
Teacher and Leader Preparation Implementation Committee

January 11-12, 2012



Florida Department of Education

Using Student Performance Data in the Evaluation of Teacher and School Leader Preparation Programs

- Recap of the measure adopted by Florida to measure student learning growth – Value-Added analysis
- Discussion of additional data requested at November 9-10 TLPIC meeting

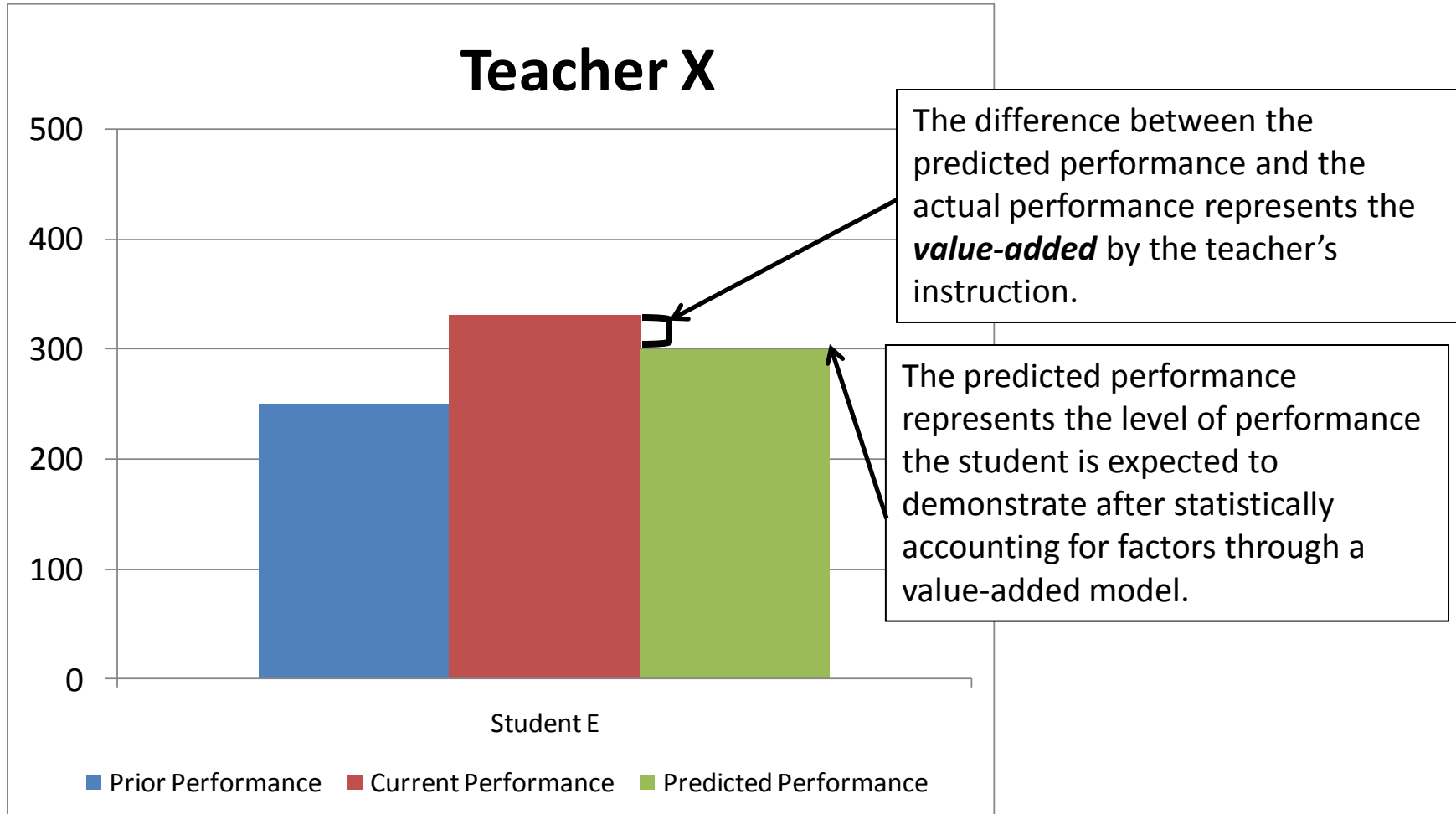


The Measure: Value-Added Analysis

- 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



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.



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



What does a teacher's value-added score represent?

- An estimate of a teacher's impact on student learning, after accounting for other factors that may impact learning.
 - A score of "0" indicates that students performed no better or worse than expected based on the factors in the model
 - A positive score indicates that students performed better than expected
 - A negative score indicates that students performed worse than expected



Value-added results – Additional Considerations

- To account for differences in the FCAT vertical scale across grade levels, subject areas, and years, value-added scores can be combined into one measure.
- Combining the scores into one common measure can provide the scores some context
- Approach used in the analyses both in November and here is to represent the score as proportion of an “average year’s growth”



Value-added results – Additional Considerations

- For example, it has been established that a score of “0” means typical performance
- ***However, what does a score of 20 points mean?***
- It means that students, on average, performed 20 scale points higher than typical
- Transforming that score into a proportion of an “average year’s growth” provides more context and helps describe the magnitude of the gain



Value-added results – Additional Considerations

- Thus, if the average amount of growth in a given grade, subject, and year is 40 scale score points, transforming a score of 20 points into a proportion yields a score of 0.50 (20 divided by 40)
- Now one can interpret the raw value-added score of 20 to say that on average students performed 50% higher than an average year's growth
- These analyses (and those provided in November) use this metric of a proportion of an “average year's growth”



Value-added results – Additional Considerations

- 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



The Standard Error of the Teacher's Value-Added Score

- An estimate of a teacher's impact on student learning contains some variability
- The ***standard error*** is a statistical term that describes the variability
- Standard errors can be used to construct confidence intervals around value-added scores
- These confidence intervals can be used when classifying teachers, schools, and preparation programs into performance categories



Use of the Standard Error in Classification of Teachers

- 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



Review of Additional Data – Student Performance



Florida Department of Education

Student Performance Data – Additional Requested Analyses

- The committee identified additional data requests in reference to the student performance of program completers
- The subsequent slides focus on these requests:
 - Focus on completers teaching in fields they were trained in
 - Investigation of additional thresholds for programs comparison (i.e., various other state averages)
 - Performance data of student subgroups taught by program completers
 - Comparison of student performance data of program completers before and after program completion



Student Performance Data – Additional Requested Analyses

Request #1

- Data with alignment to the program completers teaching in the appropriate subject area



Student Performance Data – Request #1 – Focus on “in field” completers

- These data were provided during AIR’s November 9-10 TLPIC presentation
- There are specific challenges with solely focusing on this subset of completers:
 - Data indicating field of study is currently only available for ITP completers; there is currently no method to discern “in-field” for EPI and DACP completers
 - Solely focusing on these completers further reduces the number of completers on which to base the program evaluation



Student Performance Data – Request #1 – Focus on “in field” completers

ITP		2007-08	2008-09	2009-10
Completers		7,025	7,328	6,493
Completers with VAM Data in Reading and/or Math	#	1,337	1,348	1,299
	%	19.0	18.4	20.0
“In-Field” Completers with VAM Data in Reading	#	707	661	705
	%	10.1	9.0	10.9
“In-Field” Completers with VAM Data in Math	#	554	494	530
	%	7.9	6.7	8.2



Student Performance Data – Request #1 – Focus on “in field” completers

ITP		2007-08	2008-09	2009-10
Reading Program Completers		3,835	3,870	3,568
Reading Program Completers with VAM Data in Reading	#	707	661	705
	%	18.4	17.1	19.8
Math Program Completers		3,456	3,444	3,227
Math Program Completers with VAM Data in Math	#	554	494	530
	%	16.0	14.3	16.4



Student Performance Data – Request #1 – Focus on “in field” completers

Questions to explore

- Should this data be used in the evaluation of ITP programs?
 - Supplemental information?
- Can this data be captured for EPI and DACP programs?
- Overarching challenge of small “n” size



Student Performance Data – Additional Requested Analyses

Request #2

- Could we see the value-added scores of institutions/districts compared against other state averages?



Student Performance Data – Request #2 – Various State Averages

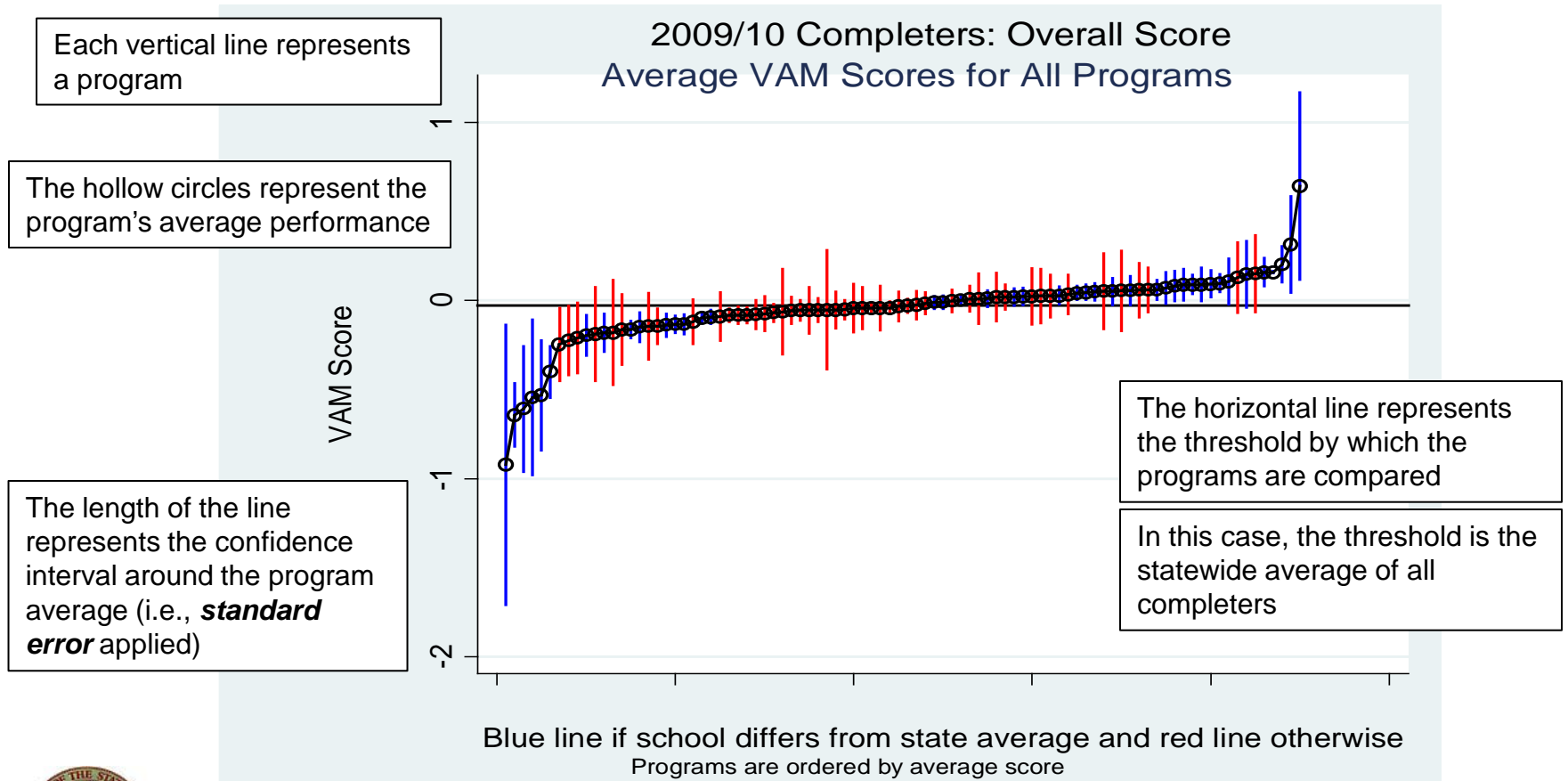
Review

- In November, AIR provided data comparing average VAM performance for institutions/districts to the overall statewide average for all completers in the state
- AIR applied different thresholds of statistical confidence to classify programs in terms of performance ***against the statewide average of all completers***



Student Performance Data – Request #2 – Various State Averages

Review – “Caterpillar” Chart



Student Performance Data – Request #2 – Various State Averages

- The committee requested to investigate different thresholds (i.e., the horizontal line in the “caterpillar” chart) of comparison
- Six different thresholds were identified
 - Performance of all teachers with 0-1 year of experience
 - Performance of all teachers with 0-1 year of experience and advanced degrees
 - Performance of all teachers with less than 5 years of experience
 - Performance of all teachers with less than 5 years of experience and advanced degrees
 - Performance of all teachers with greater than 5 years of experience
 - Performance of all teachers with greater than 5 years of experience and advanced degrees



Various Statewide Averages of Overall (Reading and Math combined) VAM Scores

State Averages – VAM data expressed as a proportion of an average year's growth	Across 3 years (2007-08 to 2009-10)
All Completers	-0.024
Teachers with 0-1 year of experience	-0.024
Teachers with 0-1 year of experience and advanced degrees	0.016
Teachers with less than 5 years of experience	-0.012
Teachers with less than 5 years of experience and advanced degrees	0.008
Teachers with 5 or more years of experience	0.018
Teachers with 5 or more years of experience and advanced degrees	0.026



Various Statewide Averages of Reading VAM Scores

State Averages – VAM data expressed as a proportion of an average year's growth	Across 3 years (2007-08 to 2009-10)
All Completers	-0.027
Teachers with 0-1 year of experience	-0.019
Teachers with 0-1 year of experience and advanced degrees	0.028
Teachers with less than 5 years of experience	-0.013
Teachers with less than 5 years of experience and advanced degrees	0.012
Teachers with 5 or more years of experience	0.019
Teachers with 5 or more years of experience and advanced degrees	0.031



Various Statewide Averages of Math VAM Scores

State Averages – VAM data expressed as a proportion of an average year's growth	Across 3 years (2007-08 to 2009-10)
All Completers	-0.026
Teachers with 0-1 year of experience	-0.033
Teachers with 0-1 year of experience and advanced degrees	-0.015
Teachers with less than 5 years of experience	-0.011
Teachers with less than 5 years of experience and advanced degrees	-0.001
Teachers with 5 or more years of experience	0.015
Teachers with 5 or more years of experience and advanced degrees	0.014



Student Performance Data – Request #2 – Various State Averages

Summary

- Of the six additional thresholds explored, measuring program performance against the standard of experienced teachers with advanced degrees offers the highest threshold (i.e., average of 2.6 percent above an average year's growth (reading and math combined))
- The range of standards explored is tight – ranging from an average of 2.4 percent below an average year's growth (average of program completers) to 2.6 percent above an average year's growth (average of experienced teachers with advanced degrees).



Student Performance Data – Request #2 – Various State Averages

Questions to explore

- On what standard of student performance should programs be evaluated?
 - The average performance of new teachers?
 - The average performance of experienced teachers?
 - Others?
- Should multiple standards for student performance be used in developing a program evaluation system?



Student Performance Data – Additional Requested Analyses

Request #3

- Student performance data – through the value-added model – by student subgroups.
- Compare the performance of student subgroups taught by program completers by institution/district



Student Subgroup Performance – Percent Meeting/Exceeding Expectations – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)

Student Subgroup	Reading	Math
White	50.0	48.9
African American	44.7	46.4
Hispanic	50.6	49.2
Asian	53.7	55.1
Native American	46.7	51.8
Multiracial	49.7	48.4
Free/Reduced Lunch	47.3	48.0
Students with Disabilities	47.8	48.0
English Language Learners	48.1	49.9

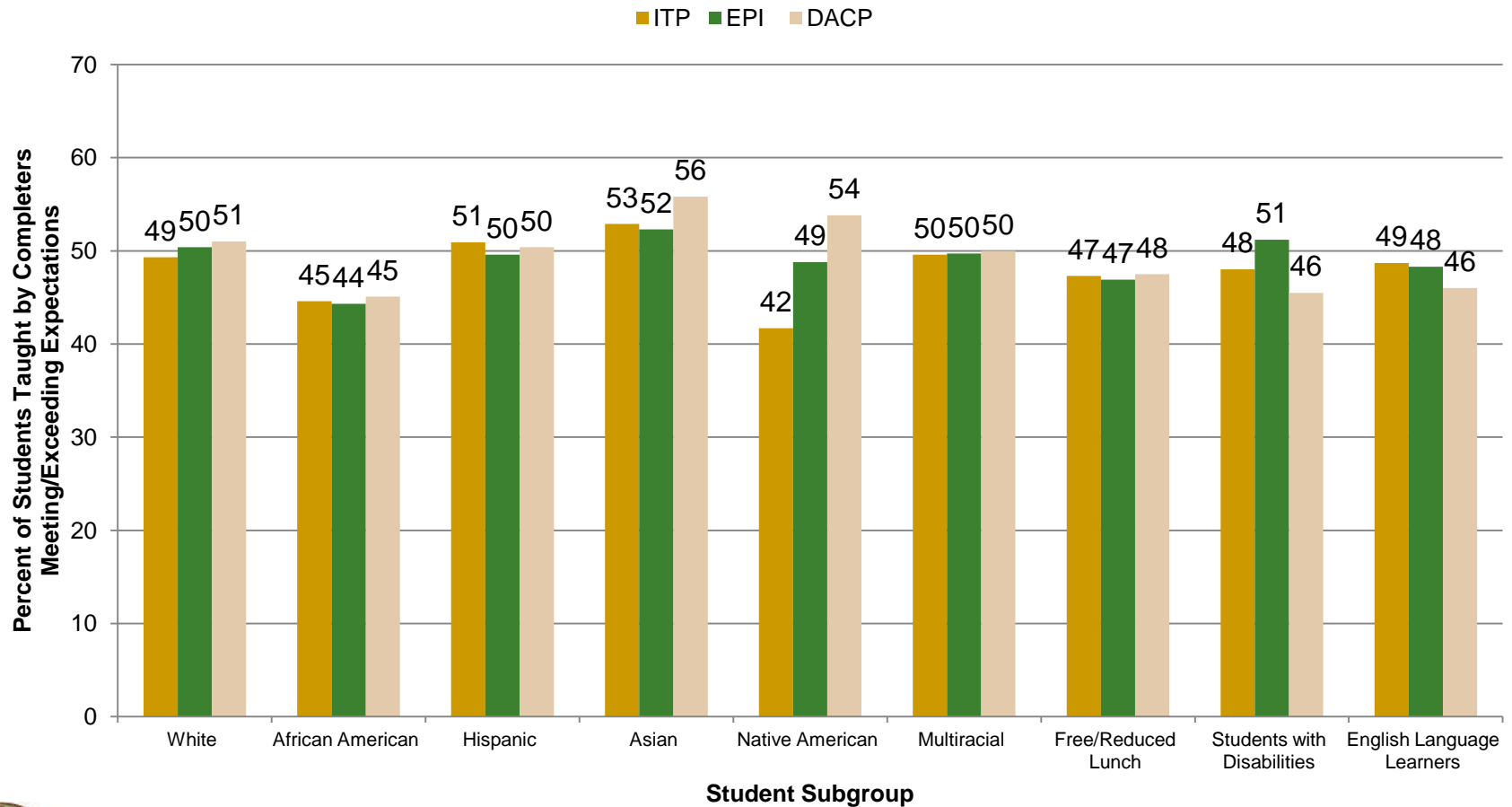


Student Subgroup Performance – Percent of Students Taught by Completers Meeting/Exceeding Expectations in *READING* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)

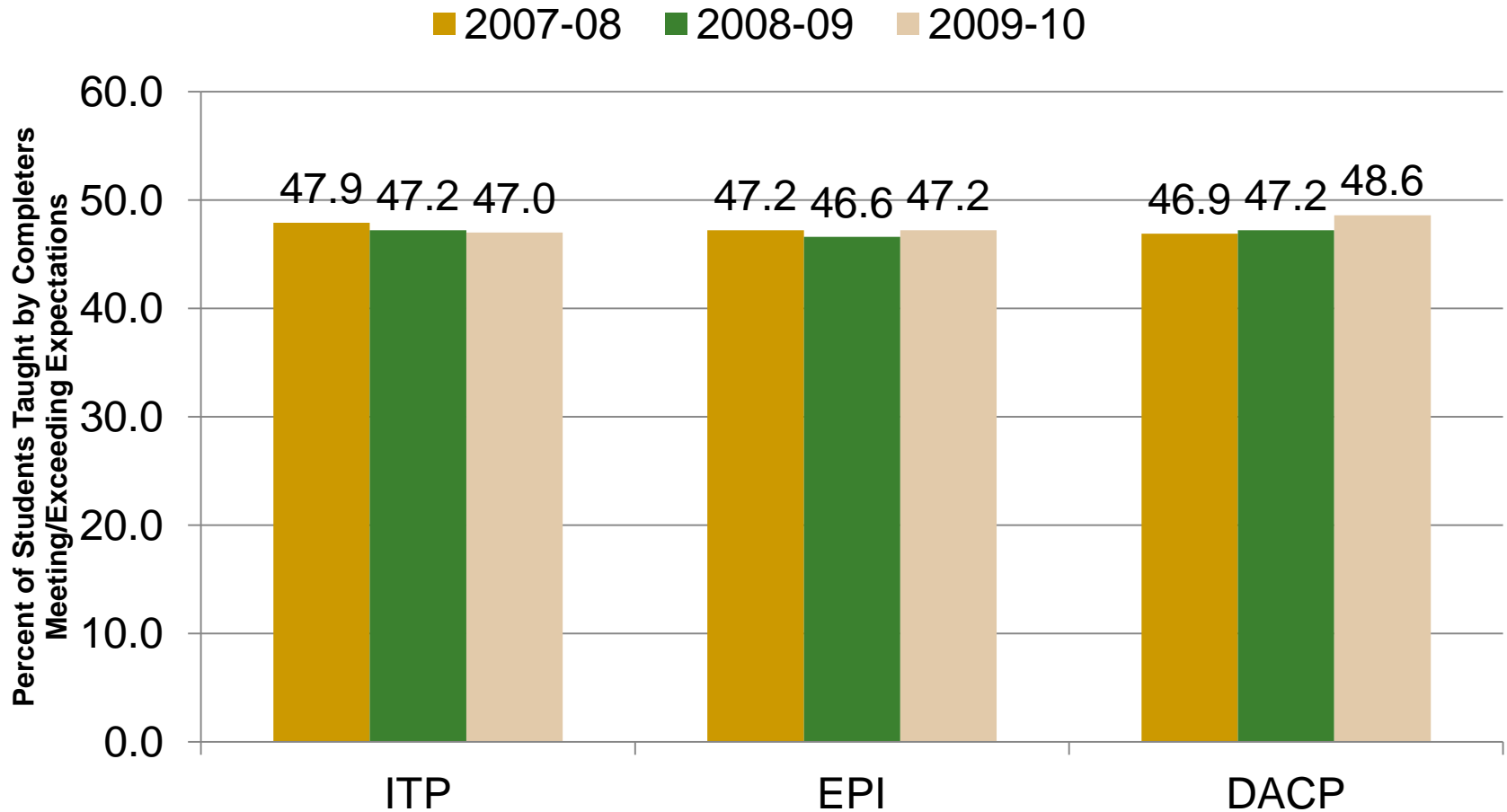
Student Subgroup	ITP	EPI	DACP
White	49.3	50.4	51.0
African American	44.6	44.3	45.1
Hispanic	50.9	49.6	50.4
Asian	52.9	52.3	55.8
Native American	41.7	48.8	53.8
Multiracial	49.6	49.7	50.0
Free/Reduced Lunch	47.3	46.9	47.5
Students with Disabilities	48.0	51.2	45.5
English Language Learners	48.7	48.3	46.0



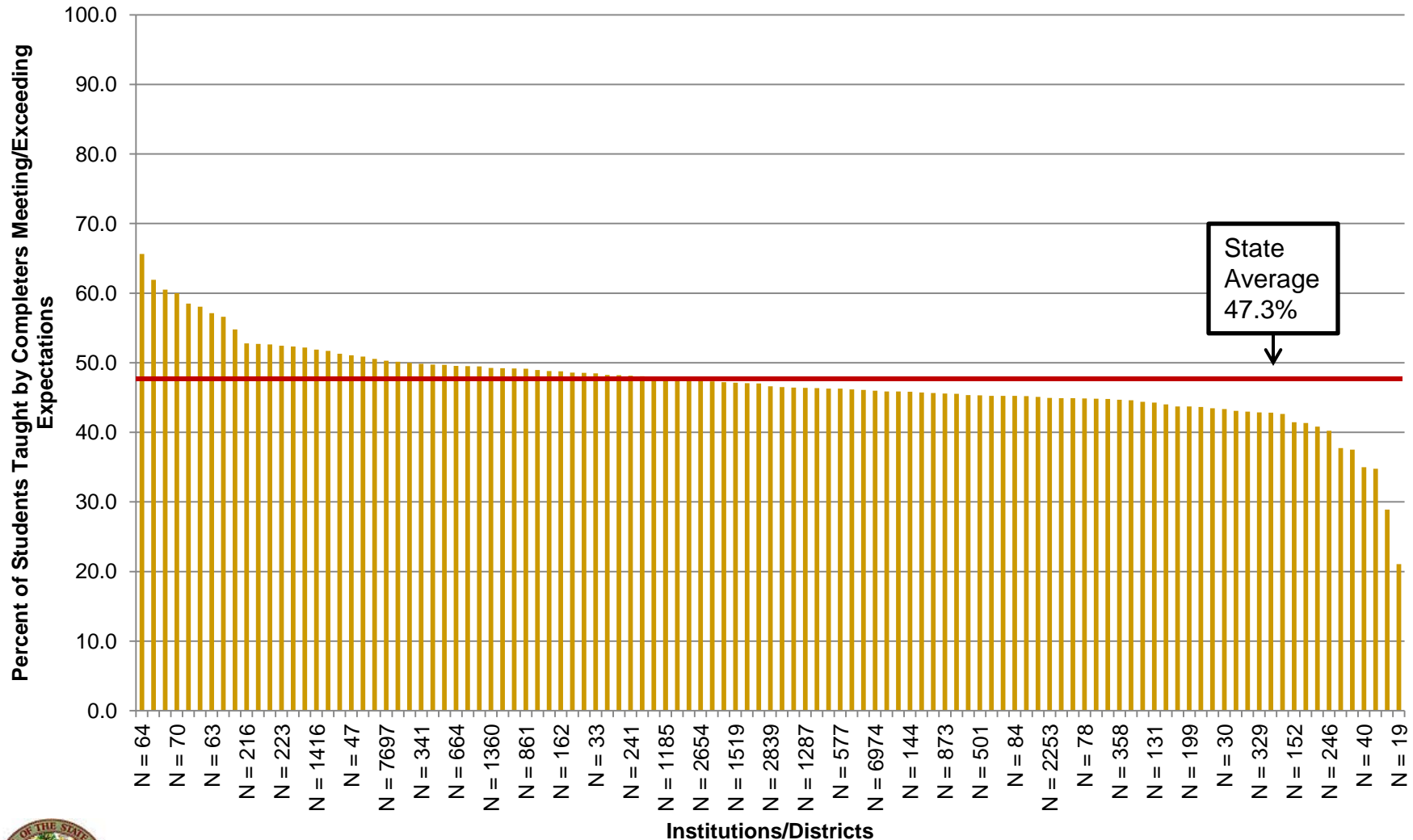
Student Subgroup Performance – Percent of Students Taught by Completers Meeting/Exceeding Expectations in *READING* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)



**Student Subgroup Performance –
Percent of FREE/REDUCED LUNCH Students Taught by Completers
Meeting/Exceeding Expectations in *READING* –
Performance OVER TIME (2007-08, 2008-09, and 2009-10 Completers)**



Percent of *FREE/REDUCED LUNCH* Students Taught by Completers Meeting/Exceeding Expectations in *READING* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)

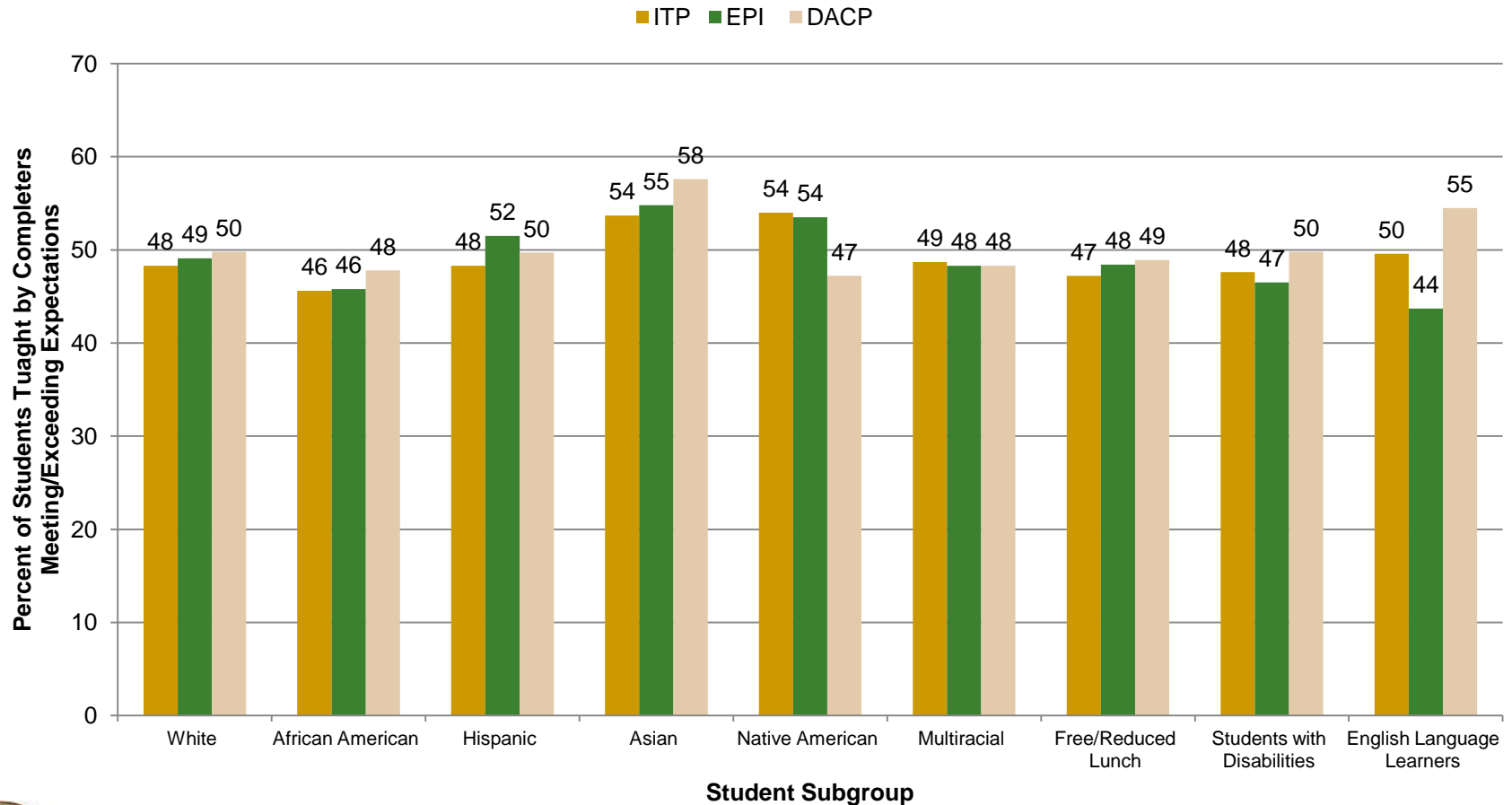


Student Subgroup Performance – Percent of Students Taught by Completers Meeting/Exceeding Expectations in *MATH* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)

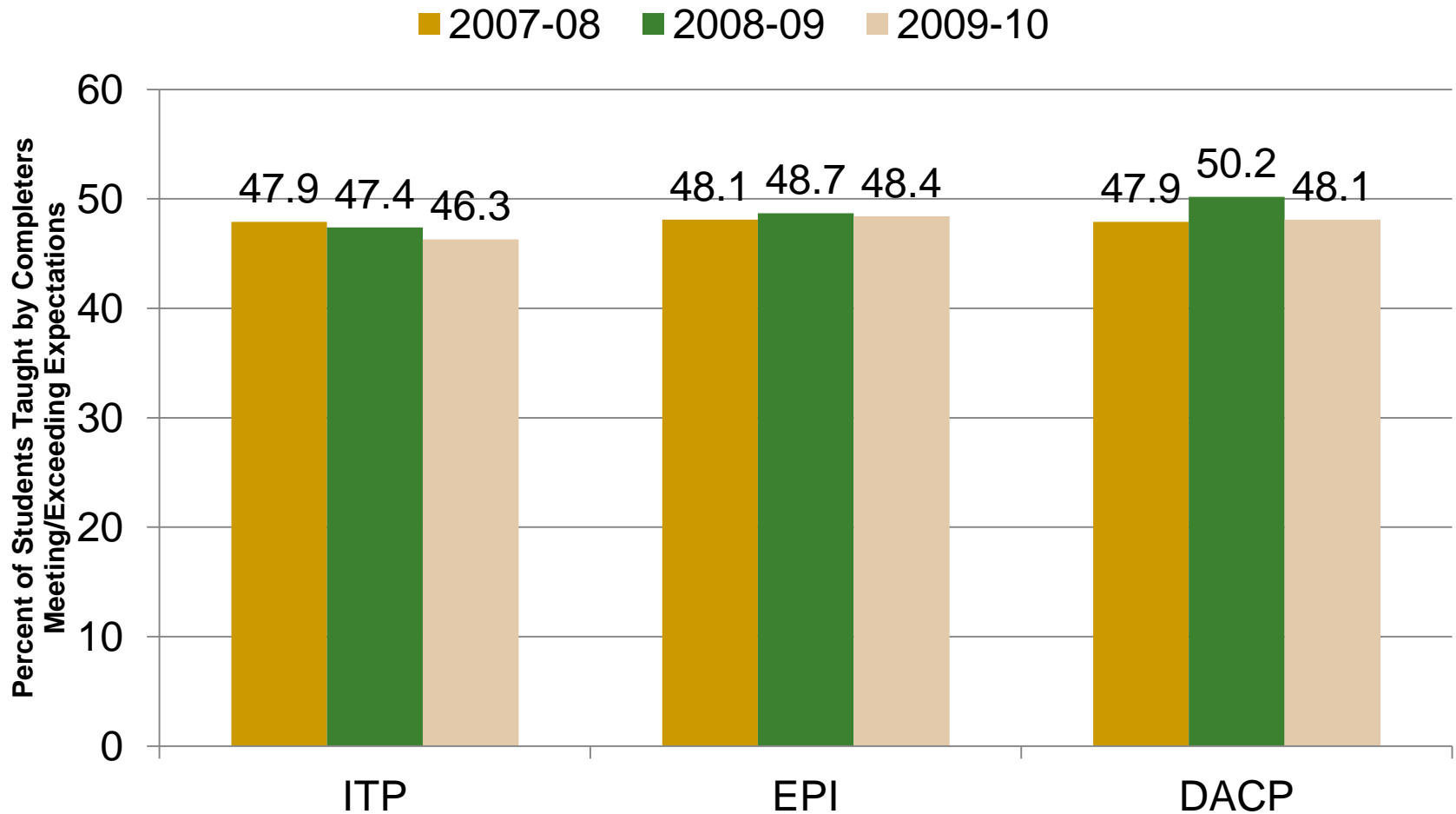
Student Subgroup	ITP	EPI	DACP
White	48.3	49.1	49.8
African American	45.6	45.8	47.8
Hispanic	48.3	51.5	49.7
Asian	53.7	54.8	57.6
Native American	54.0	53.5	47.2
Multiracial	48.7	48.3	48.3
Free/Reduced Lunch	47.2	48.4	48.9
Students with Disabilities	47.6	46.5	49.8
English Language Learners	49.6	43.7	54.5



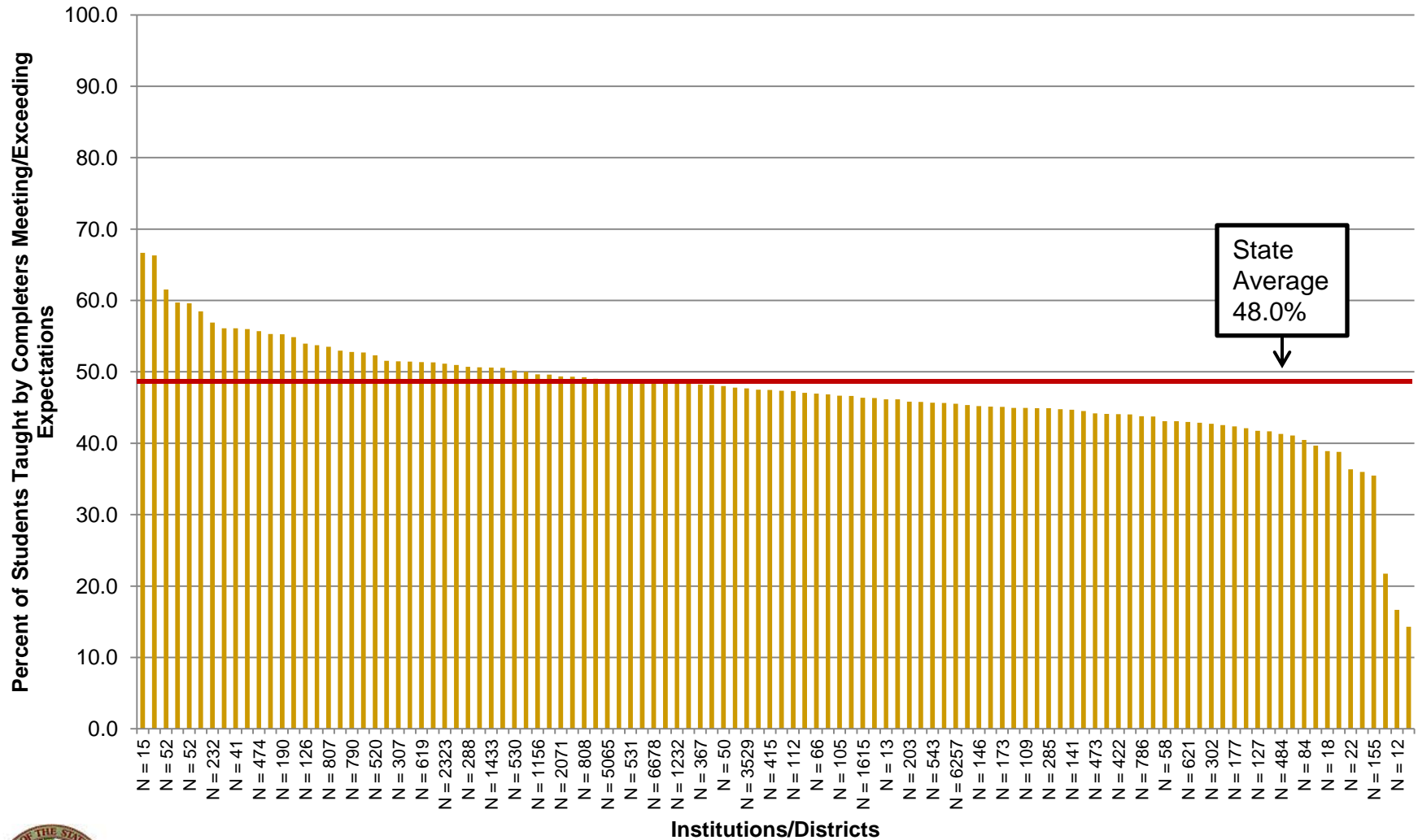
Student Subgroup Performance – Percent of Students Taught by Completers Meeting/Exceeding Expectations in *MATH* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)



**Student Subgroup Performance –
Percent of FREE/REDUCED LUNCH Students Taught by Completers
Meeting/Exceeding Expectations in *MATH* –
Performance OVER TIME (2007-08, 2008-09, and 2009-10 Completers)**



Percent of *FREE/REDUCED LUNCH* Students Taught by Completers Meeting/Exceeding Expectations in *MATH* – All Completers Across Three Years of Performance Data (2007-08 to 2009-10)



Student Performance Data – Request #3 – Student Subgroup Performance

Summary

- Though not providing information on the magnitude of gain, this measure provides insight regarding the quantity of students taught by program completers that meet or exceed expectations, by student subgroup
- Small “n” sizes are less of a concern with this analysis since it is based on the overall number of students taught by program completers, not the number of program completers



Student Performance Data – Request #3 – Student Subgroup Performance

Questions to explore

- Should, and if so, how could this data be included in the evaluation of teacher preparation program?
 - Performance targets?
 - Annual snapshots?
 - Improvement over time?
- Should focus be placed on particular subgroups?



Student Performance Data – Additional Requested Analyses

Request #4

- How are ITP, EPI, and DACP candidates who are already teaching performing prior to or during the program compared to how they perform after program completion?



Comparison of Student Performance Data Before and After Program Completion

2009-10 Program Completers who Student Performance Data in **READING AND/OR MATH** Both Before and After Program Completion

		ITP	EPI	DACP
Total Completers in 2009-10		6,493	1,773	1,236
Completers with VAM data in Reading and/or Math in 2010-11	#	1,299	378	463
	%	20.0	21.3	37.5
Completers with VAM data both before and after program completion	#	381	255	401
	%	29.3	67.5	86.6
Average VAM Score before program completion		-0.025	-0.028	0.008
Average VAM Score after program completion		0.005	-0.005	-0.029



Comparison of Student Performance Data Before and After Program Completion

2009-10 Program Completers who Student Performance Data in READING Both Before and After Program Completion

		ITP	EPI	DACP
Total Completers		6,493	1,773	1,236
Completers with VAM data in Reading in 2010-11	#	1,052	245	301
	%	16.2	13.8	24.4
Completers with VAM data both before and after program completion	#	298	168	265
	%	28.3	68.6	88.0
Average VAM Score before program completion		-0.013	-0.043	0.019
Average VAM Score after program completion		-0.010	-0.043	-0.051



Comparison of Student Performance Data Before and After Program Completion

2009-10 Program Completers who Student Performance Data in MATH Both Before and After Program Completion

		ITP	EPI	DACP
Total Completers		6,493	1,773	1,236
Completers with VAM data in Math in 2010-11	#	829	219	239
	%	12.8	12.4	19.3
Completers with VAM data both before and after program completion	#	203	145	196
	%	24.2	66.2	82.0
Average VAM Score before program completion		-0.032	-0.019	-0.015
Average VAM Score after program completion		0.037	0.058	0.030



Student Performance Data – Request #4 – Performance Before and After Program Completion

Summary

- In nearly all instances, average student performance data improved after program completion
- These data demonstrate the variation in students served by each type of program
 - About 80% of DACP completers with VAM data had student performance data both before and after program completion; whereas about 25-30% of ITP completers with VAM data had student performance data both before and after program completion
- These data demonstrate the challenges of small “n” sizes – with analyses at the pathway level reduced to 200-400 completers



Student Performance Data – Request #4 – Performance Before and After Program Completion

Questions to explore

- Is this a pertinent piece of data for the purpose of program level evaluation?
- Given the different nature of the students served by each program, does it make sense to include this indicator in an evaluation system for all programs, or as supplemental data?

