Incorporating English Learner Progress into State Accountability Systems

Pete Goldschmidt
California State University Northridge
Kenji Hakuta
Stanford University

September 13, 2016

Purpose of Session

- Provide background on English Learners in accountability
- Review current requirements
- Understand the nature of developing English Language Proficiency (ELP)
  - Developing English language vs. English language content
  - Correspondence between ELP and content performance
- Considerations for a coherent and comprehensive accountability system

History

- Civil rights protection for English Learners (previously referred to as Limited English Proficient students) dates back to 1974 through the U.S. Supreme Court decision Lau v. Nichols as well as the Equal Educational Opportunities Act of 1974.
- While being a newcomer to the English language by nature of national origin is the defining feature, **equal opportunity for academic content achievement** is the end goal for English Learners.
- Thus, both English Language Proficiency as well as academic achievement in the content areas such as language arts, literacy, and mathematics are important to address and assess.

History

- From the beginning of the standards-based framework in *Improving America's Schools Act of 1994*, there has been a requirement for inclusion of ELs in state assessments with appropriate accommodations to maximize reliability and validity of the assessments.

History

- *No Child Left Behind (2001)* introduced strict subgroup reporting for ELs within Title I but added separate assessment and accountability provisions within Title III.
- Through NCLB, states were required to develop English language proficiency (ELP) assessments aligned to state ELP standards in the domains of speaking, listening, reading and writing. Accountability requirements for Title III required districts (not schools) to report progress (AMAO 1) and status (AMAO 2) for students in attaining English language proficiency.
- It is important to recognize that because ELP was part of Title III, accountability through the ELP assessment was not applicable to students in Title I districts not receiving Title III funding.

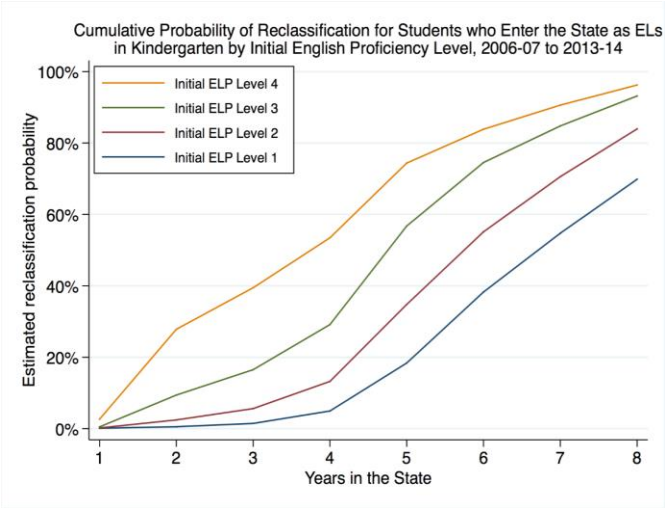
History

- The Every Student Succeeds Act (ESSA) shifts accountability to all Title I students, and requires it at the school (not district) level. Title I now requires ELP assessment to be a school-level indicator for EL students.
- This shift puts a spotlight on English language proficiency as a second dimension of English Learner achievement in addition to their subgroup performance in the academic content assessments.
  - EL subgroup performance can also be monitored using growth.
- ESSA requires that EL students are monitored on their progress towards English language proficiency.
  - Using an appropriate ELP assessment that corresponds with state ELP standards and with English content performance

Notable Features of EL Development

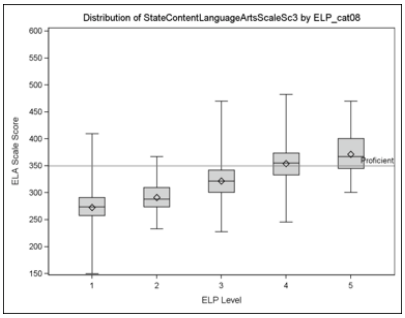
- Academic status and progress of ELs is related to (1) initial ELP level and (2) time in the system.
- English Proficiency Level is related to academic proficiency.
- The number of years reclassified ELs are kept in the subgroup for accountability purposes will determine the performance of the subgroup (the "Ever EL" demonstration).

EL Progress and Status are related to (1) initial ELP level and (2) time in the system



Courtesy: Karen Thompson, Oregon State University

English Proficiency Level is related to academic proficiency



Highest Score

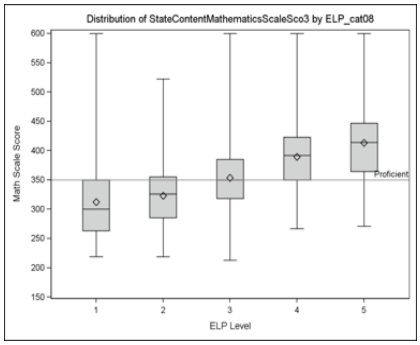
Q3 – 75th percentile rank

Mean

Median

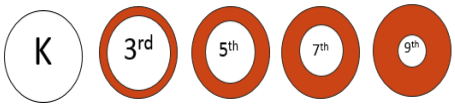
Q1 – 25th percentile rank

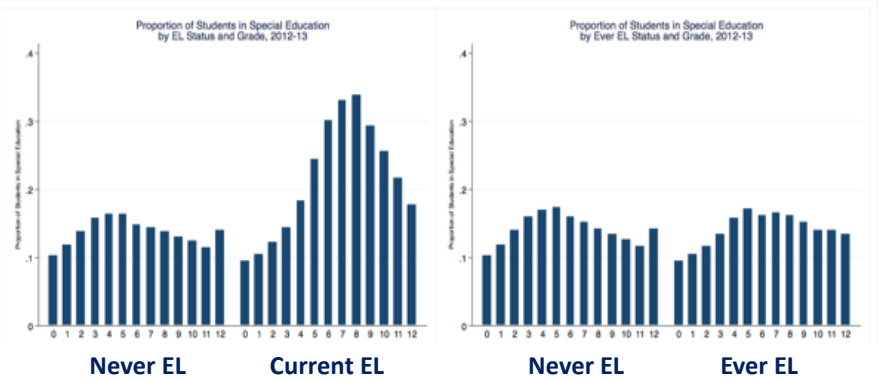
Lowest Score



(Cook, Linquanti, Chinen, & Jung, 2012)

The Ever EL Demonstration Applied to Special Education Participation





Courtesy: Karen Thompson, Oregon State University

Building a Coherent System

- In order to have a coherent accountability system, it is important to determine *a priori* what it is schools ought to be held accountable for.
- What should the EL progress portion be monitoring?

Aggregate Student Performance and School Accountability Scores

- How do schools earn points in the accountability system?
- What score prompts targeted or comprehensive intervention?
- Is expected = "meets", "C", , or ?
  - For the ELP progress component
  - Overall School performance
- Is Average a "C" (etc.)?
- Is "C" related to progress towards English Language Proficiency?
- Does a "D" initiate concern? How about an "F"?
- Should the ELP progress indicator function the same as the other elements of the system?

Table Discussion
Building a Coherent System

- What is the nature of ELs in your state?
  - When begin formal education in State?
  - Initial ELD level?
  - Distribution among schools?
- What is the nature of EL progress in your state?
  - How long does reclassification generally take?
  - What does ELD look like (steady, decelerating)?
- What is your state's "adequate" or "expected" progress?

English Language Proficiency Indicator

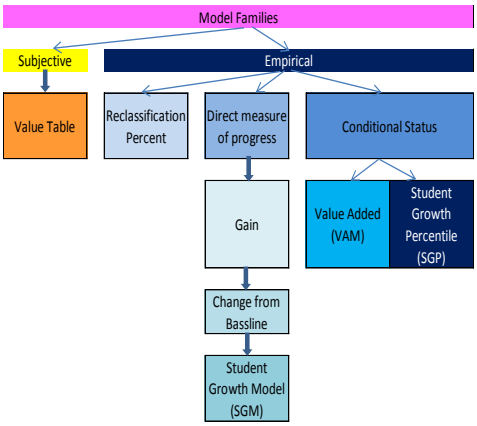
Similarly, proposed § 200.14(b)(4) would clarify how a State measures progress in achieving English language proficiency for all English learners for annual meaningful differentiation. The proposed regulation would provide States flexibility to develop a specific measure for this purpose, while ensuring that States use objective, valid, and consistent measures of student progress. Critically, the proposed regulations would require an objective and valid measure that English learners are attaining, or are on track to attain, English language proficiency in a reasonable time period, consistent with the State-determined timeline in proposed § 200.13. As the Progress in Achieving English Language Proficiency indicator would receive substantial weight in annual meaningful differentiation under proposed § 200.18 and could affect which schools are identified for support, it is important for States to design this indicator in ways that are valid and reliable and provide an accurate determination of English learners' progress toward achieving proficiency in English. Finally, the indicator chosen by the State must include a student's English language proficiency level, as well as additional student characteristics that are used, at a State's discretion, in the English learner-specific long-term goals and measurements of interim progress, for the reasons discussed previously in proposed 200.13(c) and to provide consistency across the components of State accountability systems.

Key Components

- Meaningfully measure progress
  - Objective and valid measure
- Differentiation
- English learners are attaining, or are on track to attain, English language proficiency in a reasonable time period.
- Consider English Language Development Level

Meaningfully Measure Progress

- Change in ELD level from one year to the next
  - Value Table/Index
- Reclassification Rate
- Change in ELP Assessment score from one year to the next
  - Change from baseline (two year)
- Student Growth Model
- Value Added Model
- Student Growth Percentile (SGP) – i.e., school aggregate (mean, median)
- Growth to Standard
- Other



Value Table

Change in ELD level

Translation of Year to Year Changes in ELD Level into "Points"

		Year 1					
	ELD	1	2	3	4	5	6
Year 0	1	0	1	2	3	4	5
	2	-1	0	1	2	3	4
	3	-2	-1	0	1	2	3
	4	-3	-2	-1	0	1	2
	5	-4	-3	-2	-1	0	1
	6	-5	-4	-3	-2	-1	0

Transformed Table (eliminates negative values)

		Year 1					
	ELD	1	2	3	4	5	6
Year 0	1	5	6	7	8	9	10
	2	4	5	6	7	8	9
	3	3	4	5	6	7	8
	4	2	3	4	5	6	7
	5	1	2	3	4	5	6
	6	0	1	2	3	4	5

A value table allows states to explicitly link changes in ELD levels to school accountabili ty points.

Value Table

Change in ELD level

Table Adjusted to Reflect State's Conception of Expected

		Year 1					
	ELD	1	2	3	4	5	6
Year 0	1	7	8	9	10	10	10
	2	6	7	8	9	10	10
	3	5	6	7	8	9	10
	4	4	5	6	7	8	9
	5	3	4	5	6	7	8
	6	2	3	4	5	6	7

A Value Table can explicitly or implicitly take time to English Language Proficiency into account.

Table Adjusted to Reflect State's Conception of Expected

		Year 1					
	ELD	1	2	3	4	5	6
Year 0	1	7	8	9	10	10	10
	2	0	7	8	9	10	10
	3	0	0	7	8	9	10
	4	0	0	0	7	8	9
	5	0	0	0	0	7	8
	6	0	0	0	0	0	7

As Well as A state's Theory of Action – e.g. no value to students who decrease ELD levels from one year to the next.

Reclassification Rate

Reclassification Percent

- Percent of EL students reclassified
  - Percent of ELs in grade?
  - Assuming reclassification occurs at ELD level 5, percent of previous year ELD level 4?
  - Does not monitor students who are not level 4 or higher.
  - Does not give schools credit for progress before level 4.
  - Might disproportionally benefit Middle Schools.
    - Or might create incentives to reclassify in elementary school.

Change in ELP Assessment Scores from One Year to the Next

Gain

- Year over year change
- Change = $SS_{yr1} – SS_{yr0}$
  - Provide unbiased estimate of change
  - Gains tend to be inversely related to initial performance (due to measurement error in the pre-test) – High scorers in year 0 would be related to low gains.
  - Although there is some evidence that this is less of an issue with ELP assessments, gains tend to be unstable over time.

Change in ELP Assessment Scores

Gain

- To reduce the impact of measurement error can use a Kelley True score for year 0.
- $KTSS_{yr0i} = SS_{yr0i}(r) + SS_{yr0.}(1-r)$
  - Under classical test theory *r* is the reliability of the assessment (could use IRT and CSEM).
- Then Change = $SS_{yr1} – KTSS_{yr0}$

Change in ELP Assessment Scores

Change from Bassline

- Another option to improve upon simple gain scores is to:
  - Create a Kelley True score for Year 0 and
  - Use a two year gain.
  - Change = $SS_{yr1} – KTSS_{yr0}$
    - Could use a "change from baseline" approach that uses a two year change for those with two years of data, and one year change for those with one year of data.
    - Could extend to add three years of change for students with three years of data.

Student Growth Model (SGM)

Student Growth Model

- A student growth model measures growth as a function of time, not as a series of gain scores.
- For example a student who scores:

TIME	SCORE
Year 0	100
Year 1	200
Year 2	300

  - Growth = (300-100)/(2-0) = 200/2 = 100pts/year
  - One big advantage of growth models is that a student can remain in the model with incomplete data.

Value Added (VAM)

Value Added Model

- Gain = $SS_{yr1} - SS_{yr0}$
- Or could write as:

$b = SS_{yr1} - \gamma SS_{yr0}$ where = 1.

If $SS_{yr1} = 30$ and $SS_{yr0} = 20$, then b = 10.

So,

$b + \gamma SS_{yr0} = SS_{yr1}$ (10 + 20 =30).

And rearranging

$SS_{yr1} = b + \gamma SS_{yr0}$ (30 = 10 +20).

Value Added (VAM)

Value Added Model

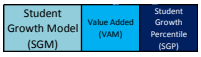
- $SS_{yr1} = b + \gamma SS_{yr0}$
- If we take away the restriction that g = 1 and allow it to be estimated by the data and
- $SS_{yr1} = b + \gamma SS_{yr0} + e$
- add an residual(error) term, e, then this is the most basic Value Added Model.

- The advantage of a VAM over a simple gain is that more variables can be added to the right hand side of the equation (e.g. additional prior test scores, student background, etc.)

Student Growth Percentile (SGP)

Student Growth Percentile

- The Value Added Model (VAM) estimates a single line through the data (although the line need not be linear).
- $SS_{yr1} = b + \gamma SS_{yr0} + e$
- A Student Growth Percentile Model (SGP) estimates the same model 99 times – one for each percentile of the distribution of scores. Depending on which percentile a student's score is in, that is the specific model in which s/he is included.



Student Growth, VAM, and SGP

- These models are more flexible because they can more closely model the shape of progress over time.
- However, VAM and SGP are both conditional status models because they are trying to estimate where a student's score is expected to be given her prior performance.
- Only a Student Growth Model specifically estimates growth over time.
- VAM and SGP are more robust to scale because they are estimating an endpoint, while a Student Growth Model generally requires a vertical scale to make sense.

Table Discussion II

- Given your earlier discussion of ELs, who and where they are, and what their growth and reclassification looks like, what seems to align with your thinking?
- Does the model you plan to use to measure ELP progress need to be the same as the models used for content?
- What model has your state used? Are there lessons learned that you can share?
- What questions do you have when trying to align model with you EL monitoring objectives?

Considering Time Explicitly

- Can set boundaries on all results that lead to reclassification in a given time frame, e.g. for gain.
- (Target Score – $Year_0$ Score)/5.
- Hence, a student who started at 200, but needs to get to 350 for English Language Proficiency needs to gain 150/5, or 30 points per year. Aka Growth to Standard.
  - Possible, reasonable?
  - How to calculate second year requirement?
    - What is shape of ELP progress curve?
      - Students that fall behind and school incentives?
      - Impact of initial level and grade.

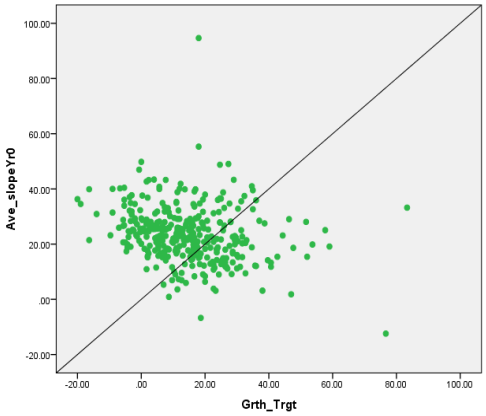
Growth to Standard and Student Growth

- Example:
  - 5th grade students
  - Time frame for reclassification is 8th grade (5 year time frame starting in 3rd grade).

Growth to Standard and Student Growth

Descriptive Statistics				
	Minimum	Maximum	Mean	Std. Deviation
Ave Student Growth	-12.4	94.7	23.6	10.2
Ave Growth Req'd	-20.0	83.3	14.5	13.9

Growth to Standard and Student Growth

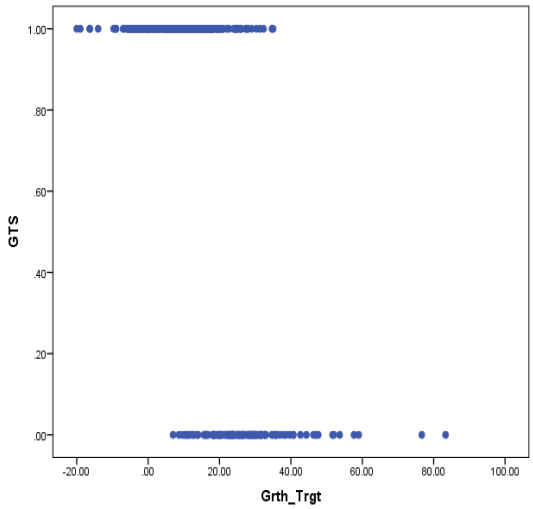


Data from 5th grade in a single district:

Students above the line meet growth requirement.

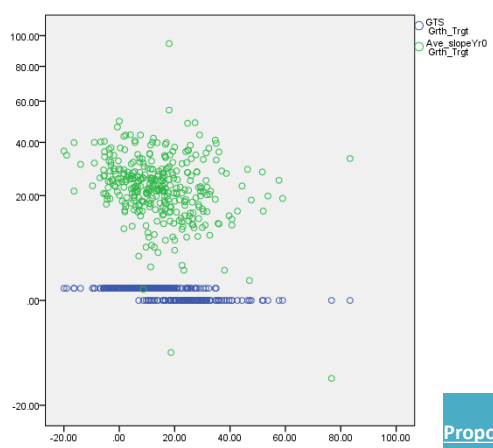
Assumes growth will continue into the future on the same trajectory.

Growth to Standard and Student Growth



Students who were above the line (1) count towards a school's performance score, those below the line (0) do not.

Growth to Standard and Student Growth



Counting students contribution as 0 or 1 is very similar to NCLB and:

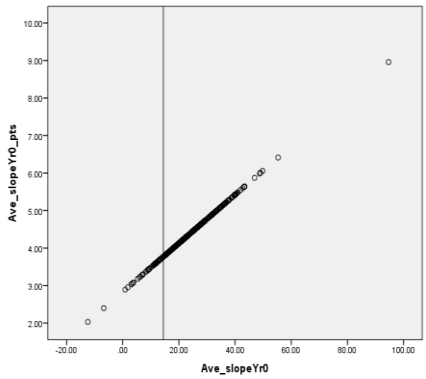
Creates incentives to focus on "bubble kids;"

Lose information about actual student growth;

Does not give credit for "good" growth.

Proportion of	Accounted for by ELP status
Growth to Standard	0.61
Ave. Student Growth	0.21

Growth to Standard and Student Growth



If there is a linear relationship between points based on growth and average growth, then can use the average required growth as a benchmark, for example.

Many options that depend on your conception of what expected, or average means or translates to in terms of accountability.

Empirical Example for Monitoring ELP Progress

School Level		Value Table	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
Elem	Mean	3.07	4.05	6.10	1.44	5.24	4.66	4.76
	N school	229	227	224	232	227	226	228
	SD	1.52	1.46	1.38	1.20	1.36	1.06	0.53
Middle	Mean	1.42	1.59	2.13	1.08	4.00	4.05	3.62
	N school	78	78	76	78	78	76	76
	SD	0.90	1.04	0.86	0.78	1.02	0.76	0.16
High	Mean	3.39	3.19	3.18	1.75	5.54	6.70	3.57
	N school	80	80	78	84	80	78	81
	SD	1.85	1.85	1.20	1.12	1.56	1.81	0.27

Value Table based on ELD changes
Gain 1 = year over year gain (including Kelly True score in Year 0
Gain 2 = two year gain (including Kelly True score in Year 0
Reclass % = Reclassification as percent of EL
SGP = aggregation (mean) of Student Growth Percentile
VAM = Value Added Model
SGM = Student Growth Model

Normalizing Results to Compare Models on ELP Progress

	Value Table	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
Elem	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Middle	46.2%	39.3%	34.9%	74.9%	76.3%	87.0%	76.1%
High	110.3%	78.7%	52.1%	120.9%	105.7%	143.8%	74.9%

The points earned within each model varies significantly by school level for some models. This happens by construction for some models.

For example: Given the decelerating of growth in ELP over time, it is expected that middle schools would earn fewer points than elementary schools because growth has slowed and the fastest growing students will have been reclassified.

Is this desirable?

How Models Compare Across School Levels on ELP Progress

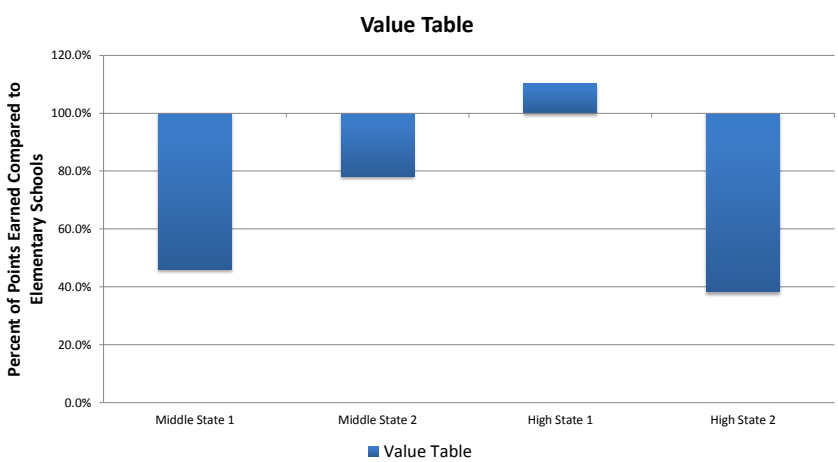
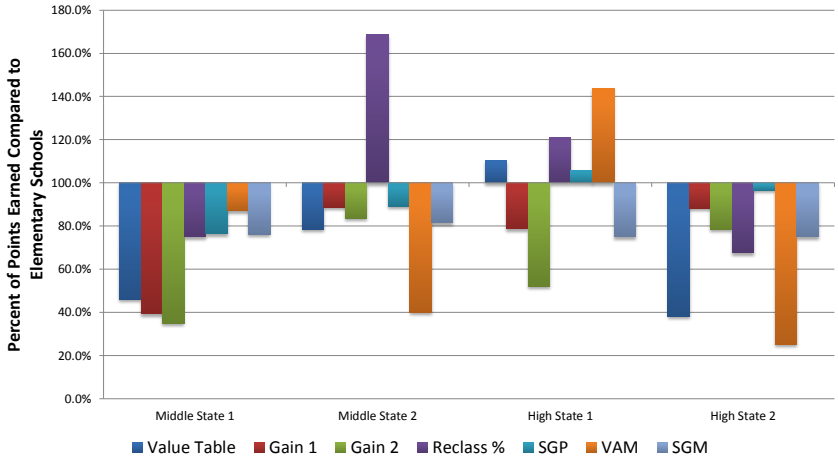
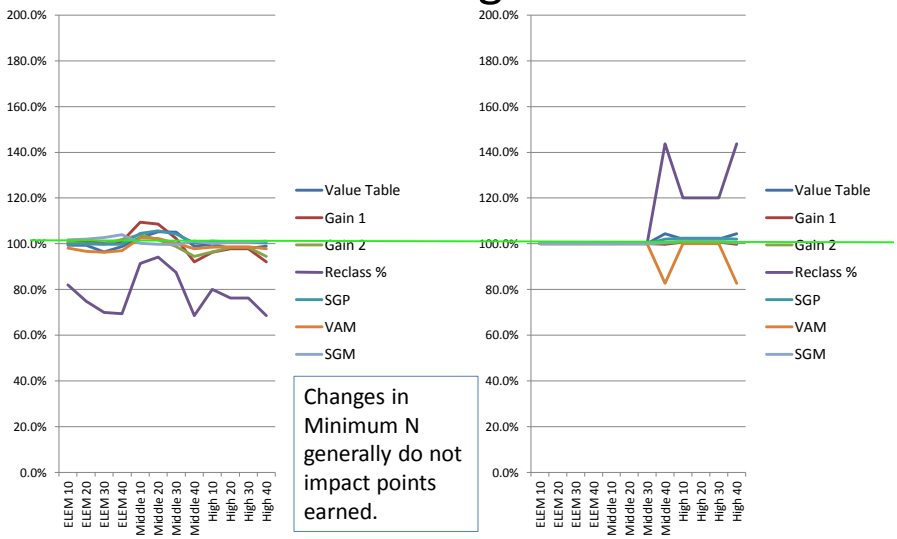


Chart summarizes the proportion of points earned by Middle and High Schools compared to Elementary schools.

How Models Compare Across School Levels on ELP Progress

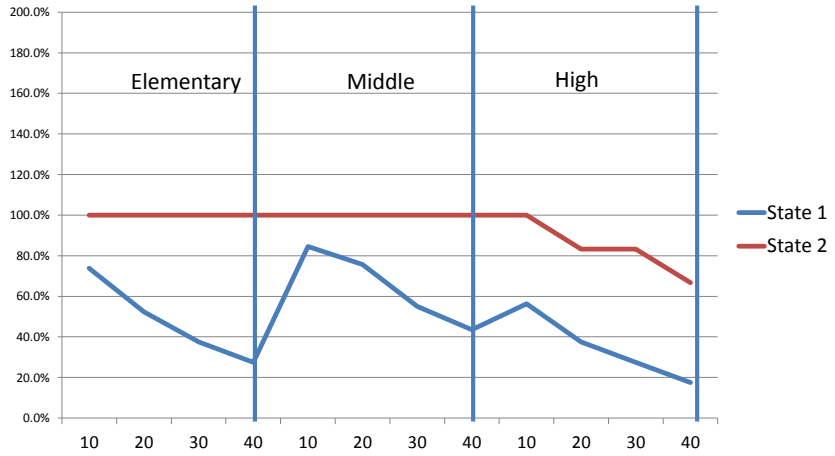


Minimum N Impact on Points for EL Progress



Changes in Minimum N generally do not impact points earned.

Minimum N Impact on Number of Schools Participating in EL Progress



Increasing the minimum N consistently decreases schools held accountable for EL progress.

How Model Results Afford Similar Inferences About Schools

State 1						
	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
Value Table	.618	.716	.245	.019	.700	.715
Gain 1		.966	-.058	.528	.649	.964
Gain 2			.073	.418	.678	.998
Reclass %				-.350	.006	.078
SGP					.271	.406
VAM						.670

State 2 (District)						
	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
Value Table	.601	.371	.264	.701	.118	.088
Gain 1		.598	.386	.815	.057	.261
Gain 2			.163	.268	.039	.861
Reclass %				.446	.079	-.132
SGP					.221	-.061
VAM						-.054

Patterns of correlations among models are consistent with expectations and vary across states

Impact of Student Background Characteristics on ELP Progress Results

State 1: Proportion of Variation in Model Results Shared with Student Background							
Model	Value Table	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
State ELA SS	0.03	0.03	0.03	0.12	0.01	0.01	0.00
State Math SS	0.03	0.05	0.08	0.12	0.01	0.00	0.04
Pct Prof ELA	0.02	0.02	0.03	0.12	0.00	0.00	0.01
Pct Prof Math	0.03	0.04	0.09	0.10	0.01	0.00	0.06
Pct SWD	0.00	0.02	0.05	0.01	0.00	0.02	0.08
Pct FRL	0.00	0.00	0.00	0.11	0.00	0.02	0.02
Pct EO	0.01	0.03	0.05	0.04	0.01	0.00	0.08
Number of EL	0.01	0.01	0.01	0.07	0.00	0.02	0.00

State 2: Proportion of Variation in Model Results Shared with Student Background							
Model	Value Table	Gain 1	Gain 2	Reclass %	SGP	VAM	SGM
State ELA SS	0.45	0.19	0.30	0.61	0.01	0.19	0.26
State Math SS	0.69	0.63	0.73	0.22	0.02	0.47	0.59
Pct Prof ELA	0.46	0.19	0.31	0.58	0.03	0.20	0.26
Pct Prof Math	0.70	0.69	0.78	0.17	0.03	0.48	0.66
Pct SWD	0.05	0.00	0.03	0.15	0.02	0.10	0.00
Pct FRL	0.01	0.16	0.16	0.02	0.06	0.01	0.09
Pct EO	0.00	0.23	0.19	0.00	0.21	0.01	0.16
Number of EL	0.05	0.00	0.00	0.15	0.00	0.13	0.00

Ideally shared variation should be less than .05. Larger values may be addressed by expanding the specification of some models.

Stability of Models over Time

Model	State 1	State 2
Value Table	0.17	0.82
Gain 1	0.25	0.80
Gain 2	0.86	0.95
SGM	0.77	0.95
Reclass %	0.27	0.93
SGP*	0.10	0.27
VAM	0.42	0.49

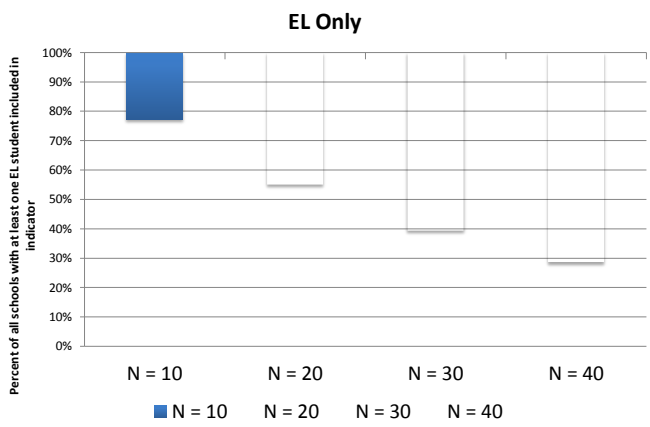
*Previous multi-state studies using state content assessments found SGP stability to range fro .32 to .46.

The results highlight the potential differences state context can make and why specific state results will vary.

Generally gains or growth based on more data are more likely to be stable than when based on less data.

Results for SGP and VAM are somewhat less stable for ELP assessments than they tend to be for state content assessments.

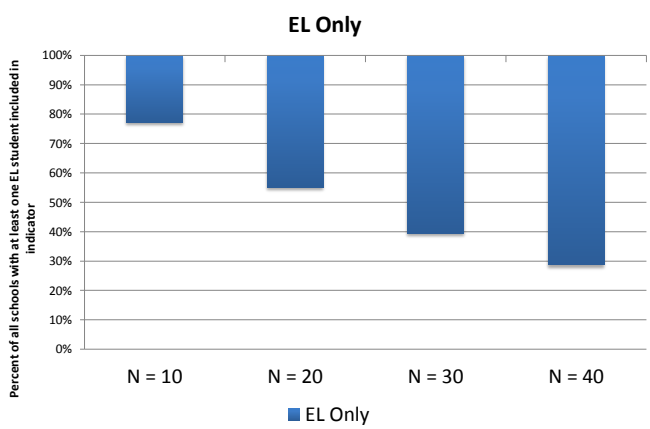
Impact of Minimum N and Inclusion of Reclassified Students in EL Content Performance Participation



Increasing the minimum N for schools to be held accountable for the EL subgroup substantively decreases the percentage of schools included compared to the universe of schools with at least one EL student.

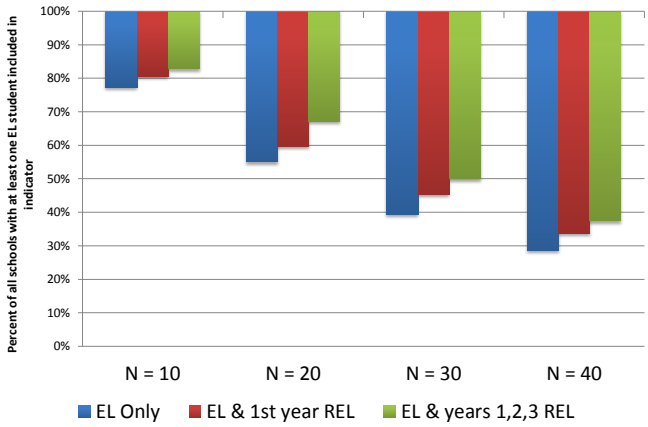
A minimum N of 10 results in about 78% of schools included.

Impact of Minimum N and Inclusion of Reclassified Students in EL Content Performance Participation



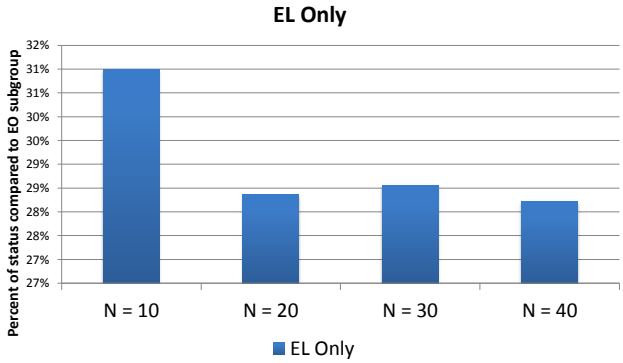
A minimum of N of 40 results in fewer than 30% of elementary schools participating in state 1.

Impact of Minimum N and Inclusion of Reclassified Students in EL Content Performance Participation



Including reclassified ELs in the EL subgroup increases the proportion of schools with sufficient N to "count", but the effect is substantially smaller than the impact of minimum N size.

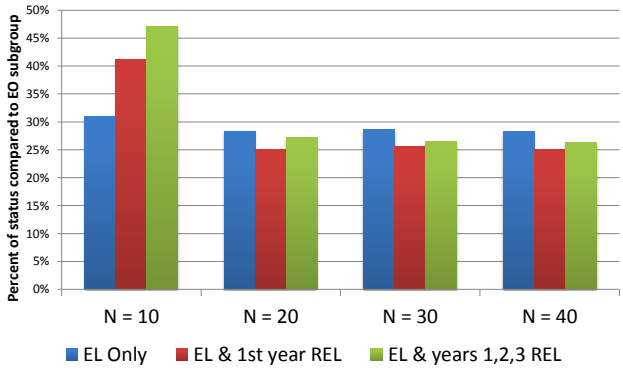
Impact on Content Status (percent proficient)



Elementary - Status

At a Minimum N of 10, EL performance (percent proficient) is about 31% of EOs. Increasing minimum N will have unknown, though small, effects on performance indicators.

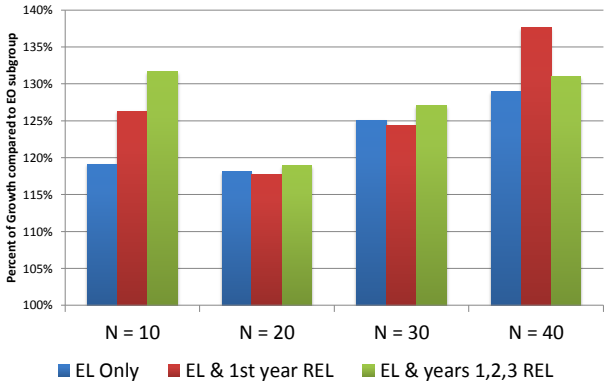
Impact on Content Status (percent proficient)



Elementary - Status

Impact on points is not straight-forward. Points earned by schools in the EL subgroup do not necessarily increase as schools are dropped due to minimum N size.

Impact on Content Growth



Elementary - Growth

Overall, EL and REL students demonstrate more content growth than EO students. Shifting REL to the EL subgroup generally increases growth for the EL subgroup, but the effect is not uniform.

Impact of Inclusion of Reclassified Students in EL Content Performance

- The impact on moving REL into the EL subgroup consistently increases the number of schools included for the EL subgroup.
- Status and growth may not necessarily increase for the EL subgroup.
  - Depends on Minimum N
  - Depends on School level
- Reporting and calculations may differ. This analysis assumed independent calculations – not a weighted composite.

Summary

- Before selecting a model it is important to understand the state context.
- What are causes for action (good or bad).
- Select model that results in inferences aligned with goals.
- No single approach is best.
  - Can use presented results to narrow choices but should use own state data to model impact.
- Attempt to include as many students/Schools as possible.
  - Increasing Minimum N greatly impact the number of schools included for the EL subgroup.
    - Some models allow all schools to be included and are less impacted by N sizes – The tradeoff between precision and exclusion. What is the impact of false positive/negative vs. no information?
      - Using Confidence Intervals can help address N size issues