Update: Transforming MSA Scores for Evaluations

Dave Volrath Ben Feldman

March 21, 2013

TRSG Matrix showing typical objectionable pattern

	B1	B2	B3	P1	P2	P3	A1	A2	A3
B1	9	11	**13**	15	17	19	21	23	25
B2	8	10	12	14	16	18	20	22	24
B3	7	9	11	13	15	17	19	21	23
P1	6	8	10	12	14	16	18	20	22
P2	5	7	9	11	13	15	17	19	21
P3	4	6	8	10	12	14	16	18	20
A1	3	5	7	9	11	13	15	17	19
A2	2	4	6	8	10	12	14	16	18
A3	1	3	5	7	9	11	**13**	15	17

Although this weighting scheme was repeatedly endorsed by the National Psychometric Council and by the July 2012 Standard Setting Panel, diverse audiences have argued TRSG will provide false positives for those beginning with high performing classes and will create false negatives for those beginning with lower performing classes. *LEA stress testing confirms these concerns.*

Four illustrative cases

N Attributable Students	Performance	TRSG Median	TRSG Decile Points	123 Mean	Principal Rating
36 Reading Grade 5	16 of 27 improve, 8 hold	14	4/20	2.48	Commendable
44 Reading Grade 4	19 of 40 improve, 9 hold	15	20/20	2.15	Satisfactory
15 Math Grade 4	13 of 14 improve, 1 holds	15	12/20	2.92	Commendable
30 Math Grade 5	5 of 29 improve, 9 hold	12	12/20	1.67	Satisfactory

Deciles are overly constrained, lack discretion, are punitive at the bottom of the scale and overly lenient at the top.

1-2-3 sorts the teachers correctly.

The principals generally have a more accurate practice rating than does TRSG.

The “Pure 1-2-3 Model”

	B1	B2	B3	P1	P2	P3	A1	A2	A3
B1	2	3	3	3	3	3	3	3	3
B2	1	2	3	3	3	3	3	3	3
B3	1	1	2	3	3	3	3	3	3
P1	1	1	1	2	3	3	3	3	3
P2	1	1	1	1	2	3	3	3	3
P3	1	1	1	1	1	2	3	3	3
A1	1	1	1	1	1	1	2	3	3
A2	1	1	1	1	1	1	1	2	3
A3	1	1	1	1	1	1	1	1	2

This is a variation of the Arkansas model and was suggested by Cecil County at the January 2013 meeting, using a -1 / 0 / +1 coding scheme.

An all-positive number scheme is preferred.

A value of 2 indicates the class held its own; any value north of 2 indicates a proportional lift.

Modified 1-2-3 Model

	B1	B2	B3	P1	P2	P3	A1	A2	A3
B1	1	3	3	3	3	3	3	3	3
B2	1	2	3	3	3	3	3	3	3
B3	1	1	2	3	3	3	3	3	3
P1	1	1	1	2	3	3	3	3	3
P2	1	1	1	1	2	3	3	3	3
P3	1	1	1	1	1	2	3	3	3
A1	1	1	1	1	1	1	2	3	3
A2	1	1	1	1	1	1	1	2	3
A3	1	1	1	1	1	1	1	1	3

This variation protects cell A3A3 which risks a ceiling effect for teachers working with highest performing students. This is a substantial population statewide.

Preserving symmetry, cell B1B1 is not construed as desirable. This is in most grades and subjects a sparsely populated cell.

Tiered Achievement Index

	B1	B2	B3	P1	P2	P3	A1	A2	A3
B1	1	3	3	3	4	4	4	4	4
B2	1	2	3	3	3	4	4	4	4
B3	1	1	2	3	3	3	4	4	4
P1	1	1	1	2	3	3	3	4	4
P2	1	1	1	1	2	3	3	3	4
P3	1	1	1	1	1	2	3	3	3
A1	0	1	1	1	1	1	2	3	3
A2	0	0	1	1	1	1	1	2	3
A3	0	0	0	1	1	1	1	1	3

This refinement of Modified 1-2-3 carves out the opposing corners, providing bonus points for teachers who accomplish dramatic growth with students who enter low, and providing a disincentive if top students fall below proficient.

Blue cells are challenging but capture a good representation statewide.

Tiered Achievement Index, continued

	B1	B2	B3	P1	P2	P3	A1	A2	A3
B1	1	3	3	3	4	4	4	4	4
B2	1	2	3	3	3	4	4	4	4
B3	1	1	2	3	3	3	4	4	4
P1	1	1	1	2	3	3	3	4	4
P2	1	1	1	1	2	3	3	3	4
P3	1	1	1	1	1	2	3	3	3
A1	0	1	1	1	1	1	2	3	3
A2	0	0	1	1	1	1	2	2	3
A3	0	0	0	1	1	1	1	2	3

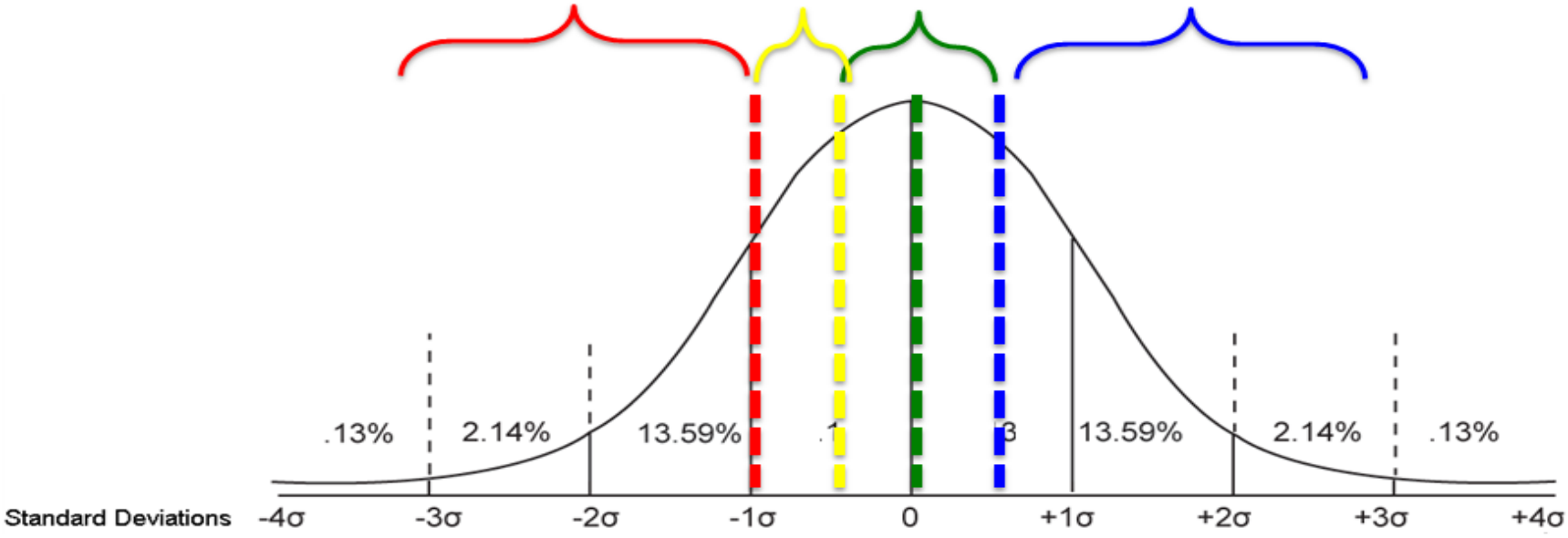
Lastly, cells A3A2 and A2A1 are treated as part of the diagonal.

This recognizes the challenge of sustaining advanced students and better reflects the statewide distribution of the data.

Aligning the MSA and adding meaning

- All models grapple with lack of vertical scale, alternating skew from year to year : “All 4th grade teachers are heroes, and all 5th grade teachers are villains.”
- Sorting 2011 TRSG medians into deciles was designed to fit a vertical scale. It did not perform.
- Relating the educator to the statewide grade-specific mean can accomplish a similar goal.
- Using the grade-specific standard deviation can make scores meaningful.

A CCPS approach to using the Standard Deviation to interpret performance



Performance spanning the grade mean by one standard deviation is considered expected and acceptable (green bracket).

Growth *more than* .5 STD above mean is beyond expected and commendable (blue bracket).

Performance .5 STD below the central range is concerning (yellow bracket); performance a full STD below mean is a significant loss and unacceptable (red bracket).

Slide borrowed from CCPS presentation, March 11, 2013

Statistical considerations

- The means and standard deviations that follow are derived from the statewide distribution of *students*.
- The performance tiers are to fit *teacher* averages.
- A purist model would divide the STD by the square root of the class N…
- But that would overly constrain the data and force many teachers out of effective.

The State Means, STDs, and Tiers

year	subject	grade	Mean	STD	Upper Limit< -1 STD	Lower Limit - 1 STD	Upper limit <-.5 STD	Lower Limit -.5 STD	Upper Limit +.5 STD	Lower Limit >+ 5. STD
2012	M	04	2.53	0.69	1.839	1.84	2.184	2.185	2.875	2.876
2012	M	05	1.66	0.74	0.919	0.92	1.289	1.29	2.03	2.031
2012	M	06	2.37	0.75	1.619	1.62	1.994	1.995	2.745	2.746
2012	M	07	1.94	0.79	1.149	1.15	1.544	1.545	2.335	2.336
2012	M	08	2.19	0.82	1.369	1.37	1.779	1.78	2.6	2.601
2012	R	04	2.36	0.73	1.629	1.63	1.994	1.995	2.725	2.726
2012	R	05	2.54	0.69	1.849	1.85	2.194	2.195	2.885	2.886
2012	R	06	1.78	0.75	1.029	1.03	1.404	1.405	2.155	2.156
2012	R	07	2.16	0.79	1.369	1.37	1.764	1.765	2.555	2.556
2012	R	08	2.04	0.77	1.269	1.27	1.654	1.655	2.425	2.426

A real example

Principal ratings were prepared a priori.

Sort based on TAI.

Data suggest anything north of 2.1 is solid for this example.

Data invite a local conversation, as they should.

Grade 4 Math actual example					
Teacher ID	Grade	N Students	TAI	Principal Rating	STD Cuts
128	4	1	3	Commendable	2.876
103	4	28	2.93		
163	4	13	2.92	Commendable	
149	4	26	2.92	Satisfactory	
66	4	16	2.88	Satisfactory	2.876
161	4	43	2.84	Commendable	2.875
158	4	71	2.82	Commendable	
17	4	34	2.79	Satisfactory	
172	4	56	2.79	Satisfactory	
23	4	27	2.78	Commendable	
83	4	40	2.78		
126	4	37	2.7	Commendable	
174	4	39	2.69		
62	4	61	2.64	Commendable	
60	4	33	2.64	Satisfactory	
58	4	36	2.61		
194	4	43	2.6	Commendable	
86	4	32	2.59	Satisfactory	
187	4	35	2.54	Satisfactory	
81	4	38	2.45		
54	4	37	2.41	Commendable	
70	4	21	2.38	Satisfactory	
34	4	37	2.27		
154	4	56	2.21	Satisfactory	2.185
166	4	36	2.11	Commendable	2.184
165	4	21	2.1	Unsatisfactory	
171	4	34	1.97	Unsatisfactory	
75	4	33	1.91	Satisfactory	1.84
94	4	34	1.32	Satisfactory	1.839

Setting performance scores

- Analysis of a test LEA sorted by grade, TAI, and principal rating indicates interleaving of scores and ratings: e.g., there are some teachers with high scores and lower ratings and vice versa.
- But, TAI mean scores within rating groups tier accurately. Commendable teachers have a higher TAI average than do satisfactory teachers, etc.
- Based on data studied so far, there are no decisive cuts, but performance north of 2 looks solid.
- MSDE asks LEAs to share their parallel analyses if they elect to test TAI.

Looking forward

- Replacement student detail records have been posted to Tumbleweed, to the attention of the LAC.
- Statewide 2012 means and standard deviations are published.
- 2012 means and standard deviations will be held constant through the life of MSA, mirroring the decile approach.
- Locals are invited to study the performance of their local means and standard deviations which MSDE can provide.

Contacts

Dave Volrath

dvolrath@msde.state.md.us

410 767 0504

Ben Feldman

bfeldman@msde.state.md.us

410 767 0142