FOIL – Educator Evaluation
November 13, 2014

Brian Dassler and Jason Gaitanis

@brian.dassler

FLORIDA DEPARTMENT OF EDUCATION
fldoe.org

"We will not evaluate ourselves to greatness but we need evaluation to help coach and support teachers to greatness."

-Commissioner Pam Stewart

Our goal is not, and must not be, evaluation for evaluation's sake. Our goal must be evaluation for the purpose of growth – for providing educators with honest and actionable information that supports their continued professional learning for the purpose of improved teaching and improved learning for all students.

Gallery Walk

Chart paper around the room invite you to indicate where your district is on two key questions

1. What progress has your district made in developing or identifying "local assessments"?

2. What progress has your district made in determining how the results from local assessments will be used for the purpose of teacher evaluation?

Our purpose is to be able to help you connect with others who may have an answer, or an approach, that meets a current need in your district.

Topics

- Brief Overview of Why We Have VAM

- Interpreting VAM

- VAM – Evaluation Comparative Analysis

- Responses to Guiding Questions

- Local Assessments

- Evaluation System Review

- The Future

Why Do We Have VAM?

Three Components of Teacher Evaluations

Section 1012.34, F.S. requires that teacher evaluations be based on sound educational principles and contemporary research in effective practices in three major areas:

- The Performance of Students

- Instructional Practice

- Professional and Job Responsibilities

Student Performance

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

*Note that this can be lowered to 40% if fewer than 3 years of data are available for the teacher.

Approved Models

Currently, the following **VAM** models are approved, and therefore required to be used, for the following subjects and grades:

Reading (4th, 5th, 6th, 7th, 8th, 9th, 10th)
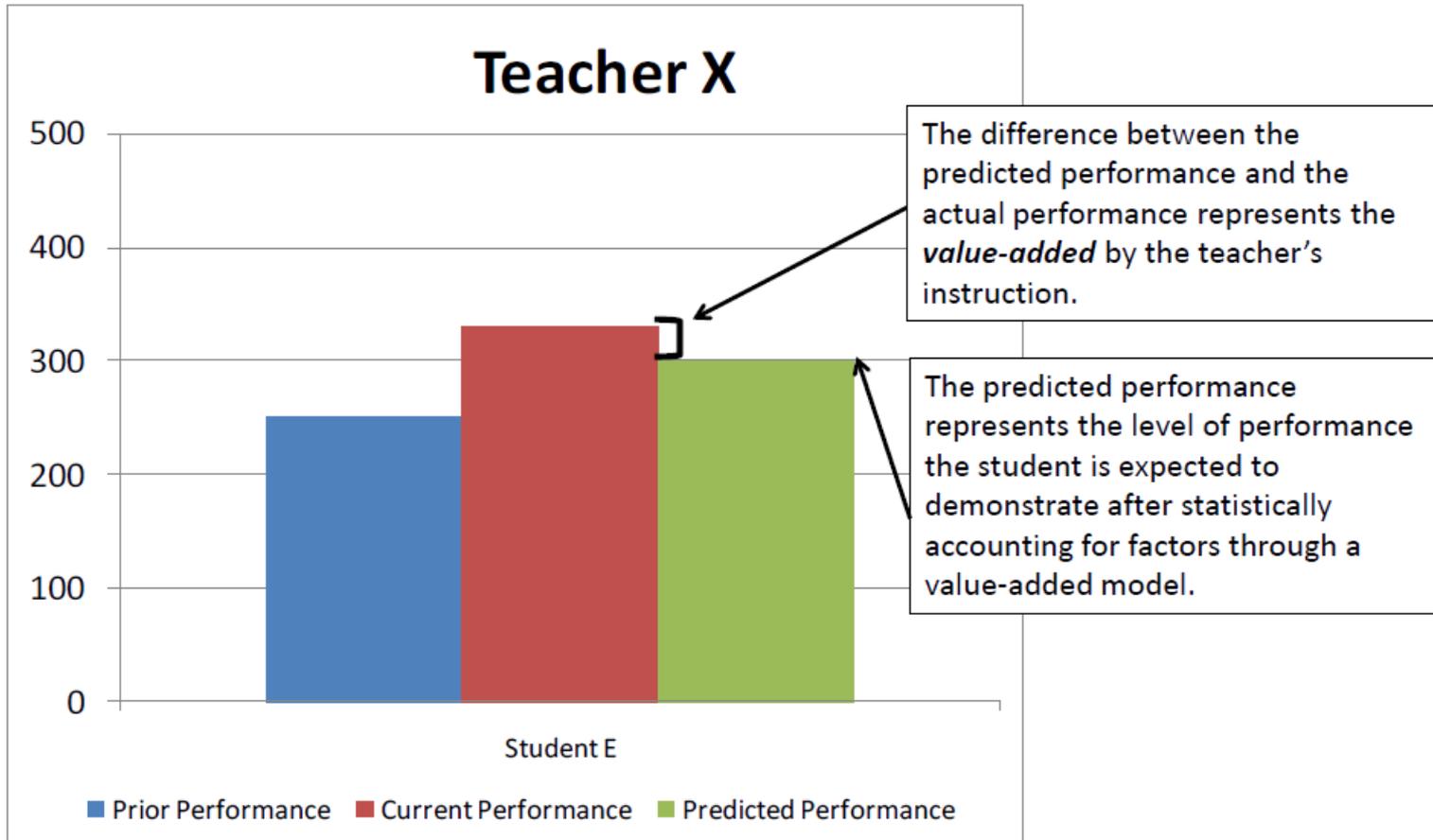Mathematics (4th, 5th, 6th, 7th, 8th)
Algebra 1 (9th)

Value-Added Models (VAM) Description

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, or

- Evaluate teachers based on simple comparison of growth from one year to the next

VAM explores the relationship among prior, predicted, and observed performance

Teacher X

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.

Student E

■ Prior Performance ■ Current Performance ■ Predicted Performance

Model Covariates

The following covariates are included in the VAM model and are used to determine the similarity of characteristics among students when calculating expected leaning growth:

- The number of subject-relevant courses in which the student was enrolled
- Up to two prior years of achievement scores
- Students with Disabilities (SWD) status
- English language learner (ELL) status
- Gifted status
- Attendance
- Mobility (number of transitions between schools)
- Difference from modal age in grade (as an indicator of retention)
- Class size
- Similarity (Homogeneity) of prior test scores among students in the class

Raw VAM Scores

Raw VAM scores are reported in a way that translates into the number of points higher or lower on the developmental scale, on average, that students taught by that teacher, in that subject at that grade level during that year scored when compared to similar students statewide.

Aggregate VAM Scores

Aggregate VAM scores combine all courses and grade-levels for a teacher and standardize the score so it represents the *percentage* higher or lower than average that the teacher's students, on average, scored when compared to similar students statewide.

Aggregate VAM Scores - Continued

Aggregate VAM scores are also combined and reported across years. This is because statute requires that VAM comprise 50% of evaluations where 3 years of data are available, but only 40% if fewer than 3 years of data are available.

Standard Error

In order to draw appropriate, statistically valid conclusions about VAM scores, it is important to also consider an additional measure, called the standard error, or SE.

The standard error is a statistical metric that is used to describe variability. It is similar to a margin of error, and allows you to construct confidence intervals to express, with a certain degree of confidence, the range of possible VAM scores that would have resulted for that teacher had they been assigned a different, but similar, group of students. Each VAM score has its own standard error to go along with it.

16

VAM Classification Methodology
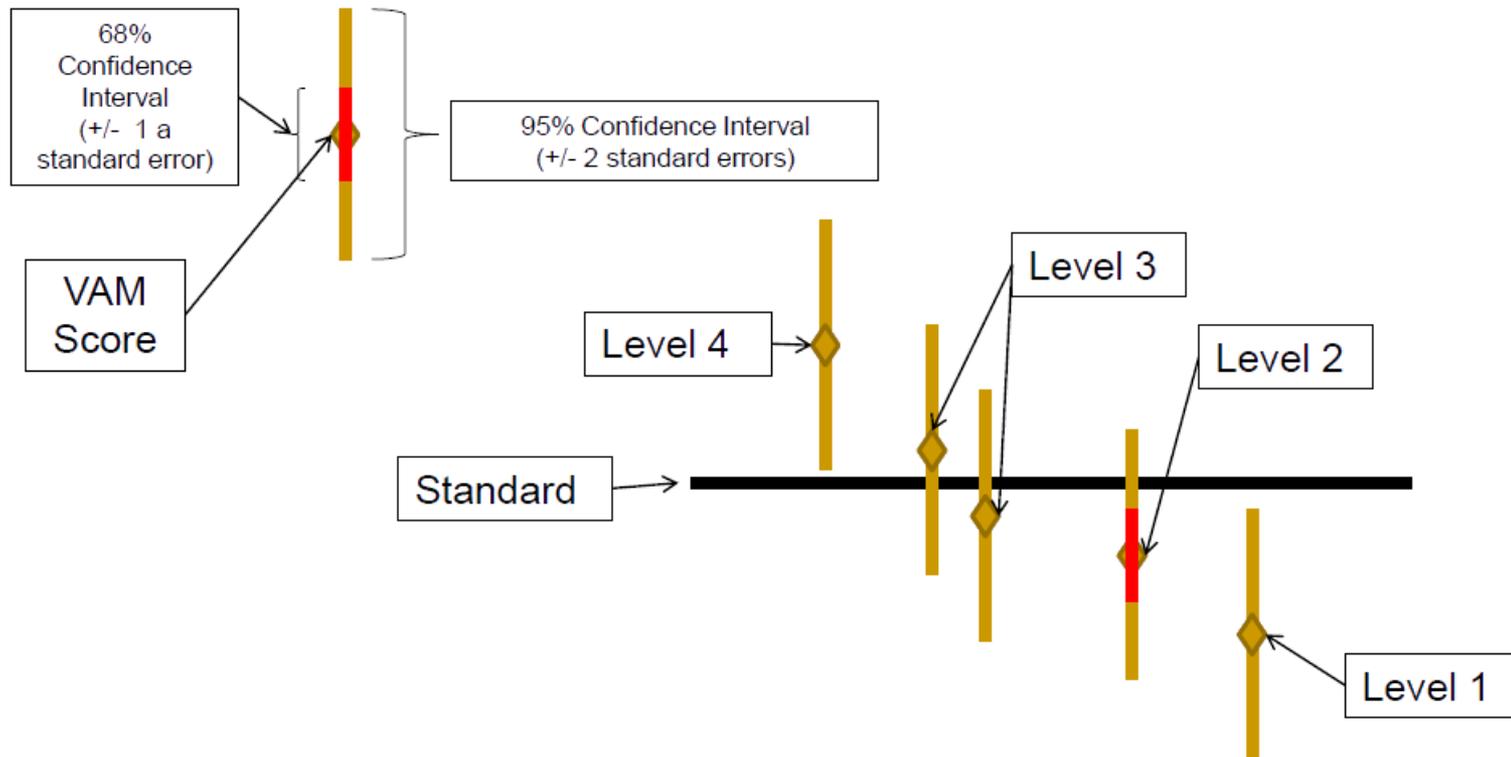
Percent of Students Meeting Expectations

The VAM model calculates an expected score for each student based on the average performance of other similar students on the assessment that year. The percent of students meeting expectations is computed by dividing the number of students whose actual score met or exceeded their expected score by the total number of students on the teacher's roster.
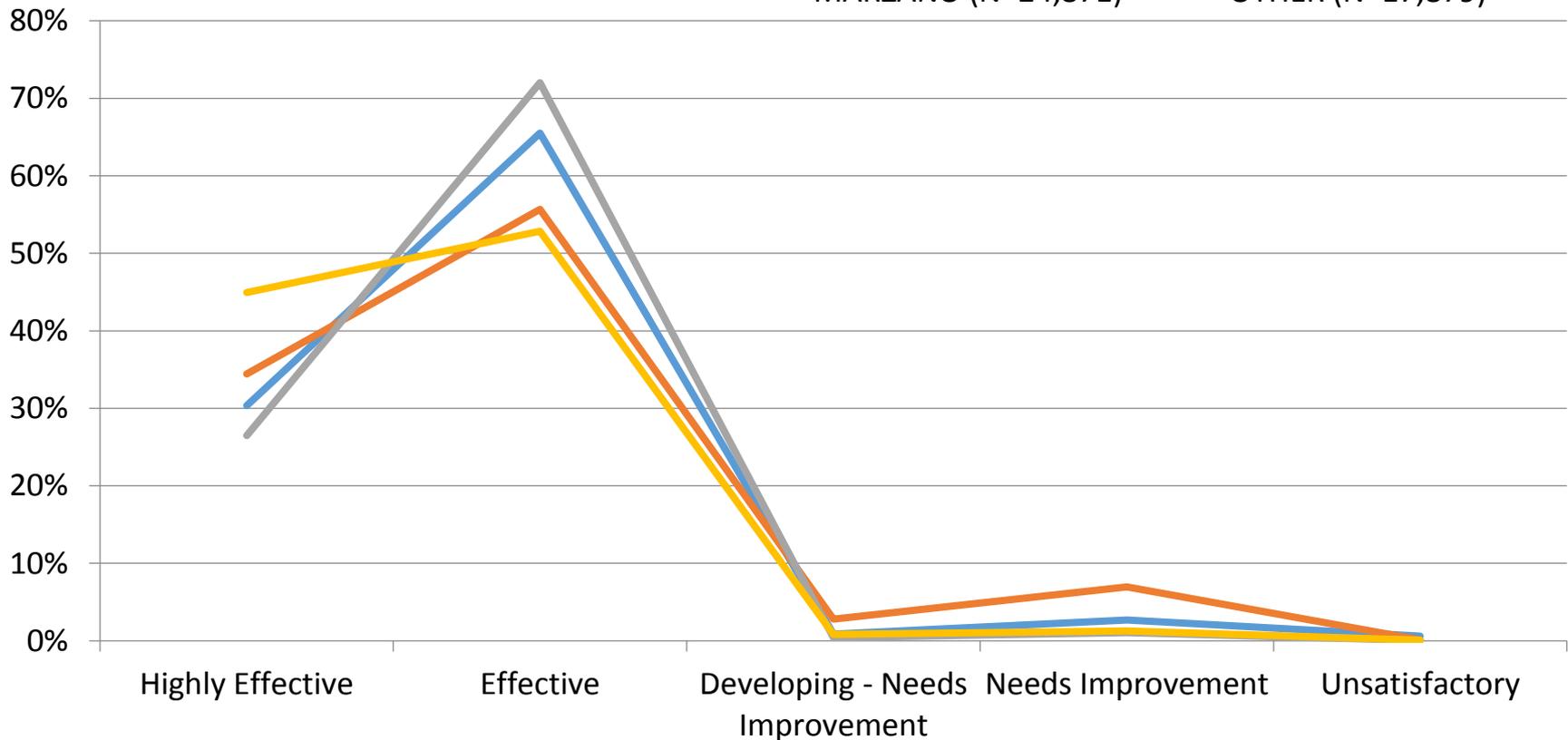
Comparison Between VAM and Evaluation Results

Statewide 2012-13 Classification Results (VAM Reading/ELA and Mathematics Teachers Only)
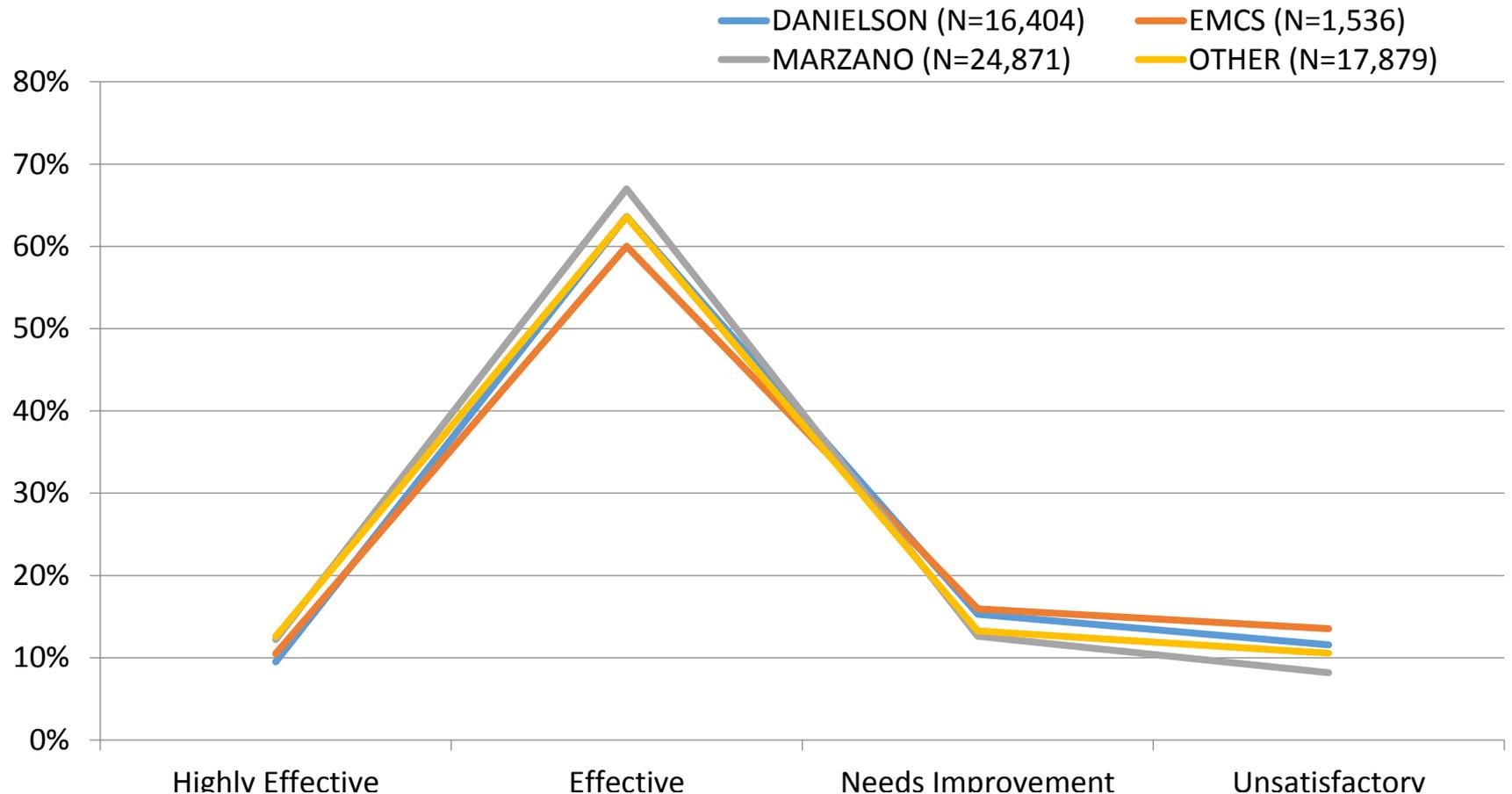
Statewide 2012-13 Final Evaluation Results (VAM Reading/ELA and Mathematics Teachers Only)


Legend:
- DANIELSON (N=16,404)
- EMCS (N=1,536)
- MARZANO (N=24,871)
- OTHER (N=17,879)

X-axis categories: Highly Effective, Effective, Developing - Needs Improvement, Needs Improvement, Unsatisfactory

Results of Possible Classification Methodology

Number	VAM FCAT 3 Year Aggregate COMBINED Score	VAM FCAT 3 Year Aggregate COMBINED Standard Error	95% Confidence Interval Lower Limit	68% Confidence Interval Lower Limit	68% Confidence Interval Upper Limit	95% Confidence Interval Upper Limit	FCAT Percent Meeting Expectation	VAM Classification	Evaluation Rating
1	0.511	0.131	0.252	0.380	0.642	0.770	66.2%	Highly Effective	Highly Effective
2	1.682	0.192	1.303	1.490	1.874	2.061	84.3%	Highly Effective	Highly Effective
3	-0.069	0.295	-0.712	-0.375	0.237	0.574	61.5%	Effective	Highly Effective
4	0.080	0.143	-0.202	-0.063	0.223	0.362	56.2%	Effective	Highly Effective
5	0.085	0.111	-0.134	-0.026	0.196	0.304	48.3%	Effective	Highly Effective
6	0.108	0.065	-0.020	0.043	0.173	0.236	57.4%	Effective	Highly Effective
7	0.121	0.123	-0.121	-0.002	0.244	0.363	47.5%	Effective	Highly Effective
8	0.218	0.170	-0.119	0.048	0.388	0.555	57.5%	Effective	Highly Effective
9	-0.268	0.281	-0.861	-0.556	0.020	0.325	38.9%	Effective	Effective
10	-0.204	0.206	-0.628	-0.413	0.005	0.220	46.2%	Effective	Effective
11	-0.179	0.186	-2.542	-0.517	0.159	2.184	0.0%	Effective	Effective
12	-0.139	0.238	-0.639	-0.382	0.104	0.361	42.1%	Effective	Effective
13	-0.129	0.343	-0.850	-0.480	0.222	0.592	42.1%	Effective	Effective
14	-0.128	0.255	-0.731	-0.401	0.145	0.475	87.5%	Effective	Effective
15	-0.115	0.199	-0.532	-0.318	0.088	0.302	35.0%	Effective	Effective
16	-0.111	0.188	-0.709	-0.335	0.113	0.487	75.0%	Effective	Effective
17	-0.073	0.120	-0.309	-0.193	0.047	0.163	44.1%	Effective	Effective
18	-0.063	0.214	-0.486	-0.277	0.151	0.360	46.0%	Effective	Effective
19	-0.027	0.201	-0.436	-0.230	0.176	0.382	38.2%	Effective	Effective
20	-0.017	0.271	-0.621	-0.300	0.266	0.587	54.5%	Effective	Effective
21	0.111	0.158	-0.202	-0.047	0.269	0.424	56.0%	Effective	Effective
22	0.135	0.308	-0.490	-0.175	0.445	0.760	45.9%	Effective	Effective
23	0.148	0.287	-0.472	-0.149	0.445	0.768	42.9%	Effective	Effective
24	-0.443	0.367	-1.209	-0.817	-0.069	0.323	33.3%	Needs Improvement	Effective
25	-0.361	0.260	-0.878	-0.621	-0.101	0.156	44.9%	Needs Improvement	Effective
26	-0.341	0.338	-1.007	-0.678	-0.004	0.325	39.4%	Needs Improvement	Effective
27	-0.287	0.160	-0.605	-0.447	-0.127	0.031	33.7%	Needs Improvement	Effective
28	-0.180	0.125	-0.438	-0.307	-0.053	0.078	32.0%	Needs Improvement	Needs Improvement
29	-0.229	0.106	-0.440	-0.335	-0.123	-0.018	34.7%	Unsatisfactory	Effective
30	-1.094	0.352	-1.810	-1.450	-0.738	-0.378	11.8%	Unsatisfactory	Needs Improvement
31	-0.531	0.171	-0.869	-0.702	-0.360	-0.193	35.9%	Unsatisfactory	Needs Improvement
32	-0.422	0.060	-0.540	-0.482	-0.362	-0.304	28.8%	Unsatisfactory	Needs Improvement
33	-0.403	0.086	-0.574	-0.489	-0.317	-0.232	23.6%	Unsatisfactory	Needs Improvement
34	-0.347	0.111	-0.569	-0.458	-0.236	-0.125	31.7%	Unsatisfactory	Needs Improvement
35	-0.231	0.081	-0.391	-0.312	-0.150	-0.071	36.2%	Unsatisfactory	Needs Improvement
36	-0.227	0.080	-0.385	-0.307	-0.147	-0.069	33.9%	Unsatisfactory	Needs Improvement
37	-0.198	0.083	-0.362	-0.281	-0.115	-0.034	43.4%	Unsatisfactory	Needs Improvement

Comparison Between VAM and Final Evaluation Results Using State VAM Classification Methodology - Reading/ELA and Mathematics Teachers Only
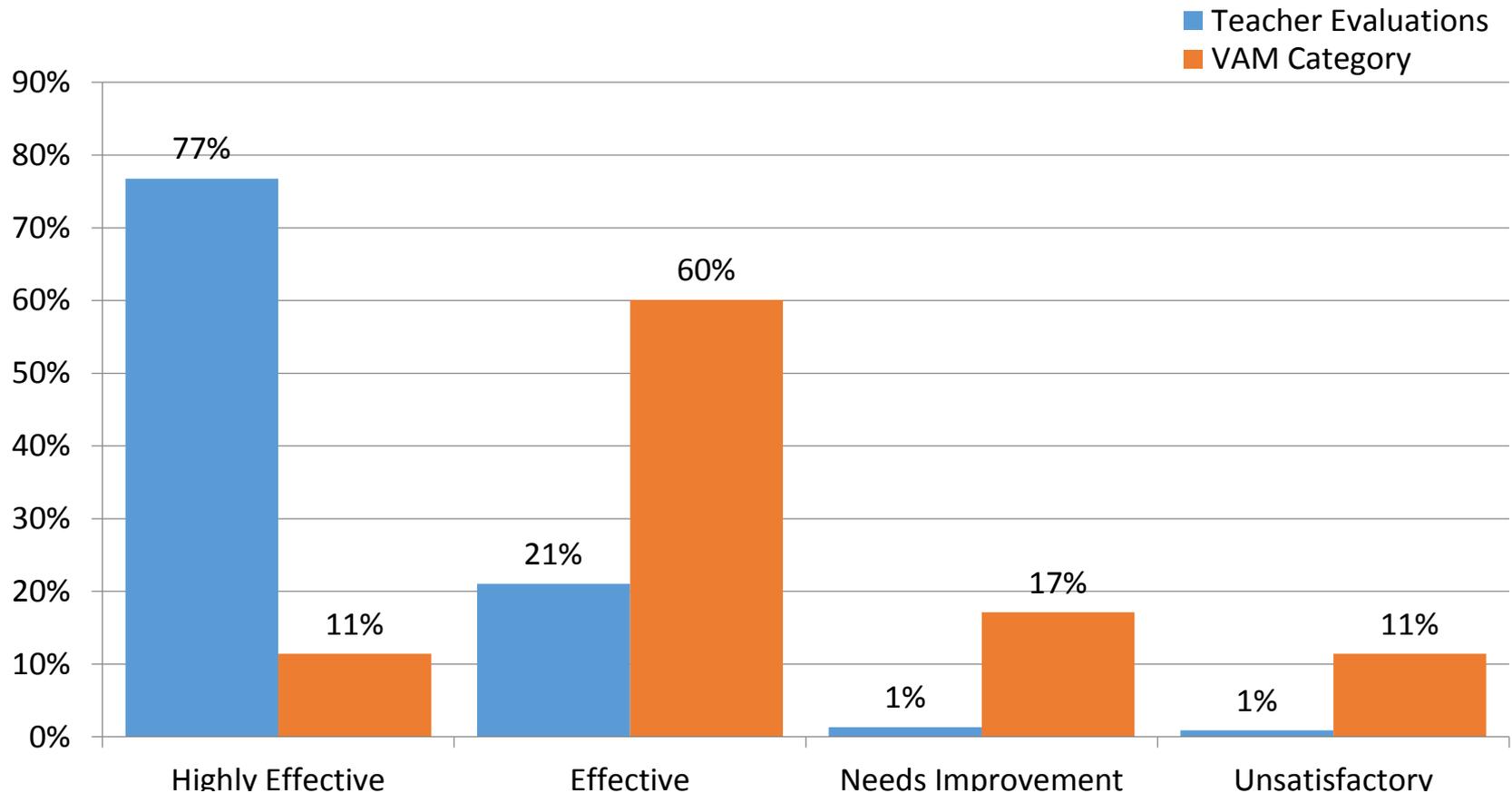
Guiding Questions

Will the non-FCAT model be produced for the 2014-15 school year?

- We have no plans to produce it at this time.

- New assessments required under 1008.22, which were not previously available in all cases, can be used.

- For districts that are not yet ready to use their new assessments for teacher evaluation, FSA results may be used according to a methodology to be determined and calculated by the district. We can provide technical support.

Will Additional Value-Added Models Be Produced for the 2014-15 School Year?

- 11th Grade ELA will be explored, and most likely provided.

- Additional EOC models will be explored by the Department, AIR and the SGIC for possible recommendation.

- May or may not be VAMs.

- Any new models that are approved will not be required to be used until the year following approval.

Can principals or teachers choose different assessments for the same course within the same district for the required 1008.22 assessment?

- Yes – If the district chooses that option and adopts a policy pursuant to Section 1008.22(6)(c), F.S, which requires that each district school board adopt policies for selection, development, administration and scoring of local assessments and for collection of assessment results.

What Criteria Will Be Used to Determine if Cut Points Need Correction?

- Until rule language is adopted by the State Board, cut points are a district decision.

- However, the law requires that evaluation systems must differentiate among four levels of performance.

Are 11th Grade ELA Results Required to Be Used in the Evaluation of ELA Teachers (AP, IB, etc.)?

- Yes - Accelerated courses being used meet graduation credit requirements must still incorporate ELA results into the teacher's evaluation. However, because these courses serve as a *substitution* for the required English III course, for evaluation purposes FSA ELA results may be combined with results from the other course-relevant assessment and weighted according to a district's own methodology.

Additional Information

Student Growth Website

http://www.fldoe.org/profdev/studentgrowth.asp

Email questions to:

VAM@fldoe.org

www.FLDOE.org

Brian Dassler and Jason Gaitanis

brian.dassler@fldoe.org and **jason.gaitanis@fldoe.org**