Reflections on M-STEP?

Dave Treder

AERA

OFFICIAL DOCUMENTS

AERA Statement on Use of Value-Added Models (VAM) for the Evaluation of Educators and Educator Preparation Programs

Approved by AERA Council, June 2015

[T]he phrase *value-added models* is used as an umbrella term to refer to a variety of "true" value added models, student growth percentiles, and certain growth models that are used for evaluation….[that are] are employed in an attempt to determine teachers' and leaders' contributions to student learning outcomes, as captured by standardized tests.

AERA Statement, on Test Transitions

(4) *VAM scores must only be calculated from scores on tests that are comparable over time.*

Many states are currently transitioning to new assessment systems and adopting new or revised performance standards.... Although such changes are to be expected, they pose a threat to the validity of the interpretations of VAM scores, especially when these scores are compared before, across, and after the transition....Transitions in student assessment not only pose difficulties for VAM, but also interrupt longitudinal trends of status measures. Such situations can lead to misinterpretations of student progress.... **Consequently, VAM scores should generally not be employed across transitions**.

Michigan, Accountability, & SGPs

English Language Learners – AMAO 1 – "Progress"

The district must demonstrate that the percentage of its students making "progress" …exceeds the current year's target. The targets change on an annual basis:

2013-14 50% WIDA
2014-15 51% WIDA
2015-16 52% WIDA

Adequate student progress is defined as obtaining an SGP of 50 or more (meaning that the student demonstrated at least the typical amount of progress for students with similar past scores).

--- SPGs are computed from ELPA test results to WIDA test results --

FAQs
9. What are the consequences if a district does not meet AMAOs?
Year 1:
-- district shall send a letter to parents of students participating in a Title III-funded language instruction educational program.

In the State's defense, the USDE was requiring that they do *something* – and USDE did reject the state's initial SPG proposal of 45%

What do you think of this chain of reasoning?

From ***Primer on Student Growth Percentiles***
Damian W. Betebenner

Consider the familiar situation from pediatrics where the interest is on measuring the height and weight of children over time. The scales on which height and weight are measured possess properties that educational assessment scales aspire towards but can never meet.

> *An infant male toddler is measured at 2 and 3 years of age and is shown to have grown 4 inches. The magnitude of increase—4 inches—is a well understood quantity that any parent can grasp and measure at home using a simple yardstick. However, parents leaving their pediatrician's office knowing only how much their child has grown would likely be wanting for more information. In this situation, parents are not interested in an absolute criterion of growth, but instead in a normative criterion locating that 4 inch increase alongside the height increases of similar children. Examining this height increase relative to the increases of similar children permits one to diagnose how (ab)normal such an increase is.*

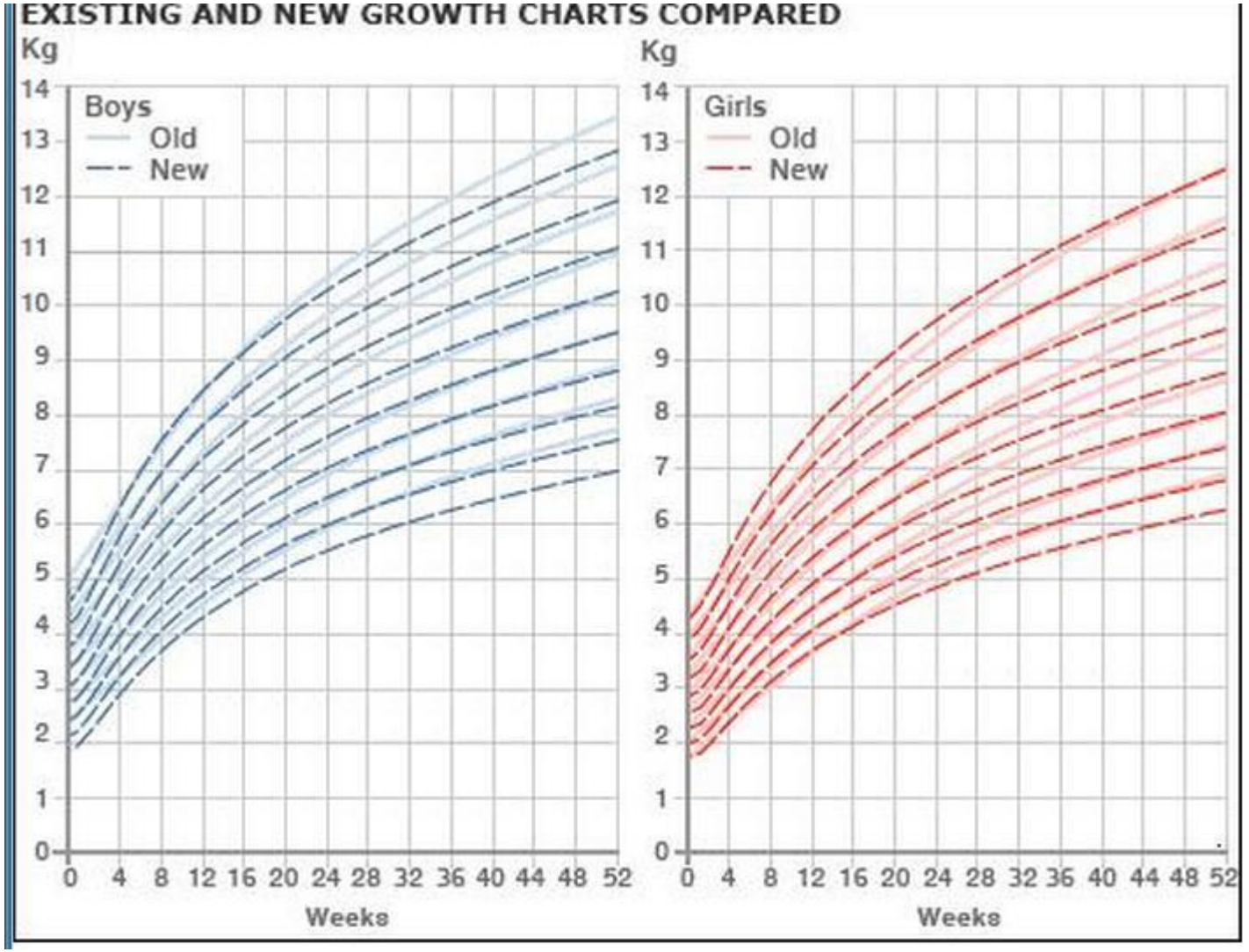
Given this reality in the examination of change where scales of measurement are perfect, it is absurd to think that in education, where scales are quasi-interval, one can/should examine growth differently.

From ***Primer on Student Growth Percentiles*** (continued)

Suppose that scales did exist in education similar to height/weight scales that permitted the calculation of absolute measures of annual academic growth for students. A parent's query about, "How much did my child progress?", would be answered with some quantity of scale score points—an answer that would leave most parents bewildered wondering whether the number of points is good or bad. As in pediatrics, the search for a description regarding change in achievement over time (i.e., growth) is best served by considering a normative quantification of student growth—a student growth percentile.

A student's growth percentile describes how (ab)normal a student's growth is by examining their current achievement relative to their academic peers—those students beginning at the same place. That is, a student growth percentile examines the current achievement of a student relative to other students who have, in the past, "walked the same achievement path". Heuristically, if the state assessment data set were extremely large (in fact, infinite) in size, one could open the infinite data set and select out those students with the exact same prior scores and **compare how the selected student's current year score compares to the current year scores of those students with the same prior year's scores—*their academic peers*.** If the student's current year score exceeded the scores of most of their academic peers, in a normative sense they have done well. If the student's current year score was less than the scores of their academic peers, in a normative sense they have not done well.

Have you ever seen a single "Growth Chart" for kids' weight/height?



What should you "Throw into the Model"?

A Comparison of Growth Percentile and Value-Added Models of Teacher Performance

Cassandra M. Guarino
Mark D. Reckase
Brian W. Stacy
Jeffrey M. Wooldridge
February 6, 2014

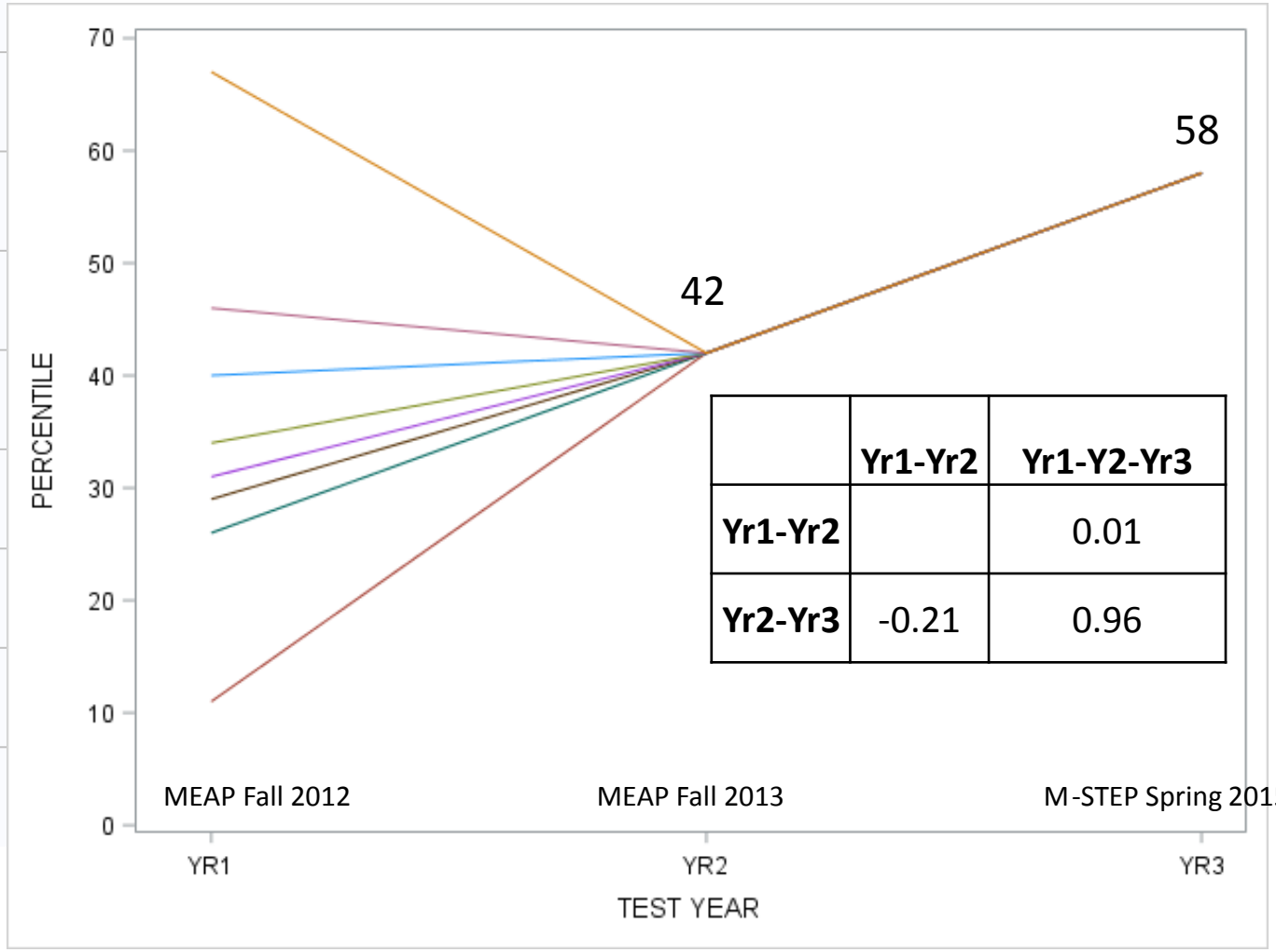
Care should be used by practitioners and researchers in evaluating teachers using these approaches **when nonrandom grouping and assignment** occurs in the school system. More generally, estimators that do not partial out teacher effects–not only growth models, but also value-added models that are relatively descriptive in nature–are less equipped to disentangle true teacher contributions to student achievement from other source of achievement than those that partial out these effects.

Student Growth Percentiles

- MEAP Fall 2012 --> MEAP Fall 2013 --> M-STEP Spring 2015
- Genesee, Oakland, Macomb, Lapeer County Data
  (25,000 5th Graders)

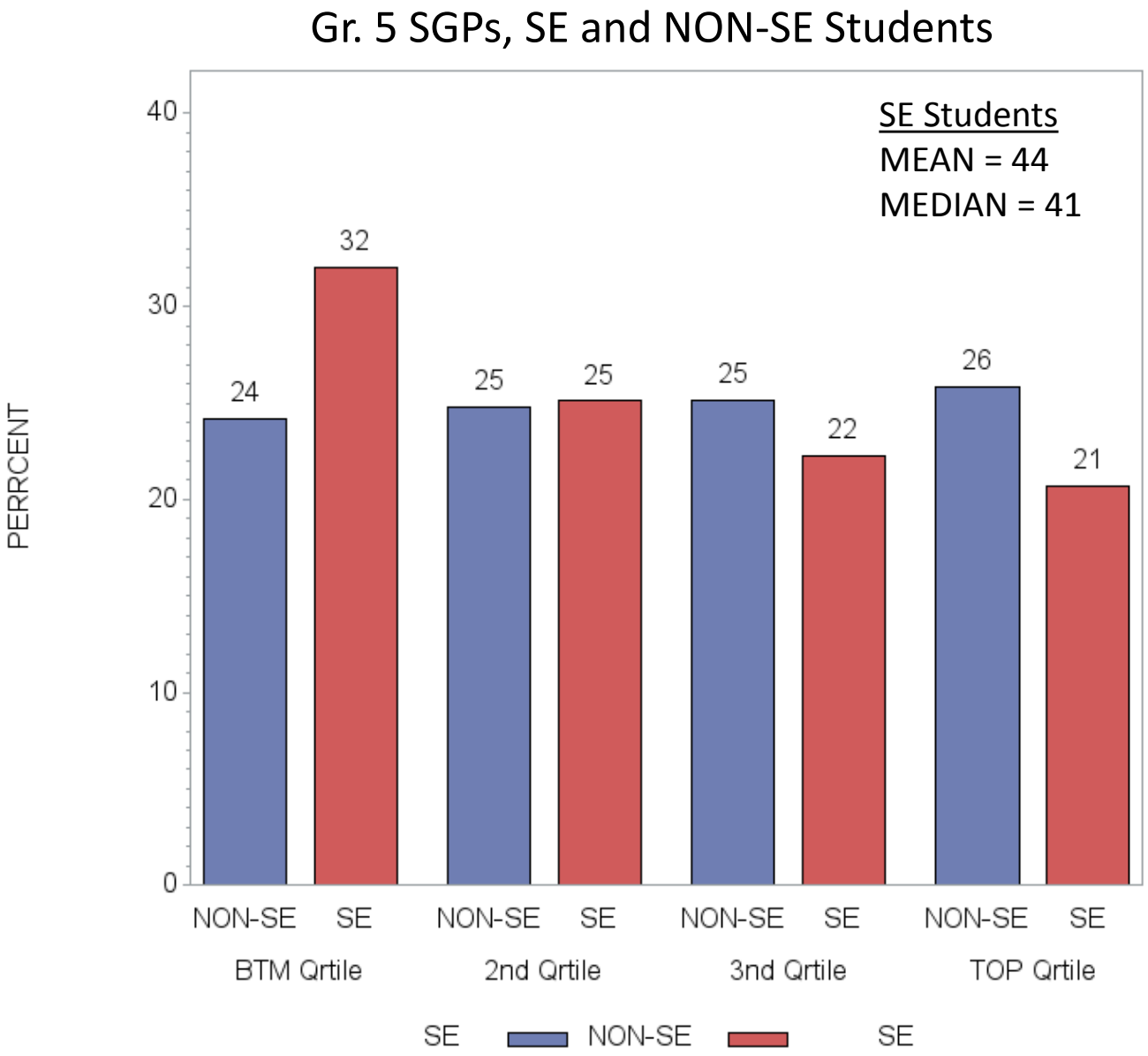
Student "Growth Trajectories"
Different Year 1 Score, Same Year 2 & Year 3 Score

YR 1 PCTILE RANK	YR 1 to YR 3 SGP
67	63
46	74
40	78
34	80
31	82
29	83
26	83
11	89



	Yr1-Yr2	Yr1-Y2-Yr3
Yr1-Yr2		0.01
Yr2-Yr3	-0.21	0.96

Back to "What to Throw into the Model"



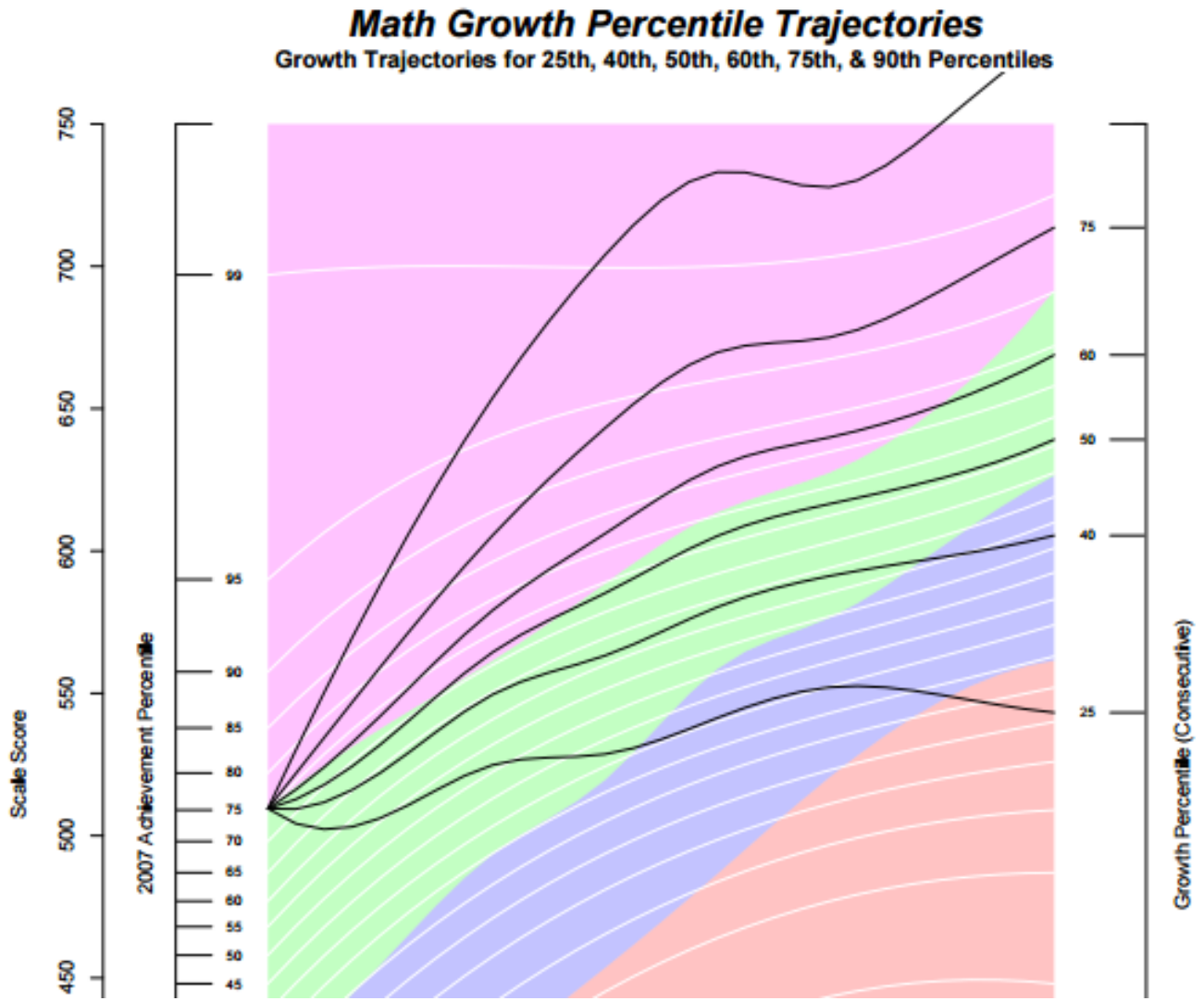


Figure 6: Growth chart depicting future math achievement conditional upon consecutive 25th, 40th, 50th, 60th, 75th, and 90th percentile growth for a student beginning the third grade at the proficient/advanced cutpoint

SPGs – Student Level Error

- MEAP Fall 2012 --> MEAP Fall 2013 --> M-STEP Spring 2015
- Genesee, Oakland, Macomb, Lapeer County Data (25000 5th Graders)

Simulation, 100 iterations:

- Within each iteration, for each student, for their Fall2013 MEAP, Fall 2013 MEAP, & Spring 2015 M-STEP score:

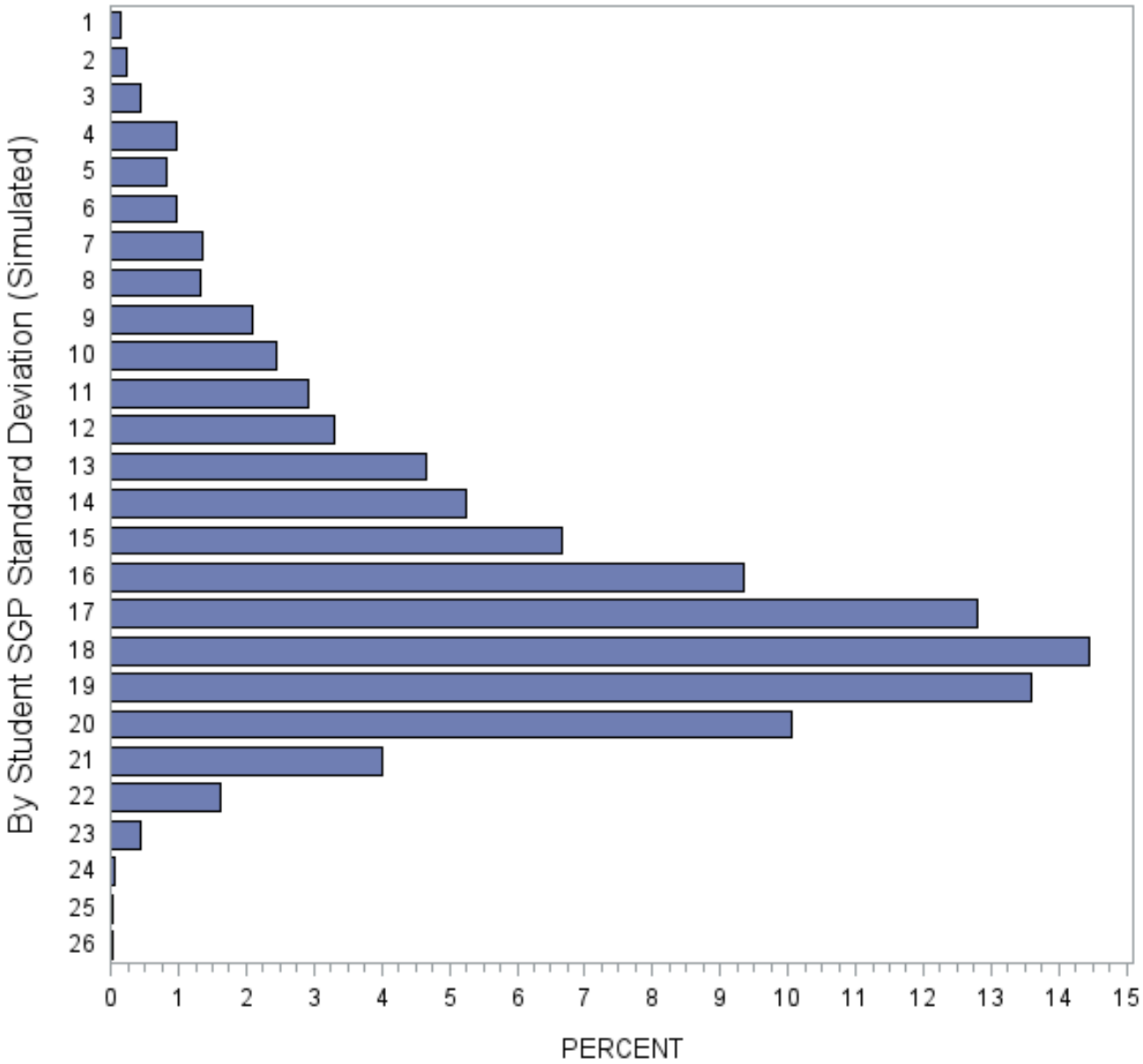
-- > Scale Score + (StandardError*Random Normal Variate)

```
data mtA; set mstep15.GEN12_13_15 mstep15.LAP12_13_15 mstep15.Mac12_13_15
mstep15.Oak12_13_15;
where gr_yr3 = 5;
if gr_yr3 - gr_yr2 ne 1 then  delete;
if gr_yr2 - gr_yr1 ne 1 then gr_yr1 = .;

%macro x;
%do i = 1 %to 100 ;
data MT&i; retain id1  g_yr1 g_yr2 g_yr3  ss_yr1 ss_yr2 ss_yr3; set mtA;
ss_yr1 = int(mathss_yr1 + (rannor(0)*mathssse_yr1));
ss_yr2 = int(mathss_yr2 + (rannnor(0)*mathssse_yr2));
ss_yr3 = int(mathss_yr3 + (rannor(0)*mathssse_yr3));
  keep id1  g_yr1 g_yr2 g_yr3  ss_yr1 ss_yr2 ss_yr3;
run;
%end;
%mend;
%x;
```

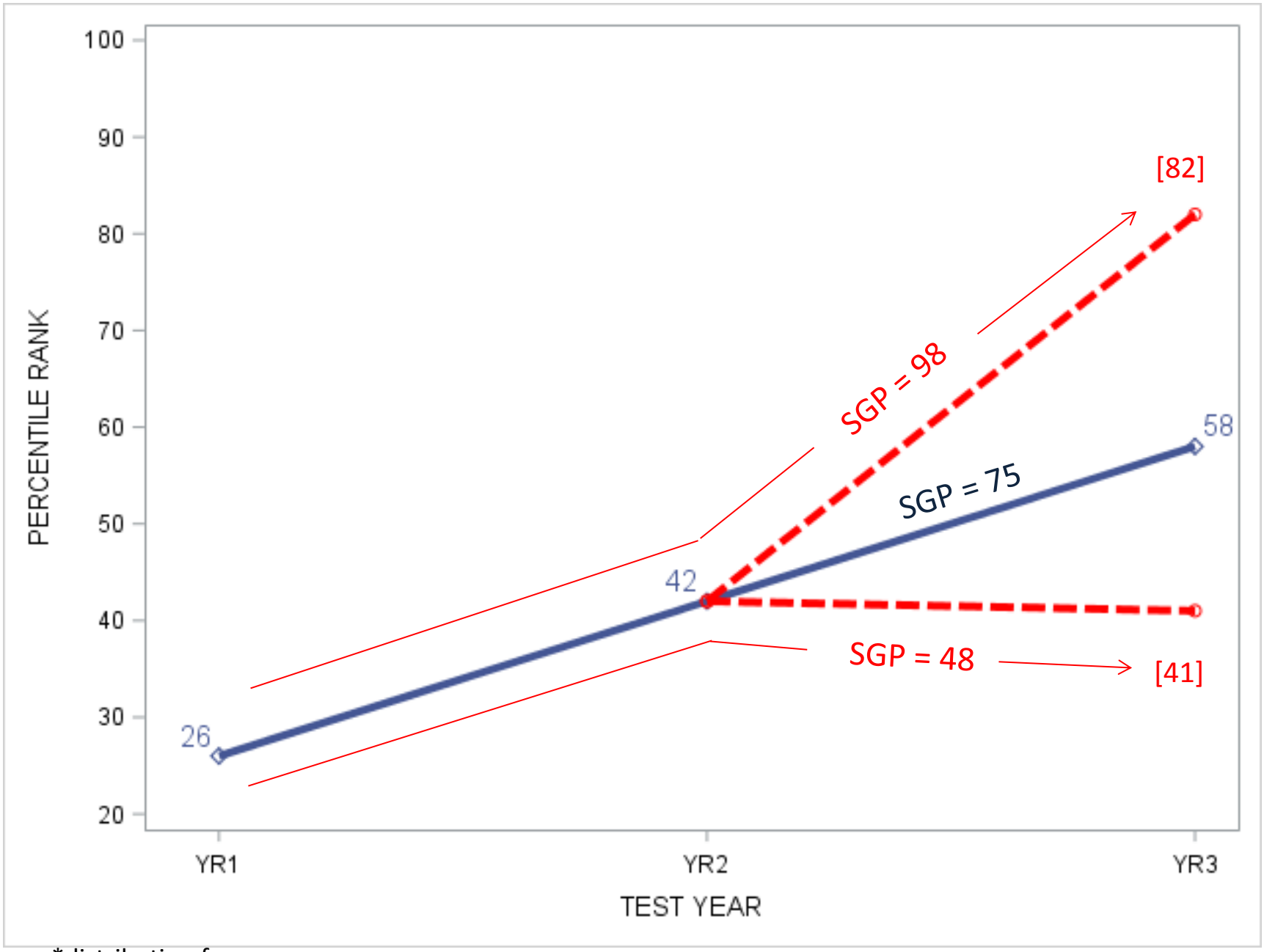
Student-Level SGP Standard Deviation (form the 100 runs)

Genesee & Lapeer County Students



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
1	8	8	0.15	0.15
2	13	21	0.24	0.40
3	23	44	0.43	0.83
4	51	95	0.96	1.79
5	43	138	0.81	2.60
6	51	189	0.96	3.56
7	72	261	1.36	4.91
8	70	331	1.32	6.23
9	111	442	2.09	8.32
10	130	572	2.45	10.77
11	155	727	2.92	13.69
12	176	903	3.31	17.00
13	247	1150	4.65	21.65
14	278	1428	5.23	26.89
15	354	1782	6.67	33.55
16	497	2279	9.36	42.91
17	679	2958	12.78	55.70
18	768	3726	14.46	70.16
19	722	4448	13.59	83.75
20	534	4982	10.05	93.81
21	213	5195	4.01	97.82
22	86	5281	1.62	99.44
23	24	5305	0.45	99.89
24	3	5308	0.06	99.94
25	1	5309	0.02	99.96
26	2	5311	0.04	100.00

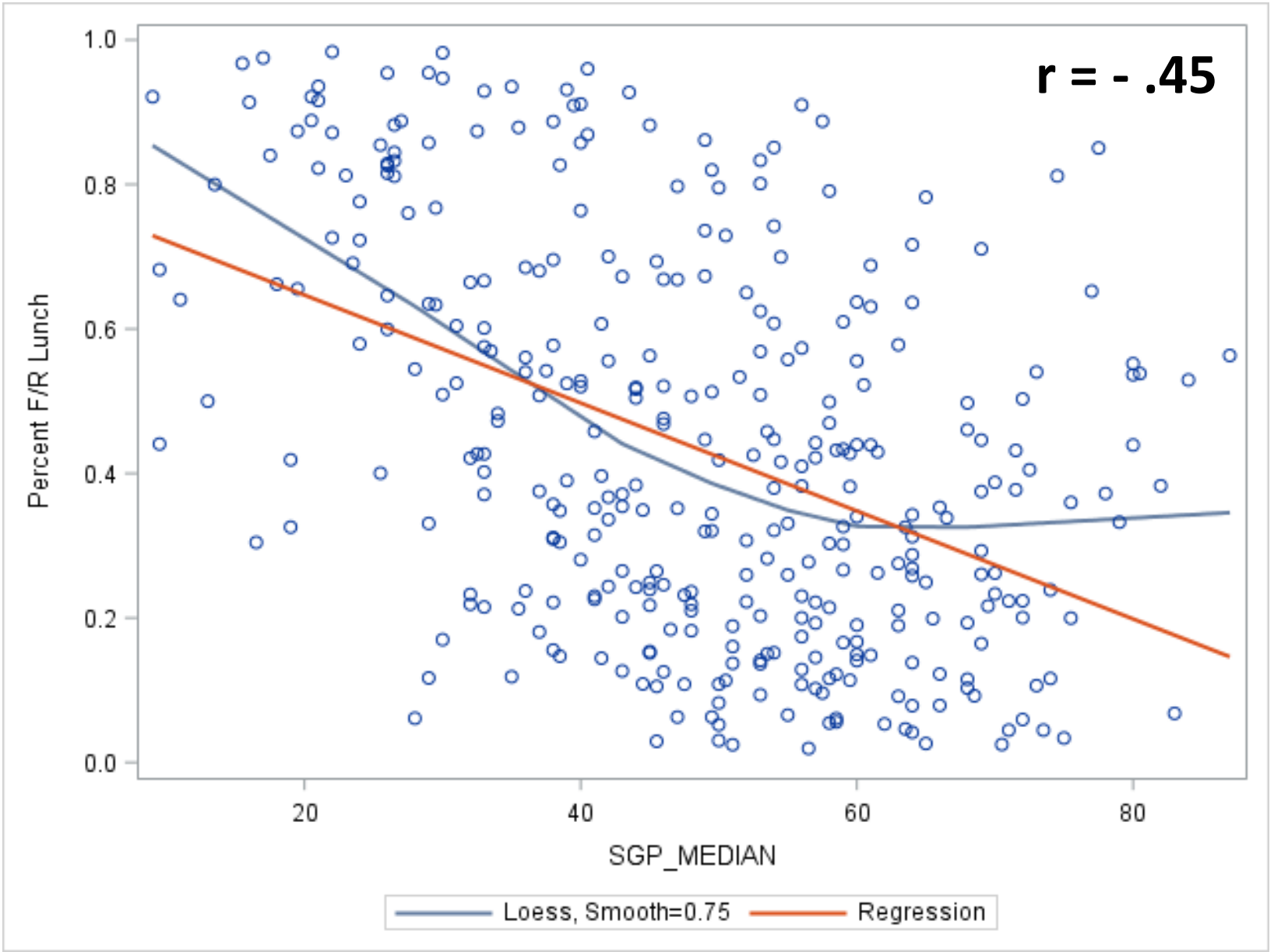
Single Student - Student Growth Percentile w/ 90% Confidence Interval*



*distribution free

Math – School-Level SGP Median (N = 349)

SGP (Yr1-Yr2- Yr3) X Percent Free/Reduced Lunch



M-STEP Paper-and-Pencil vs. On-Line Testing

To answer the question of *Form Bias*, I computed, at the kid level:
(a) the change in Percent Proficient (from Fall 2013 MEAP to Spring 2015 M-STEP)
Then, aggregated this at the building level
-- so, the resulting value, for each school, is the + or – percent of kids proficient, compared to the last test administration.

Number of Schools (used schools in Genesee & Lapeer Counties)

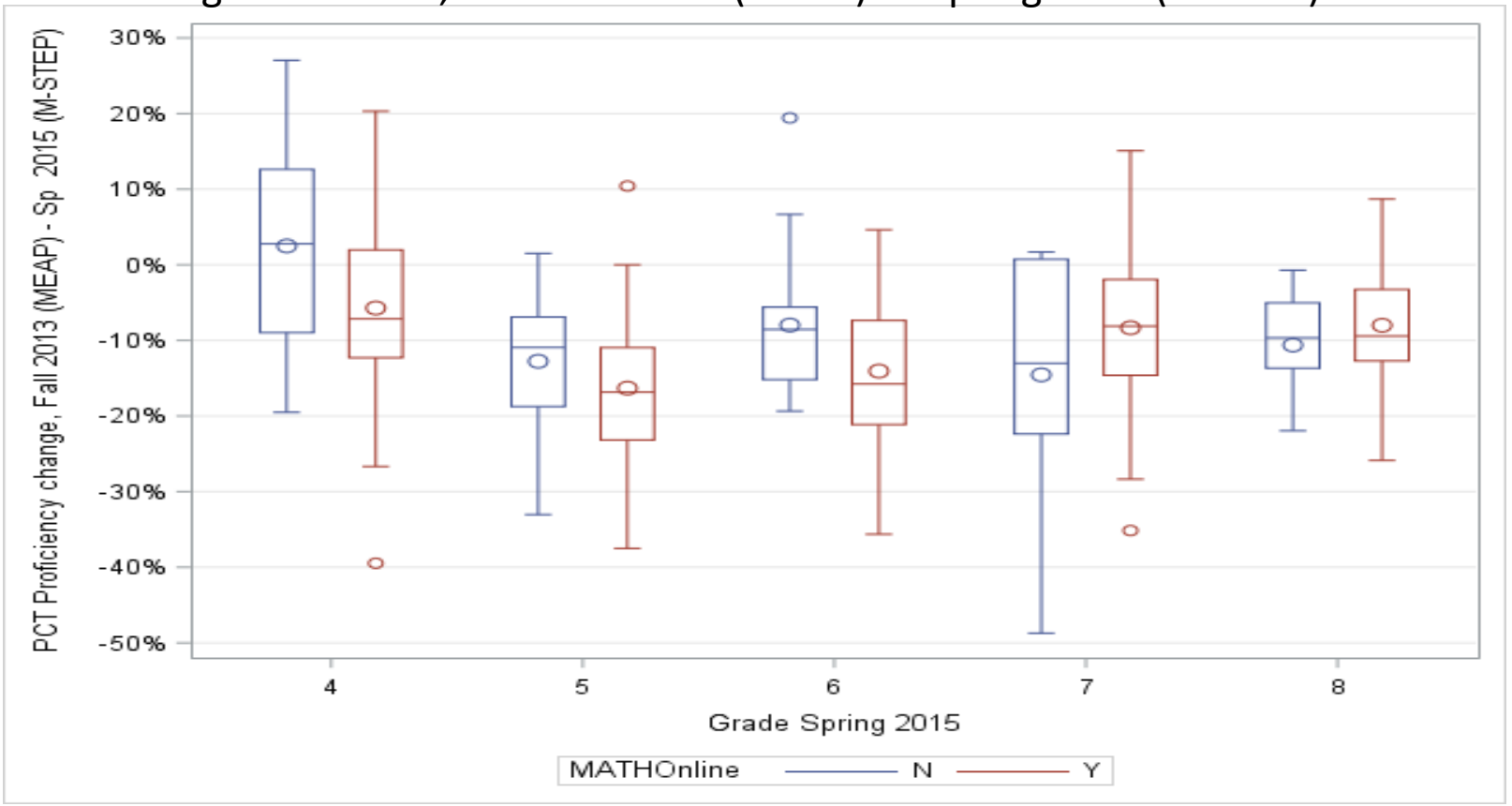
	4	**5**	**6**	**7**	**8**
P & P	21	21	18	16	11
On-Line	58	57	55	37	32

Avg Bldg. F/R Lunch Percent

P & P	69%	73%	74%	65%	65%
On-line	57%	57%	63%	62%	61%

MATH, P & P vs On-Line

Change in Pct Prof, from Fall 2013 (MEAP) to Spring 2015 (M-STEP)



Change in Pct Prof, from MEAP to M-STEP

P & P	**2%**	**-13%**	**-8%**	-15%	-11%
On-Line	-6%	-16%	-14%	-8%	-8%
P & P Advantage?	**8%***	**3%**	**6%***	**-7%**	**-3%**

Avg Bldg F/R Lunch Percent

P & P (N)	69%	73%	74%	65%	65%
On-Line (Y)	57%	57%	63%	62%	61%

(ES = .8 and.7)

Thanks!