Analysis of Paper & Pencil / On-line M-STEP Testing

MERA Fall Conference
November 22, 2016
Dave Treder

Spotlight

MICHIGAN Department of Education

on Student Assessment and Accountability

December 17, 2015

Happy Holidays!

Comparability Between Online and Paper/Pencil Modes

Please note: We are aware that there is the potential for differences between scores of schools that administered M-STEP online, and schools that did a paper/pencil administration. This is not unexpected, given both the transitional nature of this first year of assessment and two administration modes. We are going to be conducting a statewide comparability study this winter and will present our findings in a future Spotlight.

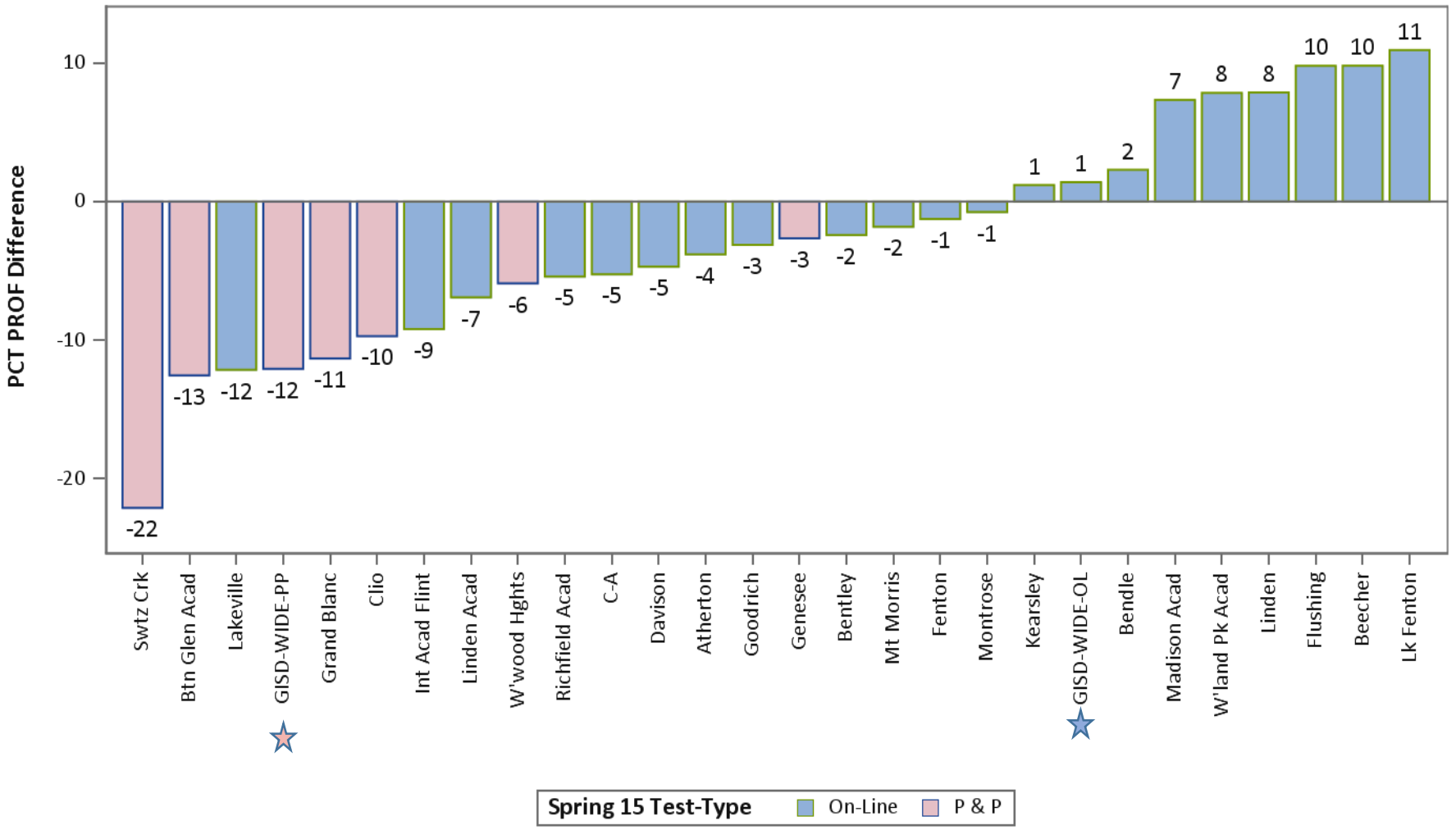
Am I the only one who missed the presentation of findings?

ELA M-STEP - Spring 2015 to Spring 2016

(DIFFERENT STUDENTS)

Difference in Percent Proficient

Grade = 8

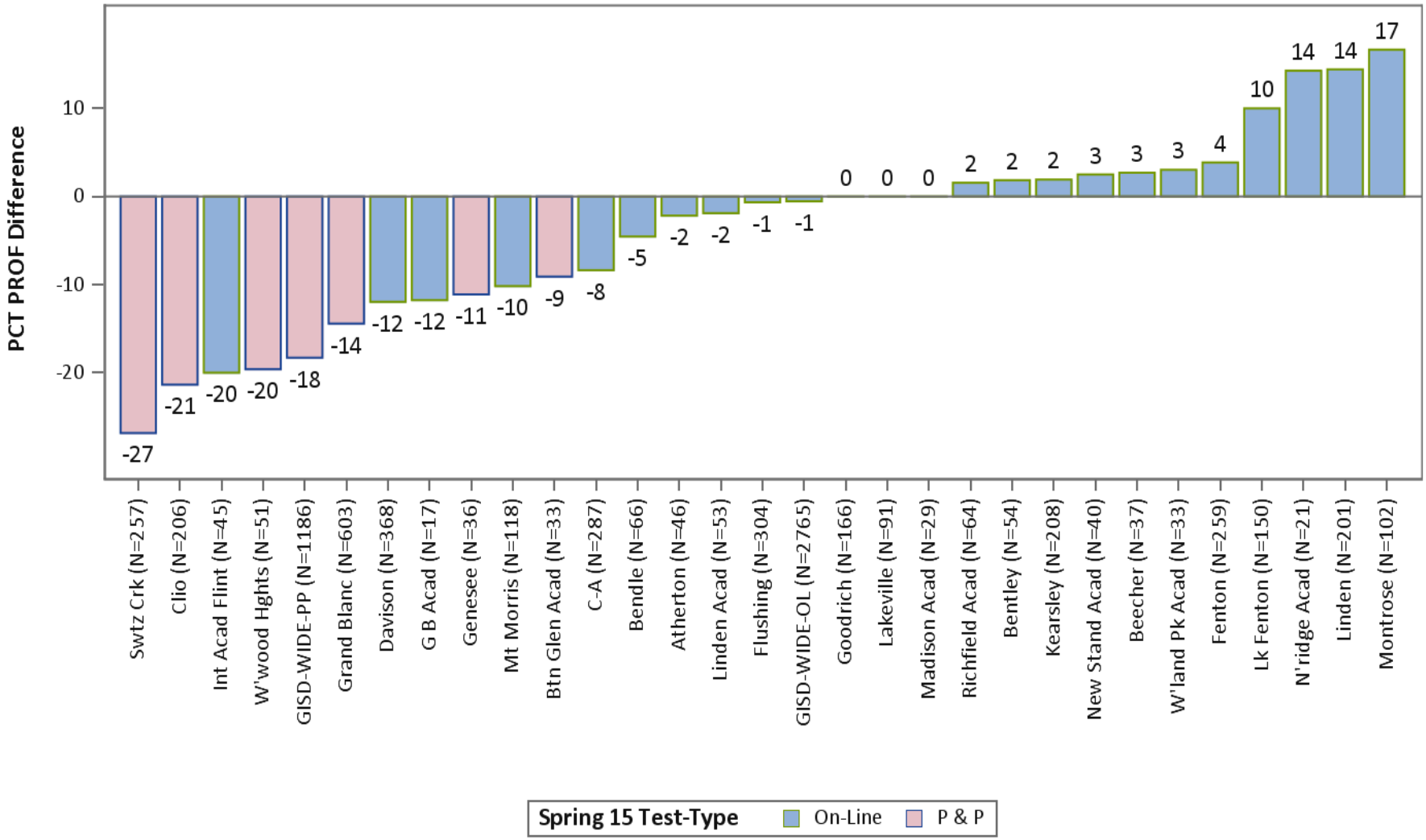


ELA M-STEP - Spring 2015 to Spring 2016

(BY COHORT - SAME STUDENTS)

Difference in Percent Proficient

Grade_Spring_2016 = 8

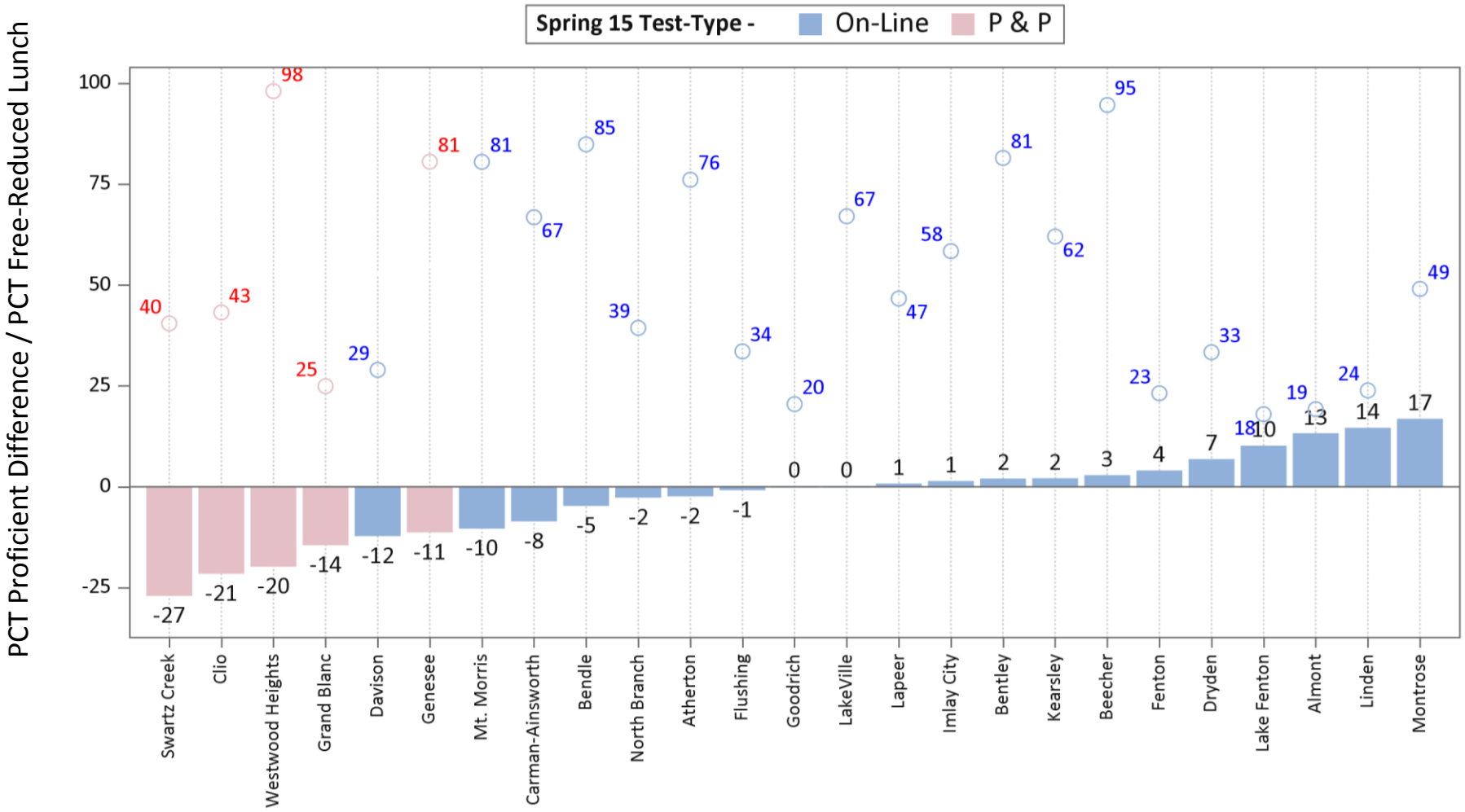


ELA M-STEP - Spring 2015 to Spring 2016

(BY COHORT - SAME STUDENTS)

Difference in Percent Proficient - With F/R Lunch Percents (tested students)

Grade_Spring_2016 = 8



So, I got more Data….

Genesee, Lapeer, Ottawa, & Oakland ISDs

- Matched Students, Gr. 7 2015 M-STEP, Gr. 8. 2016 M-STEP
- Districts who Took P & P Spring 2015, and On-line Spring 2016
  -- Kept District/Grades where students in the Districts administering:
  > On-line Sp. 15 and On-Line Sp. 16
  or
  > P & P Sp. 15 and On-Line Sp. 16
  (got rid of Districts administering (1) P & P in both Sp. 2015 and Sp. 2016, and districts that "mixed" across schools (some P & P, some On-Line)
- Counts

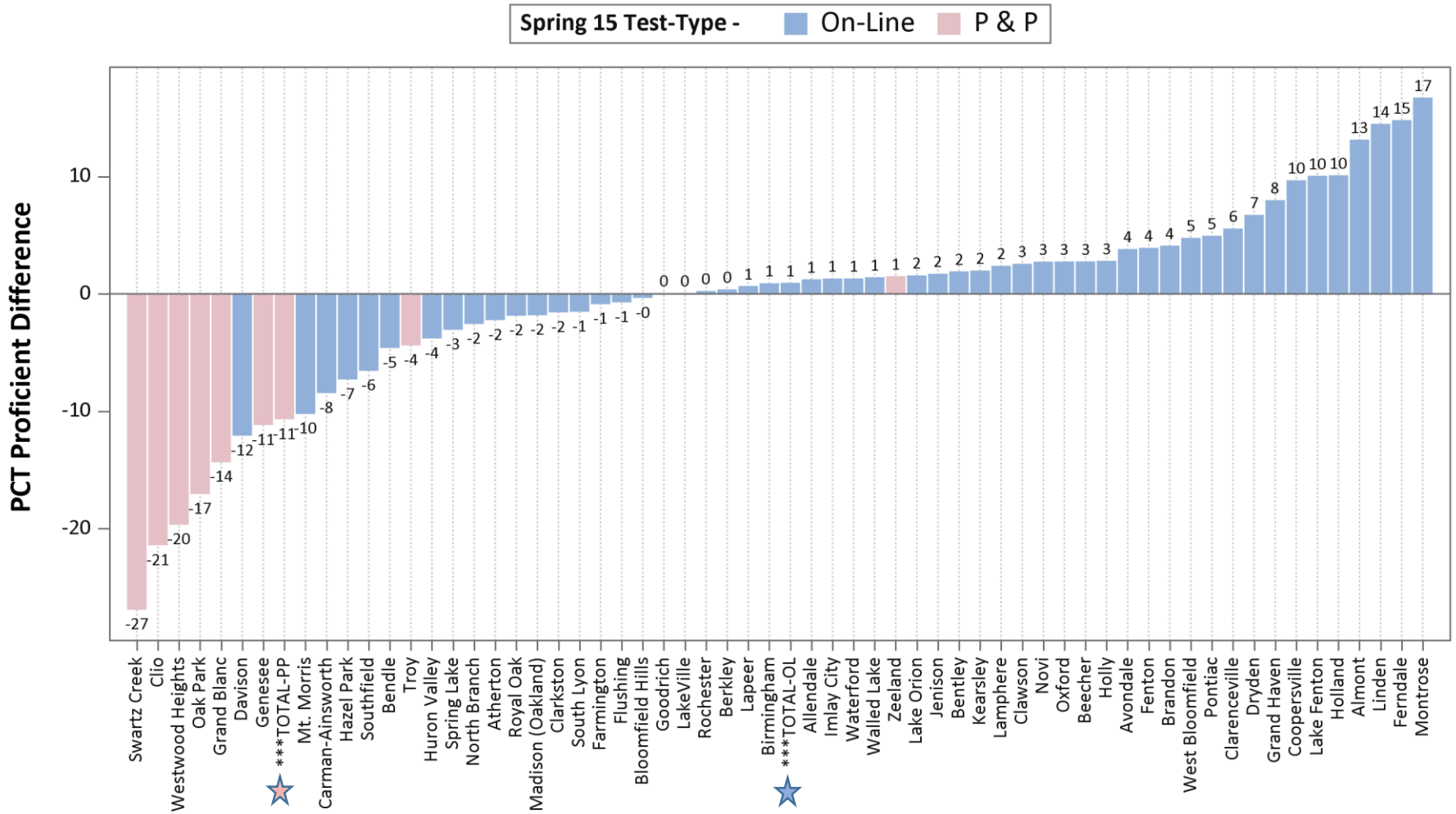
Spring 2015 Testing	Districts	Students
On-Line	52	15,451
P & P	8	2,627

ELA M-STEP - Spring 2015 to Spring 2016

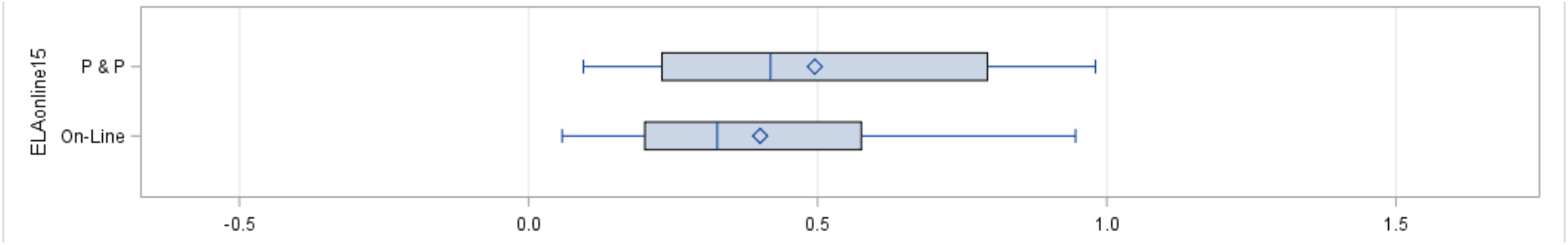
(BY COHORT - SAME STUDENTS)

Difference in Percent Proficient

Grade_Spring_2016 = 8



Free/Reduced Percent Lunch, Fall 15
By District, P & P vs. On-Line

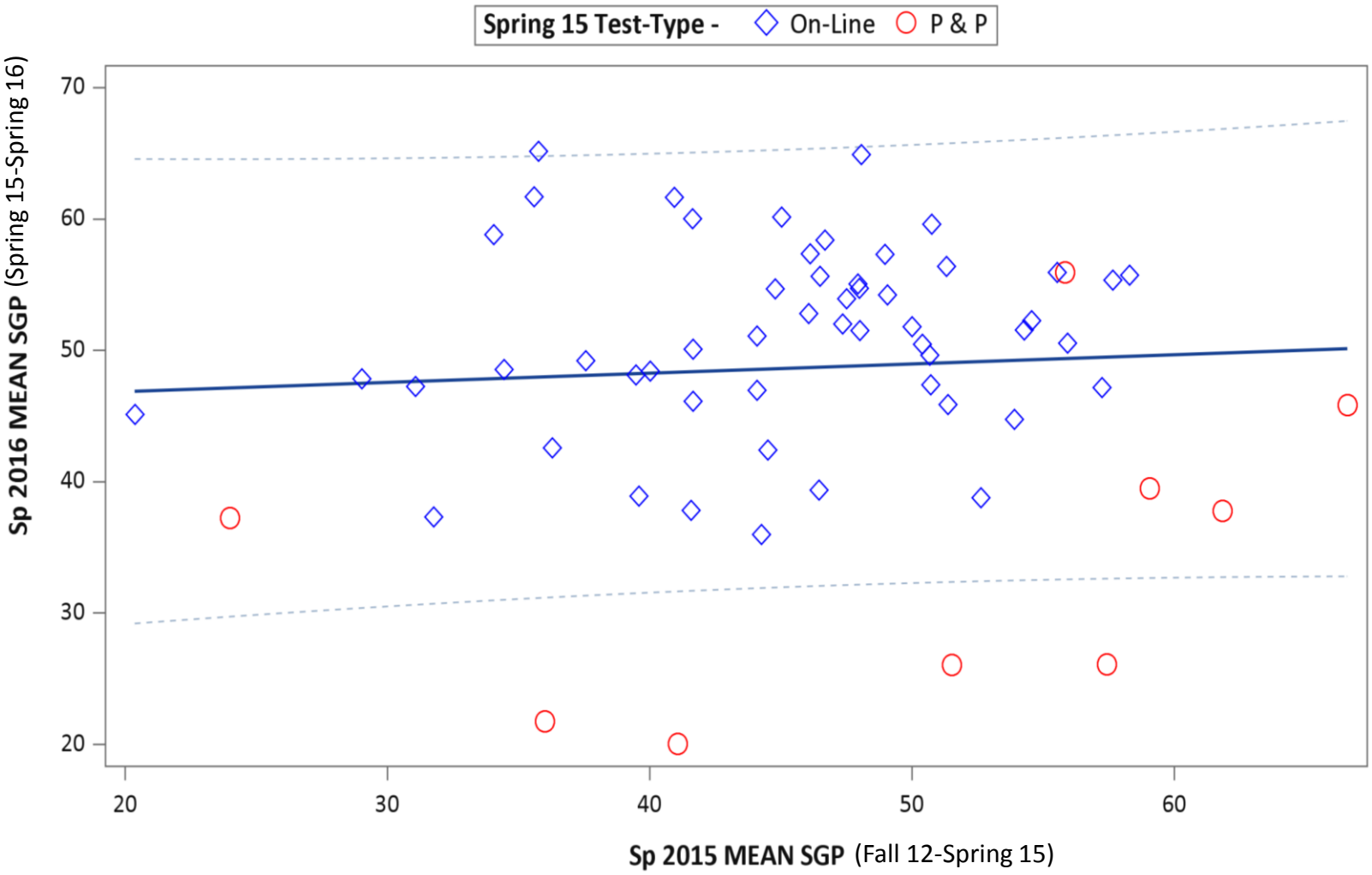


ELAonline15	N	Mean	Std Dev	Std Err	Minimum	Maximum
P & P	8	0.4950	0.3222	0.1139	0.0948	0.9804
On-Line	52	0.4005	0.2452	0.0340	0.0582	0.9459
Diff (1-2)		0.0946	0.2557	0.0971		

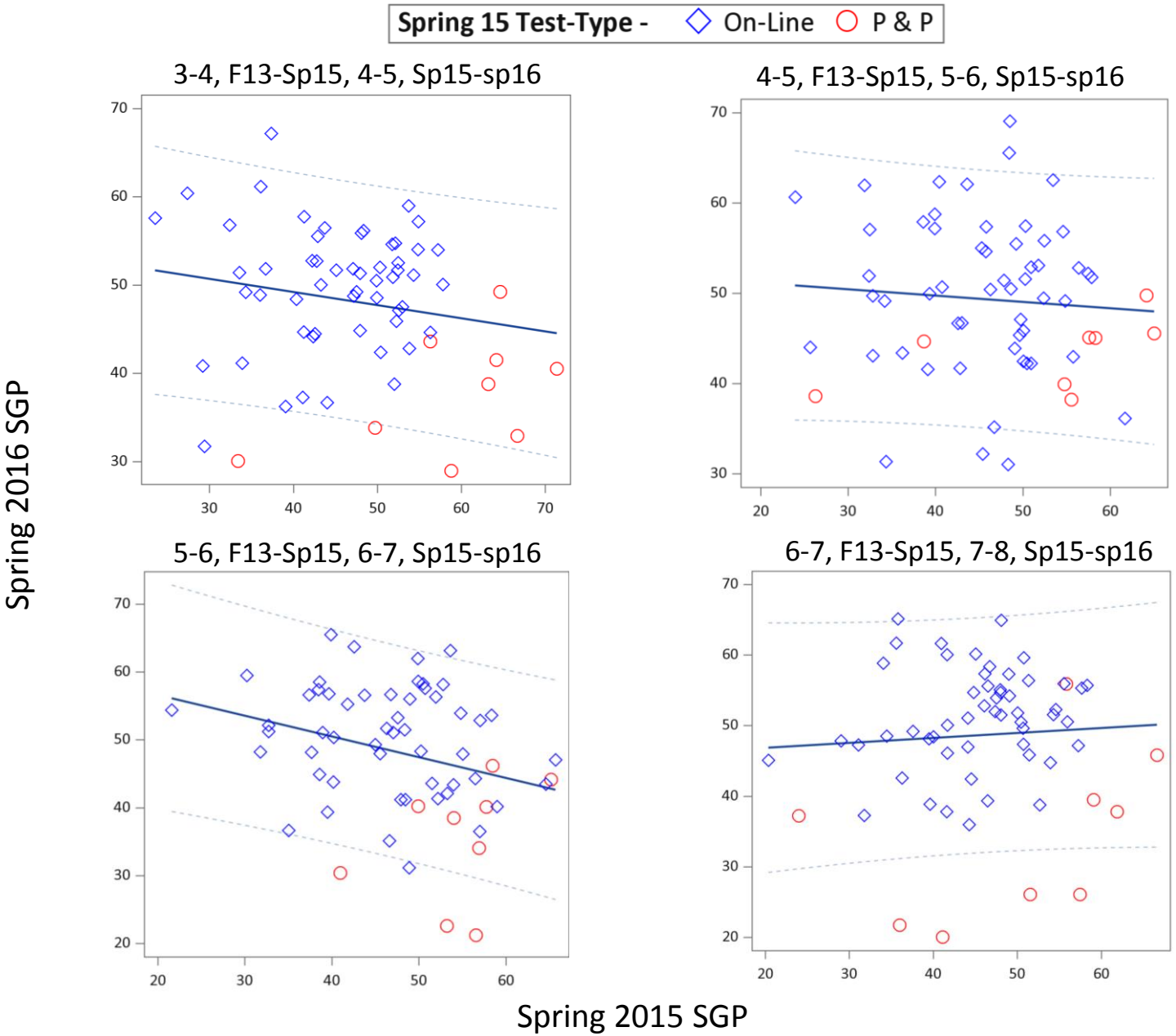
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	58	0.97	0.3343

Based onthese results of fail to reject the null hypothesis (the null hypothesis being: *there is no difference in the percent of free/reduced students, between the districts who tested on-line and those who tested P & P)*

SGP Change, Fall 12-Spring 15 to Spring 15-Spring 16
By District, P & P and On-Line



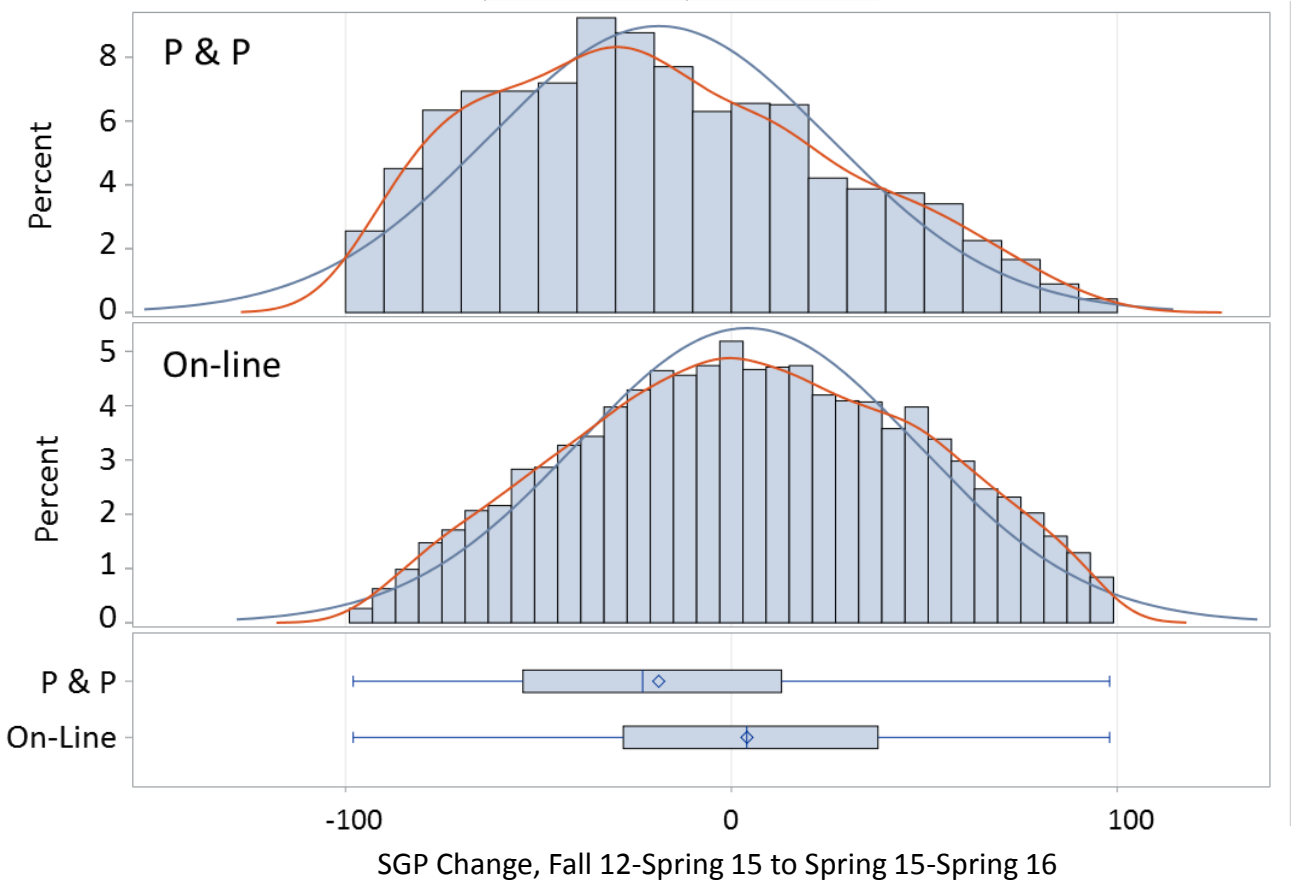
SGP Change, Fall 12-Spring 15 to Spring 15-Spring 16
By District, P & P and On-Line



SGP Change, Fall 13-Spring 15 to Spring 15-Spring 16
By Student, P & P and On-Line

ELAonline15	N	Mean	Std Dev	Std Err	Minimum	Maximum
P & P	2,350	-18.8	44.5	0.92	-98	98
On-Line	1,4034	4.0	44.1	0.37	-98	98
Diff (1-2)		-22.8	44.1	0.98		

t Value	Pr > |t|
-23.24	<.0001



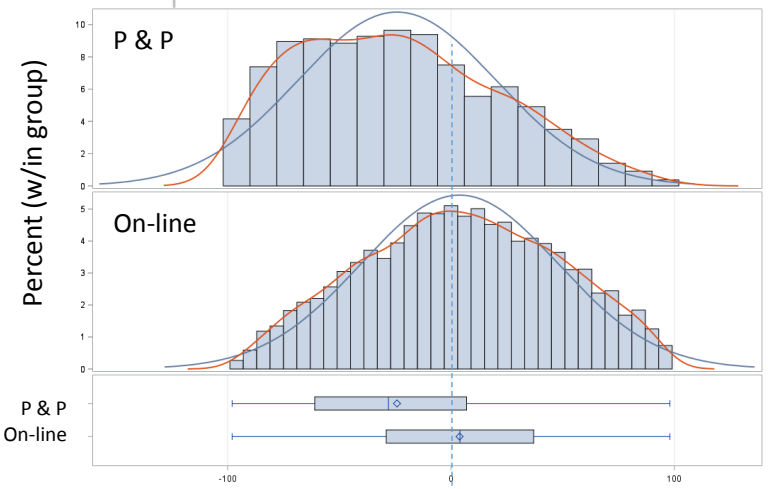
SGP Change, Fall 13-Spring 15 to Spring 15-Spring 16 By Student, P & P and On-Line

PCT F/R Lunch, Across Testing Modes (tested students)

GRADE	P & P	On-Line
4	34%	33%
5	30%	33%
6	31%	31%
7	29%	30%
8	29%	30%

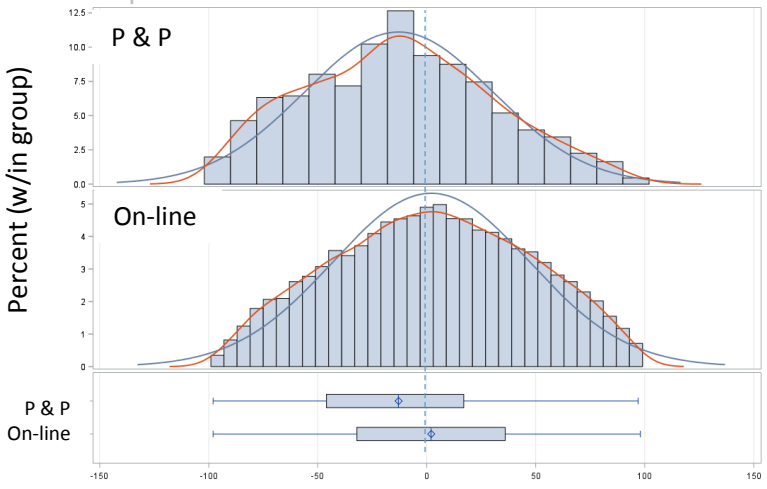
4-5, F13-Sp15 -- 5-6, Sp15-sp16

	N	Mean	Std Dev
P & P	1,854	-24.2	44.4
On-Line	13,899	3.8	44.0
Diff		-28.0	44.1



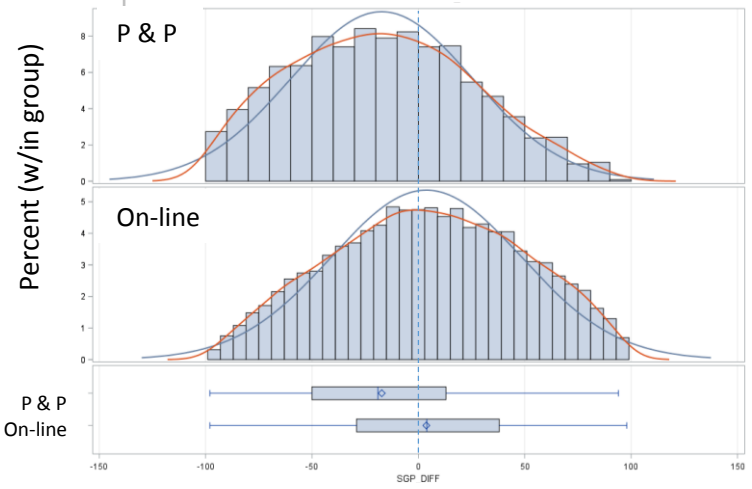
3-4, F13-Sp15 -- 4-5, Sp15-sp16

	N	Mean	Std Dev
P & P	1,770	-12.9	43.1
On-Line	14,150	2.1	44.9
Diff		-15.0	44.7



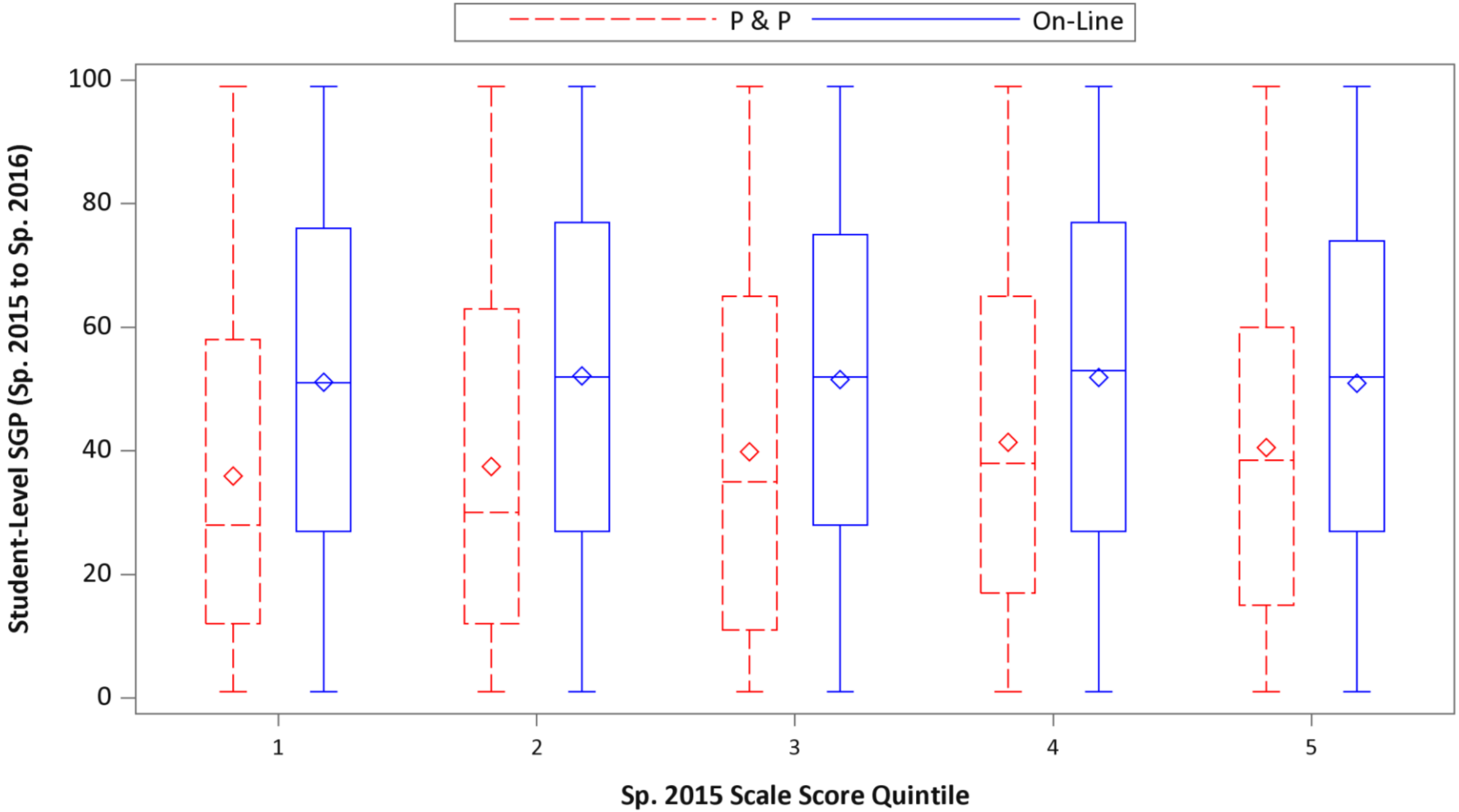
5-6, F13-Sp15 -- 6-7, Sp15-sp16

	N	Mean	Std Dev
P & P	2,305	-17.2	42.7
On-Line	13,789	3.7	44.6
Diff		-21.0	44.3



Spring 2016 SGP By Student, Comparing P & P and On-Line

Quintiles, Based on Spring 2015 Scale Scores

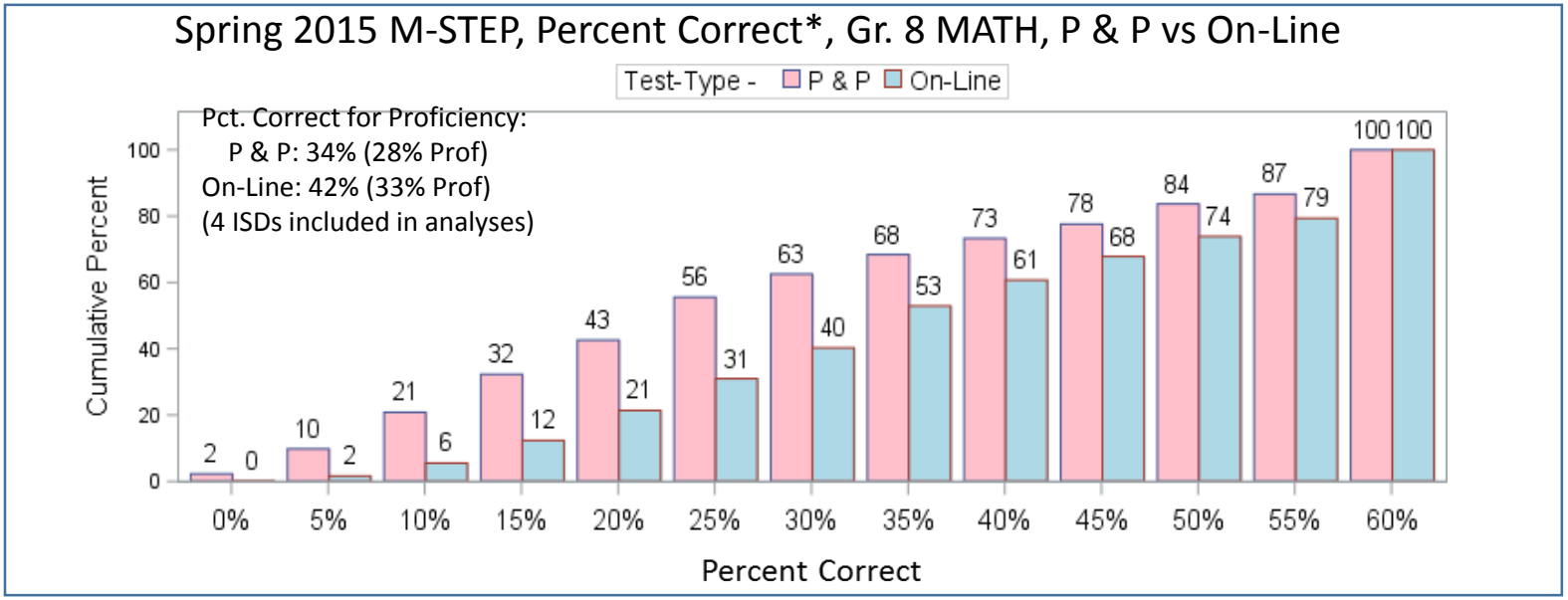


Mean SGP	36	51	37	52	40	52	41	52	41	51
ES / Prob_T	0.53 / <.0001		0.52 / <.0001		0.41 / <.0001		0.37 / <.0001		0.37 / <.0001	
N	353	3364	438	3112	502	3093	600	2927	734	2956

A number of conclusions can be reached (I'm sure more than I've listed here, and these are not meant to be either/or)

-- one which I DON'T think can be reached: *that these differences are tied to* – as are most other test related outcomes – *the Socio-economic status of the districts*

1) There was something different about the results, between the two modes of administration, which made the equating….problematic?



2) There was something pretty different in the CONSTRUCT they were actually assessing (which could well be highlighted in the previous point)

3) The districts who had tested On-line in Spring 2015 had a *practice effect* going into the Spring 2016 testing - but, to me, this begs the question (and reverts back to my previous point) – if results can be so dramatically affected by the mode of administration – what is it we are actually measuring? (Math? or Computer Skills?)

*Percent Correct = sum(ReportingLevelPts1-4) / sum(ReportingLevelPtsPossible1-4) - from the downloadable datafiles