Florida’s Value-Added Model

Overview of the Model Used to Measure Student Learning Growth on FCAT Reading and Mathematics

Juan Copa, Deputy Commissioner
Accountability, Research, and Measurement
Florida Department of Education

Florida Educational Research Association (FERA) Annual Meeting
November 14, 2012
Gainesville, FL
New Standard for Educator Evaluations in Florida

- For more than a decade, Florida has used data to measure student and school performance, and inform policy that has led to improved student outcomes.

- In 2011, the Florida Legislature enacted legislation to transform the way Florida teachers and school leaders are evaluated.
  - Placing the emphasis on data and student learning growth.
New Standard for Educator Evaluations in Florida

As set forth in the *Student Success Act* and *Race to the Top*, teacher evaluations are now:

- Designed to support effective instruction and student learning growth

Results from evaluations are to be used:

- When developing district and school level improvement plans
- To identify professional development and other human capital decisions for instructional personnel and school administrators
New Standard for Educator Evaluations in Florida

To support those objectives, the law sets forth that teacher evaluations are to be based on sound educational principles and contemporary research in effective practices in three major areas:

1. The performance of students
2. Instructional practice
3. Professional and job responsibilities
New Standard for Educator Evaluations in Florida

Performance of Students. At least 50% of a performance evaluation must be based upon data and indicators of student learning growth assessed annually and measured by statewide assessments or, for subjects and grade levels not measured by statewide assessments, by district assessments as provided in s. 1008.22(8), F.S.

- Section 1012.34(3)(a)1., Florida Statutes
New Standard for Educator Evaluations in Florida

- The performance of students represents 50% of a teacher’s evaluation, with performance based on student learning growth.
- To meet the above requirement, the development of a fair and transparent measure of student growth is essential.
- This presentation focuses on what that measure is, how it was developed, and how it is being used in teacher evaluations in Florida.
The Measure: Value-Added Analysis

- A value-added model measures the impact of a teacher on student learning, by accounting for other factors that may impact the learning process.

- These models do not:
  - Evaluate teachers based on a single year of student performance or proficiency (status model), or
  - Evaluate teachers based on simple comparison of growth from one year to the next (simple growth).
The difference between the predicted performance and the actual performance represents the **value-added** by the teacher’s instruction.

The predicted performance represents the level of performance the student is expected to demonstrate after statistically accounting for factors through a value-added model.
Advantages of Value-Added Models

- Teachers teach classes of students who enter with different levels of proficiency and possibly different student characteristics.
- Value-added models “level the playing field” by accounting for differences in the proficiency and characteristics of students assigned to teachers.
- Value-added models are designed to mitigate the influence of differences among the entering classes so that schools and teachers do not have advantages or disadvantages simply as a result of the students who attend a school or are assigned to a class.
Florida’s Value-Added Model Developed by Florida Educators

- The Department convened a committee of stakeholders (Student Growth Implementation Committee – or SGIC) to identify the type of model and the factors that should be accounted for in Florida’s value-added models.

- The SGIC’s recommended model was fully adopted by the Commissioner with no additions, deletions, or changes.

- To provide technical expertise, the Department contracted with the American Institutes for Research (AIR) to help the SGIC develop the recommended model that was adopted.
Florida’s Value-Added Model Developed by Florida Educators

- After exploring eight different types of value-added models, the SGIC recommended a model from the class of covariate adjustment models

- This model begins by establishing expected growth for each student:
  - Based on historical data each year
  - Represents the typical growth seen among students who have earned similar test scores the past two years, and share the other characteristics identified by the committee
Factors Identified by the SGIC to “Level the Playing Field”

To isolate the impact of the teacher on student learning growth, the model developed by the SGIC and approved by the Commissioner accounts for:

- Student Characteristics
- Classroom Characteristics
- School Characteristics
Factors Identified by the SGIC to “Level the Playing Field”

Student Characteristics:
- Up to two prior years of achievement scores (the strongest predictor of student growth)
- The number of subject-relevant courses in which the student is enrolled
- Students with Disabilities (SWD) status
- English Language Learner (ELL) status
- Gifted status
- Attendance
- Mobility (number of transitions)
- Difference from modal age in grade (as an indicator of retention)

Classroom characteristics:
- Class size
- Homogeneity of students’ entering test scores in the class
Factors Identified by the SGIC to “Level the Playing Field”

The model recognizes that there is an independent factor related to the school that impacts student learning – a school component.

- Statistically is simply the factors already controlled for in the model measured at the school level by grade and subject
- May represent the impact of the school’s leadership, the culture of the school, or the environment of the school on student learning
Factors Identified by the SGIC to “Level the Playing Field”

SGIC decisions on the use of the school component

- The SGIC decided to include 50% of the school component in the measurement of the teacher’s effectiveness
- By attributing a portion of the school component to the teacher in the measurement of her effectiveness, one recognizes that the teacher contributes somewhat to the overall school component, but there are factors imbedded in that component that are beyond his/her direct control and that he/she should not directly be held accountable for
Florida’s Value-Added Model

- The value-added model is one part of a multi-faceted teacher evaluation system.
- The model was developed independently by a committee of Florida educators.
- The model accounts for factors outside the teacher’s control and does not rely on a single year of data or single test score.
- The development process is an on-going process.
  - The SGIC, Department, and AIR will continue to analyze the value-added model and seek feedback to make adjustments, if necessary.
Review of State-Level Results

The next series of slides will review:

- The model fit in 2012 compared to 2011
- Comparison of the Reading and Mathematics VAM scores by Grade
- Impact analyses showing relationship of VAM score with various classroom characteristics
Model Fit: Comparison of R-Square Values for Reading

The R-Square values – which are an indicator of how well the model fits the data – are high and virtually equivalent to the values observed with the 2011 model.

The model behaves the same using FCAT 2.0 data, as it did using FCAT data.
Model Fit: Comparison of R-Square Values for Math

The R-Square values – which are an indicator of how well the model fits the data – are high and virtually equivalent to the values observed with the 2011 model.

The model behaves the same using FCAT 2.0 data, as it did using FCAT data.
Teacher VAM Score by Grade (Reading)

VAM scores by grade show no particular advantage for any given grade.
Teacher VAM Score by Grade (Math)

VAM scores by grade show no particular advantage for any given grade
Impact Analyses for VAM Scores

- The following slides show the relationship of the teacher VAM score with various classroom characteristics.
- The observed score correlations for each characteristic are reported after the scatter plots.
- In all cases, the correlations are negligible indicating no advantages or disadvantages for any group of students.
Correlation of Teacher VAM Score and Percent Students with Disabilities

**Relationship of Teacher VAM with Percent SWD in Class (Reading)**

**Relationship of Teacher VAM with Percent SWD in Class (Math)**
Correlation of Teacher VAM Score and Percent English Language Learners

Relationship of Teacher VAM with Percent ELL in Class (Reading)

Relationship of Teacher VAM with Percent ELL in Class (Math)
Correlation of Teacher VAM Score and Percent Economically Disadvantaged

Relationship of Teacher VAM with Percent ED in Class (Reading)

Relationship of Teacher VAM with Percent ED in Class (Math)
Correlation of Teacher VAM Score and Percent Gifted

Relationship of Teacher VAM with Percent Gifted in Class (Reading)

Relationship of Teacher VAM with Percent Gifted in Class (Math)
Correlation of Teacher VAM Score and Percent Non-White Students

Relationship of Teacher VAM with Percent Non-White in Class (Reading)

Relationship of Teacher VAM with Percent Non-White in Class (Math)
### Observed Correlations with Teacher VAM Score

<table>
<thead>
<tr>
<th></th>
<th>Teacher VAM (Reading)</th>
<th>Teacher VAM (Math)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent SWD</td>
<td>0.009</td>
<td>-0.03</td>
</tr>
<tr>
<td>Percent ELL</td>
<td>0.06</td>
<td>0.004</td>
</tr>
<tr>
<td>Percent ED</td>
<td>-0.08</td>
<td>-0.08</td>
</tr>
<tr>
<td>Percent Gifted</td>
<td>0.01</td>
<td>0.003</td>
</tr>
<tr>
<td>Percent Non-White</td>
<td>0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>Mean Prior</td>
<td>-0.08</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
The Use of the Value-Added Scores

- School districts in Florida were required to include data from the statewide value-added model in their 2011-12 teacher and school administrator evaluations.

- The statewide model was to be used for teachers teaching courses associated with FCAT Reading and Mathematics.

- Other measures (e.g., district developed/selected assessments, learning targets, school-wide results) could be used for non-FCAT subject area teachers.

- Districts had flexibility on how to use the value-added results in evaluations this year – including authority to set their own “cut points”.
The Use of the Value-Added Scores

- Value-added scores may need to be **aggregated**
  - Models are run by grade level, subject, and year
  - Aggregation allows those scores to be combined into one measure for use in an evaluation
  - Aggregation across years, also decreases the standard error (variability) in the measure; improving the accuracy of the score

- Scores then need to be **classified** into performance categories (highly effective, effective, needs improvement/developing, and unsatisfactory)

- Districts needed to decide on how to aggregate teacher value-added scores and classify teacher value-added scores for use in evaluations this school year

- Options were provided
Aggregation and Classification Options

Three different options for aggregating and classifying the teacher scores were provided to districts:

- **Option 1**: Transform the scores into a common metric (e.g., a proportion of a year’s average growth), combine, and then classify

- **Option 2**: Classify score for a teacher (grade/subject/year) into a performance category and then aggregate the classifications to create a single measure

- **Option 3**: Compute percentage of students in a teacher’s class who did better than statistically predicted and then classify based on that percentage
Aggregation and Classification Options

- **Option 1 Description:** Value-added scores (which represent the number of scale points, on average, students performed above/below expected) are transformed into a proportion of a year’s average growth. Those transformed scores are averaged together to produce one measure, expressing the proportion above or below average a teacher’s students grew. Teachers are classified into 4 categories based on those proportions.

- **Option 2 Description:** A teacher teaches multiple grade levels and subjects across multiple years. Those scores are independently classified on a scale of 1 to 4, and then those classifications are averaged into one value (analogous to a GPA). The overall score is then classified into one of the 4 categories.

- **Option 3 Description:** Whether or not the teacher’s students met or outperformed expectations is used to classify teachers, as measured by the quantity (the percentage) of students that met or exceeded expectations. How far students moved (or regressed) is not taken into account.
## Example Value-Added Data Provided

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Teacher Effect</th>
<th>School Component</th>
<th>Teacher VAM Score</th>
<th>Teacher VAM Score Standard Error</th>
<th>Percent of Students Meeting Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher X</td>
<td>-10</td>
<td>-6</td>
<td>-13</td>
<td>10</td>
<td>42%</td>
</tr>
<tr>
<td>Teacher Y</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>50%</td>
</tr>
</tbody>
</table>

- On average, Teacher X’s students missed their expectations by 13 scale score points
- On average, Teacher Y’s students exceeded their expectations by 6 points
- By accounting for the school component, the impact of the low growth school (Teacher X’s school) and the high growth school (Teacher Y’s school) are not fully absorbed in each teacher’s score
  - A model that does not account for a school component would fully attribute the school’s impact to the teacher
  - In such a model, Teacher X’s score would be deflated due to the low growth school and Teacher Y’s score would be inflated due to the high growth school
Use of the Standard Error in Classification of Teachers

- Using the standard error can assist in increasing the accuracy of classification decisions.
- Applying this concept, “cut points” can be established based on the teacher’s value-added score to determine dividing lines between the 4 performance categories outlined in law.
- Then some degree of the standard error can be applied to the teacher’s score to determine with some or a high degree of statistical certainty that a teacher’s score falls in one of the 4 categories.
- The aim is to use the most accurate method possible to identify a teacher’s contribution to student learning; using the standard error in classification decisions helps in that effort.
Value-Added Model Data: Example Classification Method, Applicable to Options 1&2

<table>
<thead>
<tr>
<th>Category</th>
<th>Standard (For Example – Average Teacher VAM score of 0)</th>
<th>Standard Error Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Effective</td>
<td>Above</td>
<td>2 SE (95% confidence)</td>
</tr>
<tr>
<td>Effective</td>
<td>Above/Below</td>
<td>None</td>
</tr>
<tr>
<td>Needs Improvement</td>
<td>Below</td>
<td>1 SE (68% confidence)</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>Below</td>
<td>2 SE (95% confidence)</td>
</tr>
</tbody>
</table>
Value-Added Model Data: Classification Method – Visual Example

- **VAM Score**
- **Highly Effective**
- **Effective**
- **Needs Improvement**
- **Unsatisfactory**

- **68% Confidence Interval** (+/- 1 a standard error)
- **95% Confidence Interval** (+/- 2 standard errors)
Value-Added Model Data: Classification Options Explained

- “Highly Effective” represents that the score falls above the standard for evaluation, with a high degree of confidence – 95%
- “Needs Improvement” represents that the score falls below the standard for evaluation, with some degree of statistical confidence – 68%
- “Unsatisfactory” represents that the score falls below the standard for evaluation, with a high degree of statistical confidence – 95%
- If the score falls above or below the standard for evaluation, but one cannot conclude that the score exceeds or misses the bar with any acceptable degree of statistical confidence, the score defaults to “Effective”
**Example Data Applied to Example Classification Method**

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Teacher VAM Score</th>
<th>Teacher VAM Score Standard Error</th>
<th>Performance Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher X</td>
<td>-13</td>
<td>10</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>Teacher Y</td>
<td>6</td>
<td>9</td>
<td>Effective</td>
</tr>
</tbody>
</table>

- Teacher X’s performance would be classified as “Needs Improvement” since her nominal score (-13) falls below the standard of zero, with 68% statistical confidence – but not with 95% statistical confidence.

- Teacher Y’s performance would be classified as “Effective” since her nominal score (6) falls above zero, but one cannot conclude with any acceptable degree of statistical confidence that the score falls above or below the standard; in the absence of compelling statistical evidence, the judgment errs on the side of caution.
Timeline of Activities

- **March – June 2011**: Model for FCAT developed and selected

- **August 2011**: Historical value-added data (for 2008-09 to 2010-11) shared with school districts

- **2011-12 School Year**: School districts implementing new educator evaluation systems

- **August - September 2012**: Provide 2011-12 value-added results to school districts for inclusion in their educator evaluation systems

- **December 2012**: Report baseline teacher evaluation results on a statewide basis
Development of Other Value-Added Models

- Model for Algebra I EOC developed, not yet approved
- Exploring the development of sample models for district use using AP/IB data, SAT-10 data, and FAIR (state’s progressing monitoring assessment in reading) data
- During the 2012-13 school year, model development for Florida Alternate Assessment, Biology EOC, and Geometry EOC will occur
- Guidance to school districts on building models for locally determined assessments will be made available, as well, by the conclusion of *Race to the Top*
Contact and Resources

- Information on Florida’s Value-Added Model, including presentations, meeting transcripts, and meeting recordings can be accessed at:
  - http://www.fldoe.org/committees/sg.asp

- Materials include:

- Contact me at:
  - Juan.Copa@fldoe.org