## Teacher Evaluation

Sarasota County Public Schools School Board Workshop September 18, 2018

## Objectives

- Provide historical context and information on legislative mandates for PRIDE and student growth as components of the Teacher Evaluation System (TES)
- Review the current PRIDE system
- Review the critical decisions in test/student growth model development
- Provide overview of current student growth models
- Provide information on aggregation of PRIDE with Student Growth
- Discuss final ratings and reports


## Materials



2

TES PowerPoint Slides

## Teacher Evaluation System has two major components

Initiated in 2011 with the passage of Race to the Top

Statutory
Requirement added in 2012


Use of the Current PRIDE initiated in 2008

Based on the Charlotte Danielson Model

Digitized and weighted in 2012

Both components are legislatively mandated. Sarasota TES approved by FDOE.

## Instructional Practice (PRIDE)- State Statute

- Section 1012.34(2), F.S. For instructional personnel, at least one-third of the performance evaluation must be based upon instructional practice. Evaluation criteria used when annually observing classroom teachers, as defined in s. 1012.01(2)(a), F.S.
- Section 1012.34(4), F.S. For instructional personnel, the remainder of a performance evaluation must be based on job responsibilities based on teaching practices that are consistently associated with higher student achievement, and other valid and reliable measures of instructional practice.


## Student Growth - State Statute

- Evaluation Procedures and Criteria
- Section 1012.34(3), F.S. Instructional personnel and school administrator performance evaluations must be based upon the performance of students assigned to their classrooms or schools.
- Section 1012.34(3)(1), F.S. Performance of students
- At least one-third of a performance evaluation must be based upon data and indicators of student performance, as determined by each school district.
- This portion of the evaluation must include growth or achievement data of the teacher's students or, for a school administrator, the students attending the school over the course of at least 3 years. If less than 3 years of data are available, the years for which data are available must be used.
- The proportion of growth or achievement data may be determined by instructional assignment.


## VAMS - State Statute

$\pi$
Optional use of state-generated VAMS in teacher evaluations

- Statutory references to Approved Growth Model
- Section 1004.04(4)(a)3.c, F.S. Continued approval for teacher preparation programs.
- Section 1012.56(7)(c), F.S. One-year extension of a temporary certificate based on Effective or Highly Effective VAM rating.
- Section 1012.731(3)(a)2, F.S. Beginning in 2020-21, allows a classroom teacher to qualify for the highest award tier (Best and Brightest) without an overall evaluation of Highly Effective if the teacher has a VAM score that is classified as Highly Effective.
- High Impact Teacher designation


## Decisions and Considerations

- Districts use of state VAMS was required up to SY 2016-2017
- The state only provides VAMS for FSA-ELA, FSA-Mathematics, and FSA-Algebra 1
- This is used for approximately 870 teachers (30\%)
- This leaves the majority of teachers without tests/models
- 767 (27\%) Grades PK-3 teachers
- 1018 (35\%) Subject area teachers
- 225 (8\%) IB, AICE, AP, CTE teachers


## Decisions and Considerations

- How do we group teachers so we can assign student performance scores?
- How do we ensure we have the correct students linked to each teacher?
-What do we do if a teacher only has a student part of the year?
-What if a teacher changes assignments mid-year or across years?
- What students do we assign teachers who instruct at several schools?
- What student scores do we link to teachers who do not teach in the classroom?
- What student test scores do we link to other instructional personnel (e.g. school psychologist)?


## Decisions and Considerations

- Do we develop tests to cover grades KG-3 and teachers of over 800 unique high school courses?
- What models could we use for courses with small numbers of students (i.e. ESE classrooms)?
- What assessments/models can we use for the performing arts?
- If we developed assessments, when would we give them?
- If we gave more assessments, how would we deliver them?
- If we gave more assessments, how could we score them?
- How do we control for student assignments (i.e. prior performance)? 10


## Decisions and Considerations

$\pi$

- How do we combine the PRIDE data and the Student Growth data?
- How do we account for assessments on different scales?
-What do we do with students that are not with the same teacher all year?
- Once we receive the scores, who has the expertise to analyze them?
- Where would the historical records be kept?
- Where would we get assessments of merit that all teachers would agree upon?


## Original Guiding Directions

- Develop a system that would
- Require minimal cost (recession)
- Limit the number of assessments developed and offered (time)
- Be conservative or would refrain from any over identification of teachers as Unsatisfactory or Needs Improvement
- Control for student factors, such as prior ability, ESE and ELL status
- Account for every teacher - Over 3226 teachers, both in classrooms and those not working in classrooms


## Concerns for using the Regression with LEOCs and State EOCs

- Absence of external reference for LEOCs - comparing against our own teachers
- What determines how well a child performed?
- Average score on the test (higher performing schools would benefit)?
- Control for the prior year performance of students
- What prior measurement score is used (reading for all)?
- Would we encounter more students without a prior score in traditional classes as opposed to advanced courses?
- Will ESE/ELL status be controlled adequately by just the prior year reading score?


## Decisions and Considerations - Models

- How do we differentiate scores across the rating scales to meet the statutory requirements?
- How can we be conservative and not over identify teachers rated as Unsatisfactory and Needs Improvement?
-What is Effective and how does it differ from Highly Effective?


## Cycle of Teacher Evaluation

- Annual Process
- Instructional Practice - First day of school to the last day of school every year
- Student Growth - 15 month process for analyses and data verification


## Cycle of Teacher Evaluation - PRIDE



## TES Documents Book 1(Digitized) For Classroom Instructors and Non-Classroom Instructors

- Section 1
- PRIDE Rubric
- Category 1 Teacher Checklist
- Preconference Form Category 1
- Observation Short Forms
- Individual Professional Development Plans
- Indicators for Success Form
- Ongoing Observation Forms
- Post Observation Conference Forms
- PRIDE Mid-Year Form
- PRIDE Final Form

Separate forms for
Classroom and
Non-Classroom
teachers

## TES Documents Book 1(Digitized) For Classroom Instructors, Non-Classroom Instructors, and Administrators

- Section 2
- Administrator Evaluation System (AES) Documentation
- Administrator Evaluation System (AES) Form
- Section 3
- PRIDE Rubric ‘Look Fors’
- Sarasota County Induction Program (SCIP) Documents
- Instructional Common Language Document
- SCIP Feedback Mentor Form
- New Hire Form
- Collaborative Planning Form

Cycle of Teacher Evaluation Professional Rubrics Investing \& Developing Educator Excellence (PRIDE)

Measurement of Instructional Practice


## PRIDE Rubric - 4 Domains and 24 Competencies TES Book 1, Section 1

Domain 1: Creating a Culture for Learning
6 competencies, 24 points

Domain 2: Planning for Success
6 competencies, 34.5 points
Domain 3: Instructing and Assessing for Student Achievement 8 competencies, 34.75 points

Domain 4: Communicating Professional Commitment 4 competencies, 6.75 points

## Digital PRIDE Form



## Digital Ongoing Observation Form



## PRIDE Training - Controlling for Bias

- All administrators are trained extensively on
- Teacher Evaluation Process
- PRIDE Evaluation Standards
- Use of the of the PRIDE Final Evaluation Form
- Observations/All other forms
- Instructional Contract
- Blackboard Course for new teachers which includes PRIDE evaluations
- Qualifying Event for Principals and Assistant Principals
- Administrators are required to conduct multiple observations, per year, for each teacher


## Digital Administrator Evaluation Form

School
Leader :
Evaluator
Name:
School :

|  | Score | Weight | Weighted | (2) SARASOTA <br> cOUNTY SCHOOLS |
| :---: | :---: | :---: | :---: | :---: |
| FSLA | 220 | . 80 | 176 |  |
| Deliberate Practice | 240 | . 20 | 48 |  |
|  |  | Final Ra | ating |  |
| Leadership Practice |  | Effectiver |  |  |
| $\begin{gathered} \text { Concordant } \\ \text { Score } \\ \hline \end{gathered}$ |  | 2.8 |  |  |


| Comments | nts Expand/Collapse All | Rating | $\begin{gathered} \text { Professional } \\ \text { Development Links } \\ \hline \end{gathered}$ |
| :---: | :---: | :---: | :---: |
| $P$ | Domain 1: Student Achievement ( $20 \%$ of the FSLA Score) <br> The focus is on leadership practices that impact prioritization and results for student achievement on priority learning goalsknowing what's important, understanding what's needed and taking actions that get results. |  |  |
| No Comments |  |  |  |
|  | Proficiency Area 1. Student Learning Results: <br> Effective school leaders achieve results on the school's student learning goals and direct energy, influence, and resources toward data analysis for instructional improvement, development and implementation of quality standards-based curricula. |  |  |
| No Comments |  |  |  |
| $\bigcirc{ }^{\text {L }}$ | L1. - Academic Standards: The leader demonstrates understanding of student requirements and academic standards (Common Core and NGSSS). | E V |  |
| No Comments |  |  |  |
| $\bigcirc \frac{1.2}{\text { and }}$ | 1.2 - Performance Data: The leader demonstrates the user of student and adult performance data to make instructional leadership decisions. | E V |  |
| No Comments |  |  |  |
| $\bigcirc{ }^{\text {an }}$ | L.3. Planning_and Goal Settinge The leader demonstrates planning and goal setting to improve student achievement. | E V |  |
| No Comments |  |  |  |
| $\bigcirc \frac{1 .}{}$ | 1.4-Student Achievement Results: The leader demonstrates evidence of student improvement through student achievement results. | E V |  |

Cycle of Teacher Evaluation - Student Growth


## Cycle of Teacher Evaluation - Student Growth



| $\pi$ | TES Documents Book 2 <br> For Classroom and Non-Classroom Teachers |  |  |
| :---: | :---: | :---: | :---: |
|  | STEP | TASK | ARTIFACTS |
|  | 1 | Determine who is a $\mathrm{Cl}, \mathrm{NCl}$, or Administrator | Emails, Board Reports |
|  | 2 | Review Course Code Lists and update evaluation codes | Roster File, Appendix A |
|  | 3 | Roster Verification for Survey 2 and 3 | Roster Verification Documentation |
|  | 4 | Completion of all PRIDEs and Teacher Evaluation Forms | PRIDE Forms |
|  | 5 | PRIDE Raw Score Conversion | PRIDE Concordant Table, Code |
|  | 6 | Teacher/Administrator Grouping | Group Procedures, Code |
|  | 7 | Matrix Model Analyses | AP Matrix, Documentation, Codes |
|  | 8 | Concordant for Matrix Models | Documentation, Syntax |
|  | 9 | Group 1 Regression | Documentation, Syntax, State vs. Local Regression Analysis |
|  | 10 | Group 1 Regression Concordant | Documentation, Syntax, Concordant Tables |
|  | 11 | Group 2 Regresson | Documentation, Syntax |
|  |  |  | 27 |


| $\pi$ | TES Documents Book 2 <br> For Classroom and Non-Classroom Teachers |  |  |
| :---: | :---: | :---: | :---: |
|  | STEP | TASK | ARTIFACTS |
|  | 12 | Group 2 Regression Concordant | Documentation, Syntax, Concordant Tables |
|  | 13 | Z-Score Analysis | Documentation, Syntax |
|  | 14 | Z-Score Analysis Concordants | Documentation, Syntax, Concordant Tables |
|  | 15 | Aggregated VAMS | FDOE VAM Docs, Concordant Tables |
|  | 16 | Compare Local Student Growth to State Student Growth | State VAM Ratings |
|  | 17 | Compare Grouping to each Student Growth Analysis | Sample of Outcome Data |
|  | 18 | Low Cell Analysis | Low cell business rules, analysis |
|  | 19 | Student Growth Data Aggregated PRIDE and Student Growth Aggregated | SQL-partial code |
|  | 20 | Group changes due to cell size | SQL-partial code |
|  | 21 | Data Review - changes or reanalyze, if necessary | Example of Final Output |
|  | 22 | Data Transfer to IIS system | Final TES Report, SQL - partial code |
|  | 23 | Check online reports, Complete district, school reports | Final TES Reports by School |

## Teacher/Administrator Groups

, Evaluations for classroom and non-classroom instructional personnel must include the student growth component.
> In order to link teachers to student performance, it became necessary to group teachers who have similar parameters based on

- Courses they teach
- Daily function
- Assessment scores available for their students
- Number of schools they serve
- Whether or not they teach individual students over time


## TEACHER GROUP 1

$\pi$
Classroom teacher, majority of students take state or local assessment in subject area taught.

The teacher's evaluation will be based on

- 67\% on PRIDE
- $33 \%$ on the student growth scores based on their students' national, state or district assessments (e.g. FSA, LEOC, AICE)
- $51.15 \%$ of all teachers ( $n=1650$ ) fell into this group in school year 2016-2017.


## TEACHER GROUP 2

Classroom teacher, majority of students take a national or state assessment, but in other subject areas than those taught. The teacher's evaluation will be based on

- 67\% on PRIDE
- $33 \%$ on the student growth scores based on their students' performance on national or statewide ELA assessments standardized tests when available.
- $18.47 \%$ of all teachers ( $n=596$ ) were in this group in school year 2016-2017.


## TEACHER GROUP 3

$\pi$
Classroom teacher, majority of students do not take a national or state assessment.

The teacher's evaluation will be based on

- $67 \%$ on PRIDE
- 33\% on student performance on FSA-ELA school-wide data.
- $6.51 \%$ of all teachers ( $n=210$ ) were in this group in school year 2016-2017.


## TEACHER GROUP 4

$\pi$
Non-classroom teacher, assigned to 1 or 2 specific school(s).
The teacher's evaluation will be based on

- 67\% on PRIDE
- $33 \%$ on a value-added score based on state assessment scores for those students attending the specific schools to which the non-classroom teacher is assigned.
- $7.16 \%$ of all teachers ( $n=231$ ) were in this group in school year 2016-2017.


## TEACHER GROUP 5

$\pi$
Non-classroom teacher, assigned to more than 2 school(s).
The teacher's evaluation will be based on

- $67 \%$ on PRIDE
- $33 \%$ on a value-added score based on state assessment scores for the students in the district as a whole.
- 2.73\% of all teachers ( $n=88$ ) were in this group in school year 2016-2017.


## TEACHER GROUP 6

$\pi$
School-Based Administrators

- 67\% on Administrator Evaluation Form
- $33 \%$ on a value-added score based on state assessment scores for the students in the school as a whole.
- 117 principals and assistant principals were in this group in school year 2016-2017.


## ALL GROUPS

- Group 1
- $51.15 \%(n=1650)$
- Group 1a
- $9.49 \%$ ( $n=306$ )
- Group 2
- $18.47 \%$ ( $n=596$ )
- Group 3
- $6.51 \%(n=210)$
- Group P
- $4.49 \% ~(n=145)$
- Group 4

$$
\text { - } 7.16 \%(n=231)
$$

- Group 5
- $2.73 \%(n=88)$
- Group 6
- $\quad(n=117)$


## All Evaluation Totals

- Classroom Instructional (Groups 1,12,2,3,P)
- 86.96\% ( $n=2907$ )
- Non-Classroom Instructional (Groups 4,5 )
- $9.54 \%$ ( $n=319$ )
- School-Based Administrators (Groups 6)
- $3.50 \%$ ( $n=117$ )


# 4 Analytic Models: 38 Separate Analyses (Refer to TES Book 2, Steps 7-15) 

- VAMs - Covariate Adjusted Regression Model
-Z - Score Model
- Local Regression Models
- Matrix Models


## Model 1: Value Added Models (VAMs)

VAMS - Covariate Adjusted Model for FSA, based on 'Regression’

- Covariates
- Attendance, age, PY achievement, ELL/ESE/Giftedness, class size, student mobility, homogeneity of entering test scores in a class, number of subject relevant courses for which student is enrolled.
- Attendance
- Student: Number of days a student was present during the school year
- Teacher: DOE has not provided rule.

Local Rule: non-extended leave (active more than $1 / 2$ year)

## Model 1: State VAM Model

$\pi$ Models are run separately by grade, subject, and year. In its most general form, the model can be represented as follows

$$
\text { yti }=\text { XiJ }+y t-1, i y 1+y t-2, i y 2+Z 1 i 01+Z 2 i 02+e t i
$$

Where:

- $y_{t i}$ is the observed score at time $t$ for student $i$.
- $X_{i}$ is the matrix for the student and classroom demographic variables for student $i$.
- $J J$ is a vector of coefficients capturing the effect of any covariates included in the model except prior test score.
- $y_{t-r, i}$ is the prior test score at time $t-r(r \in\{1,2\}$.
- $y_{l}$ is the coefficient capturing the effects of the most recent prior test score.
- $y_{2}$ is the coefficient capturing the effects of the second prior test score. Elsewhere in this document, $y_{l}$ and
- $y_{2}$ are concatenated such that $y^{\prime}=\left\{y_{b} y_{2}\right\}$ is the coefficient vector capturing the effects of up to two prior test scores.


## Model 1: State VAM Model

## $\pi$

 and Where:- $\quad Z_{1 i}$ is a design matrix with one column for each teacher and one row for each student record in the data file. The entries in the matrix indicate the association between the student record represented in the row and the teacher represented in the column.
- $0_{1}$ is the vector of teacher random effects.
- $Z_{2 i}$ is a design matrix with one column for each school and one row for each student record in the data file. The entries in the matrix indicate the association between the student record represented in the row and the school represented in the column. Elsewhere in this document, $Z_{1}$ and $Z_{2}$ are concatenated such that $Z=\left\{Z_{1}, Z_{2}\right\}$.
- $O_{2}$ is the vector of school random effects. Corresponding to $Z=\left\{Z_{1}, Z_{2}\right\}$, define $0^{\prime}=\left(0^{1,}, 0^{1}\right)$.
- $\mathrm{e}_{\mathrm{ti}}$ captures all residual student-level factors contributing to student achievement.

Because Florida's VAM model treats these vectors of effects as random and independent from each other, it is assumed that the distributions of teacher and school effects are approximately normal about a mean of $0\left(0_{q} \sim N\left(0, a^{2}\right)\right)$ for each level of $q$ where $q \in\{1,2\}$,
with 1 referencing teacher and 2 referencing school. In the subsequent sections, the notation $\mathrm{li}^{\prime}=$ $\left\{P^{\prime}, y^{\prime}\right\}$ is used to refer to the covariate coefficient vectors collectively, and $W=\{X, y t-1, y t-2\}$ is used to refer to the covariate values collectively in order to simplify computation and explanation.

## Model 2: Local Regression

$\pi$, Uses information about a student's prior performance to make a prediction on a future test

- Must have a pre-measure and post-measure
- Not a pre and post test model

Ability to control for factors outside the teachers' control (ESE status, ELL Status, gifted, school, PY performance)

- Must have access to test and demographic data

Measures growth by comparison to a sample of other similar students
Based on this comparison, a determination is made about the 'value-added' as a result of the teachers' influence

## Model 3: Z-Score Differences

, Ability to control for factors outside the teachers' control (ESE status, ELL Status, gifted, school, PY performance)

- Must have access to test and demographic data
, Measures growth by comparison over time on a similar assessment
> Based on this comparison, a determination is made about the 'value-added' as a result of the teachers' influence


## Model 4: Matrix Model

, Ability to control for factors outside the teachers' control (ESE status, ELL Status, gifted, school, PY performance)

- Must have access to test and demographic data
, Measures growth by comparison over time on a similar assessment
, Based on this comparison, a determination is made about the 'value-added' as a result of the teachers' influence


## Calculating the Final Rating

Numerous Scales Based on Test data and analysis used
Placed on a common scale and weighted to produce a Concordant Score between 0.00 and 4.00


Scores are Combined/Weighted to produce a final rating between 0.00 and 4.00 corresponding to a Final Rating of Unsatisfactory, Needs Improvement, Effective, Highly

## Precautionary Measures for Student Growth Calculations

$\pi$

- Extensive training on PRIDE and student growth components
- Offer multiple opportunities for teacher input
- Offer assistance on interpretation all year
- Monitor small cell sizes
- Use PRIDE only at Oak Park (small cell sizes) and STC (adult students)
- All student growth analysis are conducted twice to ensure reliability
- Student growth results are monitored carefully for accuracy


## Precautionary Measures for Student Growth Calculations

- Use larger distributions when possible - look for a state, national, international criterion:
- State VAMS
- Matrix
- For local regressions: mimic the state model
- Use Z-scores so different assessments are on common scales
- Give the benefit to teachers
- Did not use the state cut scores model
- Assume teachers' performance are at the highest end of the confidence interval (i.e. the highest possible VAM)


## What is a Concordant Score?

- In order to weight and combine information from different test scales, the data must be put on a common scale.
- Additionally, in order to combine student growth data across years, individual data is put on a common scale and aggregated across years.
- In order to combine the PRIDE and student growth data, information from both measures were put on a four-point scale and then aggregated.



## Teacher Example - 1

Evaluation Components
FSLA/PRIDE
Student Growth Component
Final Score

Details
Details

Concordant Points (0-4) Possible


Weight in Total Appraisal
67\%
33\%

Number of Years of Data Included
1
1

## Evaluation Components



Final Rating $=$ 3.00 Highly Effective

| Concordant Range |  |  |
| :--- | :---: | :---: |
|  | Minimum | Maximum |
| Highly Effective | 3.00 | 4.00 |
| Effective | 2.00 | 2.99 |
| Needs Improvement | 1.00 | 1.99 |
| Unsatisfactory | 0.00 | 0.99 |

I acknowledge that I have reviewed this report.
$\square$ Electronically acknowledged by on
Evaluatee Signature:


## Teacher Example - 2

E Classroom Teacher Final Evaluation Report for 16-17
-
Number of Years of Data Included

1
Student Growth Component
oints Earned
Concordant Points (0-4) Possible
Details
Details

| 2.90 |  |
| :--- | :--- |
| 3.17 |  |
| 2.99 |  |

Weight in Total Appraisal
67\%
$33 \%$

Final Rating $=$ 2.99 Effective

| Concordant Range |  |  |
| :--- | :---: | :---: |
|  | Minimum | Maximum |
| Highly Effective | 3.00 | 4.00 |
| Effective | 2.00 | 2.99 |
| Needs Improvement | 1.00 | 1.99 |
| Unsatisfactory | 0.00 | 0.99 |

Resources

## I acknowledge that I have reviewed this report

$\square$ Electronically acknowledged by on

## Discussion/Questions/Feedback

