Teacher Leader Preparation Implementation Committee

November 27, 2012

Florida Department of Education
Bureau of Educator Recruitment, Development, and Retention

Welcome and Agenda Review

Florida Department of Education
Bureau of Educator Recruitment, Development, and Retention

Primary Goal of the TLPIC

- Provide input, feedback and recommendations to the state on the development and implementation of performance standards and targets for continued approval of state-approved teacher and school leadership preparation programs.

TLPIC Timeline

- Fall/Winter 2012/2013
  - Analyze requested teacher preparation data and recommend performance targets for pilot annual report
  - DOE produces pilot annual report
- Summer 2013
  - Recommend to Commissioner draft continued approval standards and performance targets for teacher preparation programs
  - Draft continued approval standards for teacher preparation are released for public input through rule development process
  - Rule revision workshops (6A-5.066)
  - Program data released via report card

Note : March-May 2013

Legislation will likely affect teacher preparation statutes

TLPIC Timeline

- Fall 2013
  - Program approval office conducts pilot site visits using new continued approval standards that include performance targets for teacher preparation programs

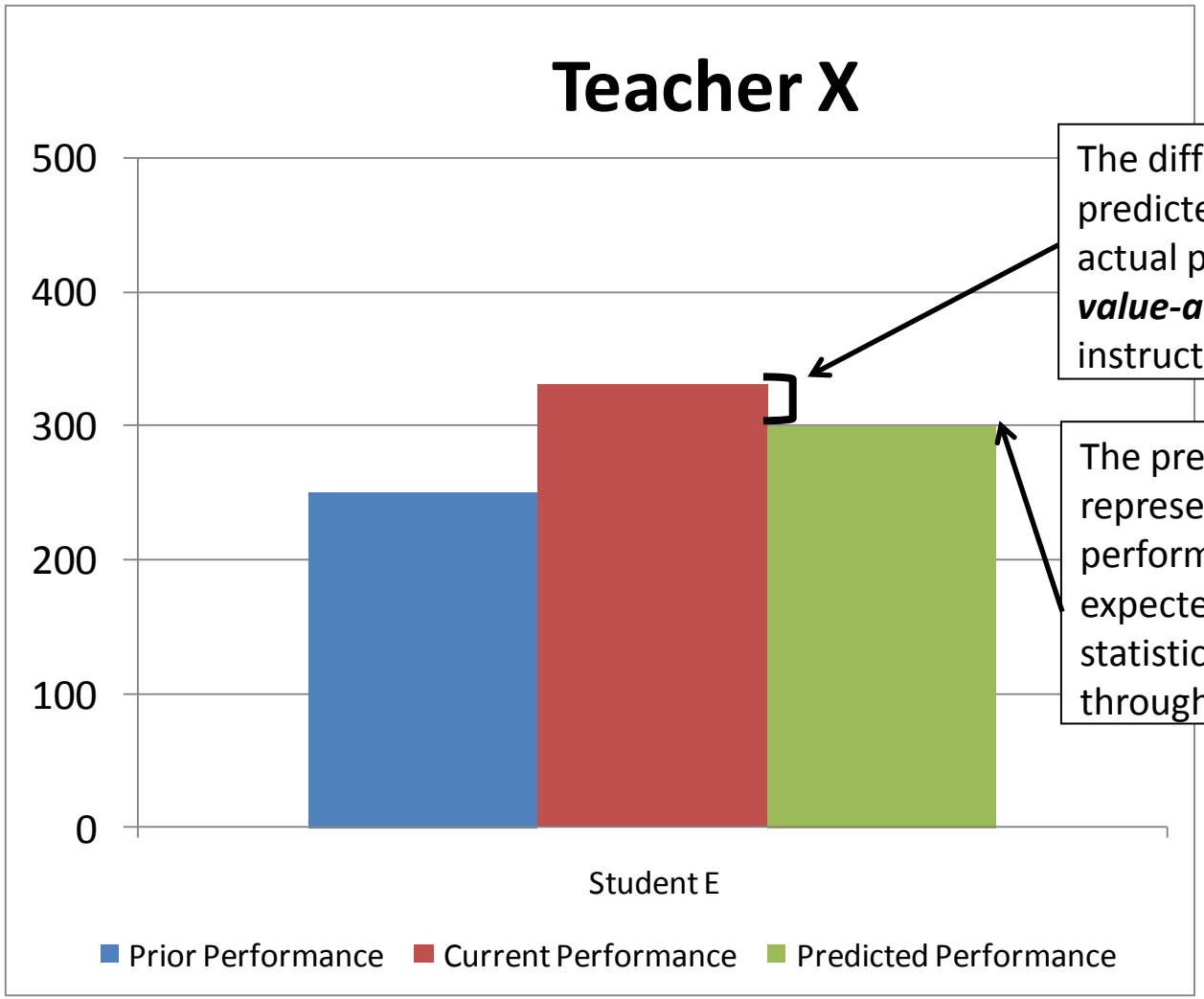
Impact on Student Learning:

- Value-added Scores
- Students Meeting/Exceeding Expectations by Subgroup
- (Teacher Evaluation Summative Results)

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)

Value-Added Example



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.

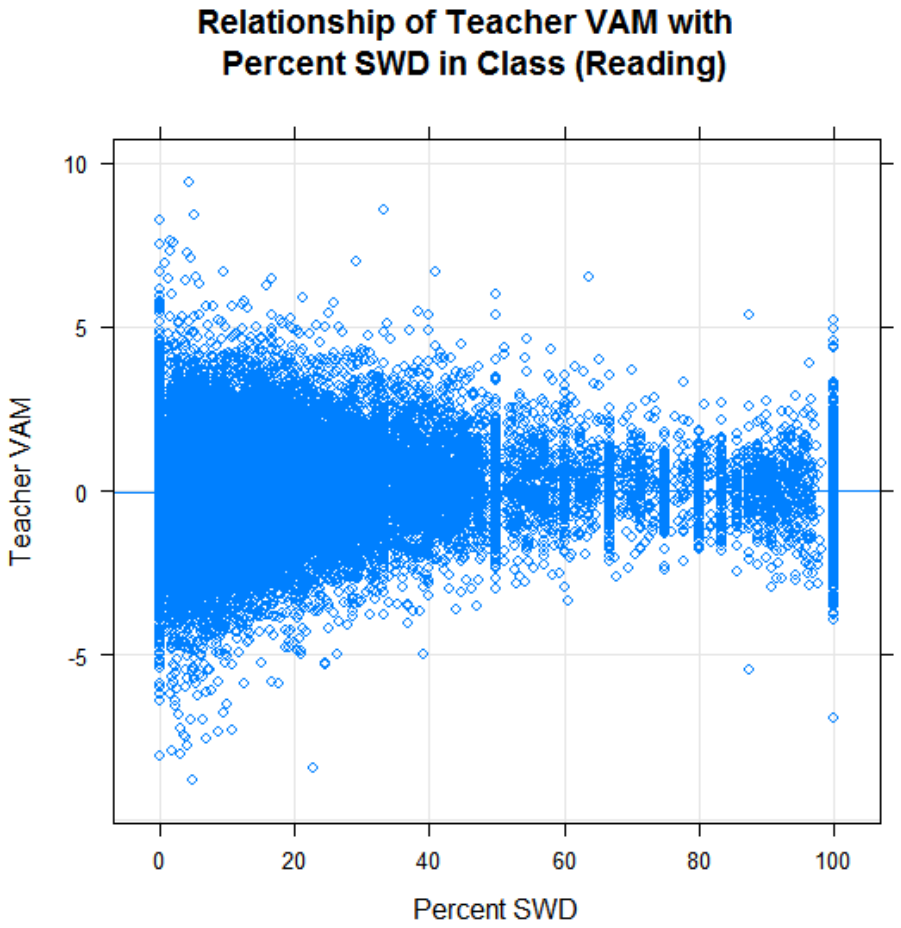
Advantages of Value-Added Models

- 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

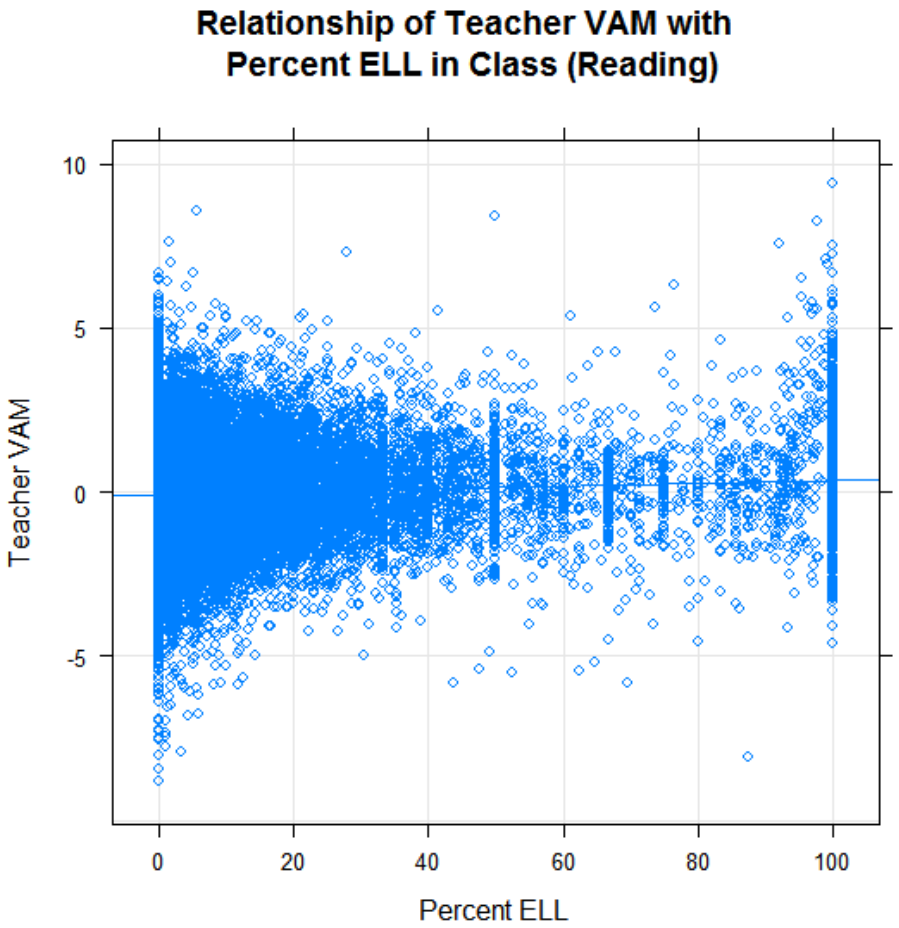
Impact Analyses for 2011-12 VAM Scores

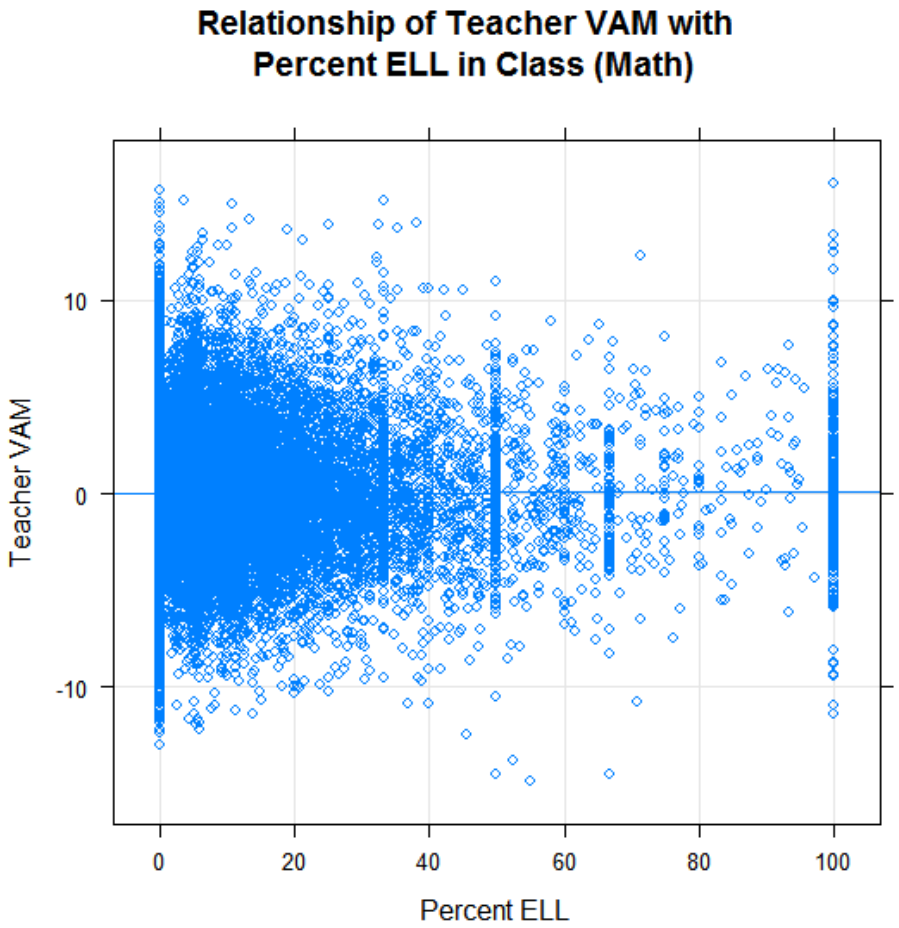
- The following slides show the relationship of the teacher VAM score with various classroom characteristics
- The observed score correlations for each characteristic are reported after the scatter plots
- In all cases, the correlations are negligible indicating no advantages or disadvantages for any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

Correlation of Teacher VAM Score and Percent Students with Disabilities

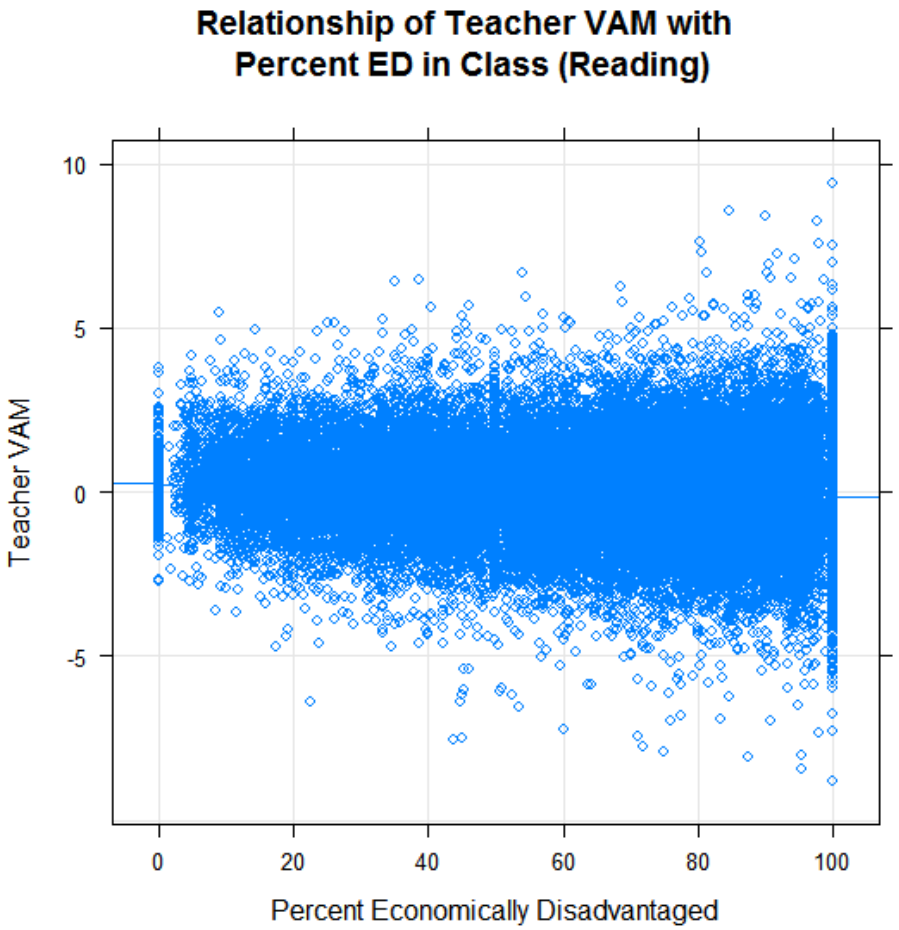


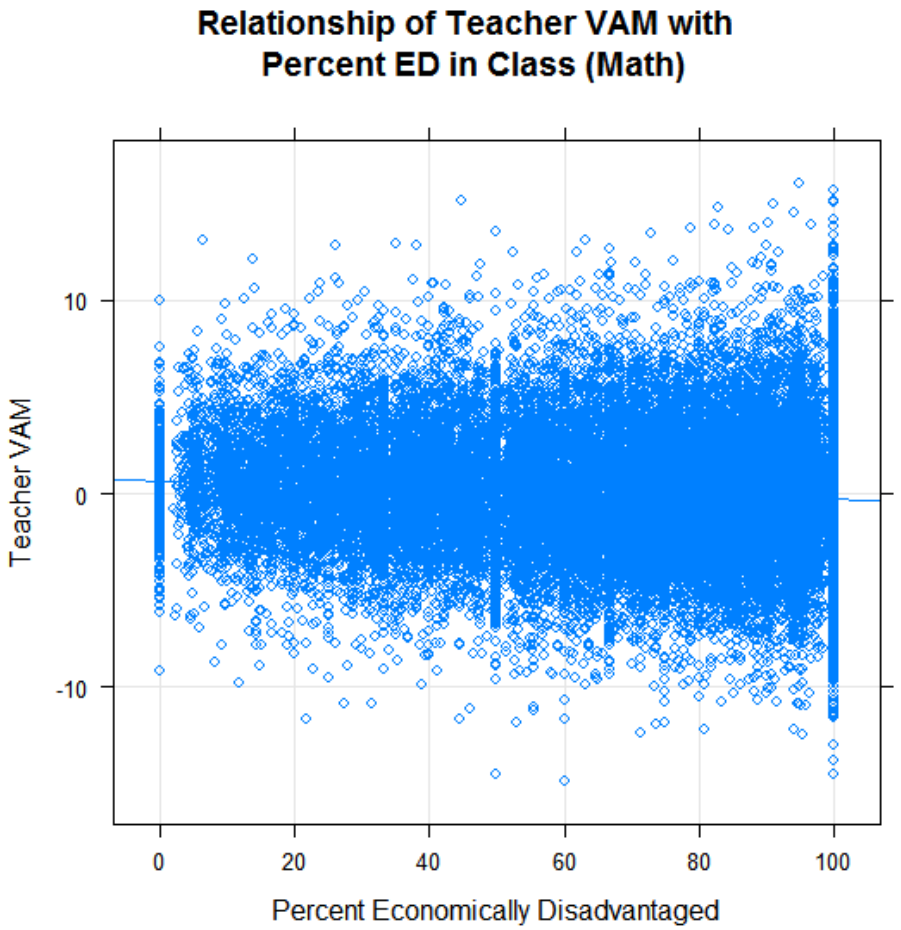
Correlation of Teacher VAM Score and Percent English Language Learners



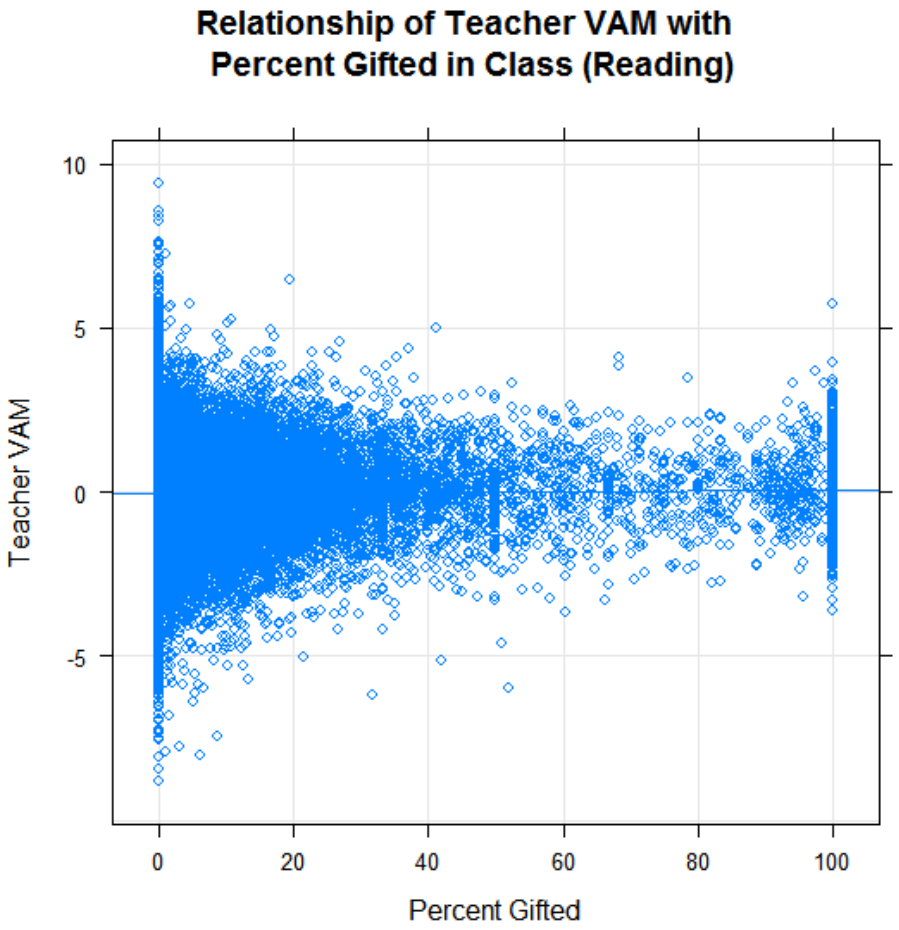


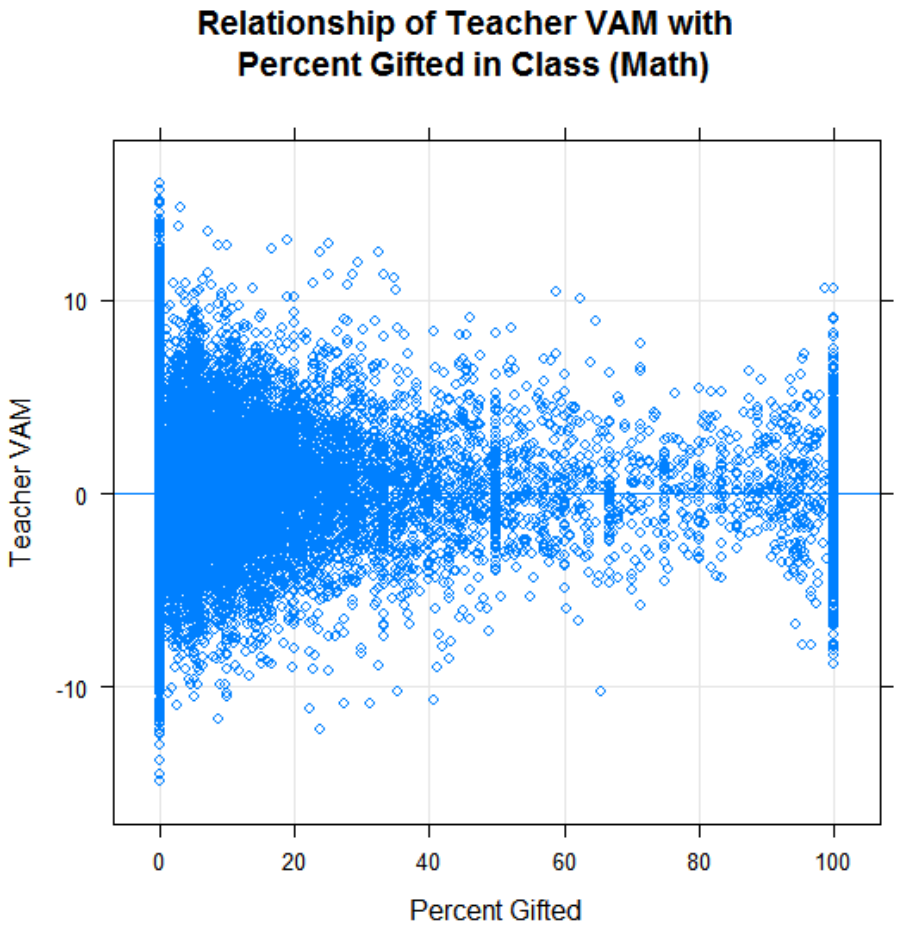
Correlation of Teacher VAM Score and Percent Economically Disadvantaged



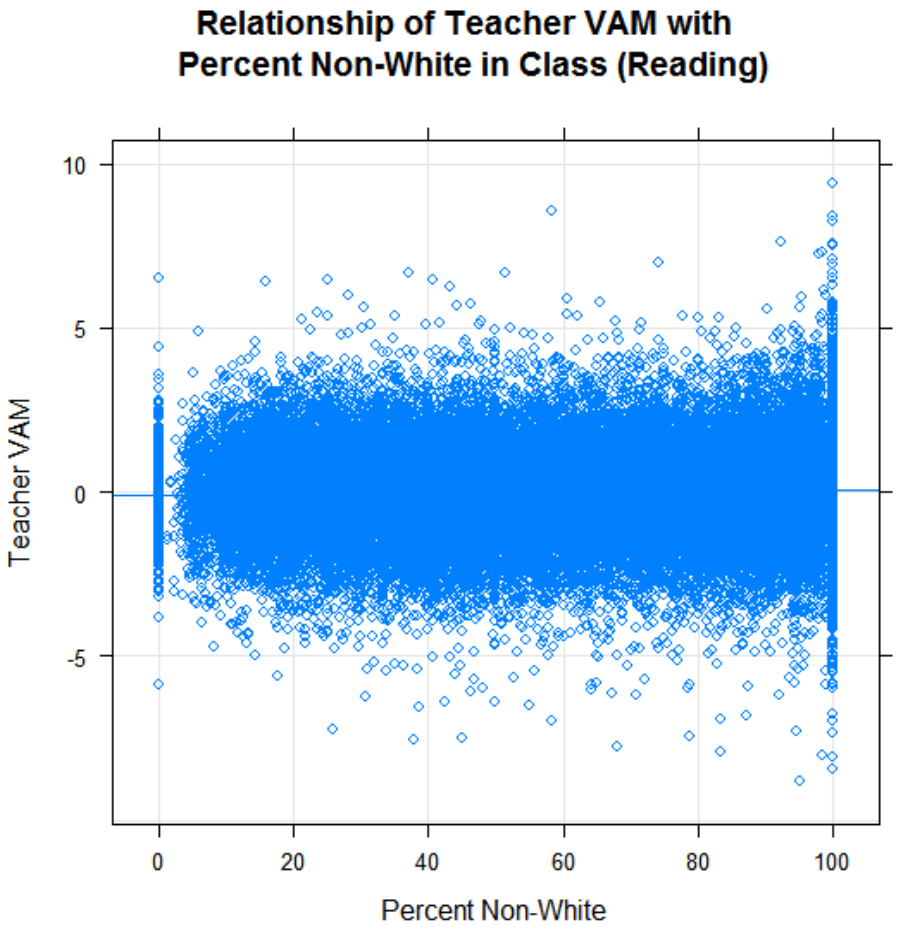


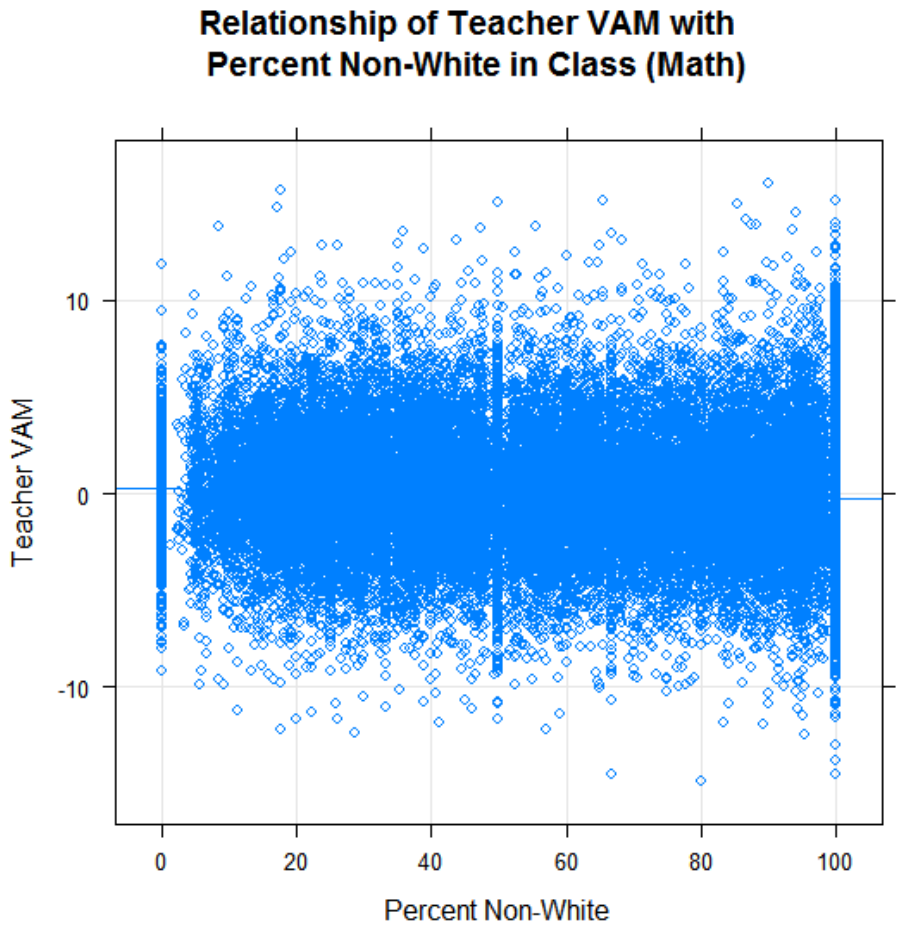
Correlation of Teacher VAM Score and Percent Gifted



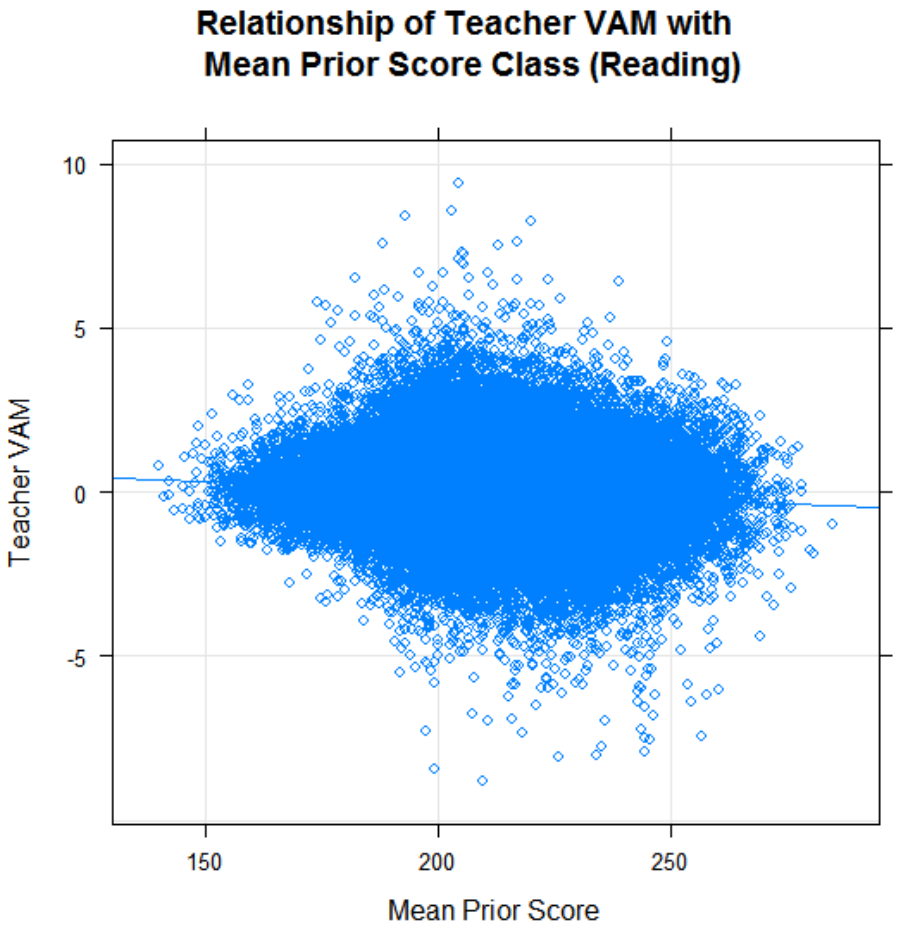


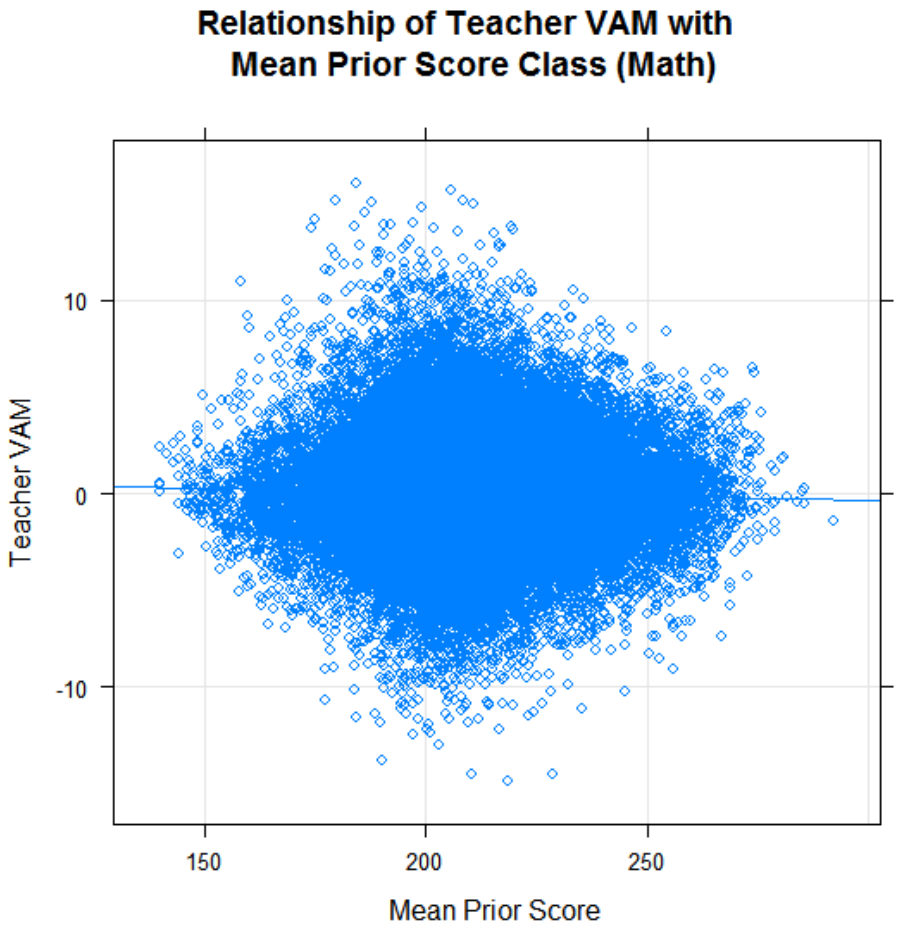
Correlation of Teacher VAM Score and Percent Non-White Students





Correlation of Teacher VAM Score and Mean Prior Achievement





Value-Added Model Data

- The Measure:
  - Average value-added model (VAM) score of completers one year following program completion
    - Aggregated across three years (i.e., three cohorts of completers)
    - Use in-program/in-field data, when possible, in evaluating programs
    - Using the standard error of the VAM score in classification decisions

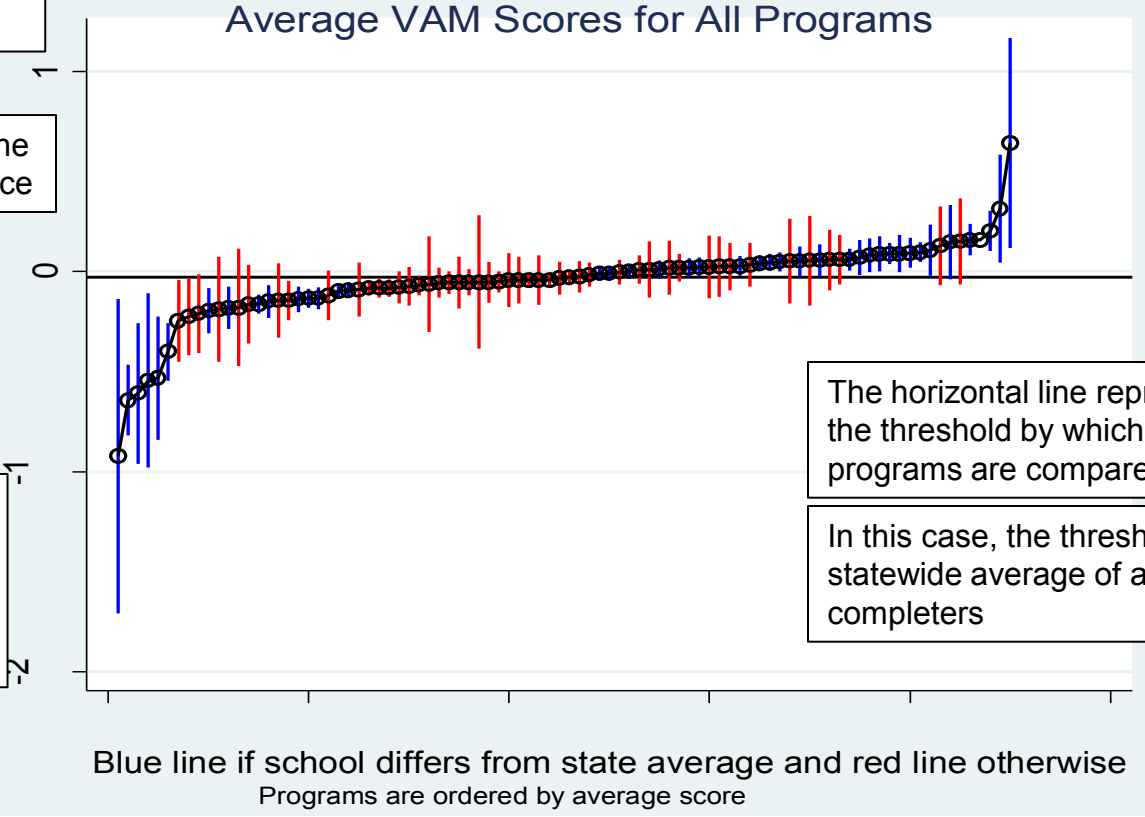
Value-Added Model Data

Review – "Caterpillar" Chart

Each vertical line represents a program

The hollow circles represent the program's average performance

The length of the line represents the confidence interval around the program average (i.e., ***standard error*** applied)



The horizontal line represents the threshold by which the programs are compared

In this case, the threshold is the statewide average of all completers

Value-Added Model Data: The Use of Standard Error

- Remember, an estimate of a teacher's impact on student learning contains some variability
- The ***standard error*** is a statistical term that describes the variability
- Using the standard error can assist in increasing the accuracy of classification decisions
- 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 value-added score meets a certain performance threshold

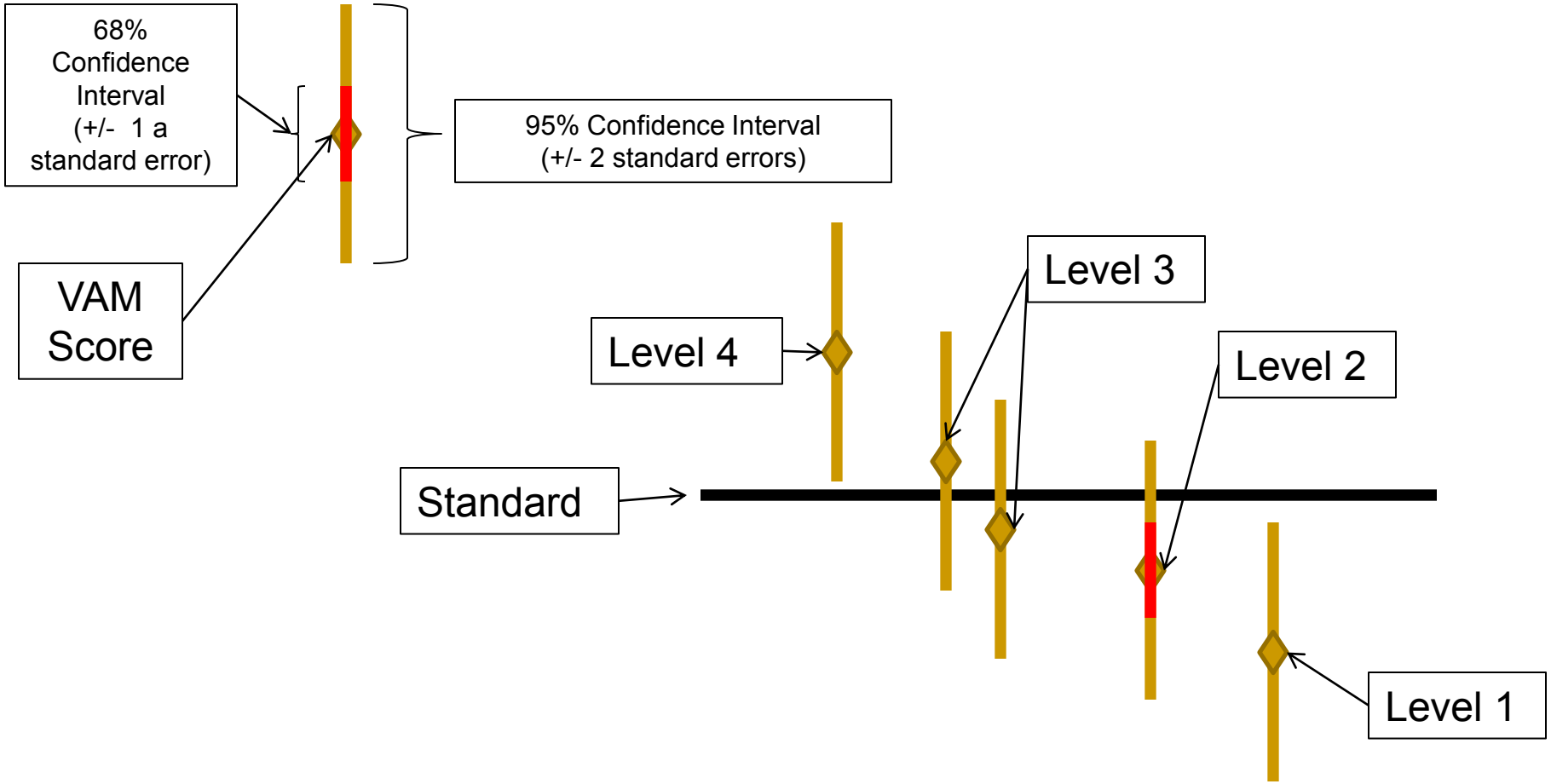
Value-Added Model Data: The Use of Standard Error

- AIR recommended that Florida use at least a 68 percent confidence (i.e., one standard error) and preferably a 90 percent level of confidence in comparing performance.
- In determining the level of confidence, consideration must be given to the ability to distinguish performance (more likely when using lower levels of confidence) and the risk of misclassifying programs (less likely when using higher levels of confidence).

Value-Added Model Data: Classification Recommendation

Category	Standard – Avg. of all Teachers (0)	Standard Error Applied
Level 4 (Highest)	Above	2 SE (95% confidence)
Level 3	Above/Below	None
Level 2	Below	1 SE (68% confidence)
Level 1 (Lowest)	Below	2 SE (95% confidence)

Value-Added Model Data: Classification Recommendation – Visual Example



Value-Added Model Data: Classification Recommendation Explained

- Level 4 represents that score falls above the standard for evaluation, with a high degree of confidence – 95%
- Level 2 represents that the score falls below the standard for evaluation, with some degree of statistical confidence – 68%
- Level 1 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 degree of statistical confidence, the score defaults to Level 3.

Value-Added Model Data: 2011-12 Impact Data
Institution Level –
Reading and Math Combined Across Three Years

- Standard, Score of 0

	Level 4	Level 3	Level 2	Level 1
All	3	45	13	16
EPI	0	19	4	4
ITP	2	13	4	11
DACP	1	13	5	1

43 institutions/districts with insufficient data

Value-Added Model Data: 2011-12 Impact Data
Program Level – Reading and Math Separately
Only Trained In-Program/Teaching In-Field Considered for ITP

- Standard, Score of 0, Reading

	Level 4	Level 3	Level 2	Level 1
All	2	46	14	19
EPI	0	15	3	5
ITP	0	21	7	13
DACP	2	10	4	1

- Standard, Score of 0, Math

	Level 4	Level 3	Level 2	Level 1
All	2	43	16	6
EPI	1	16	3	1
ITP	0	15	12	4
DACP	1	12	1	1

97 programs with insufficient data in Reading; 104 programs with insufficient data in Math

Student Performance by Subgroup

- Following May 9, 2012 Meeting, TLPIC moved to explore methods of including student subgroup performance as a metric of the evaluation system – not a bonus factor

Student Performance by Subgroup: Review

- In addition to the value-added score, the model also yields information on the number and percent of students that met their statistical performance expectations.
- Though these data do not provide information on how far students improved or declined, it does provide information on the quantity of students who met their expectations
- These data are used in analyzing the disaggregated performance of student subgroups

Student Subgroup Performance – Percent Meeting/Exceeding Expectations –

All Completers Across Three Years of Performance Data (2008-09 to 2010-11)

Student Subgroup	Reading	Math
White	49	49
African American	45	46
Hispanic	50	49
Asian	53	55
Native American	46	51
Free/Reduced Lunch	47	48
Students with Disabilities	48	47
English Language Learners	49	48

Student Performance by Subgroup:

- Compare student subgroup performance to the state average
- Determine the number of subgroups that exceed the state average for performance
- Classify programs in one of 4 performance categories, based on the percentage of subgroups meeting the standard

Student Subgroup Performance – Potential Option for Classifying Programs

Performance Level	Subgroup Criteria
Level 4 (Highest)	At least 75% of subgroups (e.g., 6 out of 8, 3 out 4, etc.) must exceed the state standard for performance
Level 3	At least 50%, but no more than 74% of subgroups must exceed the state standard for performance
Level 2	At least 25%, but no more than 50% of the subgroups must exceed the state standard for performance
Level 1 (Lowest)	Fewer than 25% of the subgroups exceed the state standard for performance

Student Subgroup Performance – Potential Option for Classifying Programs

The minimum number of subgroups needed to meet each performance level would vary based on the number of subgroups for which the program had data

Number of Subgroups	Minimum # Needed to Meet Level 4	Minimum # Needed to Meet Level 3	Minimum # Needed to Meet Level 2
1	1	N/A	N/A
2	2	1	N/A
3	3	2	1
4	3	2	1
5	4	3	2
6	5	3	2
7	6	4	2
8	6	4	2

Student Subgroup Performance – Example of Calculation

Student Subgroup	Reading, State Average	Reading, Program Performance	Beat State Average?
White	49	52	YES
African American	45	48	YES
Hispanic	50	49	NO
Asian	53	N/A	---
Native American	46	N/A	---
Free/Reduced Lunch	47	49	YES
Students with Disabilities	48	49	YES
English Language Learners	49	43	NO

Success in 4 out of 6 subgroups (67%) = Program Scores a Level 3 on this metric

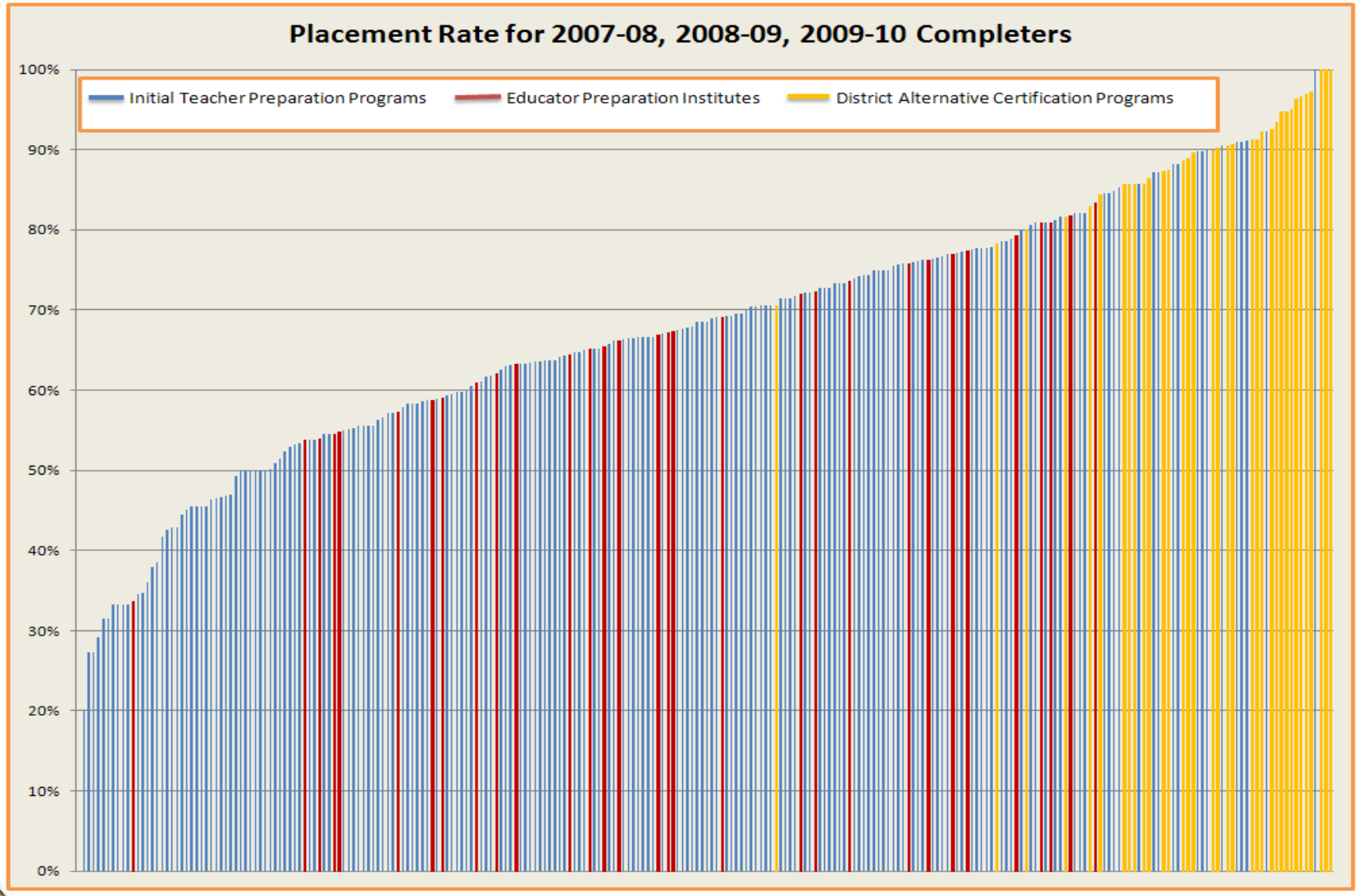
Student Subgroup Performance – 2011-12 Impact data

Performance Level	Reading	Mathematics
Level 4 (Highest) (>= 75% of subgroups)	17	14
Level 3 (50-74% of subgroups)	23	11
Level 2 (25-49% of subgroups)	24	18
Level 1 (Lowest) (<25% of subgroups)	17	24

Placement
Retention
Critical Teacher Shortage

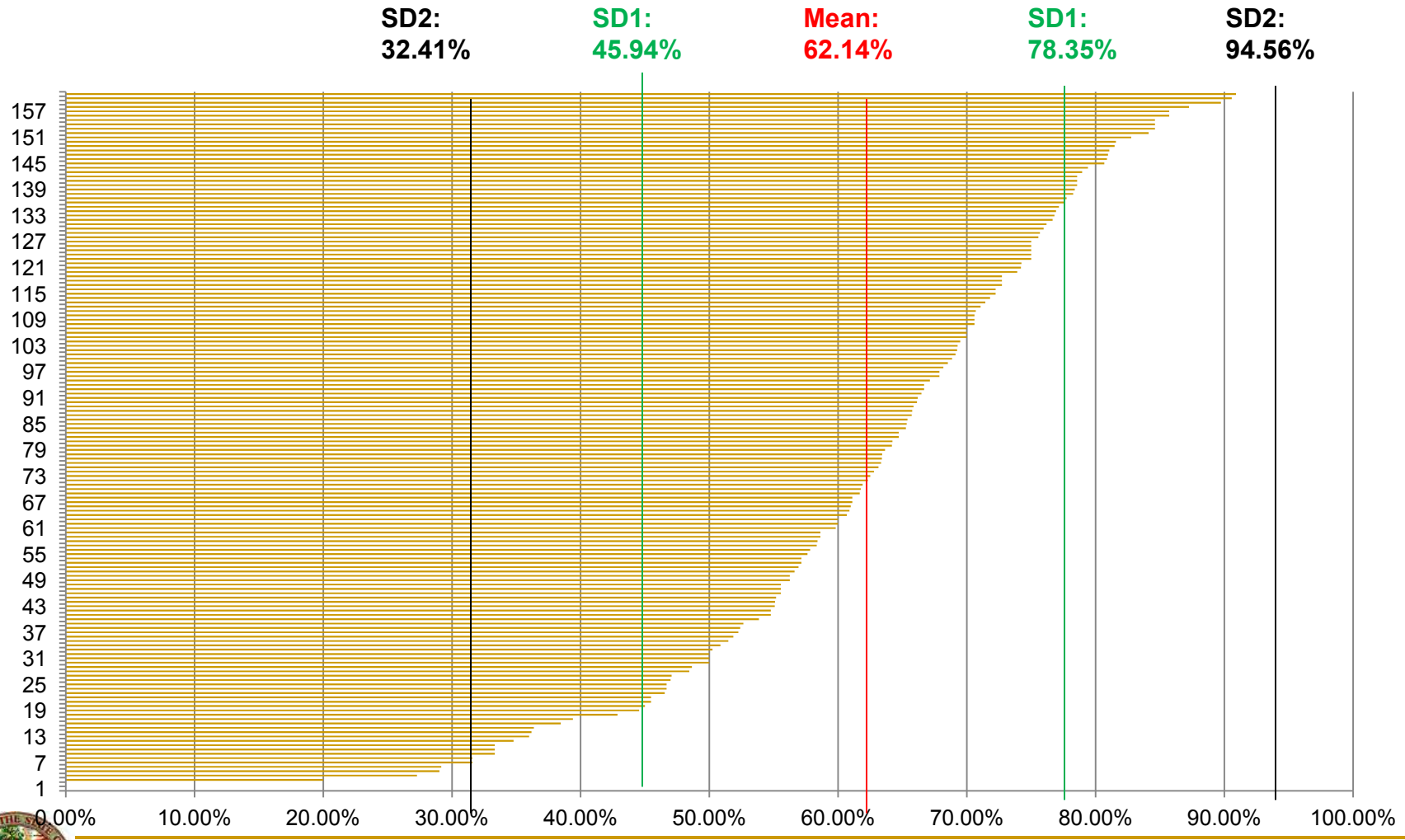
Florida Department of Education
Bureau of Educator Recruitment, Development, and Retention

Placement Data



Placement Data – ITP

2007-08, 2008-09, 2009-10 Completers employed first or second year after completion



Placement Data

- Scale Option 1: National average (60%) is the floor of Level 2/Conditional Approval

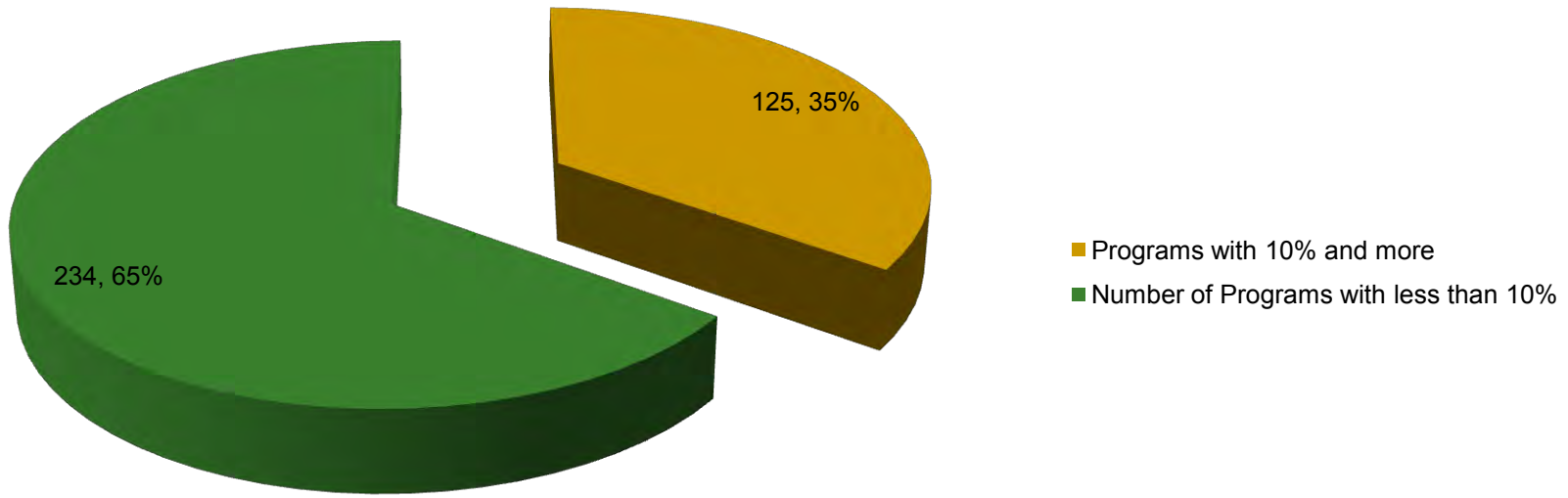
	Level 4 100-90%	Level 3 89-75%	Level 2 74-60%	Level 1 59% and below
ITP	7 (3%)	53 (24%)	80 (36%)	84 (37%)

- Scale Option 2: National average (60%) is the floor of Level 3/Full Approval

	Level 4 100-85%	Level 3 84-60%	Level 2 59-45%	Level 1 44% and below
ITP	15 (7%)	125 (56%)	61 (27%)	23 (10%)

Critical Teacher Shortage

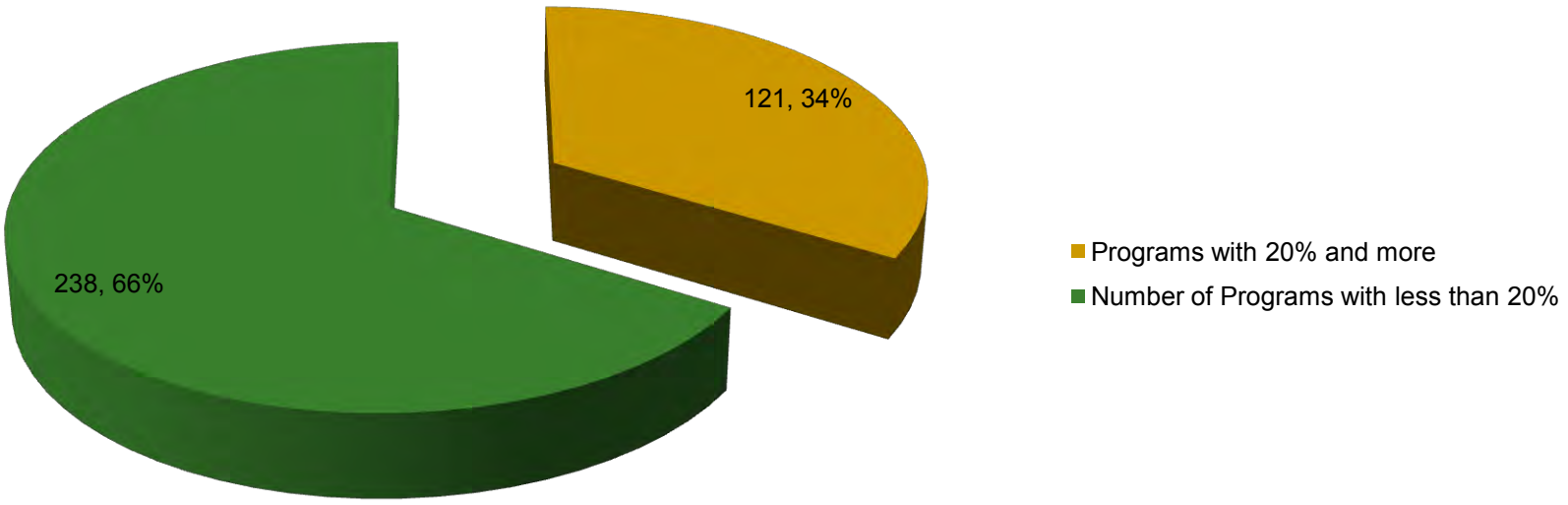
Critical Teacher Shortage Program (ITP)
10% and Above Increased Production
FY: 2009-10, 2010-11



359 Total Programs

Critical Teacher Shortage

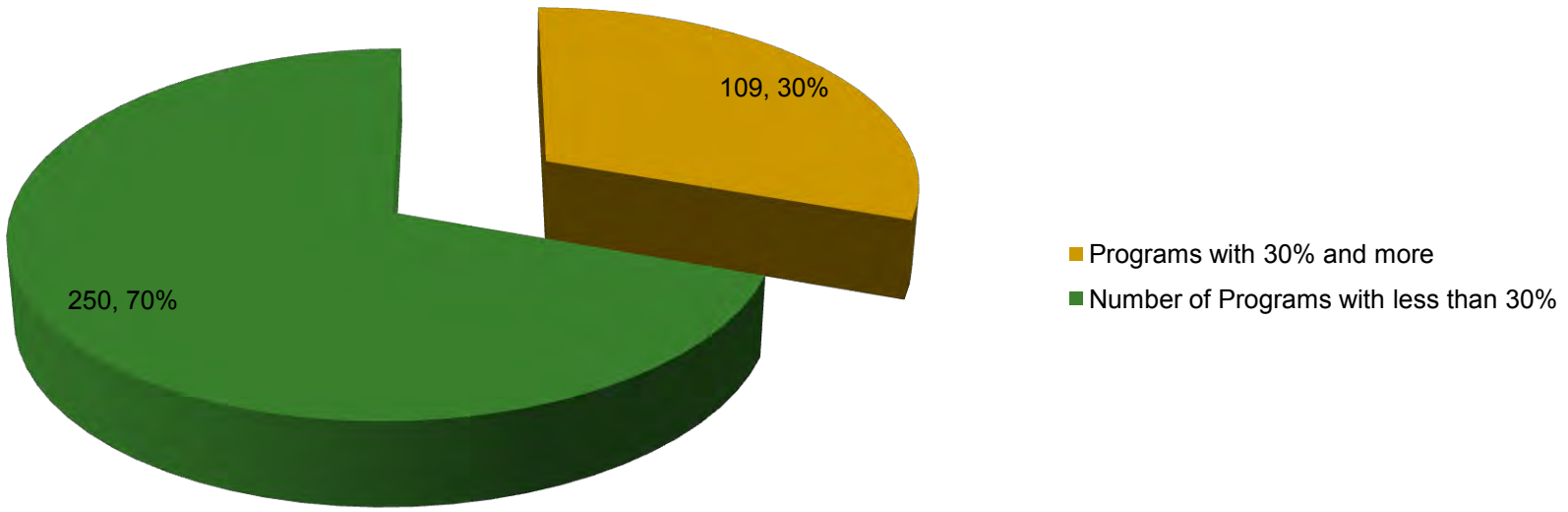
Critical Teacher Shortage Program (ITP)
20% and Above Increased Production
FY: 2009-10, 2010-11



359 Total Programs

Critical Teacher Shortage

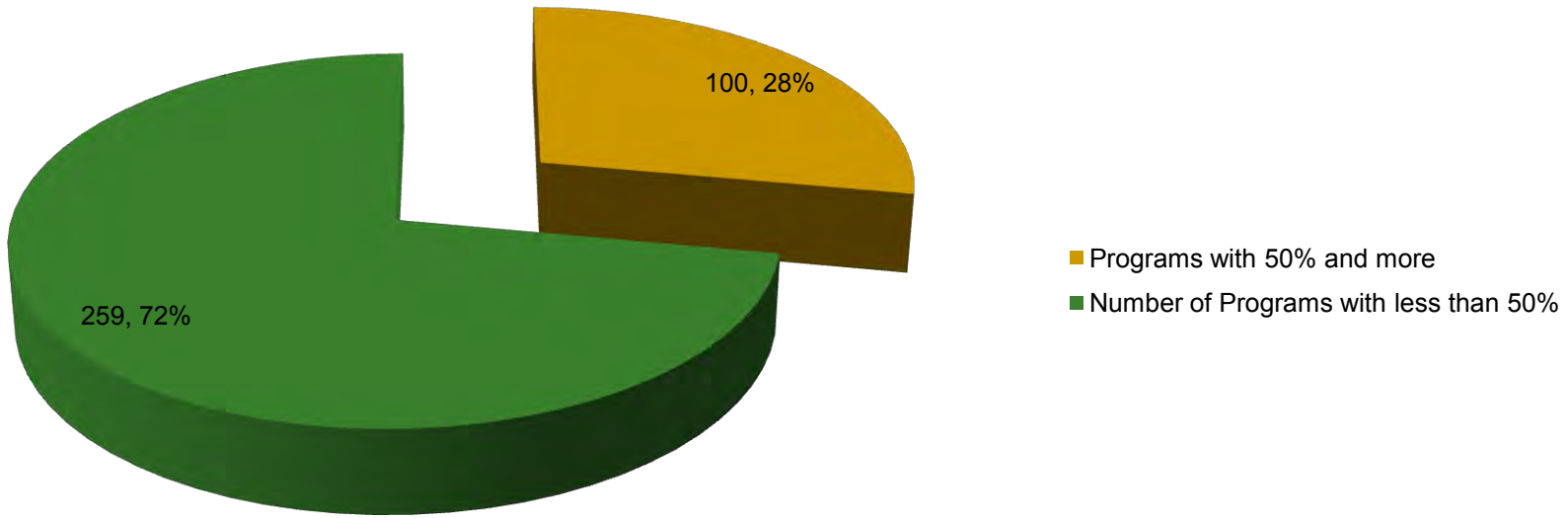
Critical Teacher Shortage Program (ITP)
30% and Above Increased Production
FY: 2009-10, 2010-11



359 Total Programs

Critical Teacher Shortage

Critical Teacher Shortage Program (ITP)
50% and Above Increased Production
FY: 2009-10, 2010-11



359 Total Programs

Accountability System Framework

- Annual Progress Report
- Continued Approval

Metrics to Include in the Accountability System

- Placement Rate Data
- Retention Data
- Value-Added Model Data
- Student Performance By Subgroup Data
- Teacher Evaluation System Data
- Bonus Area: Critical Teacher Shortage Area Data

* Agreed to not include those programs that do not meet the Rule of 10

Structure for the Accountability System

- Each Data Element has an independent scale (1-4) and an independent weight.
- A program's rating is calculated similar to a weighted GPA.
- Bonus area has one cut point. A program meets the criteria to receive a bonus or does not meet the criteria to receive a bonus.
  - If a program meets a bonus criteria, the program receives an additional 0.25 of a point added to its total score.

Structure for the Accountability System

- A program's continued approval status would be determined at the time of the site visit.
- Each year during the program's continued approval cycle, it would receive a progress report and score based upon the data elements in the accountability system.
- The annual progress scores earned during the continued approval period would be aggregated at the time of the site visit and contribute to the program's new continued approval status.

Possible Structure for the Accountability System #1

- Programs scoring 4.0 or above are at Level 4
- Programs scoring 3.0 to 3.9 are at Level 3
- Programs scoring 2.0 to 2.9 are at Level 2
- Programs scoring below 2.0 are at Level 1

Possible Structure for the Accountability System #2

- Programs scoring 3.8 or above are at Level 4
- Programs scoring 2.7 to 3.7 are at Level 3
- Programs scoring 1.6 to 2.6 are at Level 2
- Programs scoring below 1.5 are at Level 1

Sample Annual Progress Report

Introductory Information: Explanatory information about the report, definition of each metric and how calculated, total number of completers, programs and demographic data currently included in annual January 1 report.

Metrics	Scores
Placement Rate	3
Retention	2
Value-Added Model	3
Student Perf. By Subgroup	2
Teacher Evaluation	4
Subtotal / Averaged	**2.8**
Bonus: CTSA	0.25
Total for Program X	**3.05 = Level 3**

Sample Report Considerations

- Annual Program Progress Data could include?
  - Numbers and percentages
  - Notes about areas not meeting rule of 10
  - Rather than 4 continued approval ratings, use other words as descriptors for results (i.e., “on track,” “below/above expectations”)
  - Warnings for programs with two years of consecutive progress ratings below 2.7
- Aggregate information for institutions with multiple programs?

Site Visit Protocol Subcommittee Recommendations

Florida Department of Education
Bureau of Educator Recruitment, Development, and Retention

Site Visit Protocol Subcommittee Recommendations to the TLPIC

Committee Members
Gloria Artecona-Pelaez
Ana Blaine
Debbie Cooke
Erin Harrel
Mark Howse
Adriana McEachern
Vivian Posey
Lance Tomei

1. Two-phased Site Visit

The overall continued program approval site visit protocol will be a two-phased process patterned after the NCATE continuous improvement option.

Essential components of this process will include:

- Evidence required for the review will be available six months prior to the scheduled onsite visit.
- Review team members will have 4-6 weeks to review the evidence and begin drafting an offsite report.
- An offsite team meeting (web/teleconference) will be held no later than 4 ½ months prior to the onsite visit. During this meeting, team members will discuss the evidence and their preliminary findings and will continue to refine the draft offsite report.

1. Two-phased Site Visit (cont.)

- Upon conclusion of the offsite meeting, the team chair, in collaboration with team members and FDOE, will have two weeks to finalize the offsite report and make it available to the unit. The report will include:
  - a preliminary narrative summary of the team's initial (offsite) findings regarding the standards
  - a list of "areas of concern" (findings that could become weaknesses in the final report if not resolved by the time of the onsite visit)
  - requests for additional evidence
  - a list of evidence to be validated during the onsite visit

1. Two-phased Site Visit (cont.)

- The unit has one month to respond to the offsite report. The unit may also work to address/resolve any areas of concern through the time of the onsite visit.
- The onsite team (ideally the same individuals as the offsite team, and with the collective expertise appropriate for the unit being reviewed) will focus their onsite review based on (but not absolutely limited to) the findings of the offsite team.
- The onsite team will prepare the final site review recommendations to FDOE.

2. Required Evidence

The off-site phase of the periodic continued program approval review will be based on the most recent e-IPEP submitted at least 6 months prior to the scheduled on-site visit plus an e-IPEP addendum incorporating any additional evidence (TBD) deemed necessary in support of the review. These electronic submissions would replace the previously required program self-studies.

3. Reviewer Qualifications

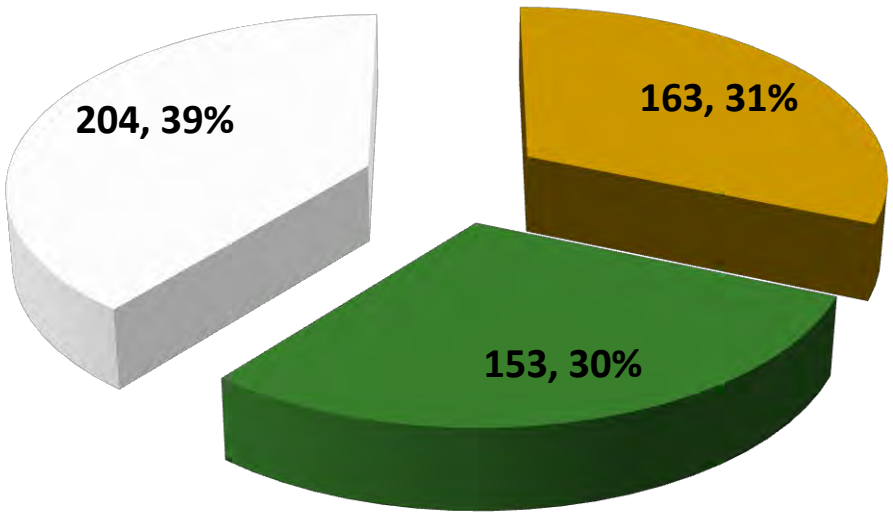
Minimum qualifications requirements should be established for individuals serving on all program approval review teams (initial/folio reviews and continuing program approval reviews), additional training should be required for team chairs, and periodic professional development should training be required to maintain currency.

Next Steps

- Provide missing data (teacher evaluation, etc.)
- Issue pilot annual progress report
- Receive feedback on report from programs
- Work through issue of “small programs” with input from approved programs
- Execute remainder of timeline

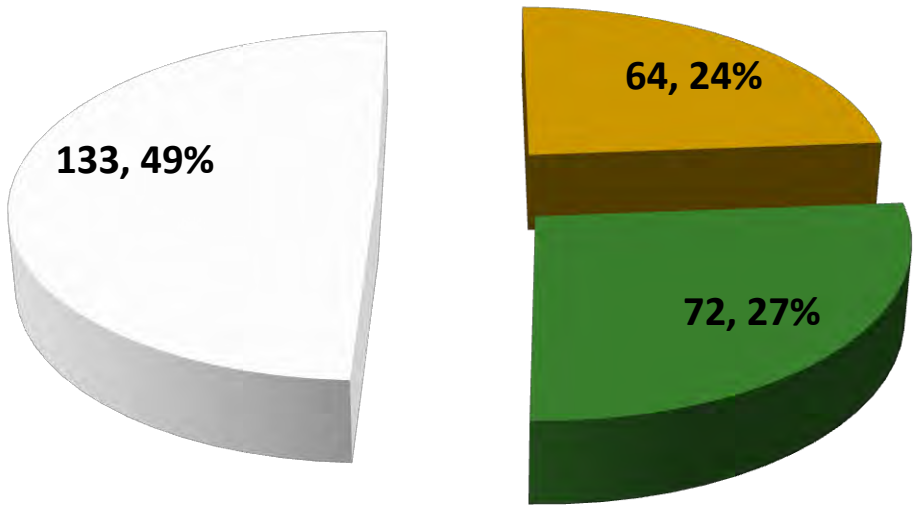
ITP Programs & Rule of 10

Total Number of ITP Programs = 518



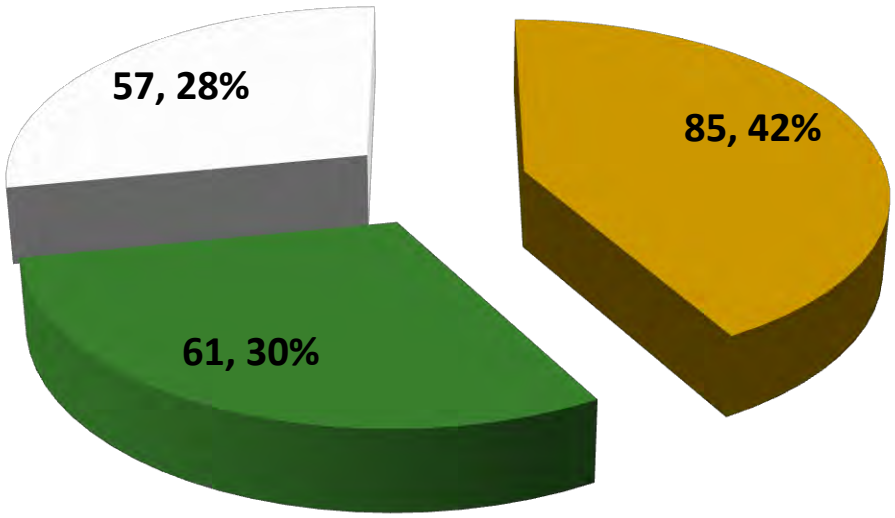
SUS ITP Programs & Rule of 10

Total Number of State University System (SUS) ITP Programs = 269



ICUF Programs & Rule of 10

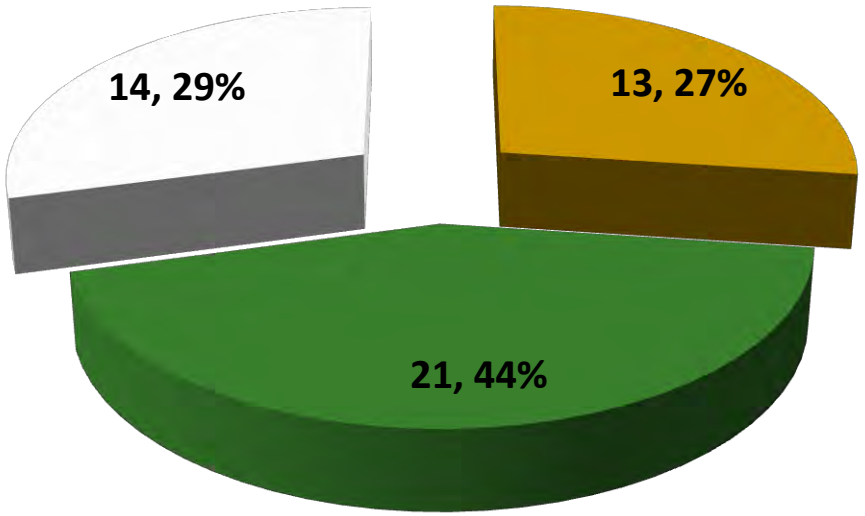
Total Number of ICUF/Private ITP Programs = 203



- Total Number of ICUF/Private Programs w/ Less than 10 (not including 0)
- Total Number of ICUF/Private Programs w/ Less than 10 (0 only)
- Total Number of ICUF/Private Programs w/ More than 10

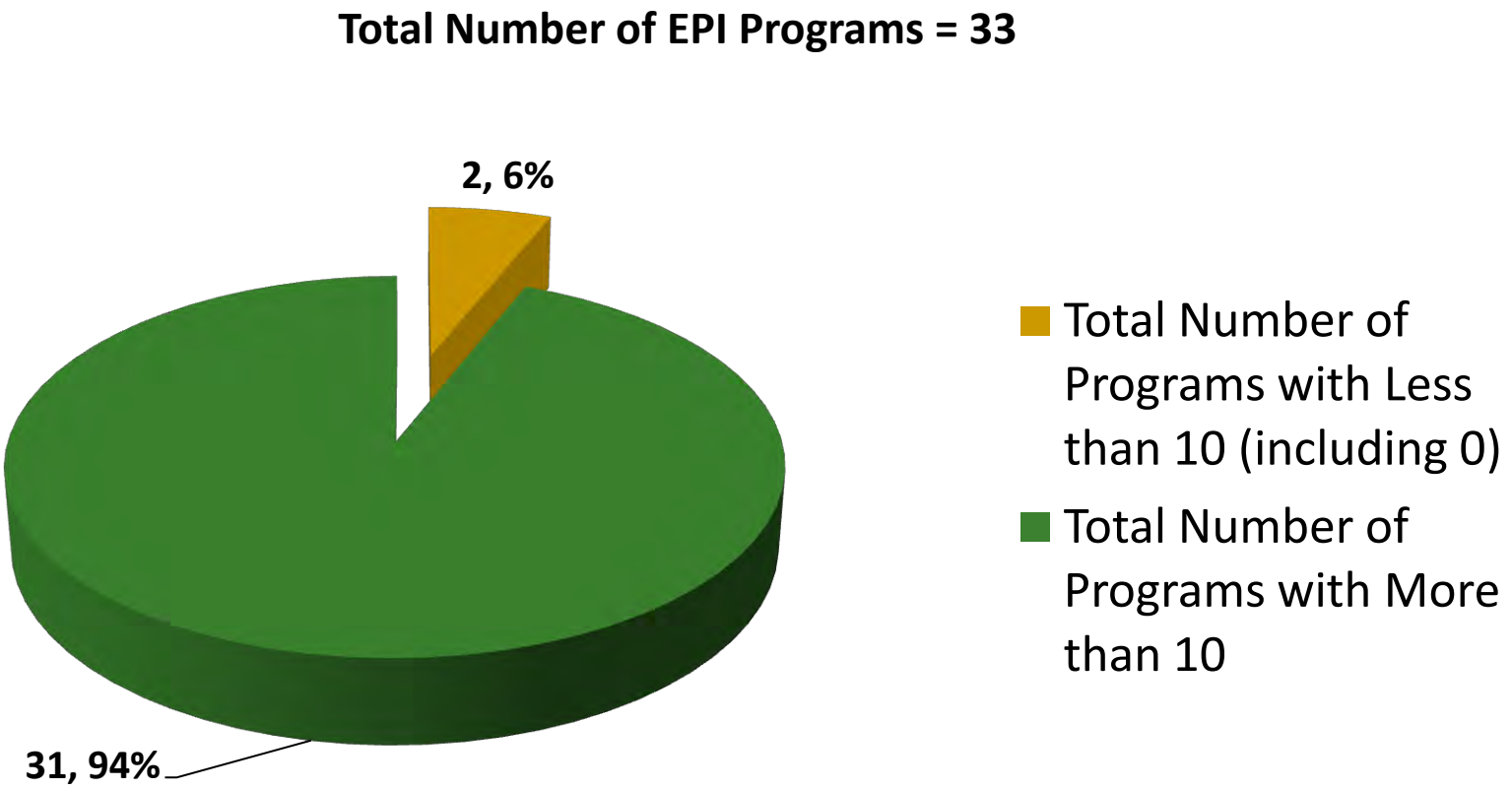
State College ITP Programs & Rule of 10

Total Number of State Colleges ITP Programs = 48

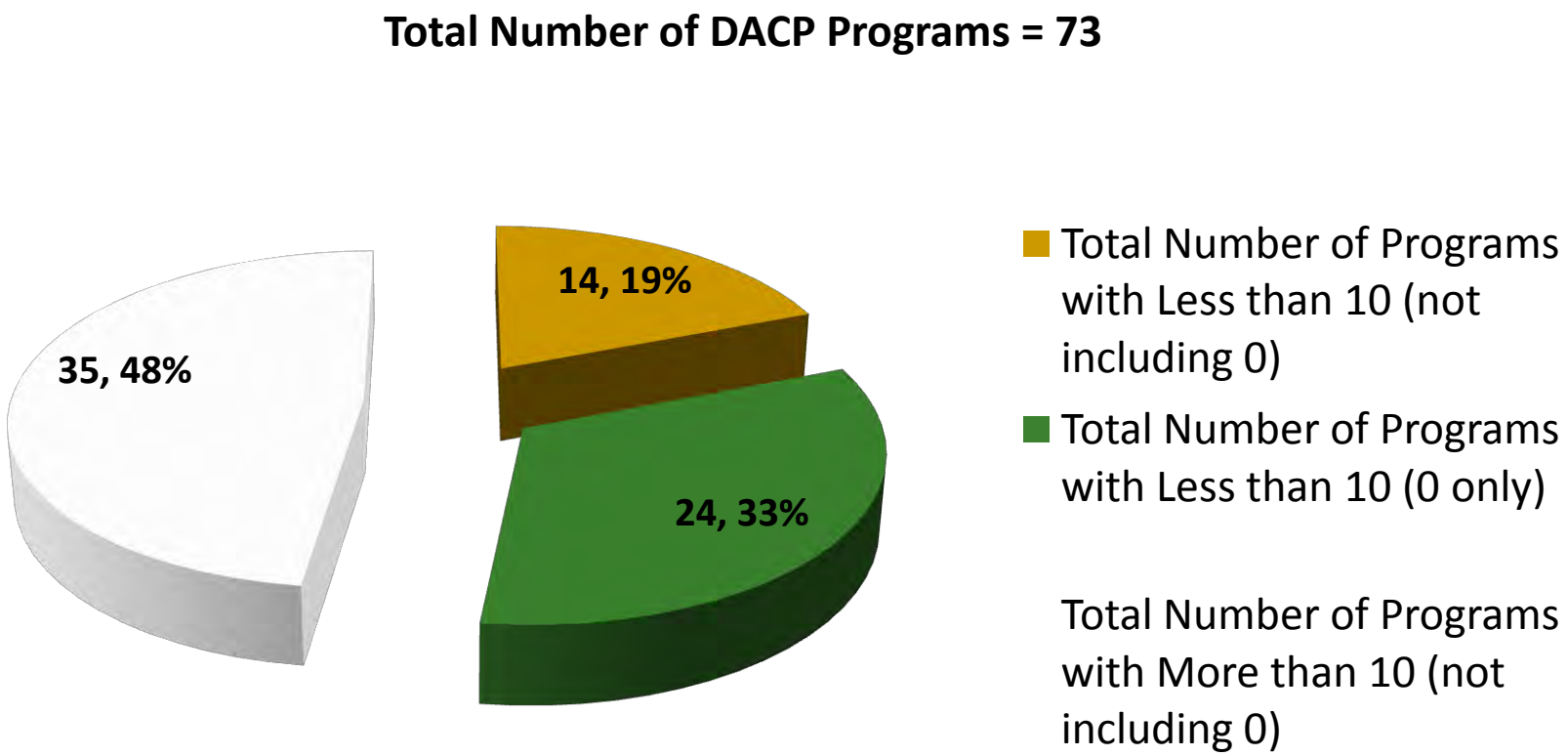


- Total Number of State College Programs w/ Less than 10 (not including 0)
- Total Number of State College Programs w/ Less than 10 (0 only)
- Total Number of State College Programs w/ More than 10

EPI Programs & Rule of 10



DACP Programs & Rule of 10



Total *Completers* &Rule of 10