grouped by LEA are averaged. While average residuals can be interpreted in z-score units—the scale of measure that results from the z-score conversion described in Step 1—for MSIP 5 purposes, a simple transformation is used to make the number easier to understand for the non-statistician. This transformation results in a **normal curve equivalent**, or NCE, and is calculated as follows:

NCE = 50 + (21.06 x [raw growth effect estimate])[1]

The NCE is a way of expressing a quantity and *resembles* a percentile. NCEs provide a way of comparing LEAs with similar prior achievement. NCEs around 50 indicate a contribution to student growth that is more or less typical, and that score predictions were generally met without either beating them or falling short. If the NCE is above 50, it means the students of a given LEA, viewed in the aggregate, outperformed predictions; if the NCE is below 50, it means the students of a given LEA—again, viewed in the aggregate—fell short of predictions. The graphic shown below highlights the average residual for two LEAs, relative to the default assumption that the LEAs met model predictions exactly:

Figure 2



Note that student-level residuals can also be converted from z-score units to NCE units.

[1] The multiplier 21.06 is used because it allows an NCE of 99 to represent the 99th percentile of the normal curve.

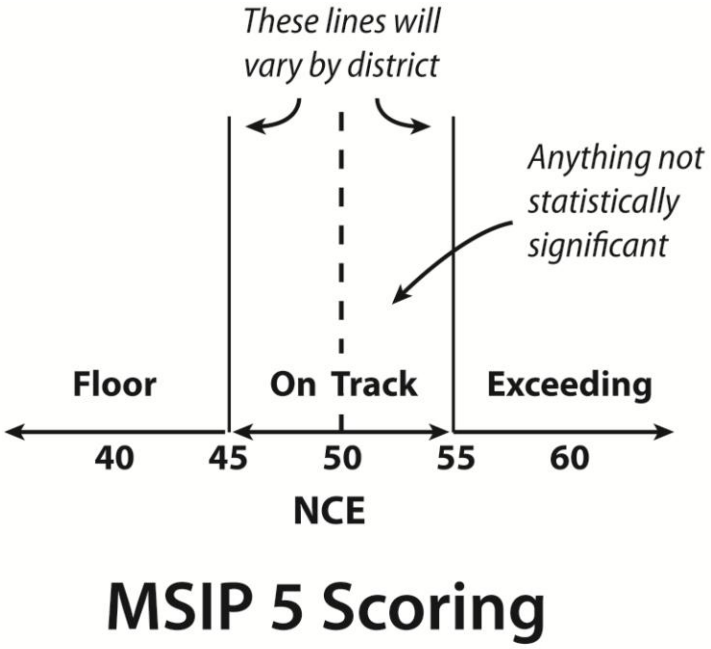
 ©2013 Missouri Department of Elementary and Secondary Education

Once the growth measures have been generated for each LEA and converted into NCE units, they must be tested for **statistical significance** prior to being used in accountability decisions. Statistical significance refers to the ability to make an impartial judgment, based purely on the data, about whether two quantities are different. The following concepts will help explain why results may or may not be significant—

- If the student-level residuals are generally very close to 50, it means the model is unable to find much to differentiate the LEA's impact on growth from the state as a whole because assessment scores matched statewide trends. At the LEA level, an average residual with an NCE very close to 50 may be evidence of this effect.
- The variability in students' test scores affects statistical significance. If students' test scores are "all over the map," the growth measure is less likely to be statistically significant.
- The number of score pairs attributed to an LEA affects statistical significance. The fewer pairs that are available to the model, the less likely the growth measure is to be statistically significant.

With both an NCE and an indication of whether the growth measure is statistically significant, an accountability decision for the Standard 1 growth measure is generated. The figure below illustrates how that decision is made:

Figure 3



Growth in MSIP 5 Standard 2

For MSIP 5 Standard 2, the focus shifts to a population of students known as the **super subgroup**. The super subgroup is comprised of historically low-achieving groups and includes students who are Black/African American, Hispanic, Free or Reduced Lunch, have disabilities, or are English

 ©2013 Missouri Department of Elementary and Secondary Education

language learners. The Standard 2 growth measure is designed to hold LEAs accountable for their efforts to close the achievement gap for these students.

While it is very important to help all students reach their full potential, achievement gaps will always remain if all students were to learn at the same pace. To demonstrate this, consider two groups with different levels of average achievement at a given time point. Mark these two starting points, one lower than the other. Mark two end points somewhere to the right as well. Then, draw a straight line between the starting points and the ending points to represent the average growth for the two groups, making sure the lines are parallel. When the lines appear to be parallel, notice that the end points are as far apart as the starting points. While both groups improved, an issue of basic fairness remains because one group is more prepared than the other in the end.

The data indicate that the super subgroup is at this lower starting point, so closing the gap will require more aggressive learning gains for these students than would be required for students who are not in the super subgroup.

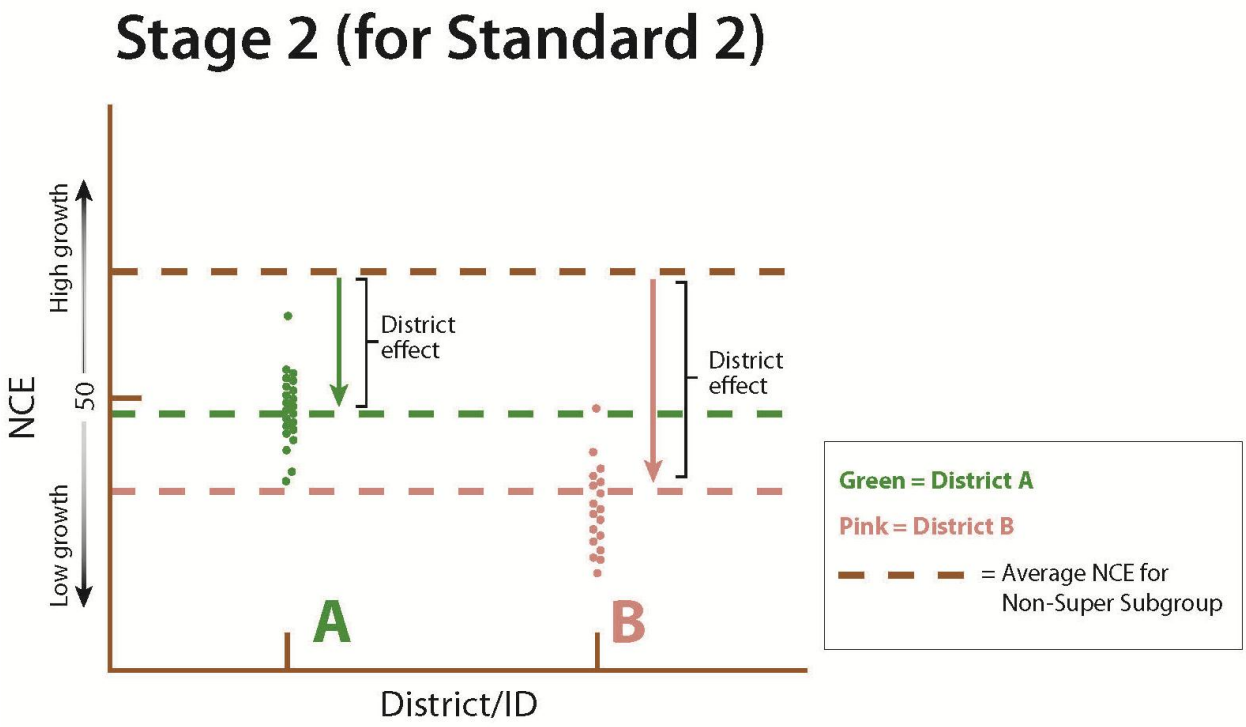
With this conceptual framework in mind, refer back to the conclusion of "stage one," discussed earlier in this document. By this point, residuals have already been calculated for each score pair and are ready to be examined for patterns. For Standard 2, the pattern of interest centers on the residuals for score pairs tied to students who make up the super subgroup. Each LEA's super subgroup is identified, and the residuals specific to those students are picked out from among all residuals generated for an LEA.

Figure 4

Then, in "stage two," the super subgroup residuals are averaged to generate a raw growth effect estimate. The estimate reflects the LEA's impact on assessment scores for this population relative to the state as a whole—i.e., all students.

 ©2013 Missouri Department of Elementary and Secondary Education

Figure 5



Finally, as was the case for Standard 1, the growth effect estimate is converted to an NCE and tested for statistical significance. However, whereas the object of Standard 1 is to look for any differences in the way specific LEAs impact assessment scores compared to statewide trends, for Standard 2, the object of the statistical test is to determine whether the LEA impacted the assessment scores of its super subgroup differently than the way the public education system as a whole impacted the scores of non-super subgroup students.

For Standard 2, when the growth measure is expressed as an NCE above 50 *and* is statistically significant, it means that the LEA is contributing enough to the education of its most disadvantaged students that, in concert with similar successes across the state, there is hope of closing the achievement gap in Missouri. On the other hand, if the NCE is less than 50 *and* statistically significant, it means that the LEA is widening the achievement gap. When the result is not statistically significant, it means that the super subgroup is learning at about the same pace as everyone else—a good start, but not enough to really contribute to closing the achievement gap.

Once the growth measure is calculated and tested for statistical significance, the result is evaluated using the same scoring framework as Standard 1 (see Figure 3). Further information is available from the *Comprehensive Guide to the Missouri School Improvement Program*.

 ©2013 Missouri Department of Elementary and Secondary Education

Incorporating Additional Assessments

As indicated previously, only the MAP assessment results for grades 3 through 8 in English Language Arts and Mathematics are currently included in LEA and school growth estimates. This limitation is important to note because most LEAs also serve grades 9-12, and all LEAs teach more subjects than just English Language Arts and Mathematics. As such, state officials are committed to exploring ways to adapt the Missouri Growth Model to incorporate additional assessments.

The Department of Elementary and Secondary Education (DESE) continues to collaborate with technical experts to develop viable options for generating growth estimates that include the required end-of-course (EOC) assessments. For example, extensions of the model to include Algebra I scores are currently under active consideration by state officials, and more studies will follow to investigate the use of the other required assessments—i.e., English II, Biology, and American Government.

More generally, as new exams, such as the Mathematics and English end-of-high school (EOHS) exams, are mandated as part of the state assessment program, DESE will conduct analysis to explore ways to include those exams in the Missouri Growth Model. The state's assessment timeline is available from http://dese.mo.gov/divimprove/assess/documents/asmt-eoc-hs-assess-plan.pdf

 ©2013 Missouri Department of Elementary and Secondary Education