

Putting in all Together: Annual Differentiation under ESSA

Scott Marion & Jenn Dunn
Center for Assessment

New York Regents Retreat

What does the law say?

“(C) ANNUAL MEANINGFUL DIFFERENTIATION.—Establish a system of meaningfully differentiating, on an annual basis, all public schools in the State, which shall—

- “(i) be based on all indicators in the State’s accountability system under subparagraph (B), for all students and for each of subgroup of students, consistent with the requirements of such subparagraph;
- “(ii) with respect to the indicators described in clauses (i) through (iv) of subparagraph (B) afford—
 - “(I) substantial weight to each such indicator; and
 - “(II) in the aggregate, much greater weight than is afforded to the indicator or indicators utilized by the State and described in subparagraph (B)(v), in the aggregate; and
- “(iii) include differentiation of any such school in which any subgroup of students is consistently underperforming, as determined by the State, based on all indicators under subparagraph (B) and the system established under this subparagraph.

What does that mean?

- Subsequent sections of the law indicate that states must be able to produce three categories of schools:
 - Identified for comprehensive support and improvement
 - Identified for targeted support and improvement
 - Non-identified schools

- State can add other categories of schools, but it is not required!

- So how should we do this?

First order question

- Do we want to produce an overall rating for each school (e.g., Level 1-4, 1-5 stars, A-F)?

OR

- Do we want to avoid a single overall determination?
 - We could also produce an overall score (e.g., 560) without producing a rating?
- No matter which you choose, there are still a host of decisions that follow from this first decision...

Single Overall Rating

Take a few minutes to jot down some of the potential advantages and disadvantages of a single rating

Advantages

Disadvantages

No Overall Rating

Take a few minutes to jot down some of the potential advantages and disadvantages of not having a single overall rating

Advantages

Disadvantages

Single Overall Rating

Advantages

- Can be used to clearly identify lowest (or lower) performing schools
- Many stakeholders want just a quick overall indication of school quality
- Make it more likely that the media will report the state designations rather than creating their own overall ranking

Disadvantages

- May oversimplify a complex system
- The overall rating on its own does not provide actionable information for school personnel
- May mask specific areas of concern or recognition

No Overall Rating

Advantages

- Presents a picture of the strengths and weakness of the school
- Allows stakeholders to hone in on the areas in need of improvement
- May allow for more accurate identification of schools based on specific needs
- Avoids overly simplistic ranking of schools

Disadvantages

- Not as intuitively understandable for stakeholders wanting a sense of overall quality
- Leaves the inferences about the quality of the school open to multiple interpretations (may be an advantage too)
- May mask specific areas of concern or recognition
- Still need a way to ID schools for CSI and TSI

Additional issues

- The ways in which we decide to produce the overall rating, including weighting of the indicators, could lead to noticeably different results
- If we decide not to produce a single overall determination, we still cannot avoid aggregation decisions...

How to decide...

- How do the type of reports we produce fit with your theory of action?
 - What approach will have the higher utility value?
- What do the stakeholder want?
 - We should play out some used cases for the variety of stakeholders?

Use Cases

- Think about the various stakeholders and how they might use the accountability results. For each stakeholder group, indicate whether they would want an overall rating or summative score, indicator ratings only, or both and WHY?
 - Parents/students
 - Teachers
 - School leaders
 - District administrators
 - State education leaders
 - Business and community members
 - Politicians

Methods for producing overall determinations

If the desire is to produce overall determinations, there are three general classes of methods for doing so

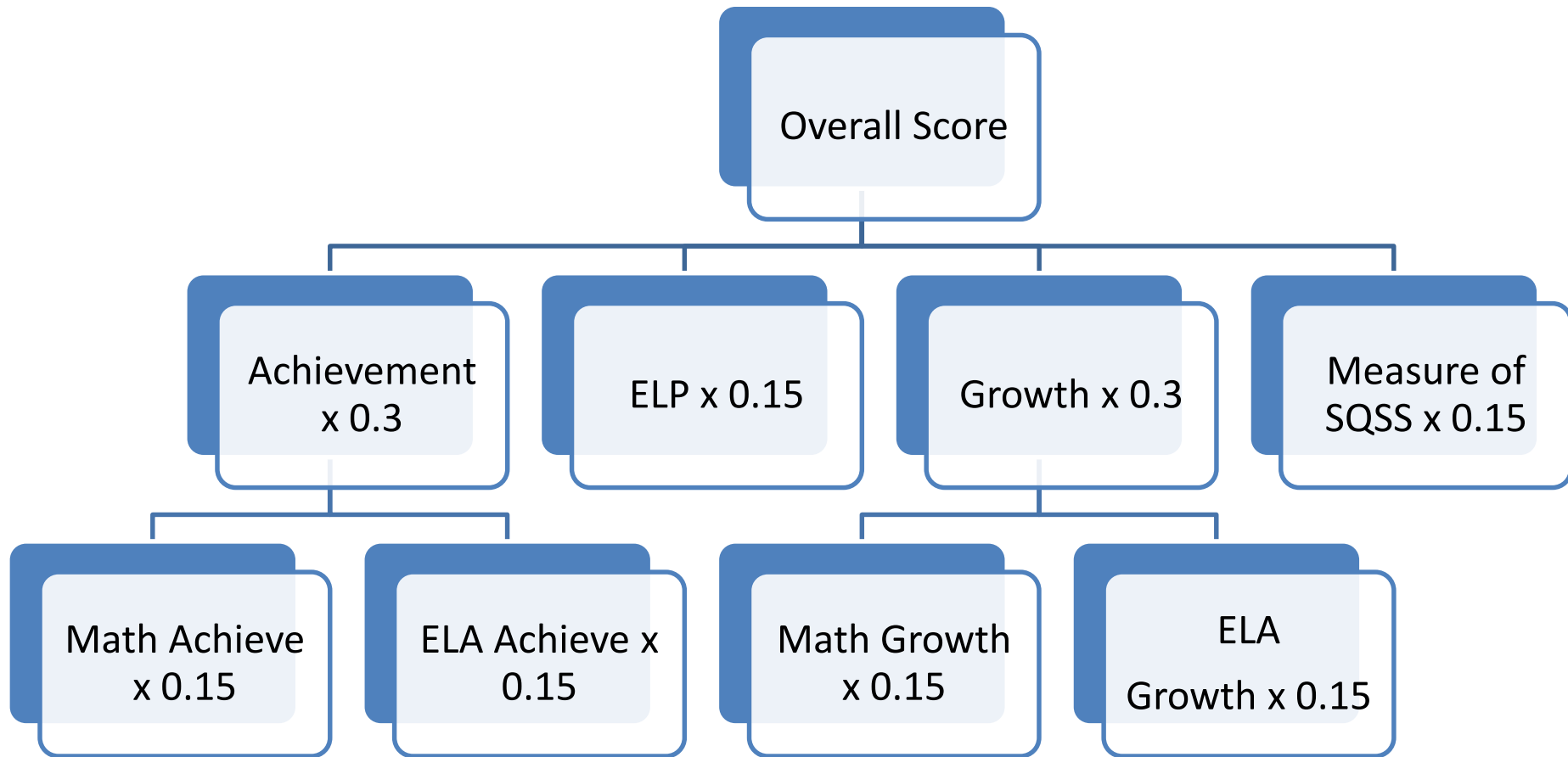
- Weighted Index or Composite
 - Decision Tables or Matrices
 - Profiles or Decision Rules
-
- As you have likely guessed, each approach has strengths and shortcomings..

Weighted Composite

- Most **commonly** used method among states right now
- Relatively easy to implement
- Results in a total score is often translated into an overall rating (but does not necessarily have to be)
- Assumes that the weights assigned (“nominal”) are the same as when the composite is calculated (“effective”)
 - This is usually wrong!
- Should employ a deliberative process (e.g., standard setting) to convert scores to ratings

- The following slides provides a typical example...

Weighted Composite



Decision table/matrix

- Easy to implement with two or three indicators; requires additional decision matrices with more than three indicators
- Explicitly reveals the values associated with any combination of indicators
- Stakeholders never experience an unexpected result
 - If you know the indicator values, you know the overall rating
- May be seen as restricting nuance or variance, but avoids over-interpretation of small differences
- Note: What follows are just some examples...

A Simple Two Indicator Decision Table

		Growth Level			
		Level 1	Level 2	Level 3	Level 4
Achievement Level	Level 1	Needs Support	Needs Support	Developing	Developing
	Level 2	Needs Support	Developing	Strong	Strong
	Level 3	Developing	Strong	Strong	Exemplary
	Level 4	Developing	Strong	Exemplary	Exemplary

NOTE: The values in the cells are just examples. Actual values would be determined through a deliberative process.

An Example of a Three-Indicator Table

Equity	Growth	Achievement		
		Below Target	Meets Target	Exceeds Target
Below Target	Below Target	Below Expectations		
	Meets Target			
	Exceeds Target			
Meets Target	Below Target			
	Meets Target		Meets Expectations	
	Exceeds Target			
Exceeds Target	Below Target			
	Meets Target			
	Exceeds Target			Exceeds Expectations

Combining multiple indicators using multiple matrices

	Equity			
ELP	Level 1	Level 2	Level 3	Level 4
Level 1	I	I	II	III
Level 2	I	II	III	IV
Level 3	II	II	III	IV
Level 4	II	II	III	IV

	Growth			
Equity/ ELP	Level 1	Level 2	Level 3	Level 4
Level 1	I	I	II	III
Level 2	I	II	III	IV
Level 3	II	II	III	IV
Level 4	II	II	III	IV

	Achievement			
Growth/ Equity	I	II	III	IV
I	Low	Low	Develop	Good
II	Low	Develop	Develop	Good
III	Develop	Develop	Good	Exemplary
IV	Develop	Good	Exemplary	Exemplary

Profiles or decision rules

- Similar to the decision tables, but would use all indicator values at once
- A set of decision rules used to evaluate school profiles (scores on the various indicators) against narrative descriptions of performance
- By working through this process, rules are established to place schools into various overall levels based on the constellation of indicator values

Profile/Decision Rules Example

School	Achievement	Growth	Graduation	ELP	Overall
A	4	4	4	4	Exemplary
B	3	3	3	3	Good
C	2	2	2	2	Developing
D	1	1	1	1	Low
E	4	3	4	3	??
F	3	2	2	3	??

As you can see, the homogeneous profiles are easy to evaluate. The heterogeneous profiles require the work of a deliberative body (e.g. Regents) to make and evaluate decisions.

What do you value?

- Which approaches do you think will have the most credibility with district and school leaders, policymakers, and the general public?
- Sometimes it is difficult to have both transparency and high technical quality. Which feature should be prioritized?
- Should this be an empirical decision largely by (once we settle on indicators) seeing how schools fare under the different approaches to shed light on how the different approaches work with NY data?

Regents Discussion and Recommendations

1. Do the Regents recommend for SED to pursue a system of reports that produces an overall rating for each school in the state (e.g., Level 1-4, ★ ★ , A-F) **beyond the ESSA-required identification of CSI, TSI**, and not identified?
2. Do the Regents recommend for SED to pursue a system of reports that produces an overall numeric score for each school in the state?
3. Do the Regents recommend for SED to pursue a system of reports that produces **BOTH** an overall numeric score and rating for each school in the state?

Regents Discussion and Recommendations

4. If the answer to number 2 OR 3 on the previous slide is **yes** (producing a numeric overall score), then some sort of weighted index is required such as the following:

Indicator	Performance	Weighting	Score
A	50	20	100
B	60	20	120
C	40	20	80
D	30	10	30
E	70	10	70
F	80	10	80
G	90	10	90
Total			570

Regents Discussion and Recommendations

5. If you recommend to produce an overall rating beyond the federal requirements, we have at least three choices:
 - a. A weighted index that will be used to create classifications—remember, this will result in a score that may be kept in the background
 - b. A multi-step decision matrix approach to create classifications—this can work fairly well with up to 4-5 indicators, but may get unwieldy with more than 5 indicators
 - c. A profile (decision rule) approach to create classifications—this might be the best approach if no total scores are desired and there is concern about a complex decision table system (although this is essentially a decision table in flat form)

Questions/Comments

- Other questions and comments?