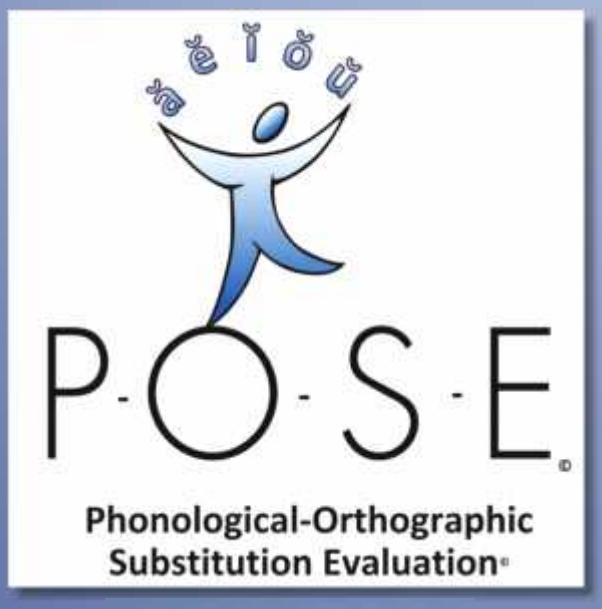


WWW.P-O-S-E.NET

A Comparative Assessment of the NYS ELA Examinations 2013, 2014: Grade 3 Mineola U.F.S.D.

Referencing outcomes for the Phonological-Orthographic Substitution Evaluation (P-O-S-E©), Fountas and & Pinnell Benchmark Assessment System (F&P BAS), Northwest Evaluation Association Measures of Academic Progress-Reading (NWEA MAP-R) and New York State English Language Arts test (NYS ELA)

Carol A. Sullivan, CCC- SLP, Consultant & Roy F. Sullivan, Ph.D. P-O-S-E© Inc.

April 12, 2015

DRAFT

SUMMARY AND CONCLUSIONS

The Phonological-Orthographic Substitution Evaluation (P-O-S-E©) is a criterion-referenced test instrument for assessing short vowel proficiency in reading and spelling, initially targeted at third grade students. Short vowel proficiency has been recognized by Common Core State Standards (CCSS) as a foundational skill for literacy, to be established by Grade 2. The P-O-S-E© was standardized at the third grade level in the Plainview-Old Bethpage Central School District (POB) of New York (NY) between years 2005 and 2010 .

In 2012-13 and 2013-14, a comprehensive program of P-O-S-E© baseline, intervention and RTI evaluation was instituted in the Mineola Union Free School District (Mineola UFSD) of NY. Twenty percent of the student population was categorized as Latino or Hispanic, 12% Asian, etc. and 3% Black or African-American.

At the end of the 2012-13 academic year, Mineola Grade 3 made significant advances in P-O-S-E© short vowel proficiency and in literacy as assessed using the Fountas and Pinnell Benchmark Assessment System (F&P BAS) and the Northwest Evaluation Association Measures of Academic Progress, Reading (NWEA MAP-R.) Grade 3 scored the highest proportion of literacy proficiency among all Mineola UFSD grades 3-8 on the 2013 New York State English Language Arts examination (NYS ELA), newly configured to conform to Common Core State Standards (CCSS.)

At the end of 2013-14, comparable RTI gains were noted on the P-O-S-E©, F & P BAS and NWEA MAP. However, Grade 3 scored the *lowest* proportion of literacy proficiency among all Mineola Grades 3-8 on the 2014 NYS ELA. In addition, the Grade 3 cohort from 2012-3 scored next-to-lowest in literacy on the 2014 Grade 4 NYS ELA. According to NYS data, ELA passing proficiency scores for the entire state were comparable between 2013 and 2014: 31.1% vs. 31.0%, respectively. Long Island ELA scores showed a greater 2013-14 reduction: 39.6% to 36.8%.

The gross inconsistency between Grade 3 NYS ELA outcomes for both 2013 and 2014 and alternative measures of literacy for the same years prompted an inquiry into possible reasons for this conflict. Mineola Grade 3 test data and NYS-released ELA reading passages and scoring data were analyzed in detail for both years.

It is to be noted that, when the multiple correlational analysis among alternative measures of literacy was restricted to Grade 3 students with P-O-S-E© error scores > 25%, ALL external correlations between the NYS ELA scores and the alternative literacy assessment instruments were significantly lower in 2014 than in 2013.

Findings reveal significant issues with face validity of the NYS ELA examination as currently implemented. NYS ELA test passages for Grades 3 and 4 in 2013 and 2014 present an exaggerated range of grade-inappropriate reading levels effectively rendering invalid any test questions based on these passages. Reading levels for NYS-released 2014 Grade 3 ELA passages were well above grade level, well above the level for 2013 Grade 3 passages and even higher than Grade 4 passages for 2013.

Data also suggest that reliability of the NYS ELA test outcomes may be compromised by the process of *"equating"* applied by NY State to the 2014 ELA scores This is a post-hoc application of raw-score-to-scale-score transformations and scale-score-to-performance level transformations to achieve a preferred outcome in year 2014 relative to 2013. According to NYS:

> *"The cut scores [defined boundaries of literacy proficiency categories L1-L4] did not change from 2013 to 2014. "*

In fact, the raw-to-scale score transformations were altered between 2013 - 2014 resulting in differing raw score values for each cut (scale) score. Continuing:

> *"The purpose of the 2014 equating was to maintain the level of difficulty established by the standard setting process in 2013, when 95 teachers from across the state recommended the level of difficulty necessary to achieve proficiency (Level 3) and partial proficiency (Level 2). Based on student performance on common anchor test questions (the same items used in both 2013 and 2014), the raw scores needed for each performance level were adjusted slightly to ensure that scale scores and performance levels are comparable from year to year. If the test is slightly easier, the number of raw score points needed to earn a performance level may increase slightly in order to maintain the performance standard. If the test is slightly harder, the number of raw score points needed to earn a performance level may decrease slightly in order to maintain the performance standard."* …
>
> *"…On the 2014 tests, year-to-year raw score changes for Level 3 were small and varied by grade. Raw scores went down slightly on 6 tests (indicating slightly harder tests in 2014 compared to 2013 for Grades 3, 4, and 7 ELA and Grades 3, 5, and 6 Math) and went slightly up on 4 tests (indicating slightly easier tests in 2014 compared to 2013 for Grades 5 and 6 ELA and Grades 4 and 7 math)."*

Finally, in 2014, three Grade 3 ELA test items were summarily discarded by NYS, *post hoc*. This accounted for the 6 point differential between the 55 point 2013 ELA and the 49 point 2014 ELA – an arbitrary net reduction of 11% in the 2014 scoring base.

Since 2012-13, Common Core State Standards have been foundational to the NYS ELA and to the literacy examinations of other states. CCSS seeks to impose an overarching set of theoretically-derived criteria for literacy proficiency. The ability of individual states to "tweak" the aggregate test score outcomes effectively invalidates the concept of *"Common Core".*

A minor shift of -3% was experimentally applied to the 2013-14 P2-P3 scale score cutoff boundary. This action dramatically elevated the 2014 Mineola Grade 3 P3+P4 literacy proficiency level from the reported 33.0% (~10% below 2013) to 44.4% (~2% above 2013). (q.v. Tables 29, 30) The differing, multi-modal nature of the scale score data distribution in 2013 and 2014 contributes significantly to the misinterpretation of ELA outcomes.

Despite NYS enlisting the best efforts of "95 teachers", the major functional and educational impact of this minor shift in a single ELA cutoff value, arbitrarily manipulated in the raw-to-scale-score transformation in 2014 by NY State, highlights the fragile inadequacy of the entire ELA evaluation process in its current form.

Literacy and the entire academic well-being of students and a reinforced level of motivation among their effective teachers cannot be subjected to the statistical vagaries of test designers with constrained perspectives. "Regents examination" scoring protocols have ceased to be relevant.

Given the outcome of the present detailed analysis of Grade 3 NYS ELA reading materials and scores contrasted with alternative measures of literacy proficiency for the Mineola UFSD, serious questions may be raised about the relevance of the NYS ELA as currently constructed. It would appear that the NYS ELA is not a suitable test instrument for assessing language arts proficiency or for directing data-driven curriculum development in Grade 3.

Carol A Sullivan, CCC-SLP; Roy F Sullivan, Ph.D. http://www.P-O-S-E.net April 12, 2015

ACKNOWLEDGEMENTS

The authors express appreciation to the administrative and teaching staff of Mineola Union Free School District, Mineola, New York for their commitment to the development and implementation of the P-O-S-E© project from 2012 to the present.

Mineola U.F.S.D. Superintendent of Schools Dr. Michael P. Nagler has been forward-looking and unswerving in his interdisciplinary pursuit of academic excellence for students and teaching staff. Ms. Catherine Fishman, Director of Pupil Services, introduced the P-O-S-E© project in-district and has fostered a productive interdisciplinary synchrony between assessment and intervention strategies. Ms. Patricia Burns, Assistant Superintendent for Curriculum, has facilitated a seamless integration of the P-O-S-E© project into the curriculum of Grade 3 and, more recently, Grade 2.

The dedication and excellence of the Mineola U.F.S.D. professional staff has been essential to the success of the P-O-S-E© project including teachers of Speech-Language-Pathology (SLP), Special Education; English as Second Language (ESL), Reading and General Education.

Special thanks to Roberta Beech for her keen-eyed editorial review.

Carol A. Sullivan, CCC-SLP
Roy F Sullivan, Ph.D.
www.P-O-S-E.net

INTRODUCTION

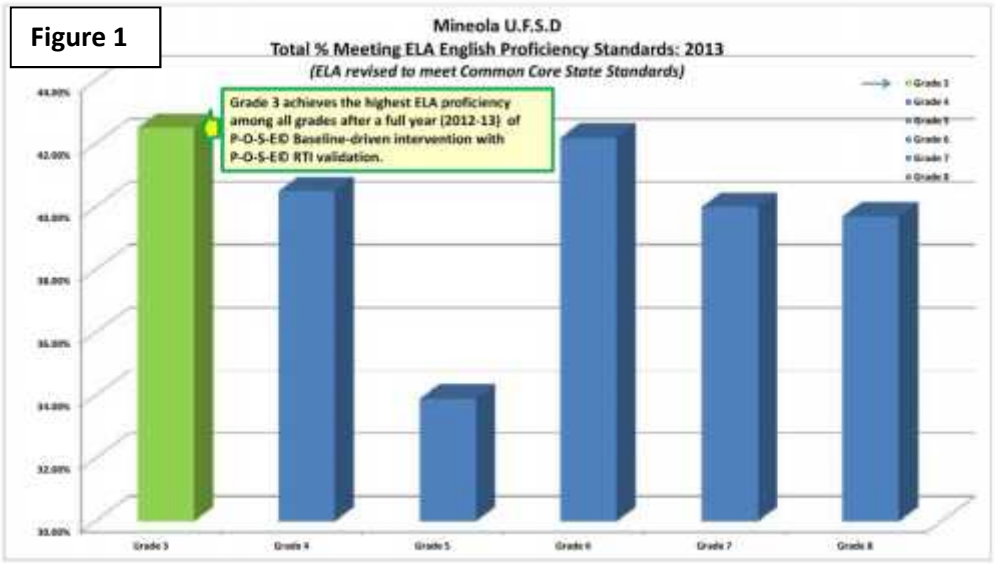
The P-O-S-E©: Phonological/Orthographic Substitution Evaluation (P-O-S-E©) is a criterion-referenced assessment instrument, designed to probe for substitution errors in a child's phonological (spoken) and orthographic (written, scored as equivalent phonology) representations of target short vowels presented in monosyllabic non-word and real word spelling and reading tasks. I.e. an incorrect phoneme is substituted for the target phoneme. Silent /e/ rule test items are incorporated as a cross-check and validation of the depth of short vowel proficiency. Outcomes provide prescriptive interventional direction when indicated. RTI outcomes are assessed at end of the school year. Common Core State Standards (CCSS) present a goal of short vowel proficiency by Grade 2. (http://www.p-o-s-e.net/#!cssi/ctlq**)**

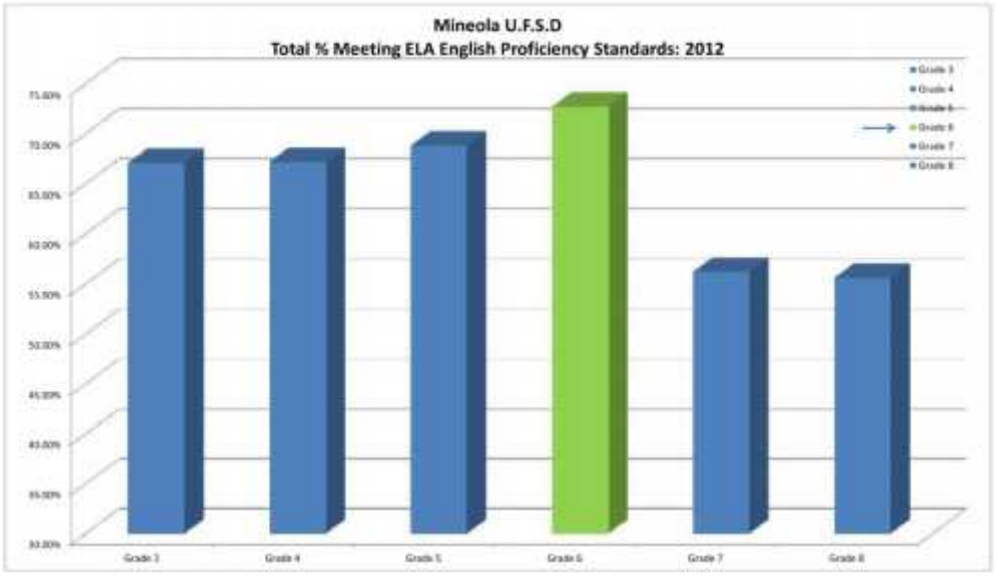
Since the baseline study in the Plainview-Old Bethpage Central School District (POB) of New York (NY) in 2006-7 (http://www.p-o-s-e.net/#!research-menu/c22u9**), the P-O-S-E© has been applied to thousands of third grade students in two major Long Island, NY school districts. In year 2012-13, the P-O-S-E© program was instituted in the Mineola Union Free School District (Mineola UFSD) of NY. End-of-year, matched pair response-to-intervention (RTI) testing demonstrated significant reductions in P-O-S-E© error scores. (**http://www.p-o-s-e.net/#!2012-13-mineola-rti-study/c2k3**).**

Concurrently, a**lthough not necessarily causally, Mineola UFSD 2013 New York State English Language Arts (NYS ELA) scores presented by Grade 3 achieved the highest L3+L4 proficiency among all Mineola UFSD Grades 3-8. Figure 1 illustrates the Mineola U.S.F.D. ELA outcomes for years 2011-2013. The overall reduction in Grade 3-8 NYS ELA literacy proficiency scores from 2012 to 2013 is a reflection of the newly applied Common Core State Standards (CCSE) template for NYS ELA literacy assessment and NYS scoring criteria.**

Table 2 shows the 2012-13 baseline and RTI scores for Grade 3 Fountas and Pinnell Benchmark Assessment System (F&P BAS), **Northwest Evaluation Association Measures of Academic Progress, Reading (NWEA MAP-R.) and the Phonological-Orthographic Substitution Evaluation (P-O-S-E©.) Grade-appropriate advances were experienced on all three assessment instruments.**

Figure 1


Mineola U.F.S.D
Total % Meeting ELA English Proficiency Standards: 2013
(ELA revised to meet Common Core State Standards)
Grade 3 achieves the highest ELA proficiency among all grades after a full year (2012-13) of P-O-S-E© Baseline-driven intervention with P-O-S-E© RTI validation.
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8


Mineola U.F.S.D
Total % Meeting ELA English Proficiency Standards: 2012

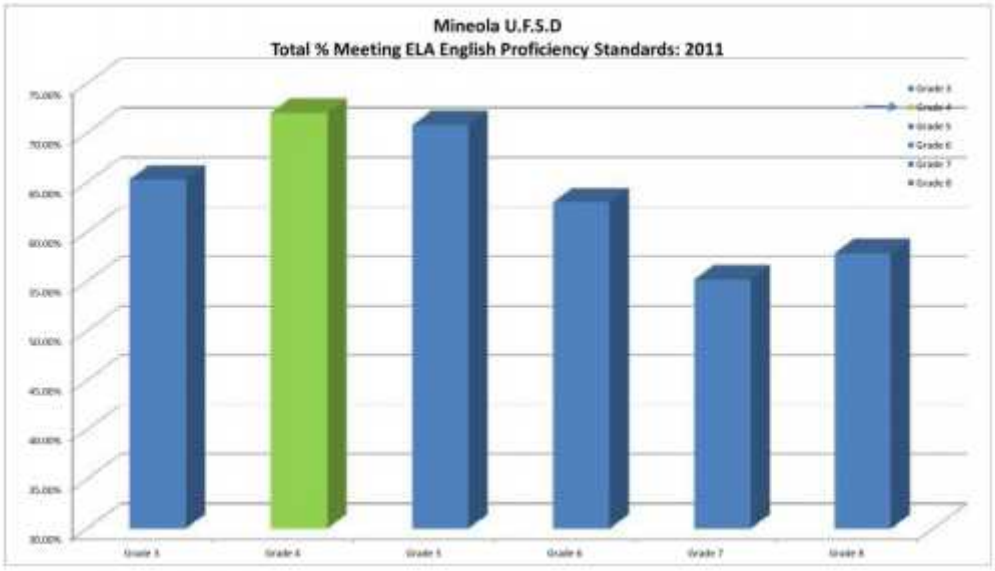

Mineola U.F.S.D
Total % Meeting ELA English Proficiency Standards: 2011

Figure 2 presents the Mineola UFSD NYS ELA outcomes for the years 2011-2014. Ranked first in 2013, Grade 3 L3+L4 literacy proficiency inexplicably dropped 9.5% from 42.5% to 33% in 2014 (red arrow), ranking last among grades 3-8. Equally puzzling was the 6.5% reduction in L3+L4 proficiency of 2014 Grade 4, tying for next-to-last ranking among grades 3-8 (yellow arrow). This very same cohort scored highest in NYS ELA literacy proficiency as Grade 3 in 2013. According to NYS data, ELA passing proficiency scores for the entire state were comparable between 2013 and 2014: 31.1% vs. 31.0%, respectively. Long Island ELA scores showed a greater 2013-14 reduction: 39.6% to 36.8%.

Figure 2

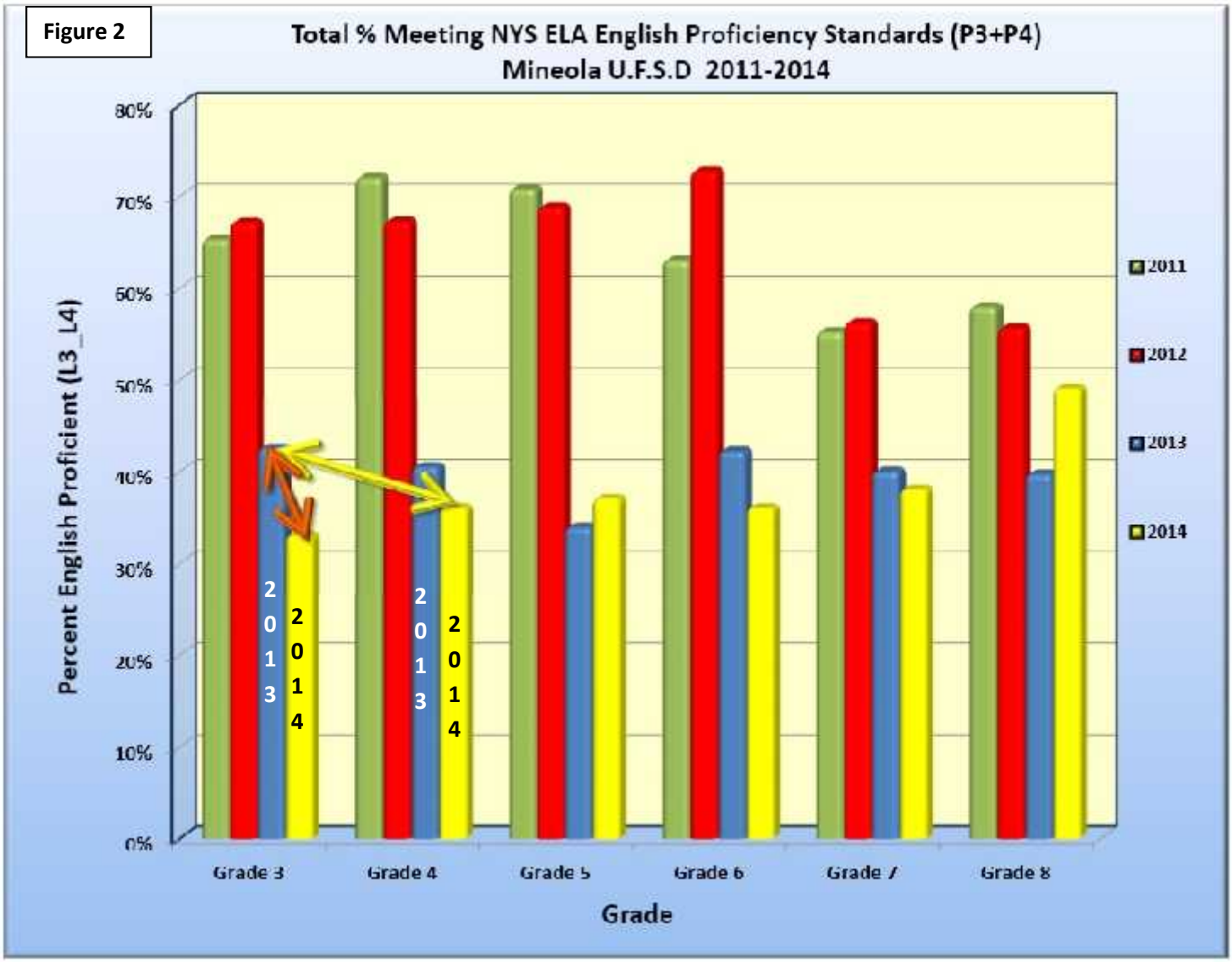


Table 1 presents the same data as above in tabular format. The dramatic reduction in NYS ELA-scored literacy proficiency between 2012 and 2013 is an artifact of the NYS contractor's revision of the ELA examination to ostensibly conform to Common Core State Standards in 2013, 2014. q.v. http://www.fairtest.org/pearsons-history-testing-problems; http://www.whec.com/article/stories/s3709812.shtml**. The 9.5% reduction in Grade 3 ELA proficiency from 2013 to 2014, given the same 319-320 scale cut score in both years, is not explained be inter-year differences in scale scores.**

Table 2 provides descriptive statistics for the distribution of NYS ELA scale scores in 2013 and 2014. Note that the means and medians do not differ significantly but the modes, inter-year, are notably disparate suggesting a non-normal or multimodal distribution of the underlying data. The implications placing a pass-fail, literate-illiterate cut score, in disregard of the fundamental data distribution, can lead to untoward outcomes and interpretations. A detailed analysis of this issue is presented the last section of this report.

Table 1

Mineola U.F.S.D

Total % Meeting ELA English Proficiency Standards 2011-2014

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2014	33.0%	36.0%	37.0%	36.0%	38.0%	49.0%
2013	42.5%	40.5%	33.9%	42.2%	40.0%	39.7%
2012	67.1%	67.2%	68.8%	72.7%	56.2%	55.6%
2011	65.3%	72.0%	70.8%	63.0%	55.2%	57.9%

Table 2

Grade 3 ELA Scale Scores

Mineola U.F.S.D. 2013, 2014

	2013 all	*2014 all*
Mean	306.8	304.4
Standard Error	2.28	2.33
Median	311	305.5
Mode	338	300
Standard Deviation	31.49	31.20
Sample Variance	991.66	973.27
Kurtosis	0.14	0.17
Skewness	0.49	0.36
Range	170	180
Minimum	212	196
Maximum	382	376
Sum	58600	54784
Count	191	180

Figure 3 presents a flow chart of sequential processes, as constructs in literacy assessment, using the NYS ELA. One or more items in the sequence could potentially account for literacy proficiency differences between academic years 2013 and 2014.

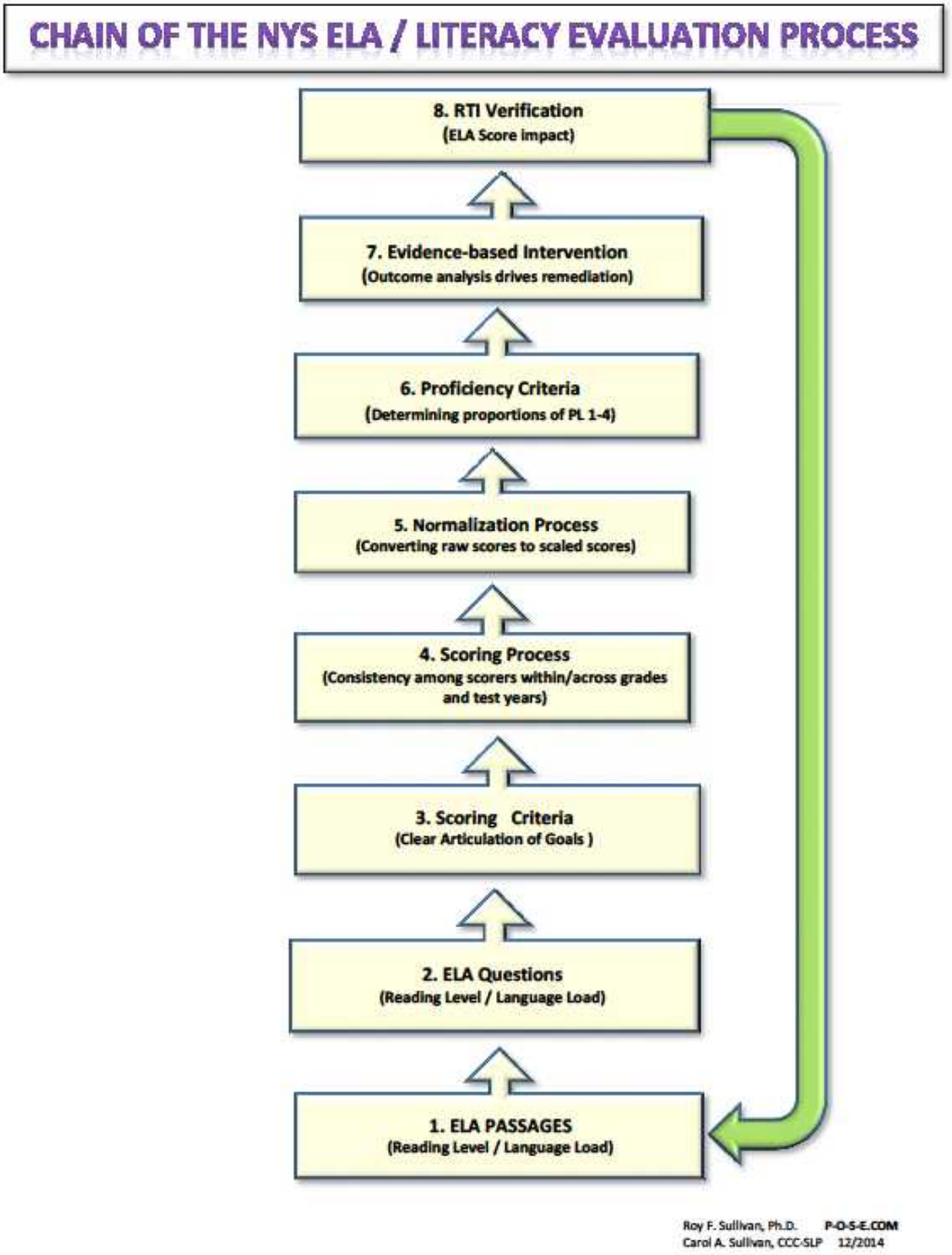
Items 1 and 2, test passage and question construction, are nominally controlled by the New York State Item Review Criteria for Grade 3-8 English Language Arts and the contracted publisher of the NYS ELA.

Item 3 represents NY State-defined criteria for test item scoring. Item 4 represents the physical process of applying those state-defined criteria to actual scoring of the individual test item responses. Raw data scoring may be performed using district personnel or an external, independent scoring service.

Items 5 and 6 reside, *post hoc*, with the State of New York, providing statistical transformations for scaling the raw scores with subsequent partitioning the scaled scores into four nominal categories of literacy proficiency.

Items 7 and 8 reside with the school district to derive information from the outcome analysis ostensibly driving intervention. At present (7), other than a single proficiency level per student, no individualized diagnostic information can be derived from the NYS ELA outcomes as reported to specify intervention. Tests and individual outcomes are sequestered by NYS. If the NYS ELA is a valid and reliability measuring instrument, scores for the following year should reflect data-driven changes in RTI derived from NYS ELA outcomes for the prior year.

Figure 3


CHAIN OF THE NYS ELA / LITERACY EVALUATION PROCESS
8. RTI Verification
(ELA Score impact)
7. Evidence-based Intervention
(Outcome analysis drives remediation)
6. Proficiency Criteria
(Determining proportions of PL 1-4)
5. Normalization Process
(Converting raw scores to scaled scores)
4. Scoring Process
(Consistency among scorers within/across grades and test years)
3. Scoring Criteria
(Clear Articulation of Goals)
2. ELA Questions
(Reading Level / Language Load)
1. ELA PASSAGES
(Reading Level / Language Load)
Roy F. Sullivan, Ph.D. P-O-S-E.COM
Carol A. Sullivan, CCC-SLP 12/2014

THE CONUNDRUM

The apparent reduction in Mineola Grade 3 Grade 3 NYS ELA literacy proficiency between school years 2012-13 and 2012-14 (Table 1, Figure 2) may be ascribed to a number of possible reasons:

1. ***The 2012-13 Grade 2 cohort, upon becoming Grade 3 in 2013-14, may have been less NYS ELA-literacy proficient at the outset, given the same applied level of Grade 3 educational intervention as in the prior academic year. It is to be noted that the 2012-13 Grade 3 cohort experienced reduced NYS ELA literacy proficiency in 2013-14 on becoming Grade 4, as did grades 3, 6, and 7 in 2013-14.***

2. ***The NYS ELA test item review criteria may have differed between school years 2012-13 and 2013-14.***

3. ***The NYS ELA raw data scoring practices may have differed between school years 2012-13 and 2013-14.***

4. ***The Grade 3 NYS ELA test instrument for 2013-14 may have differed in reading level of test passages or questions from that of 2012-13. As the NYS ELA is administered near the end of school year, the nominal Grade 3 reading level should be 3.9 or 3.10 (3rd grade, 9th or 10th month). A post-hoc analysis of the reading level of publicly released NYS ELA test content for school years 2012-13 and 2013-14 has been systematically applied to address this critical variable.***

5. ***The Grade 3 NYS ELA raw-to-scale score polynomial transformation and scale-score-to-literacy-proficiency-level boundaries or conditions underlying those boundaries may have differed significantly between school years 2012-13 and 2013-14.***

1. ***The 2012-13 Grade 2 cohort, upon becoming Grade 3 in 2013-14, may have been less NYS ELA-literacy proficient at the outset, given the same applied level of Grade 3 educational intervention as in the prior academic year. It is to be noted that the 2012-13 Grade 3 cohort experienced reduced NYS ELA literacy proficiency in 2013-14 on becoming Grade 4, as did grades 3, 6, and 7 in 2013-14.***

ANALYSIS 1a: Fountas and Pinnell Benchmark Assessment System (F&P BAS)

In order to compare beginning Mineola U.F.S.D. Grade 3 reading levels between 2012-13 and 2013-14, F&P BAS baseline data were analyzed for each group of ten classes in both academic years. Table 3 demonstrates a statistically significant difference between Grade 3 reading levels at the start of the two successive academic years. However, the data indicate that mean and median F&P BAS baseline Grade 3 2013-14 reading levels were HIGHER than Grade 3 2012-13 by one full letter category. Figure 4 presents a graphic distribution of the full range of the same F&P BAS baseline data for both academic years, demonstrated equivalent baseline scores. Figures 5 and 6 demonstrate the improvements in RTI F&P BAS scores over baselines for 2012-13 and 2013-14.

Table 3 — **Grade 3 F&P Benchmark Baseline Scores Mineola U.F.S.D. 2012-13 & 2013-14**

	2012-13 Baseline		*2013-14 Baseline*		t-Test: Two-Sample Assuming Equal Variances		
Mean	13.32	M	14.34	N		*13-14 Baseline*	*12-13 Baseline*
Standard Error	0.19		0.21		Mean	14.34	13.32
Median	14	N	15	O	Variance	8.38	7.24
Mode	16	P	16	P	Observations	186	191
Standard Deviation	2.69		2.89		Pooled Variance	7.80	
Sample Variance	7.24		8.38		Hypothesized Mean	0	
Kurtosis	2.94		2.28		df	375	
Skewness	-1.43		-0.83		t Stat	3.52	
Range	16		19		P(T<=t) one-tail	0.0002	
Minimum	1	A	1	A	t Critical one-tail	1.6489	
Maximum	17	Q	20	T	P(T<=t) two-tail	0.0005	
Sum	2545		2667		t Critical two-tail	1.9663	
Count	191		186				

Figure 4

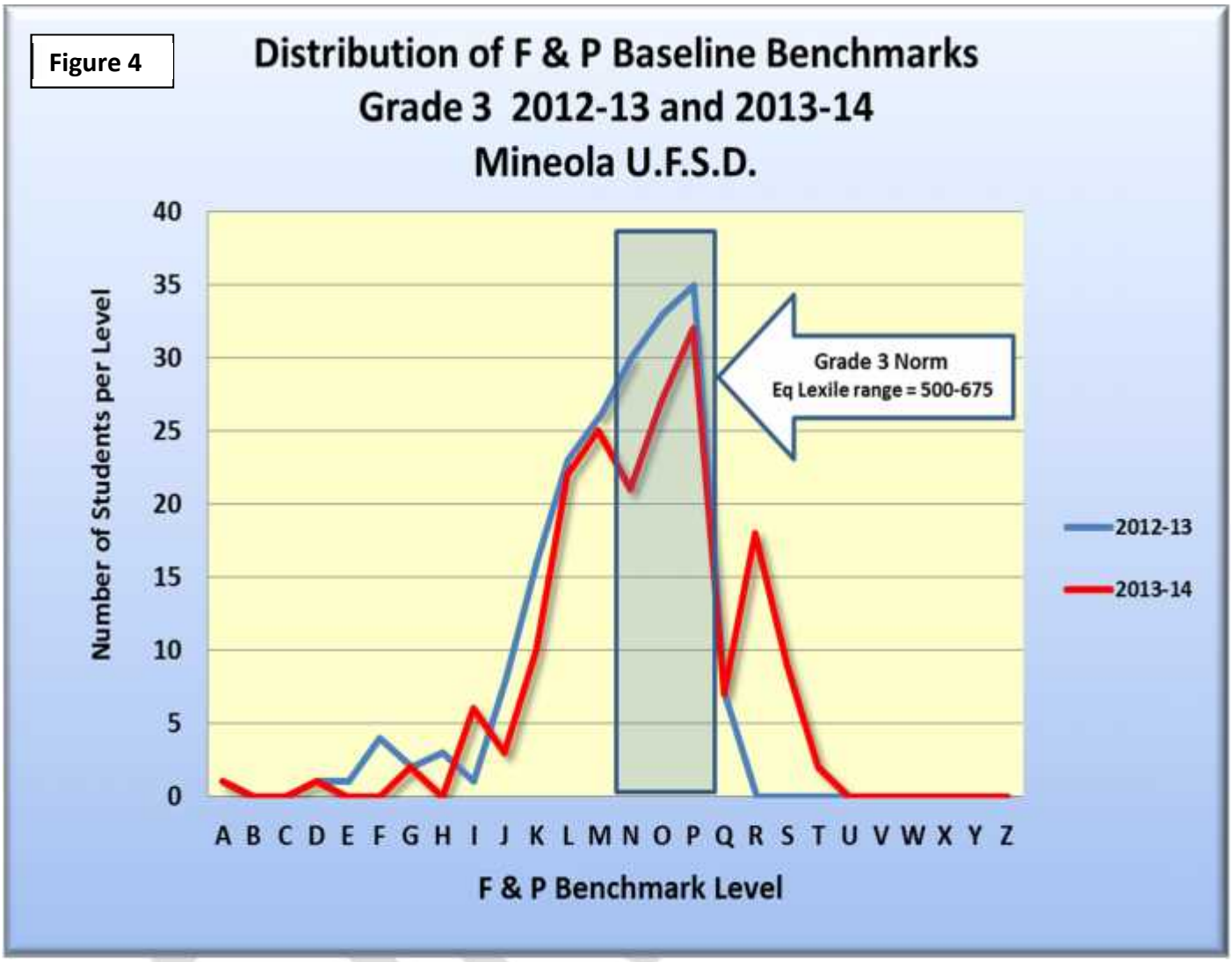

Distribution of F & P Baseline Benchmarks
Grade 3 2012-13 and 2013-14
Mineola U.F.S.D.
Number of Students per Level
40
35
30
25
20
15
10
5
0
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
F & P Benchmark Level
Grade 3 Norm
Eq Lexile range = 500-675
2012-13
2013-14

Figure 5

Distribution of F & P Baseline v. RTI Benchmarks Grade 3 2012-13 (n=191) Mineola U.F.S.D.

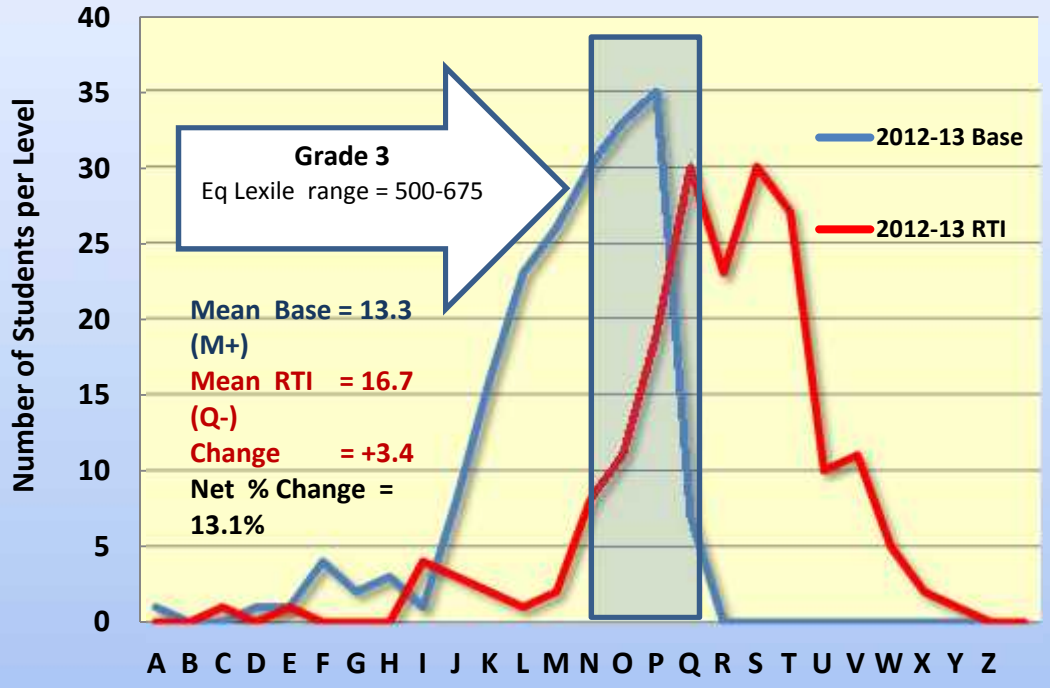
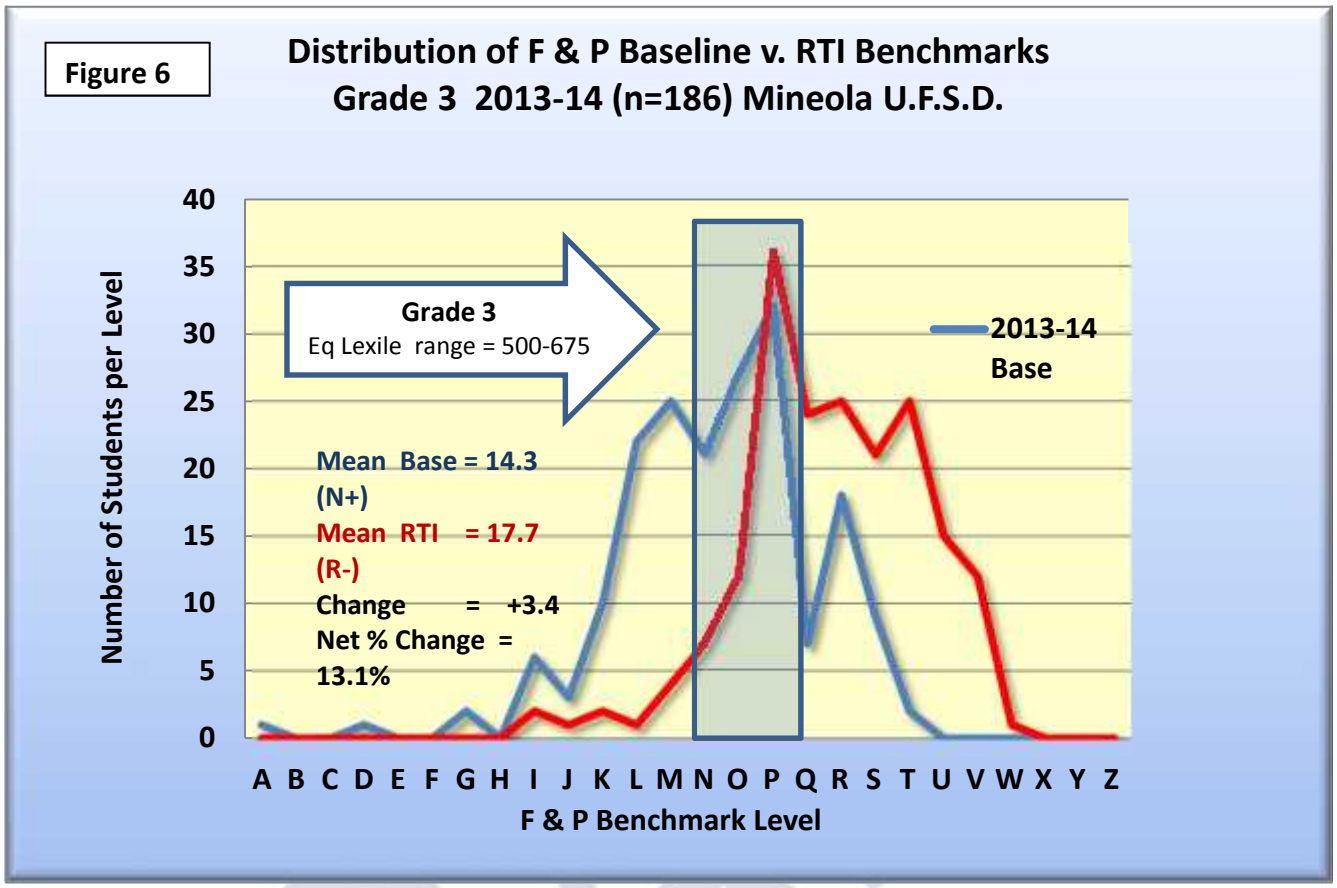


Figure 6



CONCLUSION 1b

Based on statistical equivalence of Grade 3 literacy baselines on the Fountas and Pinnell Benchmarks Assessment System (F& P BAS), the sharp decrease in Mineola Grade 3 NYS ELA proficiency score between 2012-13 and 2013-14 cannot be attributed to literacy differences at the academic year outset. In addition, the RTI gains for each of the two years are comparable for the F&P BAS.

ANALYSIS 1b: Northwest Evaluation Association Measures of Academic Progress, Reading (NWEA MAP-R.)

As a cross-check of relative literacy levels for Mineola U.S.F.D. Grade 3 2012-13 and 2013-14, a similar descriptive statistical tabulation was applied to both student populations using the Northwest Evaluation Association Measures of Academic Progress-Reading (NWEA MAP-R) as an alternative measure of literacy. Results are summarized in Table 4. Grade 3 Baseline NWEA MAP-R results do not differ significantly between 2012-13 and 2013-14. Grade 3 RTI results are 8 points higher (p <.05) in 2012-13.

Table 4

Mineola U.F.S.D. Grade 3	NWEA MAP Reading Baseline vs. RTI			
	2013-14 N=186		2012-13 N=191	
Parameters	Base (Fall)	RTI (Spring)	Base (Fall)	RTI (Spring)
Mean	**190.8**	**201.3**	**191.8**	**209.3**
2011 NWEA Norm	***189.9***	***199.2***	***189.9***	***199.3***
Standard Error	1.9	0.8	0.8	0.9
Median	191	202.5	192	208
Mode	187	204	188	203
Standard Deviation	26.31	11.63	11.02	12.00
Sample Variance	692.40	135.25	121.39	143.93
Kurtosis	82.62	0.34	0.64	-0.17
Skewness	7.37	-0.17	-0.30	0.10
Range	335	67	65	66
Minimum	148	169	151	179
Maximum	483	236	216	245
Sum	35305	37842	36630	39970
Count	185	188	191	191

CONCLUSION 1b

Based on statistical equivalence of Grade 3 literacy baselines on the F & P Benchmark and NWEA MAP-R, the sharp decrease in Mineola Grade 3 NYS ELA proficiency score between 2012-13 and 2013-14 cannot be attributed to literacy differences at the academic year outset. In addition, the RTI gains for each of the two years are comparable for the NWEA MAP-R.

ANALYSIS 1c: Phonological-Orthographic Substitution Evaluation (P-O-S-E©)

The P-O-S-E©: Phonological-Orthographic Substitution Evaluation© is a criterion-referenced assessment instrument, designed to probe for substitution errors in a child's phonological (spoken) and orthographic (written, scored as equivalent phonology) representations of target short vowels presented in monosyllabic non-word and real word spelling and reading tasks. I.e. an incorrect phoneme is substituted for the target phoneme. Silent /e/ rule test items are incorporated as a cross-check and validation of the depth of short vowel proficiency. Outcomes provide prescriptive interventional direction when indicated. Year-end response-to intervention (RTI) is assessed with the same instrument.

In Fall, 2012, the P-O-S-E© was administered to the entire Grade 3 of Mineola USFD (n=191). Based on an analysis of test outcomes, vowel training protocols were established incorporating Speech-Language Pathology (SLP), English Second Language (ESL), Reading, Special Education (SE) and General Education (GE) staff. In the spring of 2013, Grade 3 was retested. Figure 7 illustrates a histogram of baseline and RTI findings for the paired data of 191 students. Average P-O-S-E© error score was reduced from 16.3 % to 9.3%. Table 4 shows descriptive statistics for these data.

Figure 7

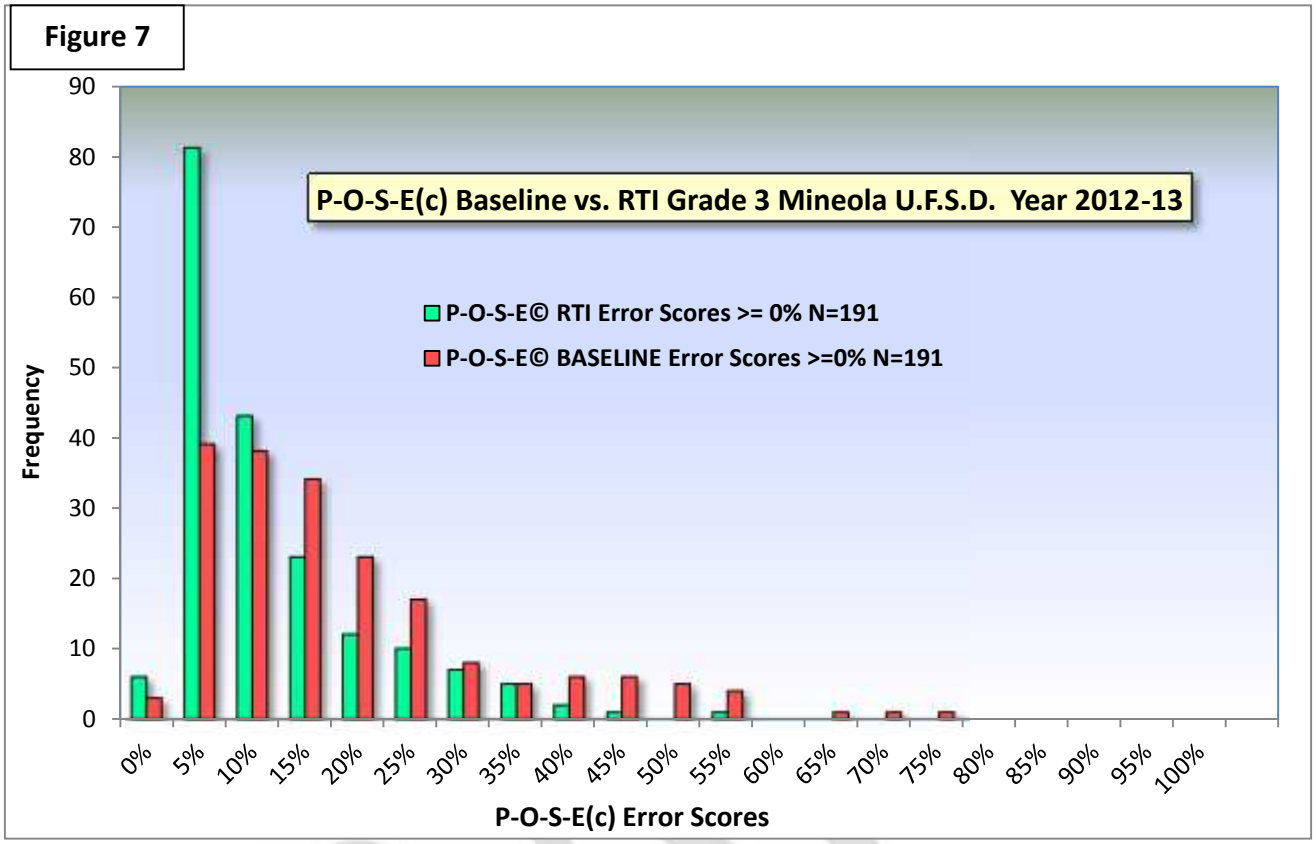


Table 5

Grade 3 P-O-S-E© Baseline vs RTI Error Scores Mineola U.F.S.D. 2012-13 year		
Descriptive Statistics	**P-O-S-E© error scores**	
	Baseline	**RTI**
Mean P-O-S-E© Error Score	**16.3%**	**9.3%**
Standard Error	1.0%	0.7%
Median	**11.7%**	**5.9%**
Mode	6.7%	0.8%
Standard Deviation	14.4%	9.6%
Sample Variance	2.1%	0.9%
Kurtosis	183.2%	267.2%
Skewness	143.1%	160.9%
Range	70.8%	52.5%
Minimum	0.0%	0.0%
Maximum	70.8%	52.5%
Sum	3120.3%	1783.4%
Count	191	191

In the Fall of 2013, the P-O-S-E© baseline was again administered to the new Grade 3, n=180. The histogram in Figure 8 compares the distribution of all Grade 3 P-O-S-E© baseline scores for 2012-13 and 2013-14. Table 6 summarizes descriptive statistics for the same data. A t-test revealed no significant difference in the distribution of P-O-S-E© error scores between the two academic years.

Figure 8

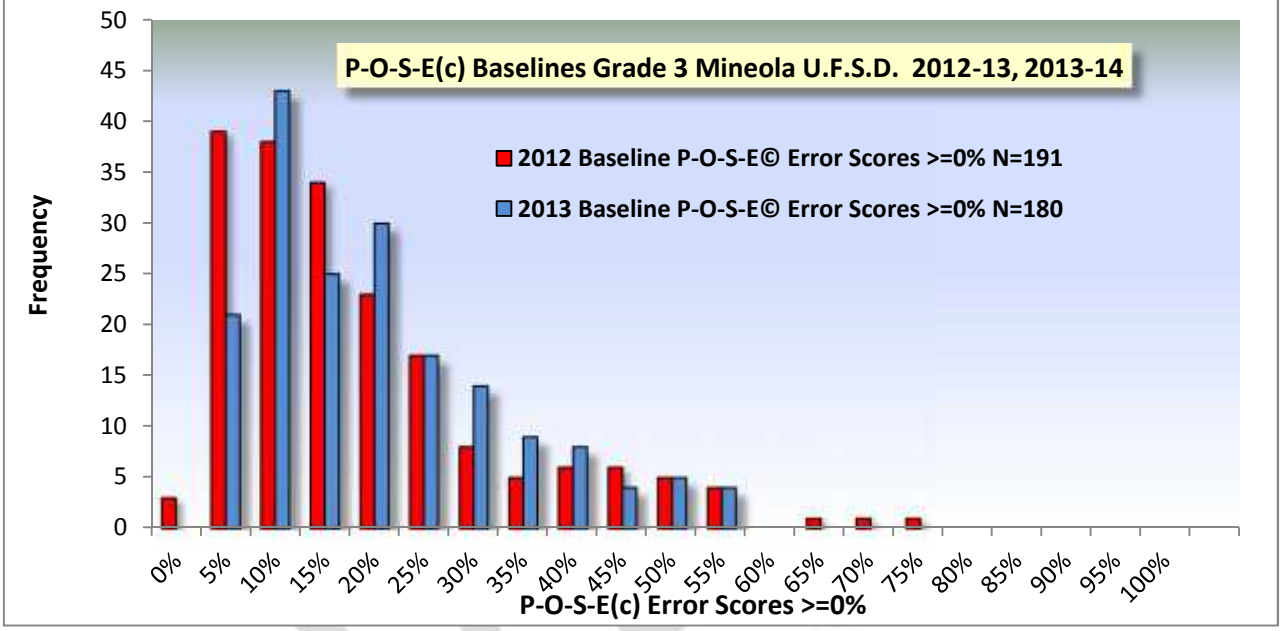


Table 6

P-O-S-E© Baseline Error Scores >= 0%
Mineola U.F.S.D. 2012-13 v. 2013-14

Descriptive Statistics	P-O-S-E© Baseline	
	2012-13	2013-14
Mean P-O-S-E© Error Score	**16.3%**	**17.4%**
Standard Error	1.0%	0.9%
Median	**11.7%**	**15.4%**
Mode	6.7%	40.0%
Standard Deviation	14.4%	12.5%
Sample Variance	2.1%	1.6%
Kurtosis	183.2%	17.4%
Skewness	143.1%	93.5%
Range	70.8%	51.7%
Minimum	0.0%	0.8%
Maximum	70.8%	52.5%
Sum	3120.3%	3133.3%
Count	191	180

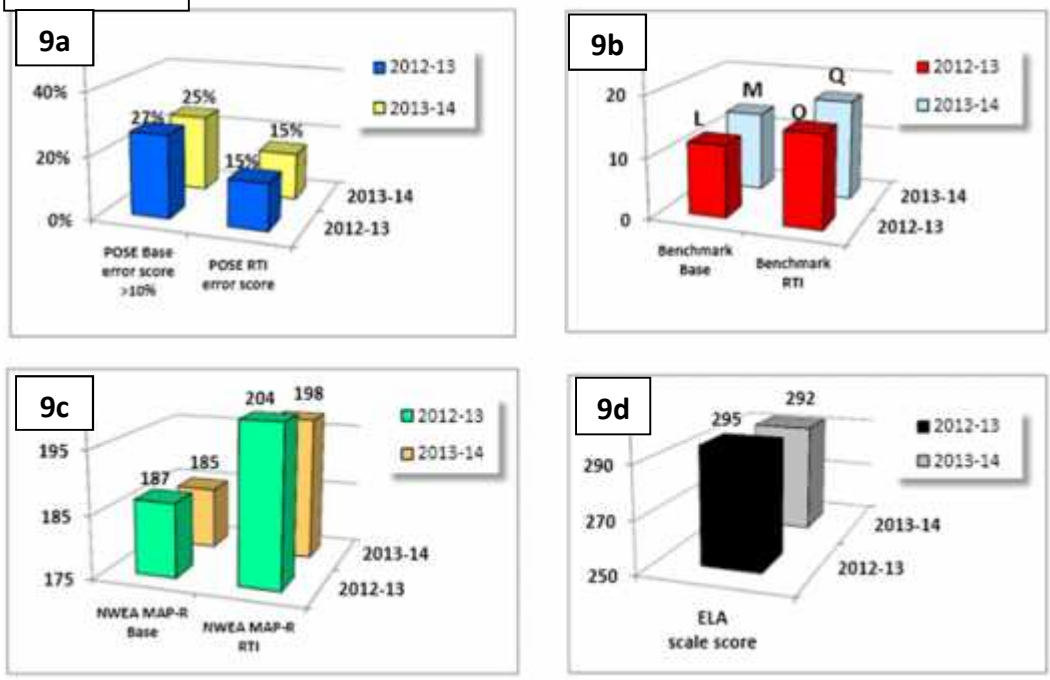
An administrative decision in 2013 limited RTI testing for that academic year to those students presenting with P-O-S-E© baseline error scores > 10%. In order to compare Grade 3 P-O-S-E© outcomes for 2012-13 with 2013-14, all literacy assessment data were re-analyzed restricting data to students with baseline P-O-S-E© error scores > 10%. This was also replicated, restricting data to students with P-O-S-E© error scores to => 25% for both years. Table 7 presents these (P-O-S-E(c) baseline > 10% error) literacy assessment instrument scores for comparison. Figures 9a-d shows the same outcomes in graphic format. Grade 3 aggregate ELA scale scores do not differ between 2013 and 2014.

Table 7

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2012-13 v. 2013-14
P-O-S-E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets (Baseline P-O-S-E©error score >10%)

Year	n	POSE Base error score	POSE RTI error score	Benchmark Base		Benchmark RTI		NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
2012-13	96	26.6%	14.8%	11.9	L	15.2	O	186.8	204.3	294.5
2013-14	96	24.8%	15.1%	13.3	M	16.7	Q	184.8	197.5	292.0

Figure 9 a-d



Tables 8 and 9 present descriptive statistics for 2012-13 and 2013-14 baseline and RTI data on the P-O-S-E©, F&P Benchmarks and the NWEA MAP-R. Grade level goals were met or exceeded on all parameters.

Table 8

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2012-13
P-O-S-E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets(Baseline P-O-S-E©error score >10% n=96)

2012-13 Parameters	POSE Base error score	POSE rti error score	Ben Base	Ben rti	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
Mean	26.60%	14.33%	11.95 L	15.23 O	186.8	204.3	294.5
Standard Error	1.40%	1.07%	0.30	0.37	1.07	1.10	3.19
Median	22.50%	12.10%	12 L	16 P	187.5	204	297
Mode	13.30%	9.20%	12 L	16 P	188	205	329
Standard Deviation	13.69%	10.49%	2.90	3.50	10.46	10.74	31.23
Sample Variance	1.87%	1.10%	8.43	12.34	109.33	115.38	975.16
Kurtosis	0.87	0.99	1.85	1.05	0.81	0.22	0.17
Skewness	1.19	1.04	-1.13	-1.02	-0.51	-0.01	-0.49
Range	60.00%	52.20%	16	21	62	49	137
Minimum	10.80%	0.30%	1 A	2 D	151	179	212
Maximum	70.80%	52.50%	17 Q	23 W	213	228	349
Sum	2553.40%	1424.00%	1147	1452	17928	19611	28276
Count	96	96	96	96	96	96	96

Table 9

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2013-14
P-O-S-E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets (Baseline P-O-S-E©error score >10% n=96)

2013-14 Parameters	POSE Base error score	POSE rti error score	Ben Base	Ben rti	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
Mean	24.80%	15.10%	13.30 M	15.70 Q	184.8	197.5	292.00
Standard Error	1.10%	0.90%	0.30	0.30	1.50	1.10	3.00
Median	20.80%	13.30%	13 M	17 Q	187	198	293
Mode	40.00%	8.30%	12 L	16 P	187	204	264
Standard Deviation	11.20%	9.30%	2.80	2.50	14.50	10.60	29.30
Sample Variance	1.30%	0.90%	7.70	6.70	210.40	111.90	855.90
Kurtosis	-0.12	0.52	0.84	0.90	-0.45	0.20	0.70
Skewness	0.86	1.07	-0.29	-0.32	-0.53	-0.41	-0.32
Range	47.50%	38.30%	16	13	62	51	167
Minimum	5.00%	2.50%	4 D	9 I	148	169	196
Maximum	52.50%	40.80%	20 T	22 V	210	220	363
Sum	2379.20%	1451.70%	1280	1606	17737	18957	28032
Count	96	96	96	96	96	96	96

Tables 10 and 11 present multiple correlations among the selected Baseline and RTI literacy assessment instruments for academic years 2012-13 and 2013-14 with students >10% baseline P-O-S-E© error scores as overarching parameter. Because the P-O-S-E© results in a percent error score, correlations bear a negative sign. For this mildly lower-scoring population, external correlations between the NYS Grade 3 ELA and the F&P BAS and NWEA MAP-R were comparable between 2014 and 2013. Correlations with ELA and the P-O-S-E© were significantly reduced in 2013.

Table 10

Multiple Correlations Grade 3 Mineola U.F.S.D. 2012-13 (P-O-S-E© Base > 10% N=96)							
2012-13 (n=96) P-O-S-E© Base >10%	P-O-S-E(c) error Base	P-O-S-E(c) error RTI	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
P O S E(c) error Base	1.00						
P O S E(c) error RTI	0.66	1.00					
Benchmark Base	-0.50	-0.40	1.00				
Benchmark RTI	0.60	0.53	0.93	1.00			
NWEA MAP-R Base	-0.47	-0.34	0.61	0.61	1.00		
NWEA MAP-R RTI	-0.36	-0.36	0.61	0.53	0.73	1.00	
ELA	-0.59	-0.50	0.66	0.55	0.65	0.67	1.00

Table 11

Multiple Correlations Grade 3 Mineola U.F.S.D. 2013-14 (P-O-S-E© Base > 10% n=96)							
2013-14 (n=96) P-O-S-E© Base –>10%	% POSE Base	% POSE RTI	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
P-O-S-E(c) error Base	1.00						
P-O-S-E(c) error RTI	0.62	1.00					
Benchmark Base	-0.62	-0.38	1.00				
Benchmark RTI	0.52	0.37	0.88	1.00			
NWEA MAP-R Base	-0.44	-0.28	0.68	0.53	1.00		
NWEA MAP R RTI	0.44	0.29	0.66	0.50	0.67	1.00	
ELA	-0.43	-0.29	0.65	0.52	0.64	0.67	1.00

Table 12 presents the literacy assessment instrument scores for students scoring >= 25% error on the P-O-S-E© baseline test. Figures 10a-d present the same data in graphic format. Note the academic progress on RTI measures relative to baselines. Mineola UFSD Grade 3 aggregate ELA scale scores to not differ between 2013 and 2014.

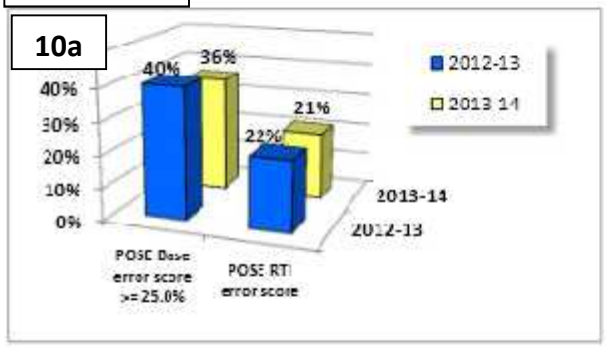
Table 12

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2012-13 v. 2013-14
P-O-S-E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets (Baseline P-O-S-E©error score >=25.0%)

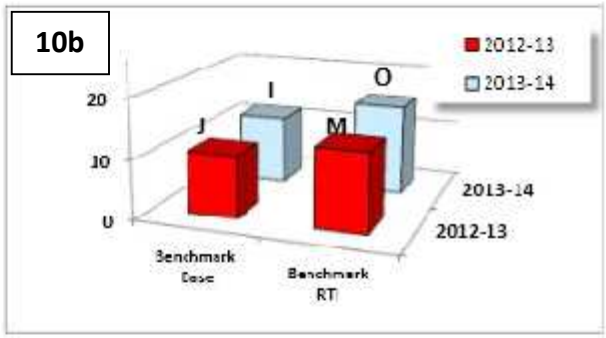
Year	n	POSE Base error score	POSE RTI error score	Benchmark Base		Benchmark RTI		NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
2012-13	38	40.4%	21.7%	10.4	J	13.2	M	181.8	199.7	275.6
2013-14	30	36.1%	21.2%	11.6	L	15.3	O	178.3	193.2	279.1

Figure 10 a-d

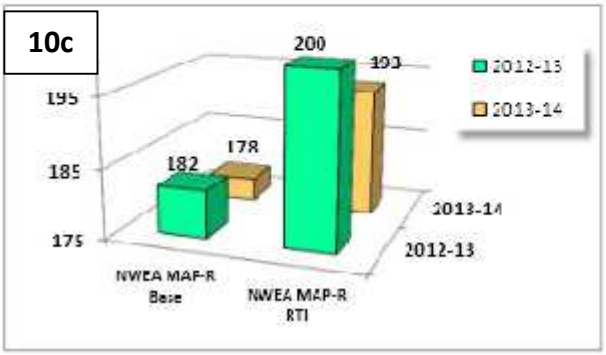
10a



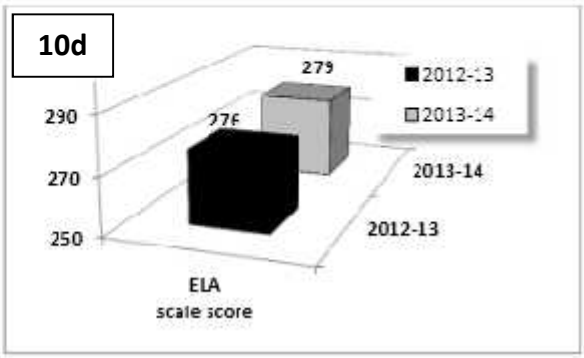
10b



10c



10d



Tables 13 and 14 provide assessment instrument outcome comparisons for both academic years in greater detail. Grade-appropriate progress is to be noted on all test instruments for both 2013 and 2013 with data matched for P-O-S-E© baseline error scores >=25.0%

Table 13

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2012-13
P O S E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets(Baseline P-O-S-E©error score >=25.0% n=58)

2012-13 Parameters	POSE Base error score	POSE RTI error score	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
Mean	40.35%	21.69%	10.39 J	13.21 M	181.8	199.7	275.61
Standard Error	1.89%	1.82%	0.44	0.53	1.78	1.59	5.01
Median	38.75%	21.65%	11 K	14 N	182.5	200	276
Mode	26.70%	9.20%	12 L	15 O	188	201	276
Standard Deviation	11.66%	11.22%	2.70	3.28	10.99	9.79	32.75
Sample Variance	1.36%	1.26%	7.27	10.77	120.73	95.89	1072.68
Kurtosis	0.39	0.15	-0.19	0.24	1.94	0.42	-0.22
Skewness	0.80	0.47	-0.70	-0.78	-0.19	0.48	0.17
Range	45.80%	51.70%	11	14	52	45	126
Minimum	25.00%	0.80%	4 D	4 D	151	180	212
Maximum	70.80%	52.50%	15 O	18 R	213	225	338
Sum	1533.70%	824.80%	395	502	6907	7589	10473
Count	38	38	38	38	38	38	38

Table 14

Mineola U.F.S.D. Grade 3 Literacy Baseline v. RTI Scores 2013-14
P-O-S-E©; F & P Benchmarks; NWEA MAP; NYS ELA
Matched student sets(Baseline P-O-S-E©error score >=25.0% n=39)

2012-13 Parameters	POSE Base error score	POSE RTI error score	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
Mean	36.13%	21.15%	11.59 L	15.28 O	178.3	195.2	279.13
Standard Error	1.33%	1.56%	0.40	0.38	2.32	1.73	4.40
Median	34.17%	20.00%	12 L	16 P	177	192	286
Mode	40.00%	#N/A	12 L	15 P	187	193	282
Standard Deviation	8.32%	9.73%	2.47	2.37	14.49	10.79	27.51
Sample Variance	0.69%	0.95%	6.09	5.63	209.84	116.47	756.59
Kurtosis	-0.98	-0.69	1.17	1.34	-0.51	-0.04	1.19
Skewness	0.58	0.42	-0.85	-0.94	0.03	-0.48	-0.94
Range	26.67%	35.83%	12	11	56	45	138
Minimum	25.83%	5.00%	4 D	9 I	148	169	196
Maximum	52.50%	40.83%	16 P	20 T	204	214	334
Sum	1409.17%	825.00%	452	596	6953	7534	10886
Count	39	39	39	39	39	39	39

Tables 15 and 16 present multiple correlations among the selected Baseline and RTI literacy assessment instruments for academic years 2012 and 2013 with a =>25% baseline P-O-S-E© error score as overarching parameter. Because the P-O-S-E© is scored as percent error, its correlations with other test instruments bear a negative sign. For this lower-scoring population subset, ALL external correlations between the NYS Grade 3 ELA and the alternative literacy assessment instruments –including the P-O-S-E© - were significantly lower for the in 2014 than in 2013.

Table 15

Multiple Correlations Grade 3 Mineola U.F.S.D. 2012-13 (P-O-S-E© Base => 25% n=38)

2012-13 (n=38) P-O-S-E© Base =>25%	P-O-S-E(c) error Base	P-O-S-E(c) error RTI	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
P-O-S-E(c) error Base	1.00						
P-O-S-E(c) error RTI	0.54	1.00					
Benchmark Base	-0.67	-0.47	1.00				
Benchmark RTI	-0.57	-0.55	0.95	1.00			
NWEA MAP-R Base	-0.35	-0.26	0.56	0.52	1.00		
NWEA MAP-R RTI	-0.24	-0.38	0.50	0.56	0.61	1.00	
ELA	-0.52	-0.50	0.76	0.76	0.62	0.68	1.00

Table 16

Multiple Correlations Grade 3 Mineola U.F.S.D. 2013-14 (P-O-S-E© Base => 25% N=39)

2013-14 (n=39) P-O-S-E© Base =>25%	P-O-S-E(c) error Base	P-O-S-E(c) error RTI	Benchmark Base	Benchmark RTI	NWEA MAP-R Base	NWEA MAP-R RTI	ELA scale score
P O S E(c) error Base	1.00						
P O S E(c) error RTI	0.38	1.00					
Benchmark Base	0.53	0.15	1.00				
Benchmark RTI	-0.34	-0.19	0.80	1.00			
NWEA MAP-R Base	-0.37	-0.05	0.75	0.55	1.00		
NWEA MAP-R RTI	-0.42	-0.25	0.66	0.62	0.51	1.00	
ELA	0.33	0.15	0.59	0.60	0.44	0.52	1.00

Tables 17 and 18 present t-tests supporting the statistical significance of grade-appropriate Mineola UFSD Grade 3 RTI advances matched on the P-O-S-E©; F&P BAS and NWEA MAP-R for both academic years 2012-13 and 2013-14. The student n is 191 for 2012-13 and 96 for 2013-14, the latter reduced because P-O-S-E© RTI testing was limited to Grade 3 students with baseline P-O-S-E© error scores greater than 10%.

Table 17

Grade 3 Mineola U.F.S.D. Literacy Baseline vs. RTI Scores 2012-13
191 matched sets of students

Grade 3 2012-13	POSE Base	POSE rti	Ben Base#	Ben rti#	NWEA Base	NWEA RTI
Mean	16.3%	9.3%	13.32	16.74	186.75	204.28
Variance	2.1%	0.9%	7.24	10.87	109.33	115.38
Net Change (Improvement)	7.0%		3.42		17.53	
Observations	191	191	191	191	96	96
Pearson Correlation	0.78		0.91		0.73	
Hypothesized Mean Difference	0.0%		0.00		0.00	
df	190		190.00		95.00	
t Stat	10.53		33.39		21.86	
P(T<=t) one-tail	0.0000		0.0000		0.0000	
t Critical one-tail	1.65		1.65		1.66	
P(T<=t) two-tail	0.0000		0.0000		0.0000	
t Critical two-tail	1.97		1.97		1.99	

Table 18

Grade 3 Mineola U.F.S.D. Literacy Baseline v. RTI Scores 2013-14
Matched student Sets (Baseline P-O-S-E© >10% t-Test of differences (n=96)

Grade 3 2013-14	% POSE Base	% POSE RTI	Benchmark Base	Benchmark RTI	NWEA Base	NWEA RTI
Mean	24.8%	15.1%	13.3 M+	16.7 Q-	184.8	197.5
Variance	1.3%	0.9%	7.7	6.7	210.4	111.9
Net Change (Improvement)	9.7%		3.40		12.7	
Observations	96	96	96	96	96	96
Pearson Correlation	0.62		0.88		0.67	
Hypothesized Mean Difference	0		0		0	
df	95		95		95	
t Stat	10.4031		-24.9490		-11.4731	
P(T<=t) one-tail	1.1402E-17		8.5643E-44		6.0172E-20	
t Critical one-tail	1.6611		1.6611		1.6611	
P(T<=t) two-tail	2.2804E-17		1.7129E-43		1.2034E-19	
t Critical two-tail	1.9853		1.9853		1.9853	

CONCLUSION 1c

Results of the Mineola UFSD Grade 3 P-O-S-E© baseline and RTI measure for 2012-13 and 2013-14, the student population error scores restricted to >10% and +>25% show consistent, grade-appropriate progress in short vowel proficiency. In addition, consistent correlations are seen across both years and defined short vowel proficiency ranges among the alternative measures of literacy used by the Mineola UFSD.

2. *The NYS ELA test item review criteria may have differed between school years 2012-13 and 2013-14.*

ANALYSIS 2

According to the NYS website: (https://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests.)

"NYSED uses the Item Review Criteria to help ensure that each item:

1. ***is clear;***
2. ***is fair;***
3. ***measures a specific Common Core standard (or standards) with fidelity; and***
4. ***conforms to the specifications for the item type…"***

CONCLUSION 2

There was no reported change in NYS ELA Item Review Criteria between 2012-13 and 2013-14.

3. *The NYS ELA raw data scoring practices may have differed between school years 2012-13 and 2013-14.*

For 2013 and 2014, in pursuit of objectivity, Mineola UFSD used the same contracted, state-approved scoring service for ELA testing in both years.

However, in 2014, NYS summarily discarded, p*ost hoc,* three Grade 3 ELA test items. This accounted for the 6 point differential between the 55 point maximum score for the 2013 ELA and the 49 point maximum score for the 2014 ELA – an arbitrary net reduction of 10.9% in the 2014 scoring base.

The original Grade 3 2014 ELA examination was comprised of 31 multiple choice items (1 point each), 8 short answer items (2 points each) and 2 open format items (4 point each). It has been reported (http://citylimits.org/2015/03/16/ny-state-must-clear-up-mystery-of-missing-test-items/) that questions #29, #30 (1 point each) (https://www.engageny.org/file/105016/download/2014gr3elaannotatedquestionsmaptothestandards.pdf) and #47 (4 points) were deleted from the final NYS scoring protocol. This reduced the maximum 2013 Grade 3 ELA scoring base by six points from the 2013 protocol.

CONCLUSION 3

There was a major, arbitrary change in NYS ELA Item scoring procedure, the state deleting three test items, after-the-fact, totaling 6 points out of 55 from the Grade 3 2104 ELA examination.

4. ***The Grade 3 NYS ELA test instrument for 2013-14 may have differed in reading level of test passages or questions from that of 2012-13. As the NYS ELA is administered near the end of school year, the nominal Grade 3 reading level should be 3.9 or 3.10 (3rd grade, 9th or 10th month). A post-hoc analysis of the reading level of publicly released NYS ELA test content for school years 2012-13 and 2013-14 has been systematically applied to address this critical variable.***

ANALYSIS 4

In response to public pressure, NY State released, *post hoc,* a total of nine "selected" reading passages from the 2013 and 2014 ELA examinations. (https://www.engageny.org/resource/new-york-state-common-core-sample-questions) Three passages from 2013:

"Copycat Elephants"	*"Go Fish"*	*"Jump"*

and six passages from 2014:

"David & the Phoenix"	*"Sea Turtles"*	*"Sugaring Time"*	*"Otter in the Cove"*	*"Snow Fun on the Run"*	*"Science Friction"*

In order for an ELA assessment of literacy proficiency to be valid, the reading level of test passages must be within the appropriate range for the given grade. To that end, the passages were evaluated in two ways. First, each passage was subjected to a Metametrics Lexile analysis (https://Lexile.com/analyzer/.) Table 19 presents the Metametrics Lexile analysis of the 2013 ELA passages compared to the 2012 CCSS and Fountas and Pinnell Lexiles recommended for Grade 3. Table 20 shows the average word count per passage for the same year ELA passages.

Table 19

2012-13
Metametrics Lexile Analysis NYS ELA Grade 3
Analysis of 3 NYS-released reading passages

	Passage 1	Passage 2	Passage 3	Average
Lexiles	*"Copycat Elephants"*	*"Go Fish"*	*"Jump"*	
Grade 3 CCSS Lexile Text Measures Recommend Range	520-820			670
Fountas & Pinnell Lexile Range	500-675			588
Metametrics Lexile Analysis	780	560	410	583
CCSS-Analysis Metametrics Lexile Midpoint Difference				87

Table 20

2012-13
Length Analysis NYS ELA Grade 3
Analysis of 3 NYS-released reading passages

	Passage 1	Passage 2	Passage 3	Average
	"Copycat Elephants"	*"Go Fish"*	*"Jump"*	
Total Word Count	367	676	512	518

Table 21 presents the Metametrics Lexile analysis of the 2014 ELA passages compared to the 2012 CCSS and the Fountas and Pinnell Lexiles recommended for Grade 3. Table 22 shows the average word count per passage for the same year ELA passages.

Table 21

2013-14
Metametrics Lexile Analysis NYS ELA Grade 3
Analysis of 6 NYS released reading passages

	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Passage 6	Average
Lexiles	*"David & the Phoenix"*	*"Sea Turtles"*	*"Sugaring Time"*	*"Otter in the Cove"*	*"Snow Fun on the Run"*	*"Science Friction"*	
Grade 3 CCSS Lexile Text Measures Recommend Range	520-820						670
Fountas & Pinnell Lexile Range	500-675						588
Metametrics Lexile Analysis	870	930	800	630	820	660	785
CCSS-Analysis Metametrics Lexile Midpoint Difference							-115

Table 22

2013-14
Length Analysis NYS ELA Grade 3
Analysis of 6 NYS-released reading passages

	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Passage 6	Average
	"David & the Phoenix"	*"Sea Turtles"*	*"Sugaring Time"*	*"Otter in the Cove"*	*"Snow Fun on the Run"*	*"Science Friction"*	
Total Word Count	571	653	439	651	556	787	610

When compared with the recommended 2012 CCSS Lexile values for Grade 3, the average Metametrics Lexile of three released 2013 Grade 3 ELA reading passage was 87 points BELOW the recommended 2012 CCSS Grade 3 Lexile (Table 19). The same criteria applied to the six released 2014 passages resulted in an average Metametrics Lexile 115 points ABOVE the recommended 2012 CCSS Grade 3 Lexile (Table 21). I.e., the 2014 released passages were 205 Lexile points more difficult than the 2013 released passages.

In addition, the three released questions from 2013 range in Metametrics Lexile level from well below grade level (410) to well above grade level (780). A similar disparate range of Metametrics Lexile level applies to the six released questions from 2014 (660-930)

A perusal of Tables 20 and 22 demonstrates an average released passage word length difference from 516 words in 2013 to 610 words in 2014, an increase of 18.2%.

In a second form of Grade 3 NYS ELA reading level analysis, twenty five different algorithms for assessing reading level (see Appendix for details) were sequentially applied and averaged for all three NYS released passages from 2013 (Table 23) and six released passages from 2014 (Table 24).

Table 23

Multiple Instrument Analyses of NYS-Released ELA Test Passages: Grade 3, 2013								
Standardized Tests of Reading Level		"Copycat Elephants"		"Go Fish"		"Jump"		Mean Grade Level Per Test
		Grade Level	Reader Age	Grade Level	Reader Age	Grade Level	Reader Age	
Automated (ARI) Readability Index		5.4	10-11	4.5	9-10	0.9	5-6	3.6
1	Bormuth Grade Placement	8.3	13-14	7.9	12-13	6.9	11-12	7.7
2	Coleman-Liau	7.5	12-13	6.4	11-12	2.7	7-8	5.5
3	Danielson-Bryan 1	5.0	10-11	4.9	9-10	4.1	9-10	4.7
4	Degrees of Reading Power (grade equivalent)	4.7	9-10	3.3	8-9	1.7	6-7	3.2
5	Easy Listening Formula	5.0	10-11	2.8	7-8	1.5	6-7	3.1
6	Flesch-Kincaid	5.9	10-11	3.3	8-9	1.4	6-7	3.5
7	Flesch-Kincaid (simplified)	5.9	10-11	3.3	8-9	1.3	6-7	3.5
8	Fry	7.0	12-13	4.0	9-10	1.0	6-7	4.0
9	Gunning Fog	8.3	13-14	4.5	9-10	3.3	8-9	5.4
10	Harris-Jacobson Wide Range Formula	4.8	9-10	4.5	9-10	3.4	8-9	4.2
11	Modified SMOG	6.0	11-12	3.0	8-9	2.0	7-8	3.7
12	New Automated Readability Index	4.5	9-10	4.0	9-10	0.0	5-6	2.8
13	New Automated Readability Index	4.1	9-10	3.6	8-9	0.0	5-6	2.6
14	New Dale-Chall	4.0	9-10	3.0	8-9	2.0	7-8	3.0
15	New Farr, Jenkins, Paterson (Kincaid)	5.0	10-11	2.0	7-8	0.0	5-6	2.3
16	New Fog Count (Kincaid)	5.1	10-11	2.8	7-8	1.9	6-7	3.3
17	Powers, Sumner, Kearl (Dale-Chall)	5.2	10-11	4.8	9-10	4.2	9-10	4.7
18	Powers, Sumner, Kearl (Farr, Jenkins, Paterson)	5.0	10-11	4.4	9-10	3.9	8-9	4.4
19	Powers, Sumner, Kearl (Flesch)	5.3	10-11	4.5	9-10	3.9	8-9	4.6
20	Powers, Sumner, Kearl (Gunning Fog)	5.0	10-11	4.1	9-10	3.8	8-9	4.3
21	Raygor Estimate	7.0	12-13	4.0	9-10	Failed*	Failed*	5.5
22	SMOG	9.5	14-15	7.4	12-13	5.1	10-11	7.3
23	SMOG (simplified)	9.0	14-15	7.0	12-13	4.0	9-10	6.7
24	Spache Revised	3.0	8-9	2.8	7-8	2.5	7-8	2.8
25	Wheeler-Smith	4+	9-10	3.0	8-9	2.0	7-8	2.5
	Average (Mean)	5.9	11.1 yrs.	4.2	9.4 yrs.	2.8	7.9 yrs.	4.2

Table 24

Multiple Instrument Analyses of NYS-Released ELA Test Passages: Grade 3, 2013-2014														
Standardized Tests of Reading Level		"David and the Phoenix"		"Otter in the Cove"		"Science Friction"		"Sea Turtles"		"Snow Fun on the Run"		"Sugaring Time"		Mean Grade Level Per Test
		Grade Level	Reader Age	Grade Level	Reader Age	Grade Level	Reader Age	Grade Level	Reader Age	Grade Level	Reader Age	Grade Level	Reader Age	
Automated (ARI) Readability Index		5.7	10-11	3.9	8-9	3.2	8-9	6.5	11-12	6.4	11-12	3.7	8-9	4.9
1	Bormuth Grade Placement	8.2	13-14	8.0	13-14	7.8	12-13	8.6	13-14	8.5	13-14	8.0	13-14	8.2
2	Coleman-Liau	6.5	11-12	6.0	11-12	5.1	10-11	8.3	13-14	7.9	12-13	5.8	10-11	6.6
3	Danielson-Bryan 1	5.0	10-11	4.7	9-10	4.6	9-10	5.3	10-11	5.3	10-11	4.6	9-10	4.9
4	Degrees of Reading Power (grade equivalent)	4.3	9-10	3.6	8-9	3.0	8-9	5.5	10-11	5.1	10-11	3.9	8-9	4.2
5	Easy Listening Formula	4.2	9-10	2.8	7-8	2.7	7-8	4.3	9-10	4.2	9-10	3.4	8-9	3.6
6	Flesch-Kincaid	5.2	10-11	3.4	8-9	3.3	8-9	5.3	10-11	5.2	10-11	4.2	9-10	4.4
7	Flesch-Kincaid (simplified)	5.2	10-11	3.4	8-9	3.2	8-9	5.2	10-11	5.1	10-11	4.1	9-10	4.4
8	Fry	6.0	11-12	4.0	9-10	3.0	8-9	6.0	11-12	6.0	11-12	5.0	10-11	5.0
9	Gunning Fog	6.6	11-12	4.2	9-10	5.6	10-11	7.1	12-13	5.6	10-11	5.4	10-11	5.8
10	Harris-Jacobson Wide Range Formula	5.5	10-11	4.7	9-10	5.4	10-11	6.9	11-12	5.9	10-11	4.8	9-10	5.5
11	Modified SMOG	5.0	10-11	3.0	8-9	5.0	10-11	5.0	10-11	5.0	10-11	3.0	8-9	4.3
12	New Automated Readability Index	4.1	9-10	3.1	8-9	2.6	7-8	5.4	10-11	5.5	10-11	2.5	7-8	3.9
13	New Automated Readability Index	3.8	8-9	2.7	7-8	2.1	7-8	5.1	10-11	5.2	10-11	2.1	7-8	3.5
14	New Dale-Chall	4.0	9-10	4.0	9-10	4.0	9-10	4.0	9-10	5.5	10-12	4.0	9-10	4.3
15	New Farr, Jenkins, Paterson (Kincaid)	4.0	9-10	3.0	8-9	2.0	7-8	4.0	9-10	4.0	9-10	4.0	9-10	3.5
16	New Fog Count (Kincaid)	5.3	10-11	3.0	8-9	2.9	7-8	5.4	10-11	3.9	8-9	4.0	9-10	4.1
17	Powers, Sumner, Kearl (Dale-Chall)	4.9	9-10	5.2	10-11	5.1	10-11	5.2	10-11	5.4	10-11	5.2	10-11	5.2
18	Powers, Sumner, Kearl (Farr, Jenkins, Paterson)	4.8	9-10	4.6	9-10	4.4	9-10	4.8	9-10	4.9	9-10	4.8	9-10	4.7
19	Powers, Sumner, Kearl (Flesch)	4.8	9-10	4.5	9-10	4.5	9-10	4.9	9-10	4.9	9-10	4.6	9-10	4.7
20	Powers, Sumner, Kearl (Gunning Fog)	4.6	9-10	4.0	9-10	4.4	9-10	4.7	9-10	4.3	9-10	4.3	9-10	4.4
21	Raygor Estimate	6.0	11-12	5.0	10-11	3.0	8-9	6.0	11-12	6.0	11-12	4.0	9-10	5.0
22	SMOG	7.6	12-13	6.0	11-12	7.3	12-13	8.1	13-14	7.8	12-13	6.6	11-12	7.2
23	SMOG (simplified)	7.0	12-13	5.0	10-11	7.0	12-13	7.0	12-13	7.0	12-13	6.0	11-12	6.5
24	Spache Revised	3.3	8-9	3.0	8-9	2.8	7-8	3.6	8-9	3.3	8-9	2.9	7-8	3.2
25	Wheeler-Smith	4+	9-10	3.0	8-9	3.0	8-9	4+	9-10	4+	9-10	4+	9-10	3.0
	Average (Mean)	5.2	10.4	4.1	9.4	4.1	9.3	5.6	10.8	5.5	10.8	4.4	9.8	4.8
	S.D.	1.22		1.24		1.58		1.24		1.26		1.30		1.24

The average reading level for the three released Grade 3 2013 ELA passages as calculated using 25 reading levels assessment algorithms was grade 4.2 ranging from 2.8 to 5.9. The average of 25 reading levels assessment algorithms for the six released Grade 3, 2014 ELA passages was at level 4.8 ranging from 4.1 to 5.6. By definition, a normative reading level value for the 9th month of grade 3 would be 3.9.

A similarly detailed analysis was executed for Mineola USFD Grade 4 NYS ELA 2013 and 2014. Table 25 presents a summary of Metametrics Lexile analysis of the NYS ELA released passages for both Grade 3 and Grade 4, 2013 and 2104. Note that the average Metametrics Lexile for the Grade 3 2014 passages is virtually identical (785) to the Grade 4 2014 passages (783). In addition, the 2014 Metametrics Lexiles for both Grade 3 and Grade 4 were significantly higher in year 2014 than year 2013. Table 26 adds the average of 25 reading level algorithms for Grade 3 and Grade 4 NYS ELA passages both 2013 and 2014. Note that the average calculated reading level for the Grade 3 2014 NYS ELA passages is identical to that for Grade 4 in the same year.

Table 25

Mineola USFD Grade 3/4 Lexile Analysis and NYS ELA % Literacy Proficiency Levels 2012-13 vs. 2013-14

NYS ELA ITEM_DESC	Number released passages	Grade 3 N Tested	Mean ELA Scale Score	P1	P2	P3	P4	P3+P4	Metametrics Lexile Analysis Released Passages	CCSS Recommended Lexile for Grade	Fontas & Pinnell Recommended Lexile for Grade
Grade 3 ELA 2012-13	3	214	307	27%	31%	40%	3%	43%	583		
Grade 3 ELA 2013-14	6	191	302	30%	37%	30%	3%	33%	785	720	588
Difference	3	-23	-5	3%	6%	-10%	0%	-10%	202		
NYS ELA ITEM_DESC		Grade 4 N Tested	Mean Score	P1	P2	P3	P4	P3+P4	Metametrics Lexile Analysis Released Passages	CCSS Recommended Lexile for Grade	Fontas & Pinnell Recommended Lexile for Grade
Grade 4 ELA 2012-13	3	205	310	15%	45%	29%	11%	41%	783		
Grade 4 ELA 2013-14	6	216	305	23%	41%	26%	10%	36%	840	808	763
Difference	3	11	-5	8%	-4%	-3%	-1%	-5%	57		

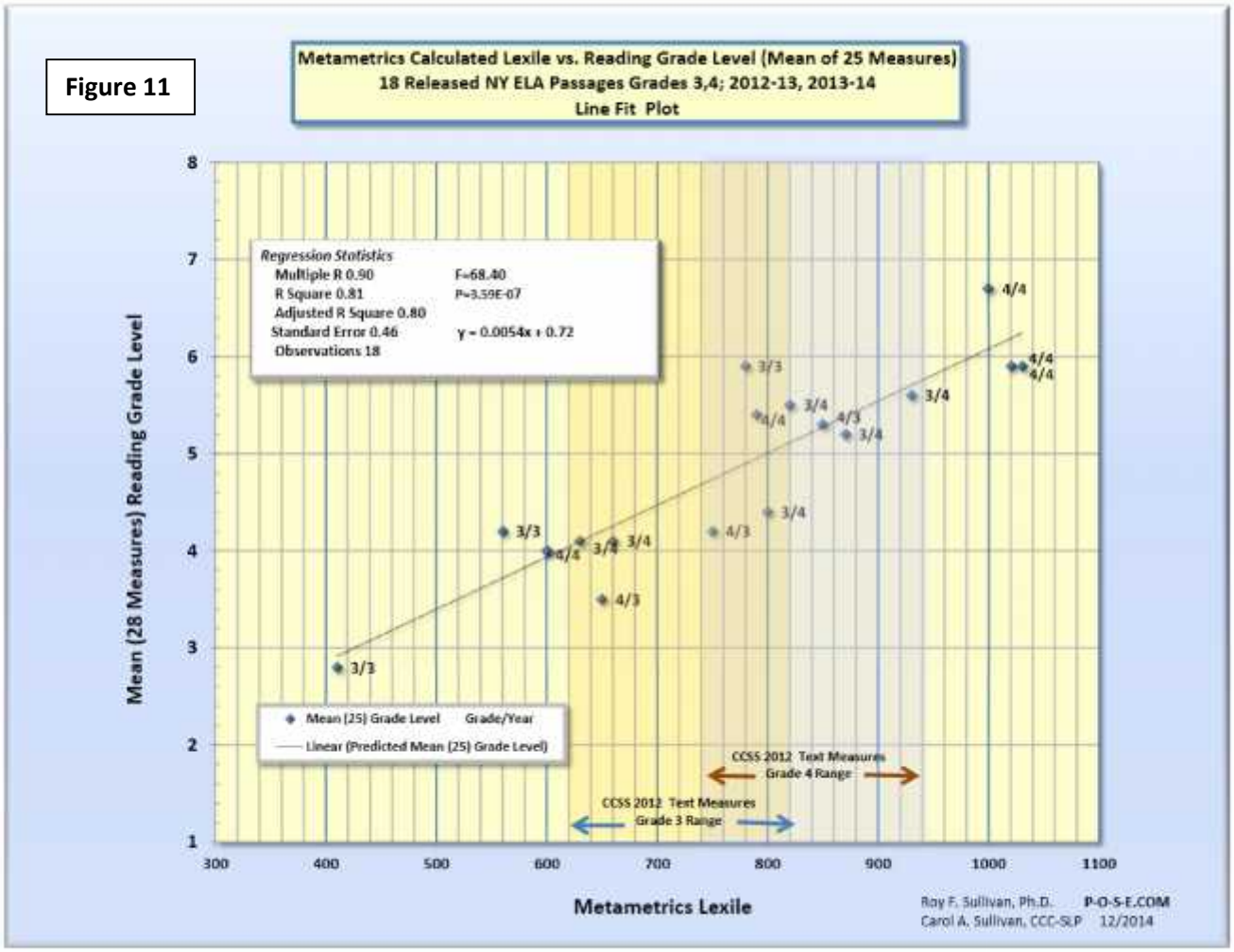
Table 26

Summary Metametrics Lexile and Average Reading Level Analyses for NYS ELA released sample passages Grade 3 and 4, 2013 and 2014.					
Grade/Year	Passage	Metatmetrix Lexile	Mean (25) Grade Level	Lexile Means	Grade Level Means
Grade 3 2012-13 ELA	"Copycat Elephants"	780	5.9	583	4.3
	"Go Fish"	560	4.2		
	"Jump"	410	2.8		
Grade 3 2013-14 ELA	"David & the Phoenix"	870	5.2	785	4.8
	"Sea Turtles"	930	5.6		
	"Sugaring Time"	800	4.4		
	"Otter in the Cove"	630	4.1		
	"Snow Fun on Run"	820	5.5		
	"Science Friction"	660	4.1		
Grade 4 2012-13	"Greeting the Sun"	650	3.5	783	4.3
	"Sitti's Secrets"	750	4.2		
	"Story of Tu-tok-a-nu-la"	950	5.3		
Grade 4 2013-14	"Cave of the Oilbird"	600	4	840	5.3
	"Pecos Bill ..."	1030	5.9		
	"When Animals..."	790	5.4		
	"Call of the Wild"	1000	6.7		
	"Lawn Boy"	600	4		
	"Elephants ...Boots"	1020	5.9		

Figure 11 presents a graph of all Grade 3 and Grade 4 NYS ELA released passage average (25 algorithms) reading levels as a function of measured Metametrics Lexile values. In the legend "3/4", for example, the numerator represents the grade, 3rd or 4th. The denominator indicates the year of ELA, 2013 OR 2014.

To the lower left, the "3/3" data point indicates a Grade 3 ELA test passage from 2013 measuring 2.8 years in reading level and 410 in Metametrics Lexile. In the upper right quadrant, one sees a group of three 2014 ELA test passages for Grade "4/4" measuring 5.4, 5.9, 5.9 and 6.7 years in reading level. Metametrics Lexiles for the same passages ranged from 1000 to 1030. There are actually Grade 3 ELA passages at higher average reading levels than some Grade 4 passages.

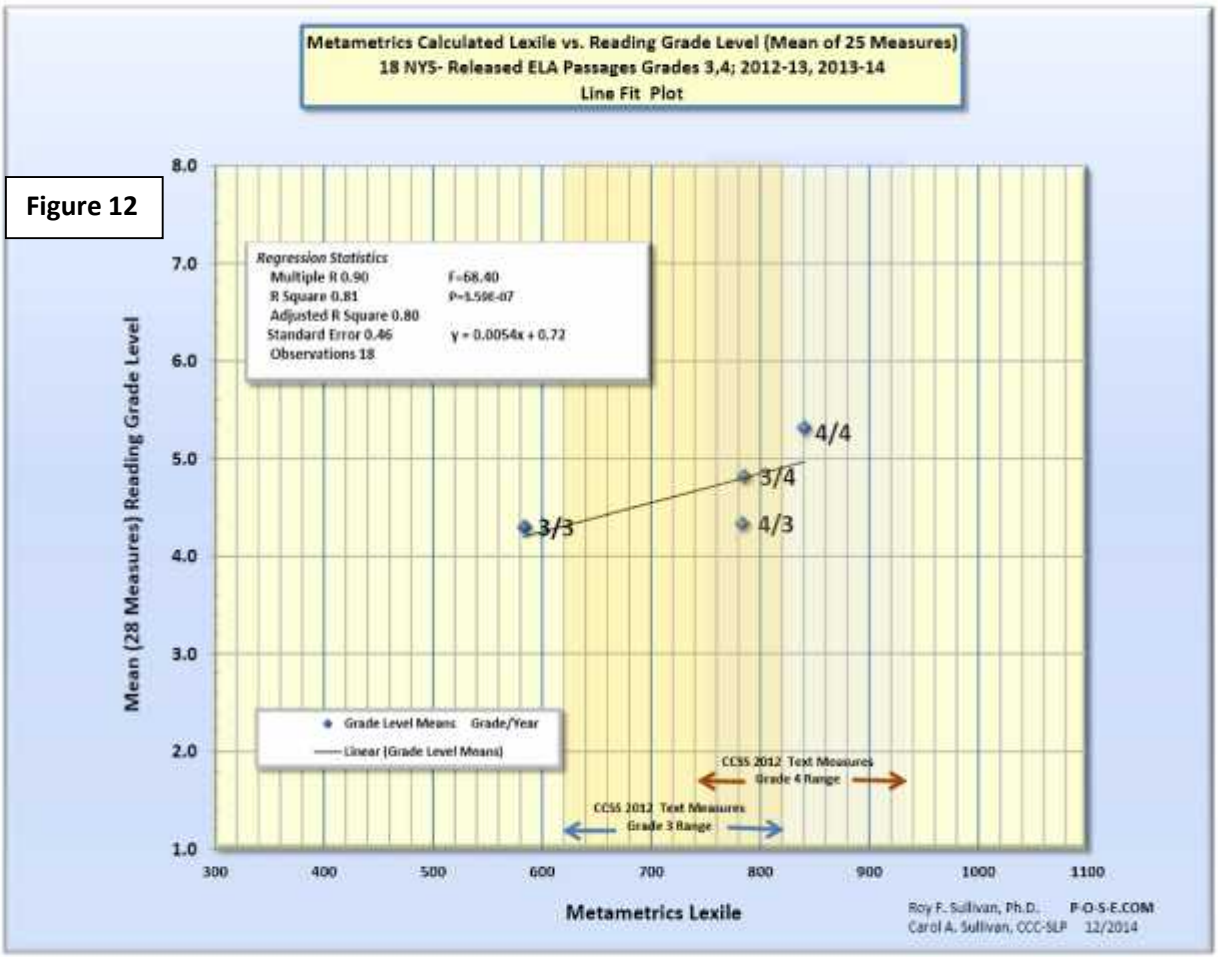
Figure 11



Smoothing the scatter of individual passage levels, figure 12 averages the NYS ELA passages reading levels for 2013 and 2014 by grade and year tested. On-target reading level for Grade 3 at the time of ELA testing would be 3.9 years for May and 3.10 for June. For Grade 4 on-target reading levels would coincide with May as 4.9 or June as 4.10. Note that the average reading levels, for the Grade 3 ELA passages analyzed, exceeded grade levels for 2013 and 2014. The average reading level for Grade 4 ELA passages was below the predicted level for 2013 and above grade for 2014. The 2014 Grade 3 passages average reading level exceeded that for Grade 4 in 2013.

When viewed as Metametrics Lexiles, the Grade 4 passages in 2013 are virtually identical with the Lexile for Grade 3 passages in 2014. There is a high correlation (.90) between Metametrics Lexile measurements and average (25 measures) reading level.

Figure 12



CONCLUSION 4

The reading level of NYS-released Grade 3 NY ELA passages was notably higher in 2014 on Metametrics Lexile analysis and in the average of 25 reading level assessment algorithms than in 2013. In both years, a significant number of NYS-released reading passages exceeded or fell below the target literacy grade level range. Reading passage word length was 18% longer in 2014 than in 2013. The disparity between Grade 3 NYS ELA P3+P4 literacy proficiency levels for Mineola UFSD in 2013 vs. 2014 is inconsistent with applied alternative measures of literacy.

Overlapping and disparate values were obtained using calculated Metametrics Lexile Levels and average calculated (25 algorithms) reading levels of the released NYS ELA 2013 and 2014 test passages. These findings provide justification for an objective re-examination of the construct validity of the entire NYS ELA test composition for Grades 3 and 4 in these years.

The significant proportion of predominantly Spanish/Hispanic students in the Mineola USFD may also be a significant variable in assessing the impact of invalid test items on literacy proficiency relative to other school districts.

5. ***The Grade 3 NYS ELA raw-to-scale score polynomial transformation and scale-score-to-literacy-proficiency-level boundaries or conditions underlying those boundaries may have differed significantly between school years 2012-13 and 2013-14.*** **See:**

http://www.p12.nysed.gov/irs/ela-math/equatingexplained.html .

> ***"On the 2014 tests, year-to-year raw score changes for Level 3 were small and varied by grade. Raw scores went down slightly on 6 tests (indicating slightly harder tests in 2014 compared to 2013 for Grades 3, 4, and 7 ELA and Grades 3, 5, and 6 Math) and went slightly up on 4 tests (indicating slightly easier tests in 2014 compared to 2013 for Grades 5 and 6 ELA and Grades 4 and 7 math). Raw scores stayed the same on two tests (Grade 8 ELA and Grade 8 Math)."***

ANALYSIS 5

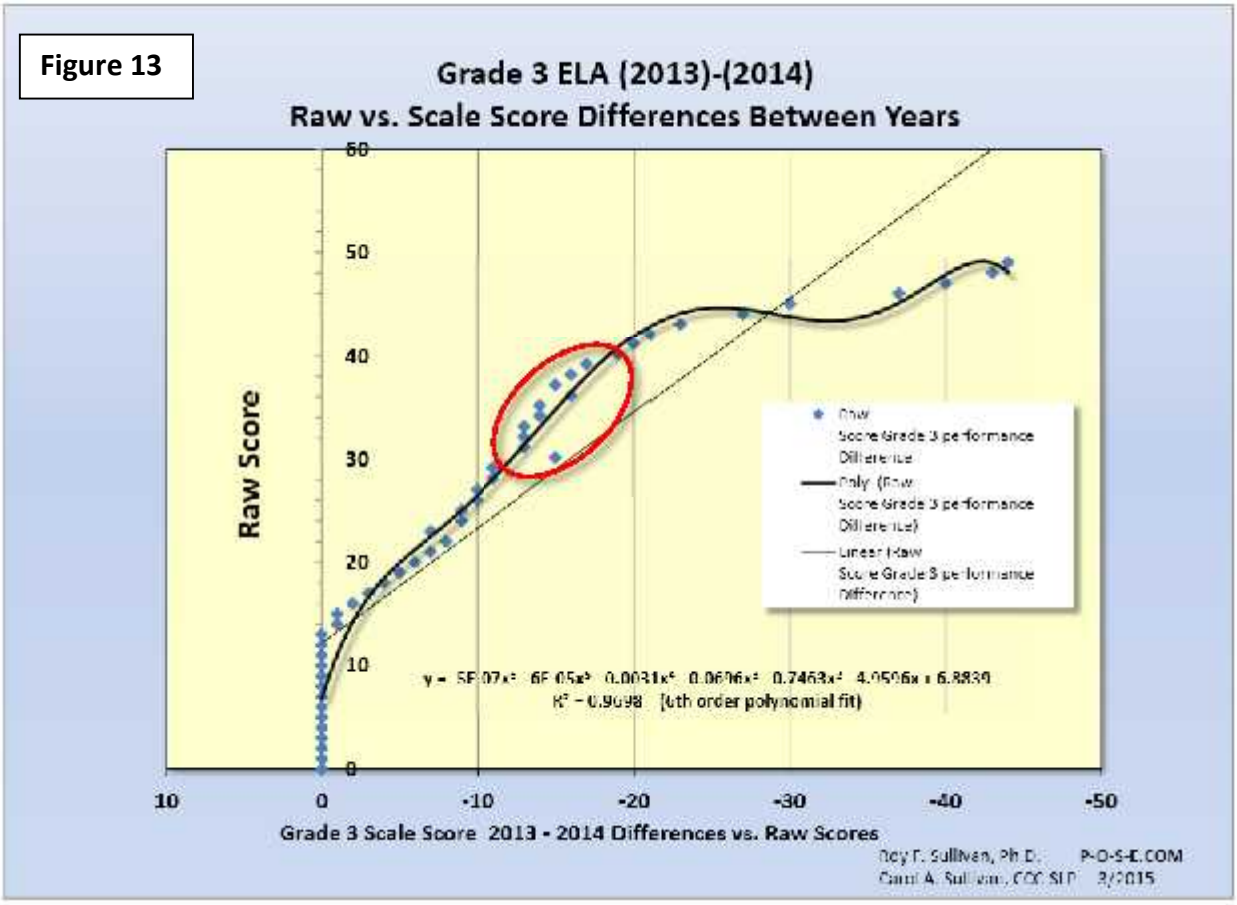
Using publicly released data, a statistical review of NYS ELA Grade 3 scoring practices for school years 2012-13 and 2013-14 addressed this issue. Figure 13 illustrates the difference between the scale score transformational values from raw ELA scores applied by NYS in 2012-13 and 2013-14. Part of the difference, at the upper end of the graph can be attributed to 55 raw score test items in 2013 and 49 raw score test items in 2014. The straight line represents a first order fit to the data (i.e multiplier plus constant). The arbitrariness of the NYS-applied 2014 raw-to-scale-score transformation is characterized by the best fit data as a 6th order polynomial.

A negative sign on the X axis indicates that NYS assigned HIGHER scale scores to the same raw score in 2014 than in 2013. The area encircled in red highlights the higher degree of arbitrary irregularity where NYS applied raw-to-scale-score "corrections" in the range immediately surrounding the NYS-assigned P2-P3 cut scale score boundary value of 319-320.

In 2014, an ELA raw score of 30 was transformed to a scale score of 320, the lower cut score boundary of P3. In 2013, an ELA raw score of 35 was required to achieve the same cut score of 320. A 2013 raw score of 30 would be transformed to a scale score of 305, a fifteen point difference.

A significant non-linearity is noted in the Grade 3 Raw/Scale score transformation between 2013 and 2014. For example, there is essentially no transform difference between years with raw scores below 15 out of 55. For raw scores > 40, major discrepancies exist between 2013 and 2014. Minimal changes in raw scores above 40 produced dramatic reductions in scale scores for 2014. For example a raw Grade 3 NYS ELA Score of 44 would result in a scale score of 376 in 2014 and 339 in 2013. A portion of this discrepancy may be ascribed to the fact that raw scores in 2013 ranged from 0-49 and from 0-55 in 2014. These variable changes between years hardly qualify as "small".

Figure 13


Grade 3 ELA (2013)-(2014)
Raw vs. Scale Score Differences Between Years
Raw Score
60
50
40
30
20
10
0
10
0
-10
-20
-30
-40
-50
Raw Score Grade 3 performance Difference
Poly. (Raw Score Grade 3 performance Difference)
Linear (Raw Score Grade 3 performance Difference)
y = 5E-07x⁵ - 6E-05x⁴ - 0.0031x³ - 0.0696x² - 0.7463x + 4.0506x + 6.8839
R² = 0.9098 (4th order polynomial fit)
Grade 3 Scale Score 2013 - 2014 Differences vs. Raw Scores
Roy F. Sullivan, Ph.D. P-O-S-E.COM
Carol A. Sullivan, CCC-SLP 3/2015

Figures 14 and 15 present the nonlinear transform functions for NYS ELA raw-score-to-scale-score conversions in 2013 and 2014, respectively. The net impact of this relatively high (6th) order polynomial data transformation is to arbitrarily enhance the relative impact of lower end performance and reduce the relative impact of higher end test performance on overall grade level performance. The Performance Level tiles (PL 1-4) on each chart list the percent of total range of raw scores (r) and of scale scores (s) in each nominally defined proficiency category.

Figure 14

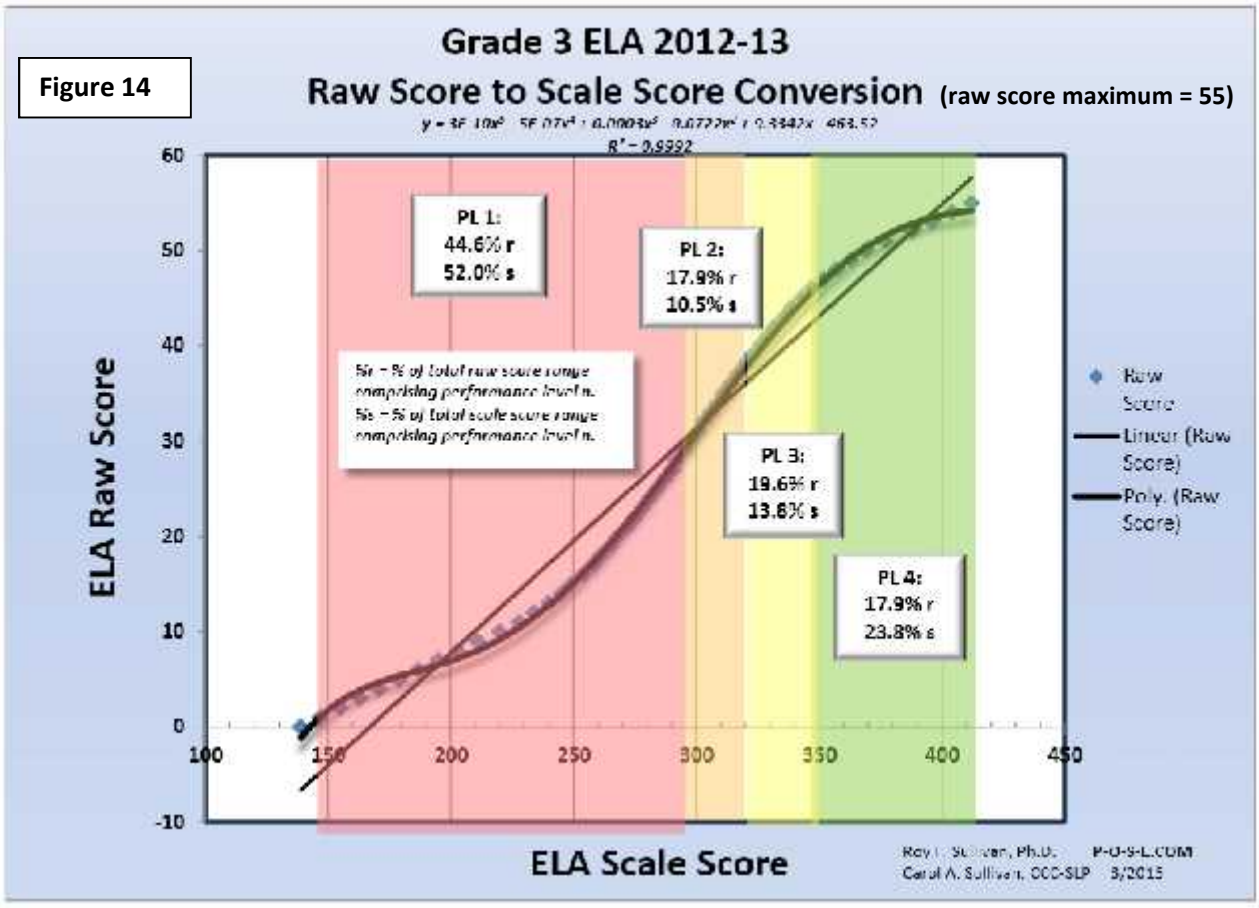


Figure 15

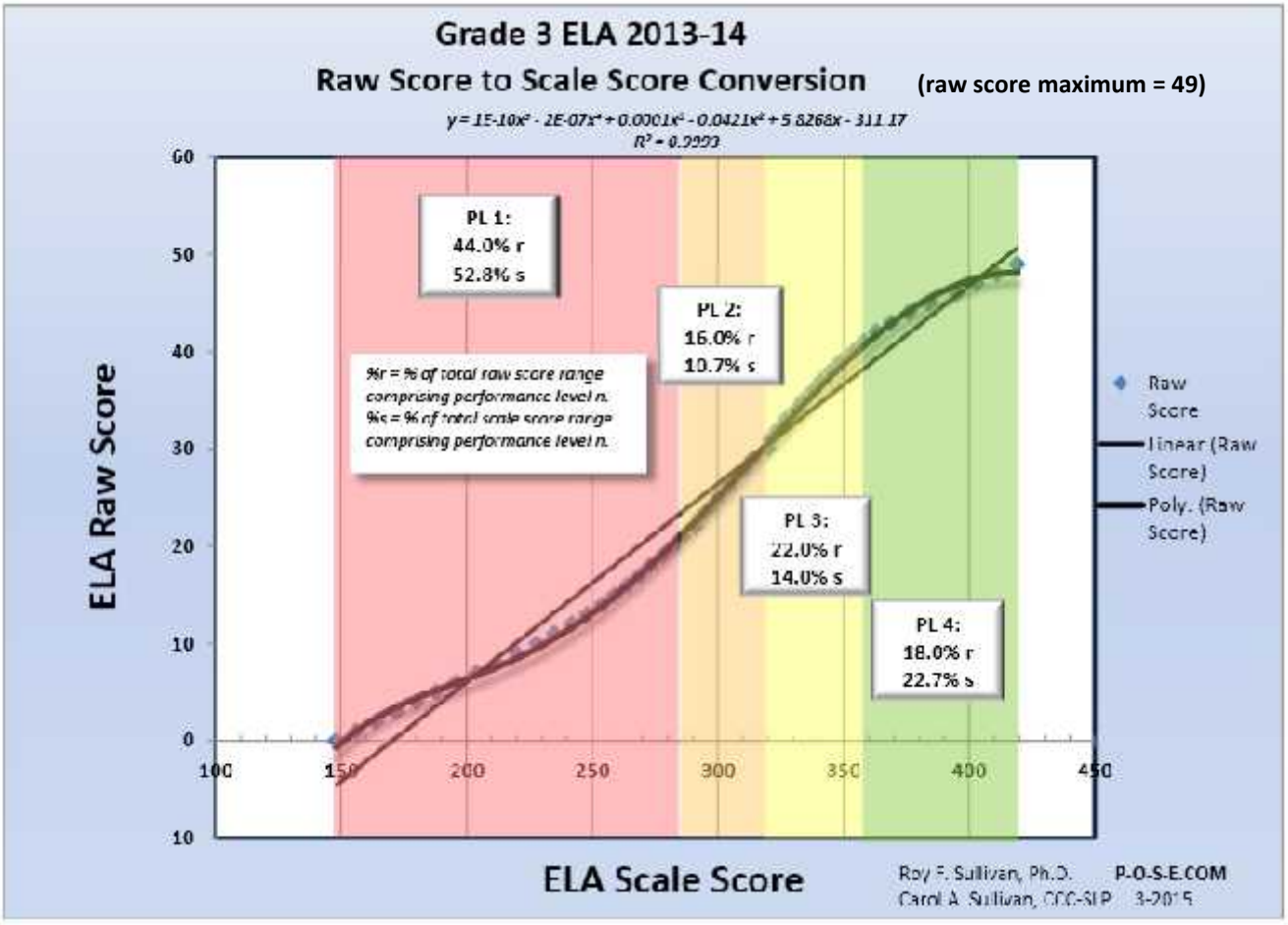

Grade 3 ELA 2013-14
Raw Score to Scale Score Conversion
(raw score maximum = 49)
y = 1E-10x⁴ - 2E-07x³ + 0.0001x² - 0.0421x² + 5.8268x - 311.17
R² = 0.9993
PL 1:
44.0% r
52.8% s
PL 2:
16.0% r
10.7% s
PL 3:
22.0% r
14.0% s
PL 4:
18.0% r
22.7% s
%r = % of total raw score range comprising performance level n.
%s = % of total scale score range comprising performance level n.
ELA Raw Score
ELA Scale Score
Raw Score
Linear (Raw Score)
Poly. (Raw Score)
Roy F. Sullivan, Ph.D.
Carol A. Sullivan, CCC-SLP
P-O-S-E.COM
3-2015

Table 27 presents the 2013 and 2014 NYS ELA data for the raw-to-scale score and scale-score-to-performance-level transformations. While the scale score cutoff values are identical from 2013 to 2014, the raw-score-to-scale-score transformational criteria differ between years.

Grade 3 cutoff values were "tweaked' in 2014 by NYS at the upper end of L4 to accommodate a reduction from a total of 55 raw score points in 2013 to 49 raw score points in 2014. Of greater significance (Figure 13) is the sizeable difference in raw-to-scale-score transformations applied *post hoc* to the raw score data by NYS between 2013 and 2014. The green columns (Table 27) enumerate these differences as differences in raw-to-scale-score conversions per item between 2013 and 2104.

Table 27


NYS ELA Raw Score to Scale Score Transformations
2013 and 2014
2013 ELA
Grade 3
2014 ELA
Grade 3
ELA (2013)-(2014)
Raw Score
Grade 3 performance Difference

Table 28 shows descriptive statistics for Mineola Grade 3 ELA scale scores for 2013 and 2014 in three contexts.

1. All available ELA scale scores for 2013 (n=191) and 2014 (n=180)

2. ELA scale scores where the P-O-S-E© error score was restricted to +> 10% for 2013 (n=96) and 2014 (n=96)

3. ELA scale scores where the P-O-S-E© error score was restricted to > 25% for 2013 (n=38) and 2014 (n=39).

Note that the mean scale scores are comparable from year to year in all three contexts. The median scale scores are comparable from year to year for contexts 1 and 2, above, differing in context 3 (with smaller n).

Of diagnostic import is the significant difference in modal scale scores (highlighted in yellow) for all Grade 3 ELA data: 338 for 2013 and 300 for 2014. This feature suggests the presence of a multi-modal data distribution.

Table 28

Grade 3 ELA Scale Scores Mineola U.F.S.D. 2013, 2014						
	2013 all	2014 all	2013 P.O.S.E=>10%	2014 P.O.S.E=>10%	2013 P.O.S.E>25%	2014 P.O.S.E>25%
Mean	306.8	304.4	294.5	292.0	275.6	279.1
Standard Error	2.28	2.33	3.19	2.99	5.31	4.40
Median	311	305.5	297	293	276	286
Mode	338	300	329	264	270	282
Standard Deviation	31.49	31.20	31.23	29.26	32.75	27.51
Sample Variance	991.66	973.27	975.16	855.92	1072.68	756.59
Kurtosis	0.14	0.17	0.17	0.70	0.22	1.19
Skewness	0.49	0.36	0.49	0.32	0.17	0.94
Range	170	180	137	167	126	138
Minimum	212	196	212	196	212	196
Maximum	382	376	349	363	338	334
Sum	58600	54784	28276	28032	10473	10886
Count	191	180	96	96	38	39

Figure 16 verifies the multimodal nature of the ELA scale score distribution for both 2013 and 2014. While the mean and median ELA scale scores are comparable from 2013 to 2014, use of the same 319-320 cutoff for the boundary of P2-P3 creates an artifactual penalty for the 2014 data based on a non-normal configuration of the scale score data distribution. The implicit equal interval nature of the raw scores is dually compromised by superimposition of an arbitrary raw-to-scale-score transformation and an arbitrary selection of critical cut scores defining nominal proficiency categories as well as the ultimate P2-P3 pass-fail cutoff.

Figure 16

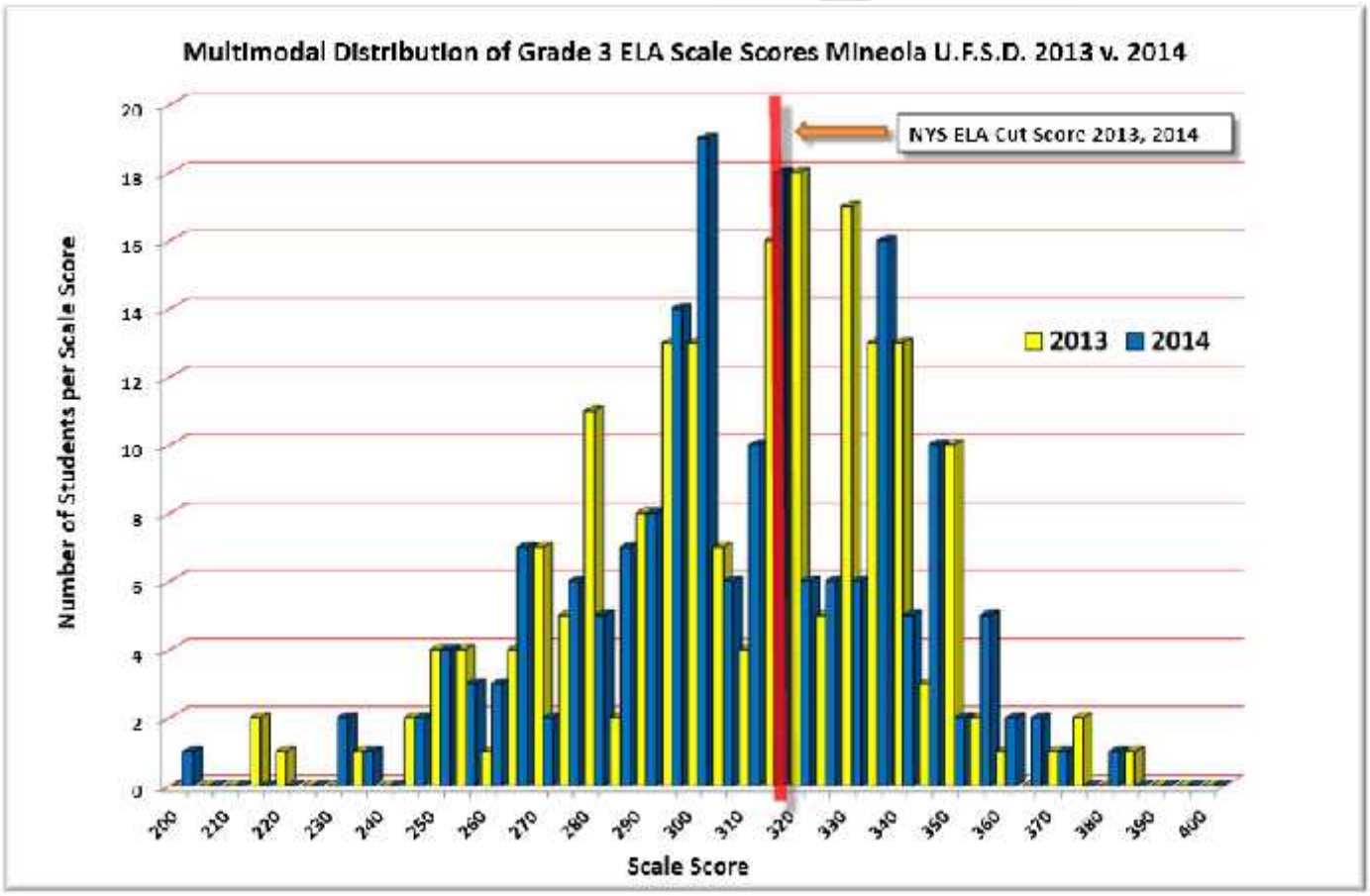

Multimodal Distribution of Grade 3 ELA Scale Scores Mineola U.F.S.D. 2013 v. 2014
NYS ELA Cut Score 2013, 2014
2013
2014
Number of Students per Scale Score
Scale Score
0
2
4
6
8
10
12
14
15
18
20
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400

Figure 17 presents a histogram of the 2014 Mineola Grade 3 raw ELA data for 180 students formatted as a linear transformation from raw score to percentage correct (raw score/49*100). The P2-P3 cut scale score of 219.5 is displayed as an equivalent percentage. The bimodal nature of the data distribution is evident.

Figure 17

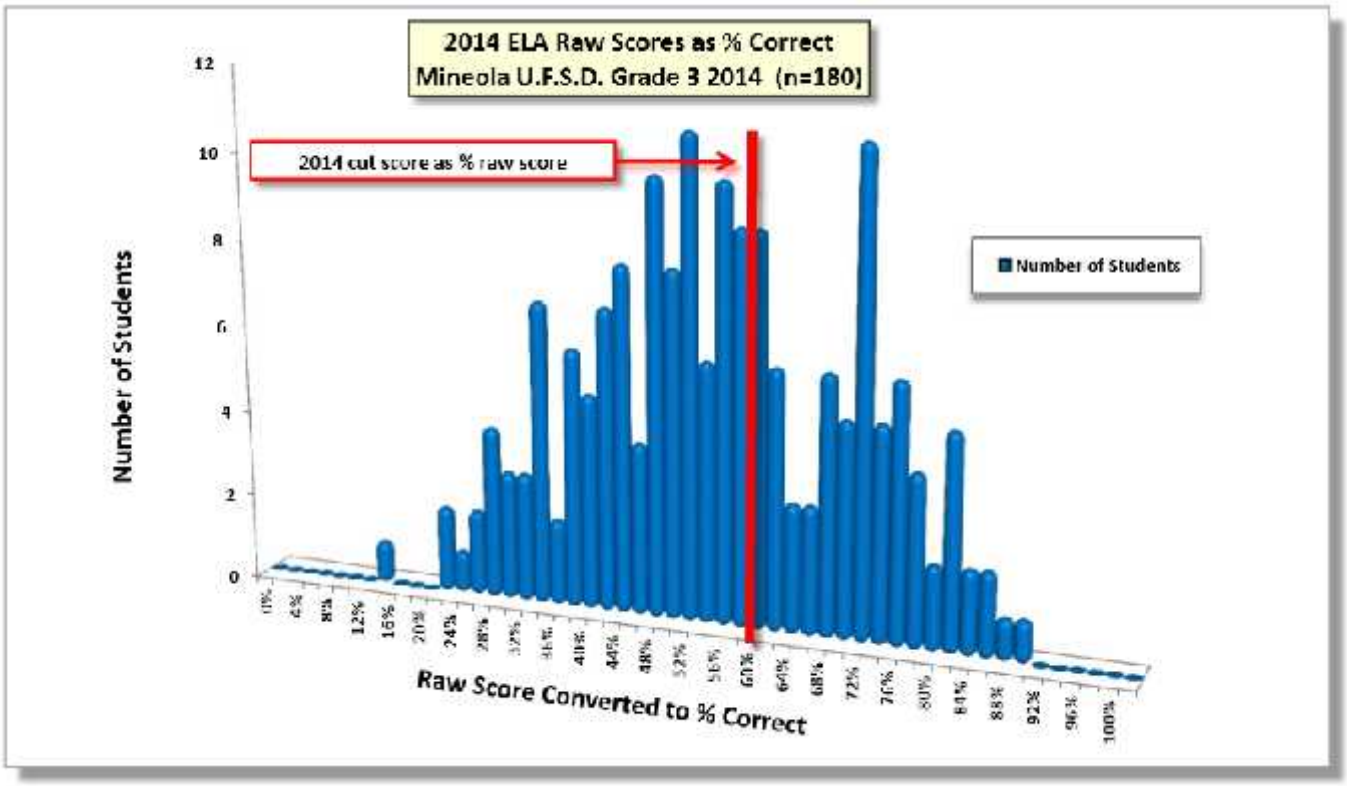


Table 29 presents the NYS ELA cut scores or scale score ranges associated with each performance level P1-4. The yellow highlighted row indicates the experimental shift by -3% (9 scale score points) applied to the 2014 ELA P2-P3 boundary between literacy non-proficiency and proficiency.

Table 30 demonstrates (yellow highlight) the effect of applying a -3% P2-P3 scale score cutoff shift to the NYS ELA 2014 Mineola U.S.F.D. Grade 3 scores. This minor adjustment to the cutoff scoring criteria created a profound, favorable impact on the Mineola U.F.S.D. 2014 Grade3 ELA P3+P4 proficiency score (33.0% > 44.4%). The simulation brings the 2014 ELA results into accord with the RTI findings of the other applied measures of literacy cited in this report: Fountas and Pinnell Benchmark Assessment System (F&P BAS), Northwest Evaluation Association Measures of Academic Progress, Reading (NWEA MAP-R.) and the Phonological-Orthographic Substitution Evaluation (P-O-S-E©.)

Table 29

NYS ELA Scale Score Ranges Associated with Each Performance Level					
Mineola UFSD Grade 3	Items	P1	P2	P3	P4
2013 NYS	55 raw	148-290	291-319	320-357	358-423
2014 NYS	49 raw	148-290	291-319	320-357	358-419
2014 P2/P3 cut -3%	49 raw	148-290	291-**310**	**311**-357	358-419

Table 30

Mineola U.F.S.D. Grade 3 ELA Proficiency Scores for 2103, 2014 and 2014 with -3% shift in P3/P4 Scale Score Cutoff						
Mineola U.F.S.D. Grade 3 NYS ELA Data vs. District Data	n students	% not meeting learning standards P1	% partially meeting learning standards P2	% meeting learning standards P3	% meeting learning standards with distinction P4	Total % meeting standards P3+P4
2014						
% reported P		30.0%	37.0%	30.0%	3.0%	33.0%
predicted n	180	54	67	54	5	59
obtained n	180	51	67	56	6	62
obtained % P		28.3%	37.2%	31.1%	3.3%	34.4%
2013						
% reported P		26.6%	30.8%	30.7%	2.8%	42.5%
predicted n	191	51	59	70	5	81
obtained n	191	52	58	70	5	81
obtained % P		27.2%	30.4%	39.8%	2.6%	42.4%
Difference 2014-2013	-11	3.4%	6.2%	-9.7%	0.2%	-9.5%
2014 -3% P2/P3a cut*						
predicted n	180	51	49	74	6	80
predicted % P		28.3%	27.2%	41.1%	3.3%	44.4%
*311/5 > 309.5						

Conclusion 5

The fragility of the critical mid-range ELA scale score cutoffs, applied by NYS with no evidence of empirical data support, combined with an undisclosed differences in data distribution configuration from year-to-year, effectively nullifies the value of the ELA as a reliable instrument for assessing literacy proficiency.

SUMMARY AND CONCLUSIONS

The Phonological-Orthographic Substitution Evaluation (P-O-S-E©) is a criterion-referenced test instrument for assessing short vowel proficiency in reading and spelling, initially targeted at third grade students. Short vowel proficiency has been recognized by Common Core State Standards (CCSS) as a foundational skill for literacy, to be established by Grade 2. The P-O-S-E© was standardized at the third grade level in the Plainview-Old Bethpage Central School District (POB) of New York (NY) between years 2005 and 2010 .

In 2012-13 and 2013-14, a comprehensive program of P-O-S-E© baseline, intervention and RTI evaluation was instituted in the Mineola Union Free School District (Mineola UFSD) of NY. Twenty percent of the student population was categorized as Latino or Hispanic, 12% Asian, etc. and 3% Black or African-American.

At the end of the 2012-13 academic year, Mineola Grade 3 made significant advances in P-O-S-E© short vowel proficiency and in literacy as assessed using the Fountas and Pinnell Benchmark Assessment System (F&P BAS) and the Northwest Evaluation Association Measures of Academic Progress, Reading (NWEA MAP-R.) Grade 3 scored the highest proportion of literacy proficiency among all Mineola UFSD grades 3-8 on the 2013 New York State English Language Arts examination (NYS ELA), newly configured to conform to Common Core State Standards (CCSS.)

At the end of 2013-14, comparable RTI gains were noted on the P-O-S-E©, F & P BAS and NWEA MAP. However, Grade 3 scored the *lowest* proportion of literacy proficiency among all Mineola Grades 3-8 on the 2014 NYS ELA. In addition, the Grade 3 cohort from 2012-3 scored next-to-lowest in literacy on the 2014 Grade 4 NYS ELA. According to NYS data, ELA passing proficiency scores for the entire state were comparable between 2013 and 2014: 31.1% vs. 31.0%, respectively. Long Island ELA scores showed a greater 2013-14 reduction: 39.6% to 36.8%.

The gross inconsistency between Grade 3 NYS ELA outcomes for both 2013 and 2014 and alternative measures of literacy for the same years prompted an inquiry into possible reasons for this conflict. Mineola Grade 3 test data and NYS-released ELA reading passages and scoring data were analyzed in detail for both years.

It is to be noted that when the multiple correlational analysis was restricted to Grade 3 students with P-O-S-E© error scores > 25%, ALL external correlations between the NYS ELA scores and the alternative literacy assessment instruments were significantly lower in 2014 than in 2013.

Findings reveal significant issues with face validity of the NYS ELA examination as currently implemented. NYS ELA test passages for Grades 3 and 4 in 2013 and 2014 present an exaggerated range of grade-inappropriate reading levels effectively rendering invalid any test questions based on these passages. Reading levels for NYS-released 2014 Grade 3 ELA passages were well above grade level, well above the level for 2013 Grade 3 passages and even higher than Grade 4 passages for 2013.

Data also suggest that reliability of the NYS ELA test outcomes may be compromised by the process of *"equating"* applied by NY State to the 2014 ELA scores This is a post-hoc application of raw-score-to-scale-score transformations and scale-score-to-performance level transformations to achieve a preferred outcome in year 2014 relative to 2013. According to NYS:

> *"The cut scores [defined boundaries of literacy proficiency categories L1-L4] did not change from 2013 to 2014. "*

In fact, the raw-to-scale score transformations were altered between 2013 - 2014 resulting in differing raw score values for each cut (scale) score. Continuing:

> *"The purpose of the 2014 equating was to maintain the level of difficulty established by the standard setting process in 2013, when 95 teachers from across the state recommended the level of difficulty necessary to achieve proficiency (Level 3) and partial proficiency (Level 2). Based on student performance on common anchor test questions (the same items used in both 2013 and 2014), the raw scores needed for each performance level were adjusted slightly to ensure that scale scores and performance levels are comparable from year to year. If the test is slightly easier, the number of raw score points needed to earn a performance level may increase slightly in order to maintain the performance standard. If the test is slightly harder, the number of raw score points needed to earn a performance level may decrease slightly in order to maintain the performance standard."* …
>
> *"…On the 2014 tests, year-to-year raw score changes for Level 3 were small and varied by grade. Raw scores went down slightly on 6 tests (indicating slightly harder tests in 2014 compared to 2013 for Grades 3, 4, and 7 ELA and Grades 3, 5, and 6 Math) and went slightly up on 4 tests (indicating slightly easier tests in 2014 compared to 2013 for Grades 5 and 6 ELA and Grades 4 and 7 math)."*

Finally, in 2014, three Grade 3 ELA test items were summarily discarded by NYS, *post hoc*. This accounted for the 6 point differential between the 55 point 2013 ELA and the 49 point 2014 ELA – an arbitrary net reduction of 11% in the 2014 scoring base.

Since 2012-13, Common Core State Standards have been foundational to the NYS ELA and to the literacy examinations of other states. CCSS seeks to impose an overarching set of theoretically-derived criteria for literacy proficiency. The ability of individual states to "tweak" the aggregate test score outcomes effectively invalidates the concept of *"Common Core".*

A minor shift of -3% was experimentally applied to the 2013-14 P2-P3 scale score cutoff boundary. This action dramatically elevated the 2014 Mineola Grade 3 P3+P4 literacy proficiency level from the reported 33.0% (~10% below 2013) to 44.4% (~2% above 2013). (q.v. Tables 29, 30) The differing, multi-modal nature of the scale score data distribution in 2013 and 2014 contributes significantly to the misinterpretation of ELA outcomes.

Despite NYS enlisting the best efforts of "95 teachers", the major functional and educational impact of this minor shift in a single ELA cutoff value, arbitrarily manipulated in the raw-to-scale-score transformation in 2014 by NY State, highlights the fragile inadequacy of the entire ELA evaluation process in its current form.

Literacy and the entire academic well-being of students and a reinforced level of motivation among their effective teachers cannot be subjected to the statistical vagaries of test designers with constrained perspectives. "Regents examination" scoring protocols have ceased to be relevant.

Given the outcome of the present detailed analysis of Grade 3 NYS ELA reading materials and scores contrasted with alternative measures of literacy proficiency for the Mineola UFSD, serious questions may be raised about the relevance of the NYS ELA as currently constructed. It would appear that the NYS ELA is not a suitable test instrument for assessing language arts proficiency or for directing data-driven curriculum development in Grade 3.

Carol A Sullivan, CCC-SLP; Roy F Sullivan, Ph.D. http://www.P-O-S-E.net April 11, 2015

APPENDIX

Readability Tests Used in the NYS ELA Grade 3 & 4 Passage Analysis (1 of 2)			
Readability Test		Type	Description
1	***Bormuth Cloze Mean***	Cloze score	Coleman-Liau is meant for secondary age (4th grade to college level) readers. This formula is based on text from the .4 to 16.3 grade level range. This test usually yields the lowest grade when applied to technical documents.
2	***Coleman-Liau***	Grade level and Cloze score	Coleman-Liau is meant for secondary age (4th grade to college level) readers. This formula is based on text from the .4 to 16.3 grade level range. This test usually yields the lowest grade when applied to technical documents.
3	***Danielson-Bryan 1***	Grade level	Danielson-Bryan 1 is designed for student materials.
4	***Degrees of Reading Power (grade equivalent)***	Grade level	Degrees of Reading Power (GE) is designed for matching documents to a student's reading ability (based on his/her DRP score). This test is a conversion of a DRP (difficulty) score into a grade level.
5	***Easy Listening Formula***	Grade level	ELF is designed for "listenability" and is meant for radio and television broadcasts.
6	***Flesch-Kincaid***	Grade level	Flesch-Kincaid is designed for technical documents and is mostly applicable to manuals and forms, rather than schoolbook text or literary works. This test is part of the Kincaid Navy Personnel collection of tests.
7	***Flesch-Kincaid (simplified)***	Grade level	Flesch-Kincaid is designed for technical documents and is mostly applicable to manuals and forms, rather than schoolbook text or literary works. This test is part of the Kincaid Navy Personnel collection of tests.
8	***Fry***	Grade level	The Fry graph is designed for most text, including literature and technical documents.
9	***Gunning Fog***	Grade level	Gunning Fog Index is generally recommended for business publications and journals.
10	***Harris-Jacobson Wide Range Formula***	Grade level	Harris-Jacobson is generally used for primary and secondary age (Kindergarten to 11th grade) readers.
11	***Modified SMOG***	Grade level	Modified SMOG is a variation of SMOG that is adjusted for primary-age materials.
12	***New Automated Readability Index (Kincaid)***	Grade level	New Automated Readability Index is a modified version of ARI created for U.S. Navy materials and was designed for technical documents and manuals. This test is part of the Kincaid Navy Personnel collection of tests.
13	***New Automated Readability Index (Kincaid, simplified)***	Grade level	New Automated Readability Index is a modified version of ARI created for U.S. Navy materials and was designed for technical documents and manuals. This test is part of the Kincaid Navy Personnel collection of tests.

APPENDIX (continued)

Readability Tests Used in the NYS ELA Grade 3 & 4 Passage Analysis (2 of 2)			
Readability Test		Type	Description
14	***New Dale-Chiall***	Grade level	New Dale-Chall is generally used for primary and secondary age readers to help classify school text books and literature.
15	***New Farr, Jenkins, Paterson (Kincaid)***	Index score	A modified version of Farr, Jenkins, Paterson designed for U.S. Navy technical manuals and forms.
16	***New Fog Count (Kincaid)***	Grade level	New Fog Count is a modified version of the Gunning Fog Index created for the U.S. Navy and was designed for technical documents and manuals. This test is part of the Kincaid Navy Personnel collection of tests.
17	***Powers, Sumner, Kearl (Dale-Chall)***	Grade level	PSK Dale-Chall is generally used for primary and secondary age readers to help classify school text books and literature.
18	***Powers, Sumner, Kearl (Farr, Jenkins, Paterson)***	Grade level	PSK Farr, Jenkins, Paterson is a variation of the Farr, Jenkins, Paterson test, which returns a grade score instead of a Flesch difficulty level.
19	***Powers, Sumner, Kearl (Flesch)***	Grade level	PSK Flesch is used for student readers.
20	***Powers, Sumner, Kearl (Gunning Fog)***	Grade level	PSK Gunning Fog Index is generally recommended for business publications and journals.
21	***Raygor Estimate***	Grade level	The Raygor estimate graph is designed for most text, including literature and technical documents.
22	***SMOG***	Grade level	SMOG (colloquially referred to as Simple Measure of Gobbledygook) is generally appropriate for secondary age (4th grade to college level) readers. SMOG tests for 100% comprehension, whereas most formulas test for around 50%-75% comprehension.
23	***SMOG (simplified)***	Grade level	SMOG (colloquially referred to as Simple Measure of Gobbledygook) is generally appropriate for secondary age (4th grade to college level) readers. SMOG tests for 100% comprehension, whereas most formulas test for around 50%-75% comprehension.
24	***Spache Revised***	Grade level	Spache is generally used for primary age (Kindergarten to 7th grade) readers to help classify school textbooks and literature.
25	***Wheeler-Smith***	Grade level	Wheeler-Smith is meant for primary-age (Kindergarten to 4th grade) reading materials.

DRAFT