Options for Measuring Achievement Growth Using Smarter Balanced Summative Test Scores

Joseph Martineau

Center for Assessment

- # Typical psychometric definition
  - ## Growth (n.)
    - ### A simple or estimated difference on the same scale from one point in time to another.

- # Some adapted relevant dictionary definitions
  - ## Growth (n.)
    1. An increase in magnitude
       Requires status to be on the same scale over time
    2. Development from one stage to another
       Allows (but does not require) different status scales

Theory of Action

- Advance document by Marion

- A relatively new resource to guide the development of a logic model/theory of action:

  - Shakman, K., & Rodriguez, S. M. (2015). *Logic models for program design, implementation, and evaluation: Workshop toolkit* (REL 2015–057). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

  - https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=401

- Many options for growth measures available to Smarter Balanced member states.

- TOA can help states identify the important characteristics of a growth measure for their purposes.

Some Typical Uses of Growth Measures

- Levels of Growth Data
  - Individual student
  - Aggregate (e.g., classroom, school, district, state)
- Typical Formal Uses of Growth Data
  - Reporting
  - Data Files
  - Accountability
  - Educator Evaluation
  - Program Evaluation
  - School Improvement
  - Policy Development/Evaluation

- Two Key Features
  - Interpreted relative to what?

    (The interpretive referent)

  - Calculated using what method?

    (The analytical model)

- ## Scale-referenced
  - Growth is interpreted relative to the achievement scale

- ## Norm-referenced
  - Growth is interpreted relative to other students/units

- ## Criterion-referenced Judgment
  - Growth is interpreted relative to judgmentally-defined growth targets (or criteria)

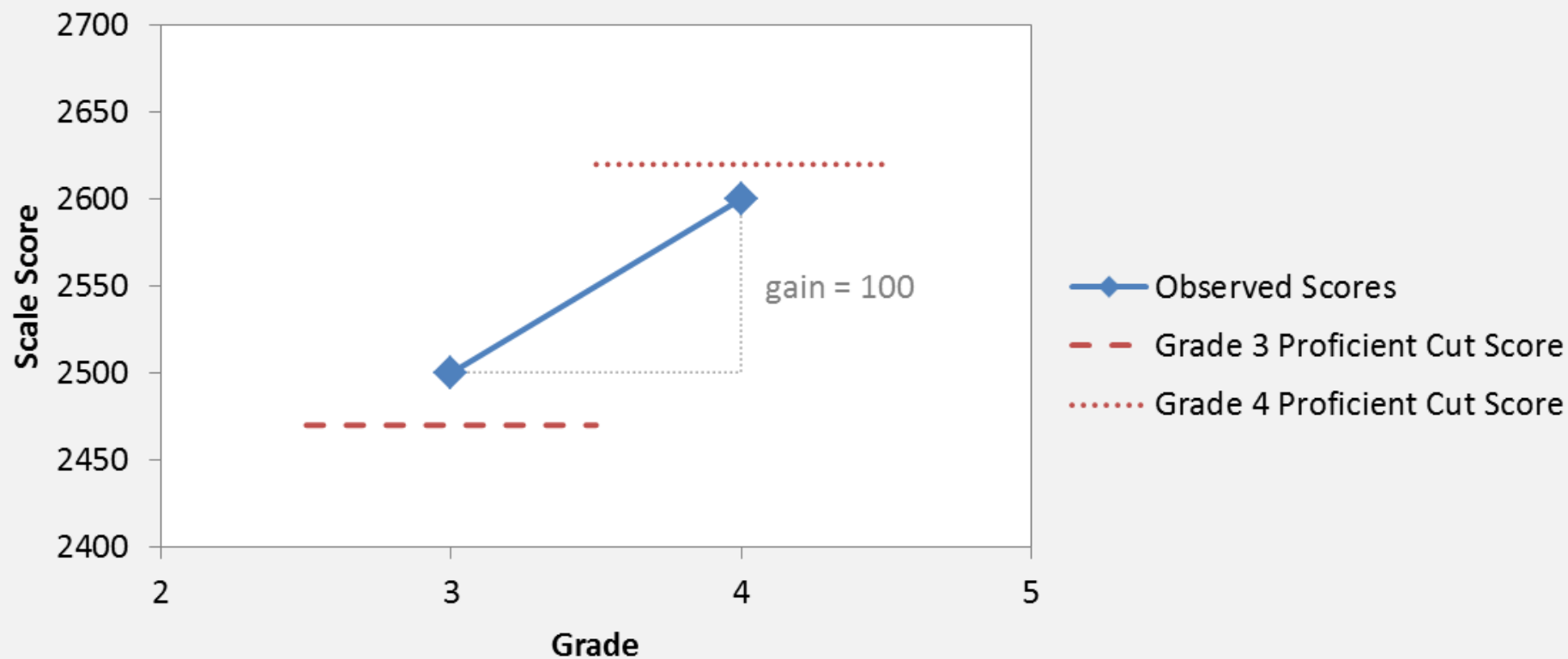# (Observed-achievement) Scale-referenced

- Cruz scored 2500 in grade 3, and 2600 in grade 4.
- He grew by 100 points on the Smarter Balanced math scale from grade 3 to grade 4
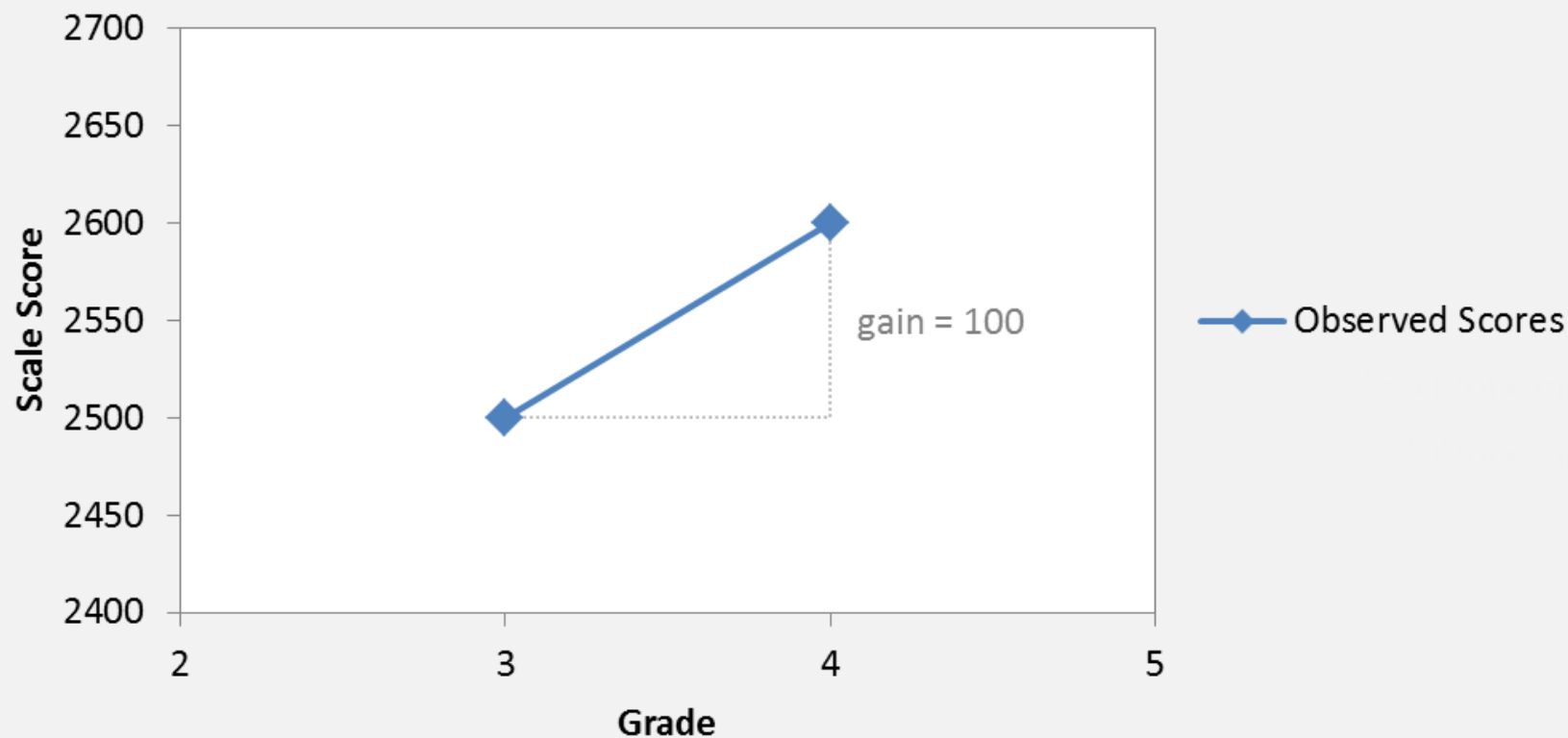
(Observed-achievement) Scale-referenced

- Need additional information about the scale to understand what a gain score means.

- Cruz's gain of 100 points from grade 3 to grade 4 was…
  - greater than only 30% of students in the state
- Alina had a Student Growth Percentile (SGP) of 78 on Smarter Balanced grade 4 math test, meaning that…
  - she scored better than 78% of her academic peers (those with the same scores on in grade 3).

- Requires an understanding of relative standing

(Observed-achievement) Criterion-referenced Judgment

- Cruz's gain score of 100 points does not count as adequate growth because the State Board of Education identified 150 points as adequate growth for students who did not score at the *proficient* level.

- Alina did not score at or above *proficient*, but her SGP of 78 counts as adequate growth because state statute identifies an SGP of 60 as adequate growth for students who do not score *proficient*.

- Requires a policy body to define what growth score counts as adequate and in what circumstances

Interpretations Focused on Future Achievement

- ## Criterion-referenced trajectory continuation
  - If the recent trajectory is continued for *N* years, the student will/will not meet an *N*-years-out status target

- ## Criterion-referenced trajectory target
  - Growth targets are calculated for the next *N* years on the basis of a future trajectory necessary to meet an *N*-years-out status target

- ## Criterion-referenced category projection
  - Based on statistical relationships among scores in earlier and later grades for a previous cohort of students, predict for a new cohort of students (from their earlier-grade scores) whether their later-grade scores will be in a desirable performance category *N* years out.

- ## Criterion-referenced probability projection
  - Based on statistical relationships among scores in earlier and later grades for a previous cohort of students, predict for a new cohort of students (from their earlier-grade scores) the probabilities that their later-grade scores will be in a desirable performance level N years out.
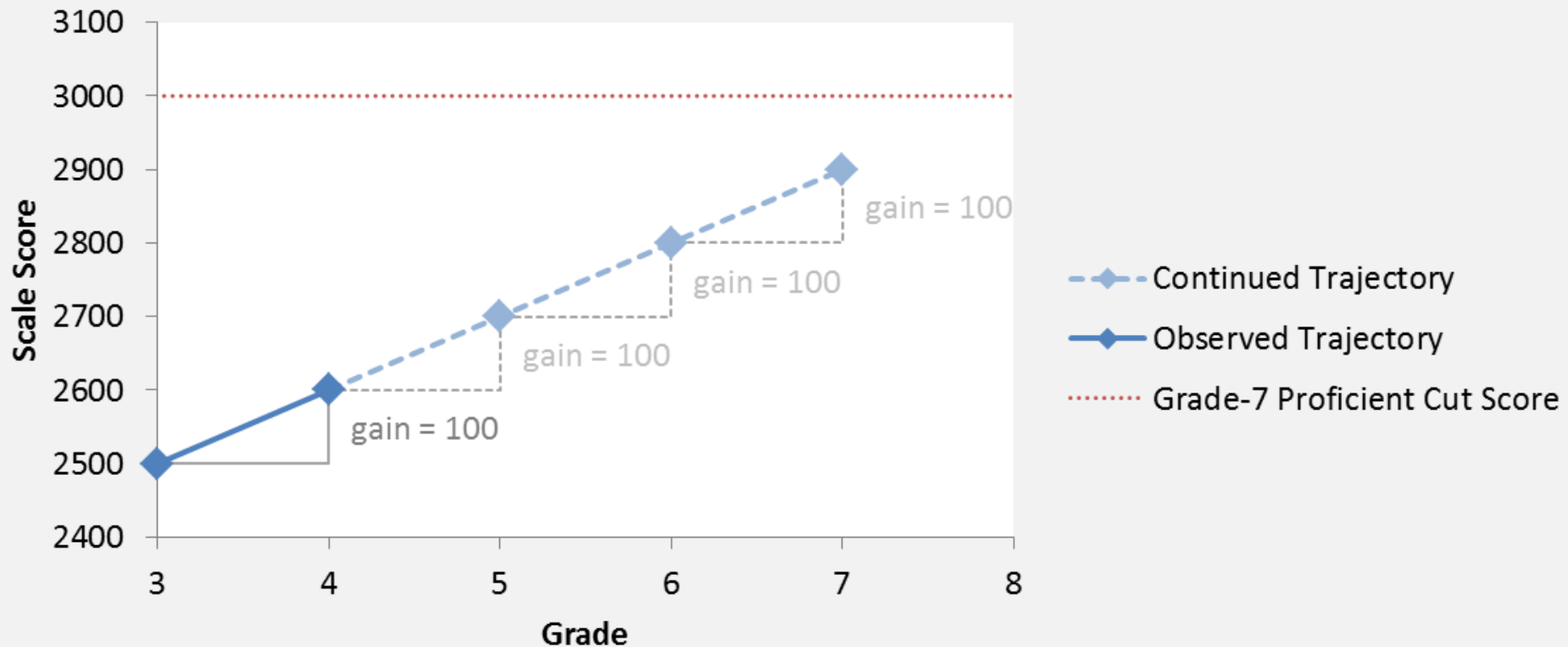
- Cruz scored 2500 in grade 3 and 2600 in grade 4, for a gain score of 100 points.

- The proficient cut score in grade 7 is 3000.
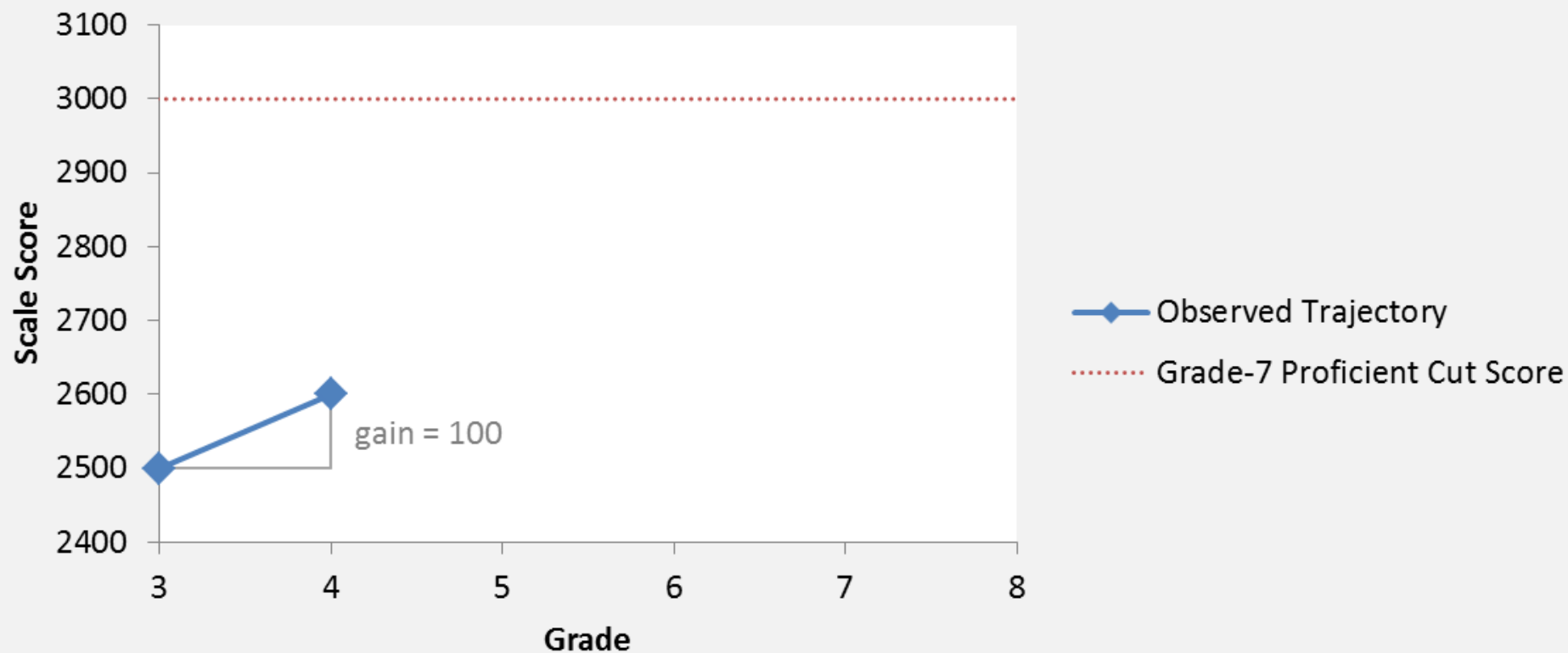
- Continuing Cruz's 100 point gain scores for the next three years (until the end of grade 7) would give him a grade 7 math score of 2900.

- This is <u>insufficient</u> to become *proficient* by the end of grade 7.

- Surya scored 2500 in grade 3 math and 2680 in grade 4 math.

- Her gain score was 180 points, giving her a student growth percentile of 65 (meaning that she outscored 65 percent of students with the same prior scores).

SGP = 65
gain = 180

Observed Trajectory
Grade-7 Proficient Cut Score

- Continuing with an SGP of 65 would give Surya the following future trajectory.

- This future trajectory of scores based on continuing an SGP of 65 is <u>sufficient</u> to become proficient by the end of grade 7.

- Cruz scored 2500 in grade 3 and 2600 in grade 4, for a gain score of 100 points.

- The proficient cut score in grade 7 is 3000.
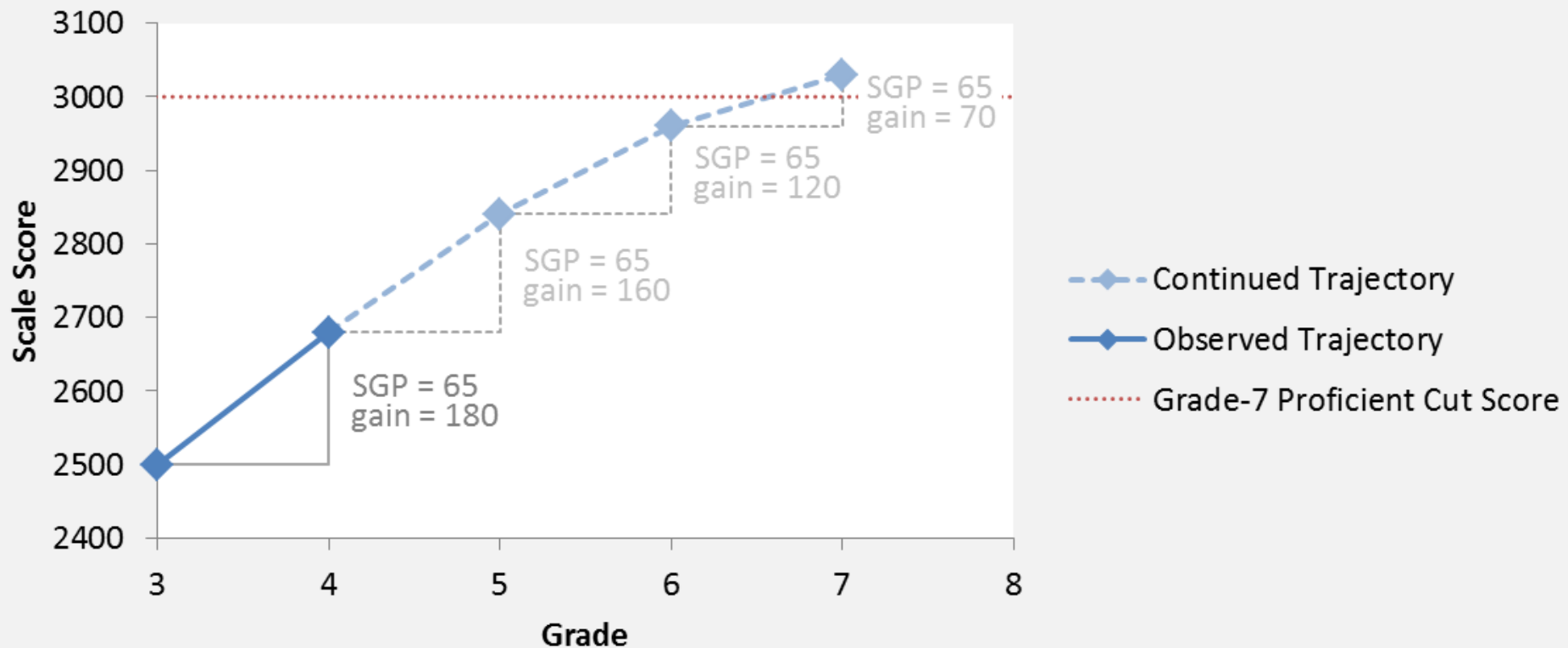
(Future Achievement) Criterion-referenced Trajectory Target Using Gain Scores

- Cruz's grade 4 math score (2600) leaves 400 points to make up by the end of 7th grade to get to the 7th grade *proficient* cut score

- Drawing a straight line from Cruz's current score (2600) to the 7th grade cut score gives the following target scores for Cruz in grades 5, 6, and 7.

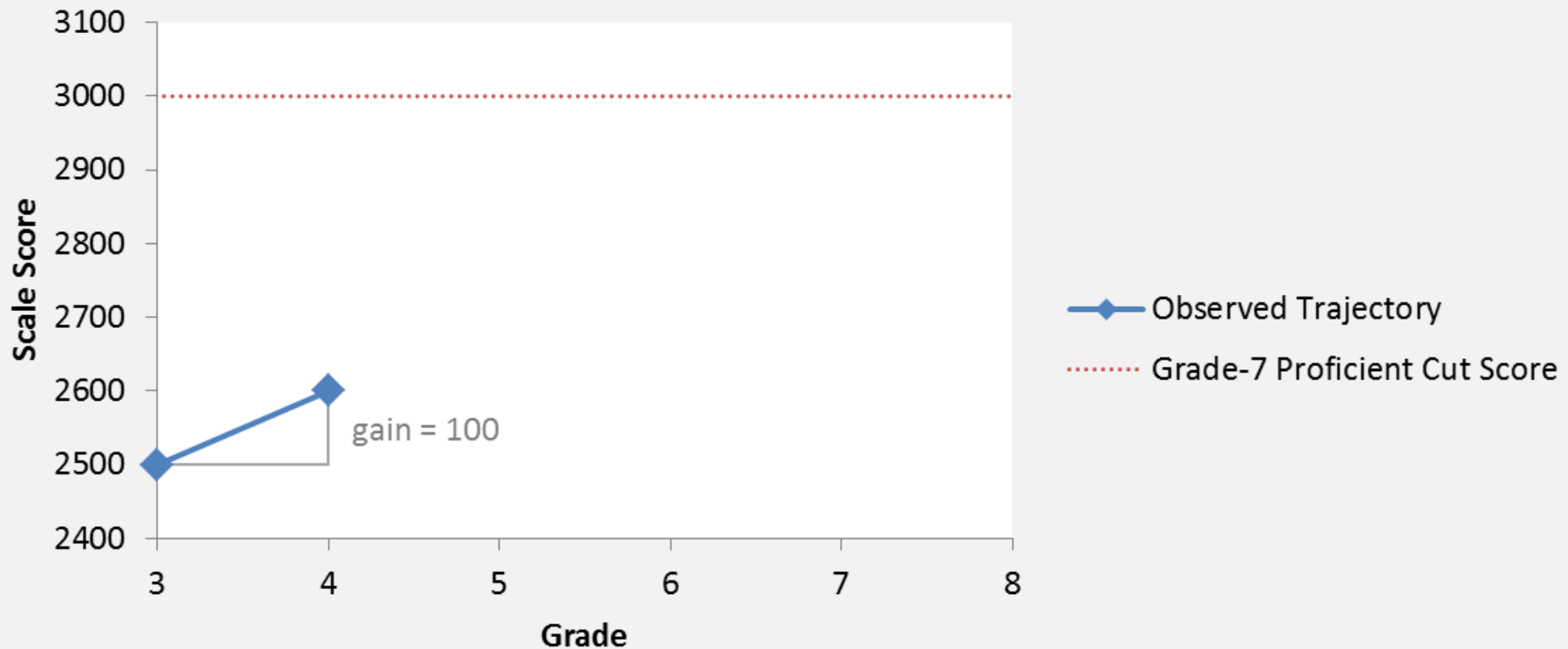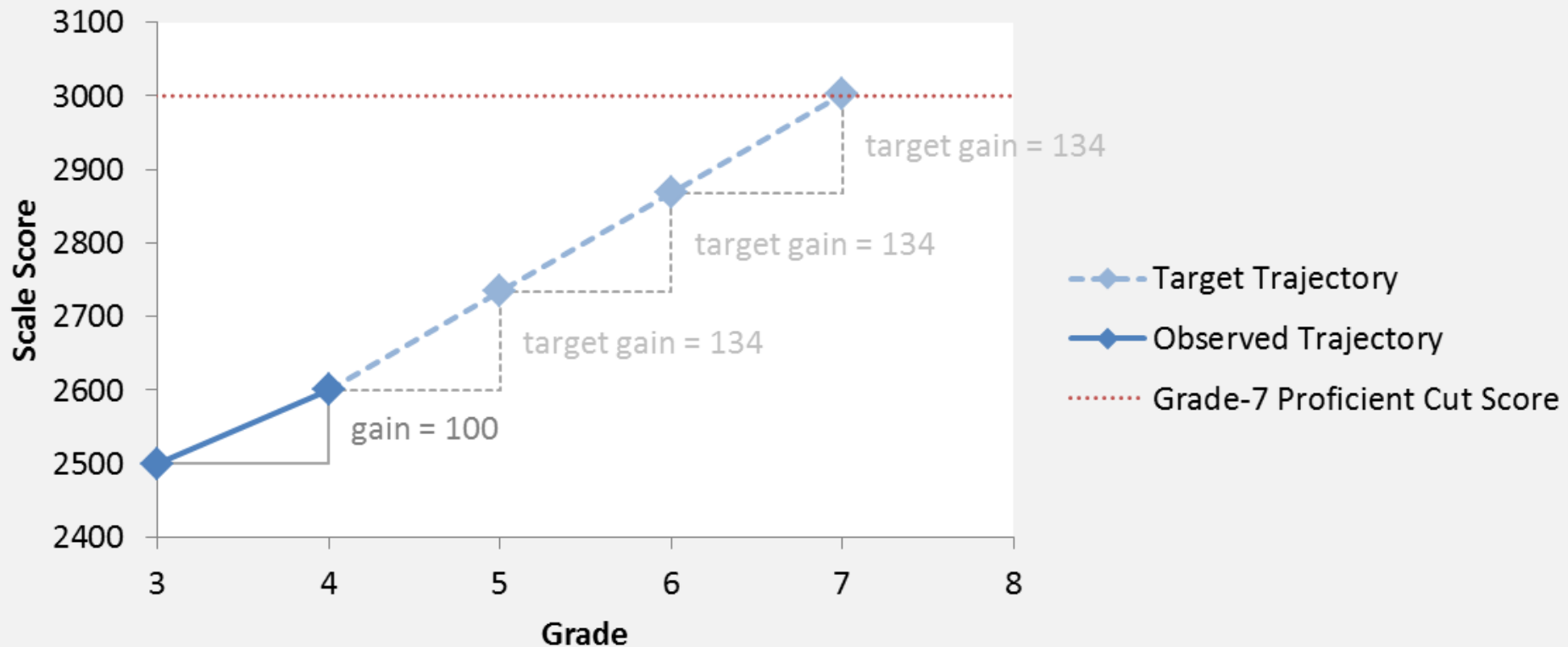# (Future Achievement) Criterion-referenced Trajectory Target Using Student Growth Percentiles

- Surya scored 2500 in grade 3 and 2680 in grade 4.
- This gave her a gain score of 180 points, and an SGP of 65 (meaning that she outscored 65% of students with the same prior scores).

- What future SGP will be adequate for her to become proficient by the end of 7th grade (what is her adequate growth percentile, or AGP)?

- By using the SGPs from current students in grades 5, 6, and 7, we find that if Sariah has an SGP of at least 50 for the next three years, she will be proficient at the end of 7th grade.
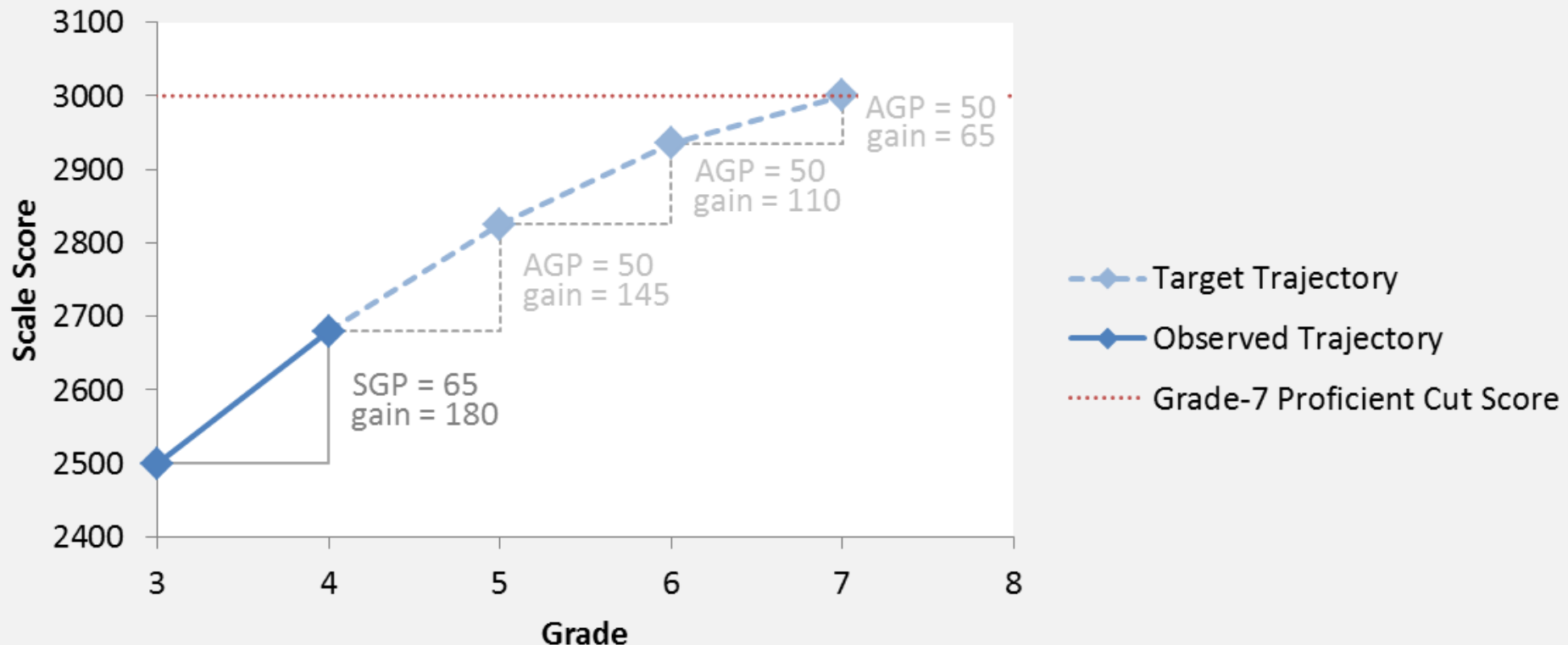
- Phillip had all possible test scores in grades 3-5.
- Need to predict his score in grade 8.
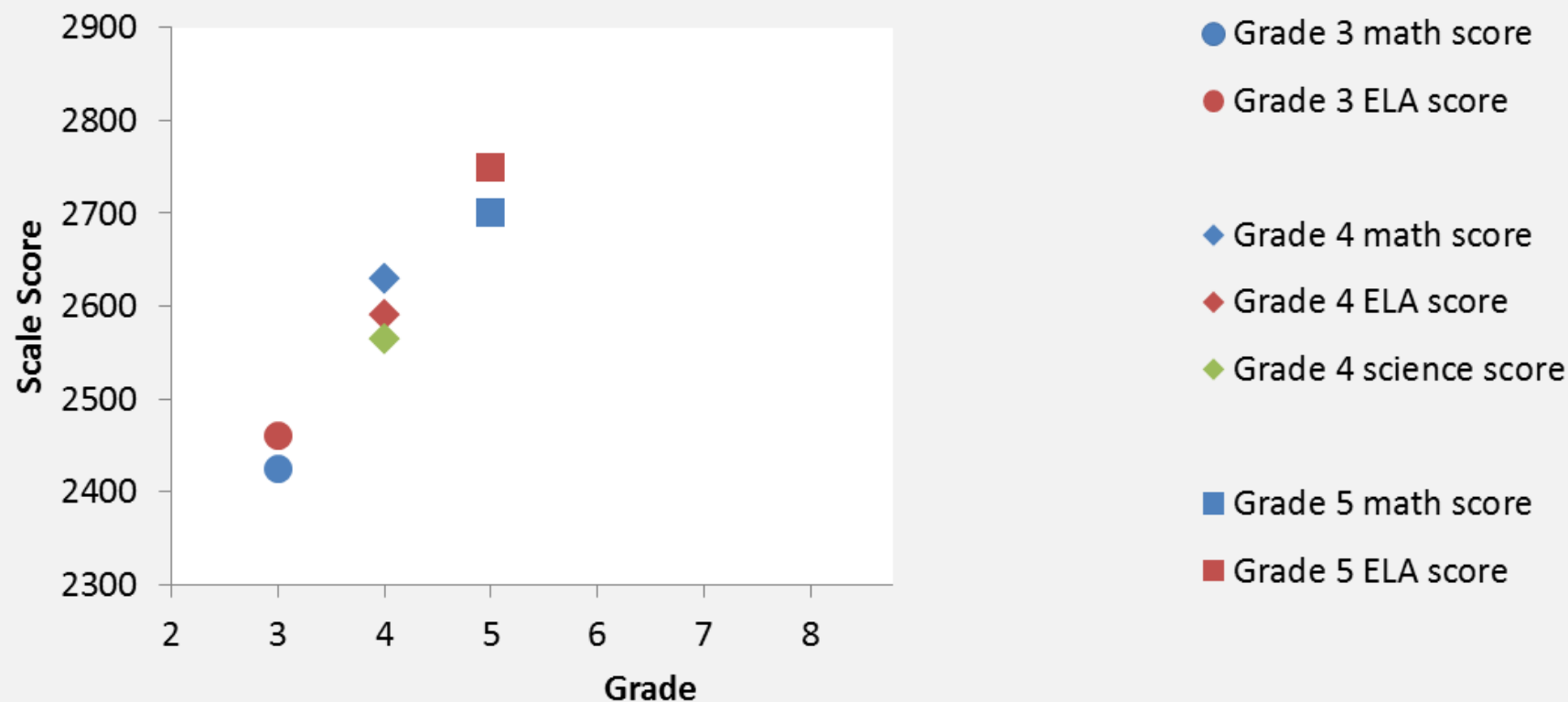
(Future-achievement) Criterion-referenced Category Projection

- With scores from grades 3-5 and grade 8 from one or more previous cohorts, we can establish statistical relationships between scores in grades 3-5 and scores in grade 8.

- Using those relationships, we can predict Phillip's grade 8 score from his grade 3-5 scores.
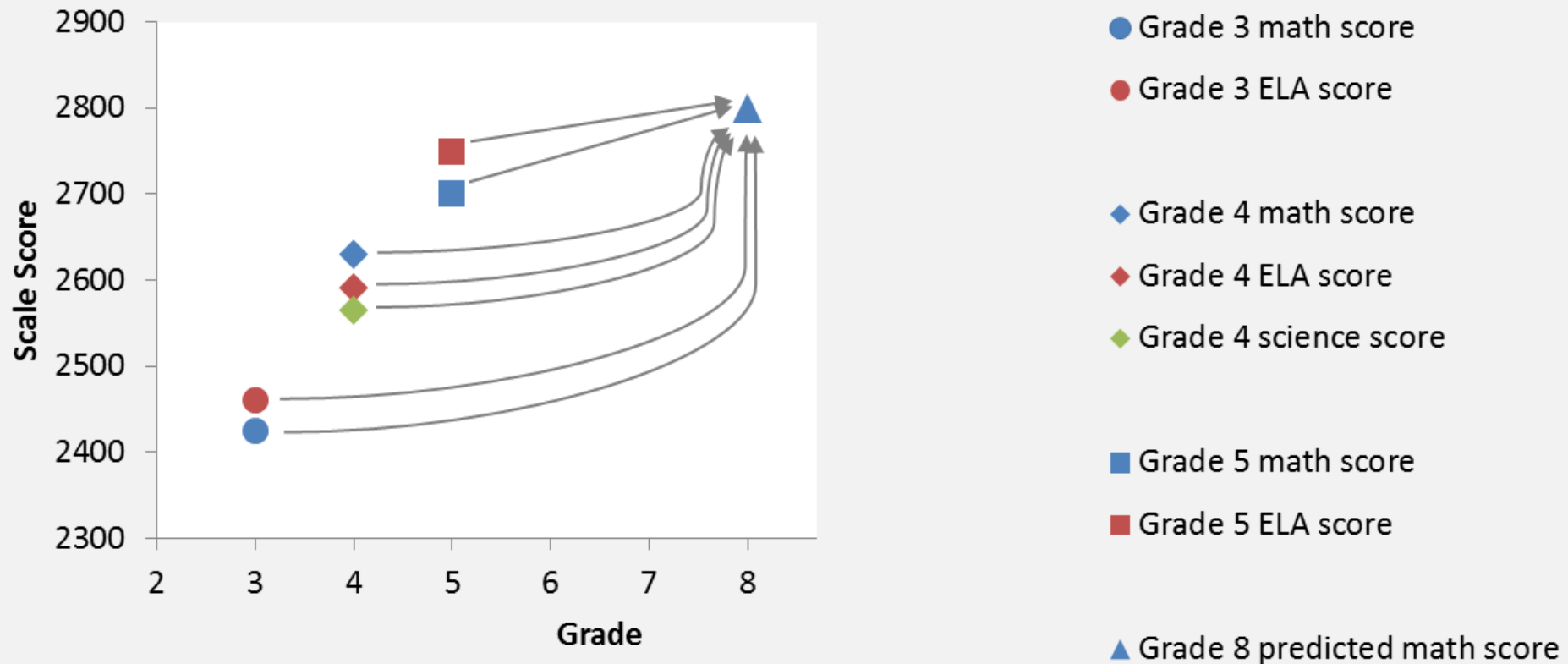
- Using the grade-8 cut scores, we can identify Phillip's predicted grade-8 achievement level.
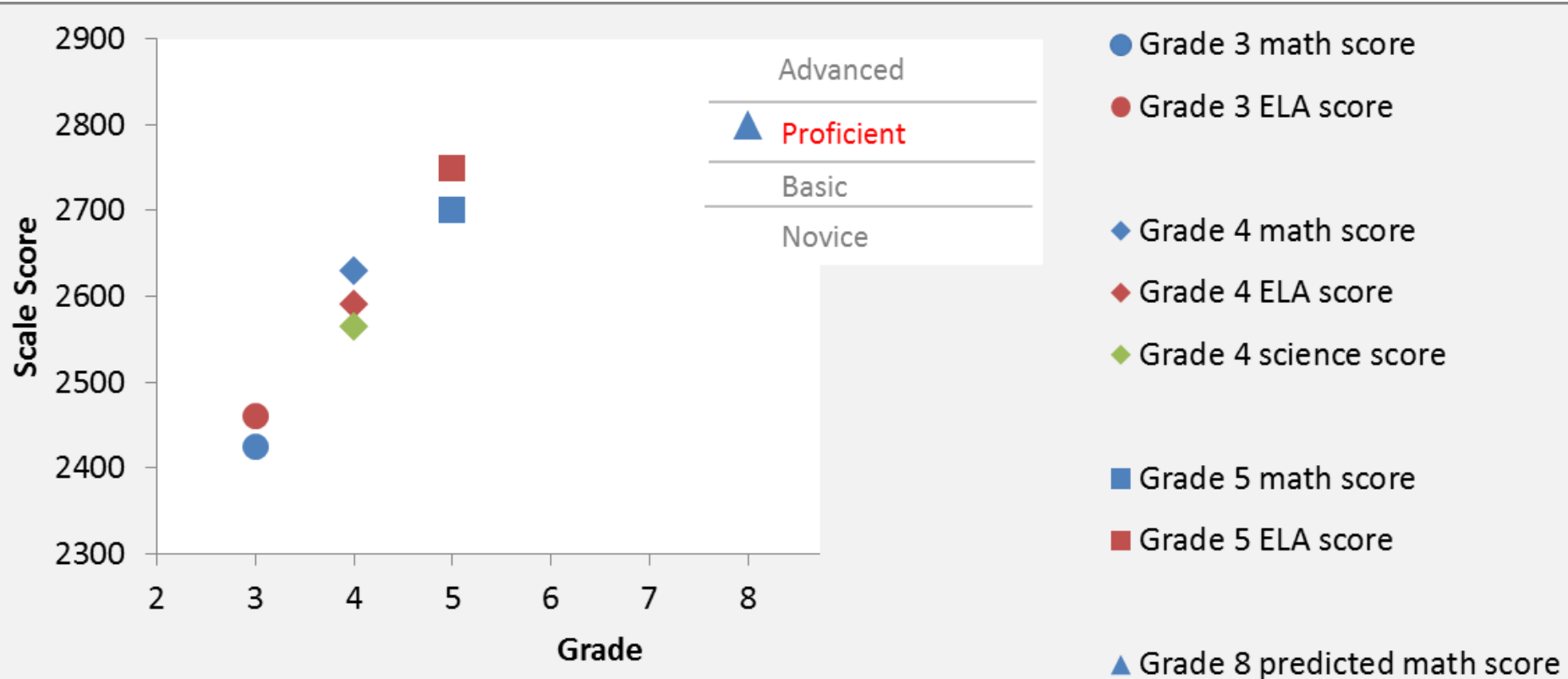
(Future-achievement) Criterion-referenced Probability Projection

- Just one more step is needed to go from the predicted category to the probability of being in each category.

- The model used to predict the grade-8 score also produces a prediction error.

- Based on Phillip's predicted score and the prediction error, we can calculate the probability that Phillip's grade-8 score will be in each achievement level.

Analytical Models
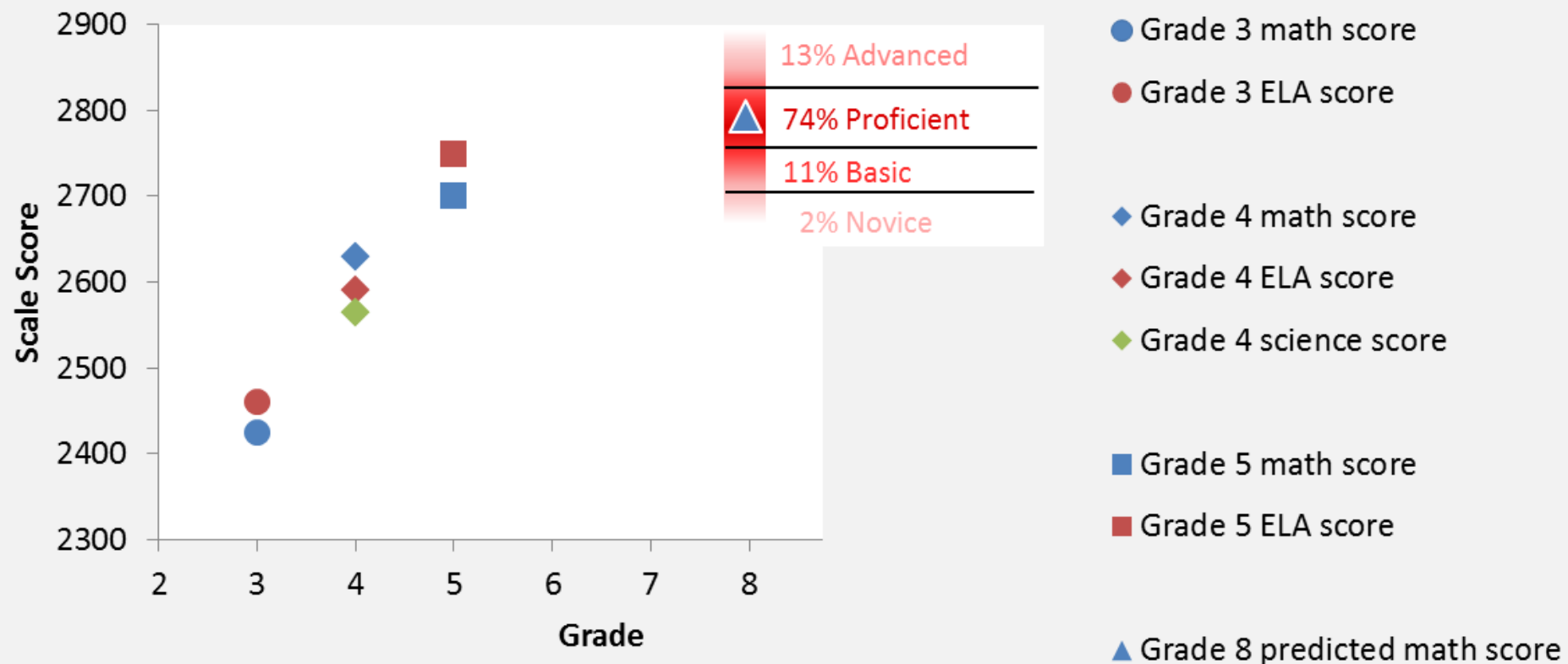
- Gain score
- Growth rates
- Z-score gains
- Baselined Z-score gains
- Student growth percentiles
- Baselined student growth percentiles
- Transitions (e.g., transition/value tables)
- Residual status-based value-added models (VAM)
- Residual gain score-based VAM
- Residual growth rate-based VAM

Gain Score

- Cruz scored 2500 in grade 3, and 2600 in grade 4.

- He grew by 100 points on the Smarter Balanced math scale from grade 3 to grade 4
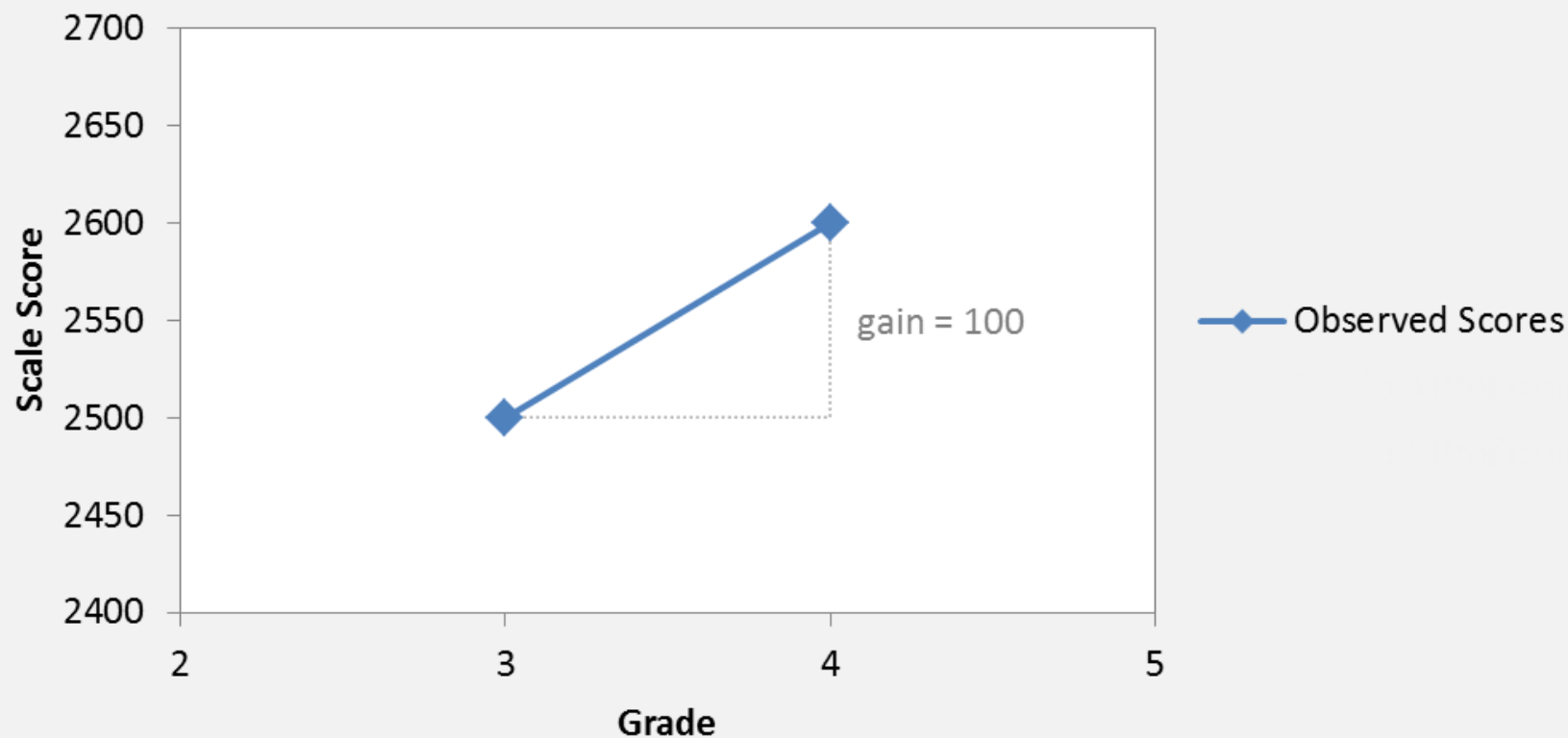
Gain Score

- Need additional information about the scale to understand what a gain score means.

Issues to Consider with Gain Scores
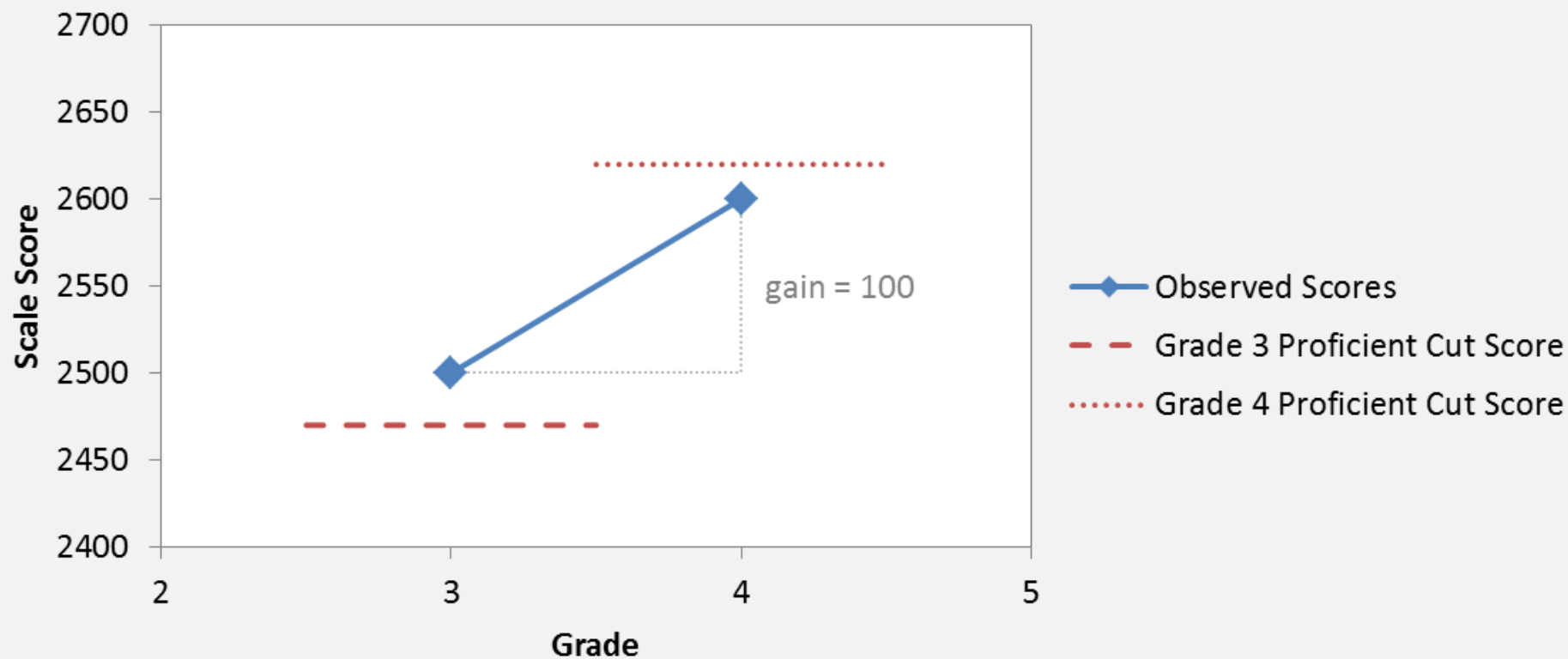
- A gain score is the simplest analytical model and may appear to have the simplest interpretation.

- However, a considerable proportion of gain scores tend to be negative, causing a challenge in communication. This arises from the following set of reasons:

  – Score variance within grades tend to be large relative to average grade-to-grade gains.

  – The correlation between pre-test scores and gain scores is negatively biased.

  – Gain scores are affected by regression to the mean (students with scores further away from the mean, either above or below, are more likely than not to score closer to the mean in the next grade).

  – Mean grade-to-grade growth tends to become smaller as grade level increases.

  – Score variance tends to increase as grade level increases.

Issues to Consider with Gain Scores

- Actual gain scores by pretest decile in five states with vertical scales

Grade 4 Mathematics

Grade 8 Reading
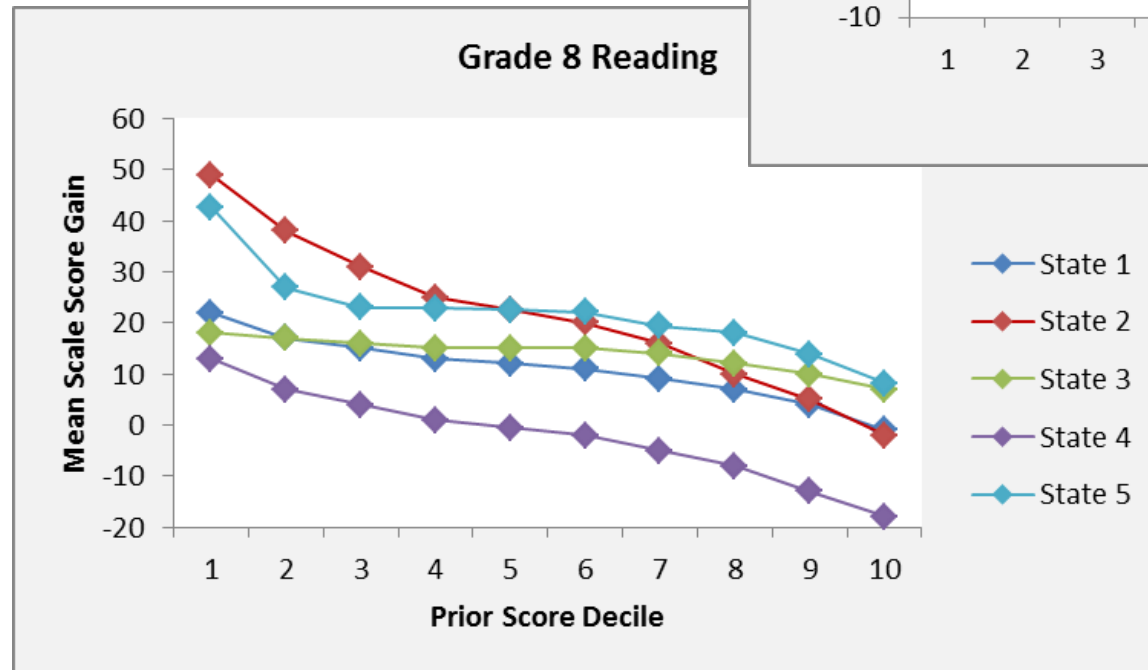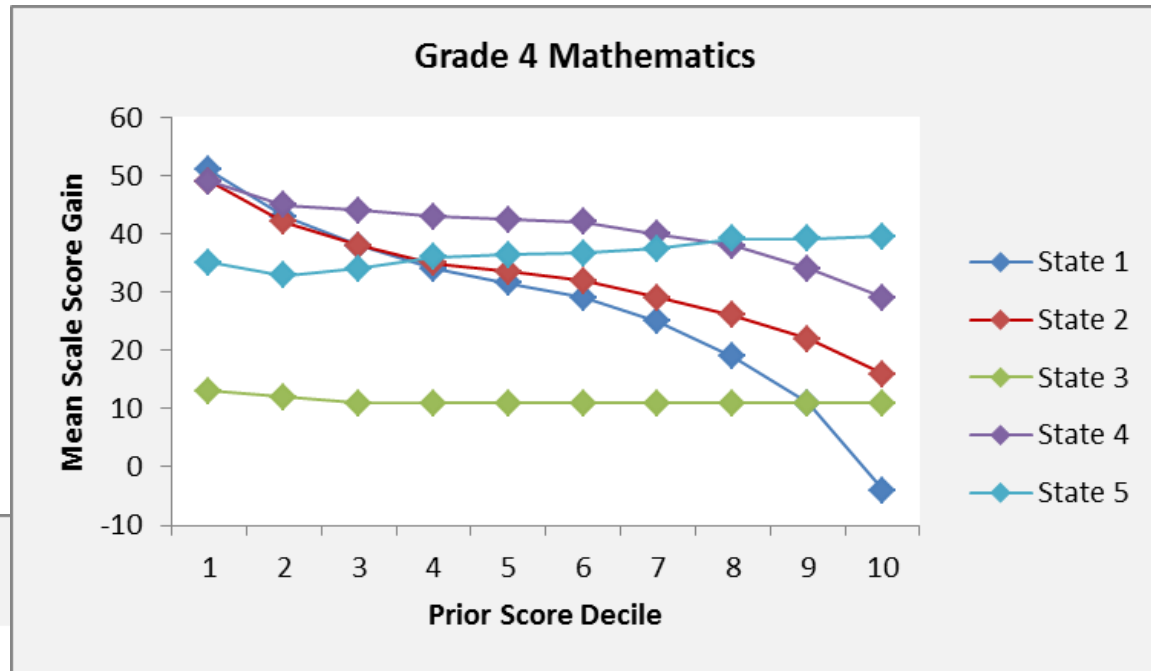
- Actual data from a state with a vertical scale

Prevalence of Negative Gain Scores
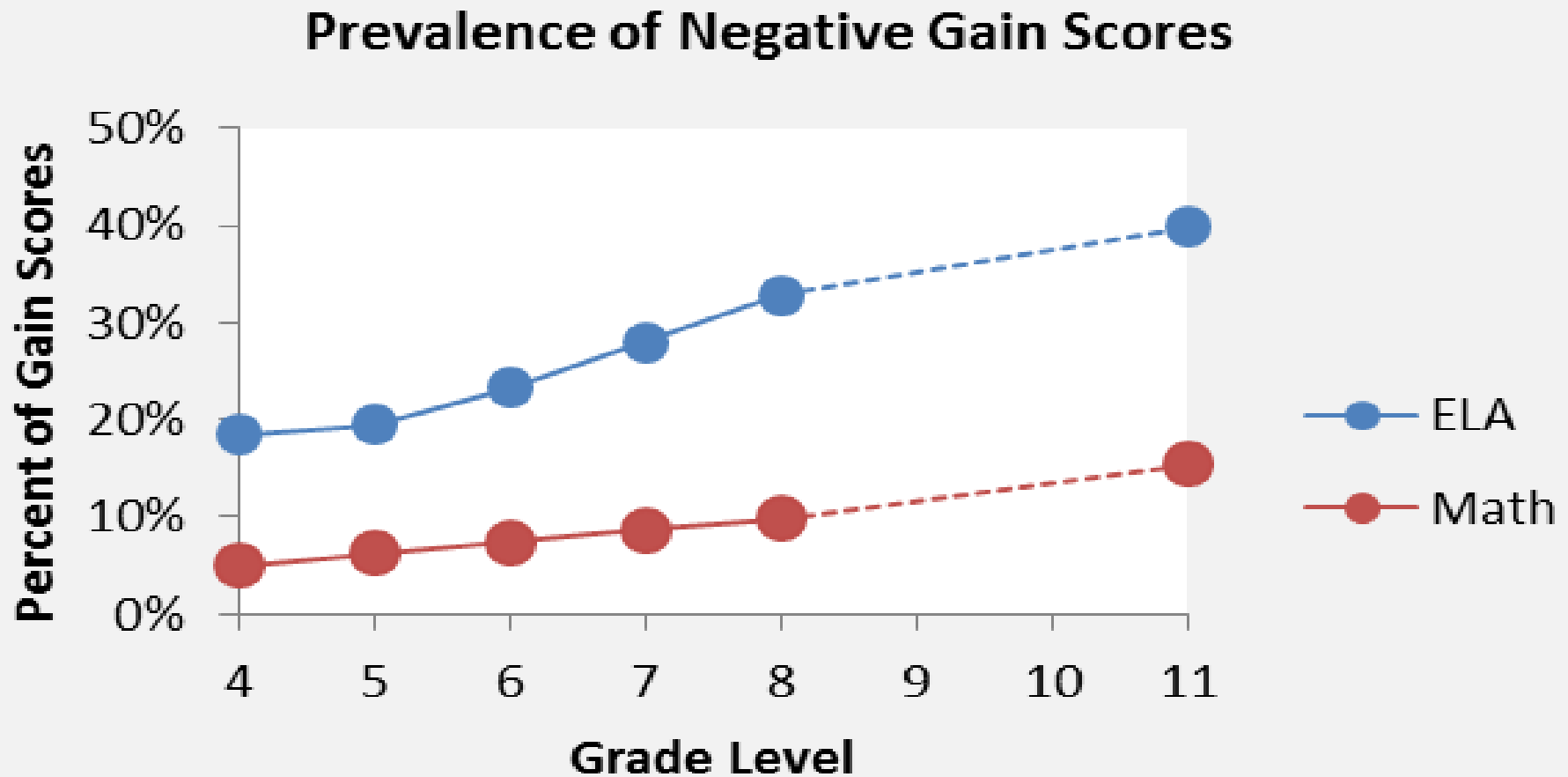
Z-score Gain

- Al's z-score was 0.2 in grade 3 (he scored 0.2 standard deviations above the mean)
- His z-score was -0.3 in grade 4 (he scored 0.3 standard deviations below the mean)
- Therefore, his relative standing decreased by 0.5 standard deviations.
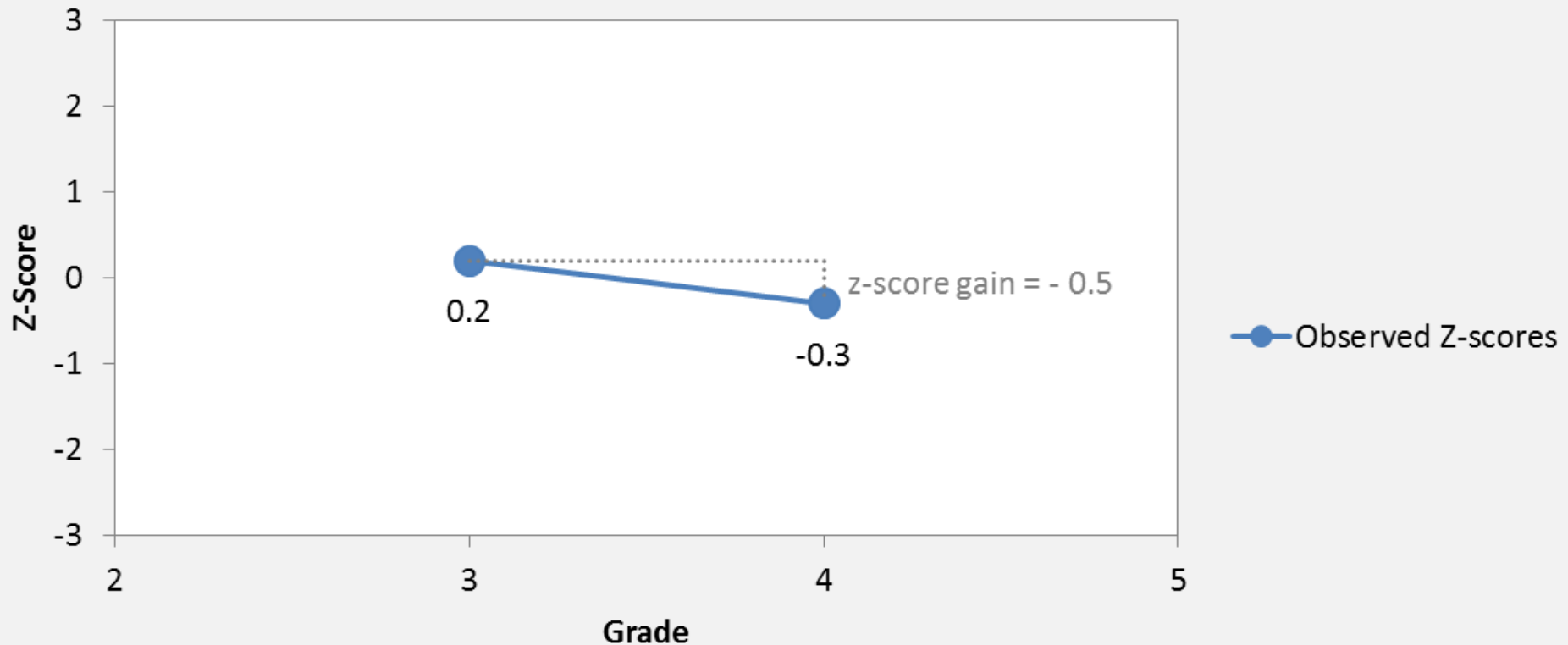
Baselined Z-score Gain

- Similar to a regular z-score gain, but the means and standard deviations used to create z-scores were taken from a baseline year.

- Baselining avoids a zero-sum game.

- Baselining introduces susceptibility to drift.

- Marc's observed scores in grade 3-6 are shown below.
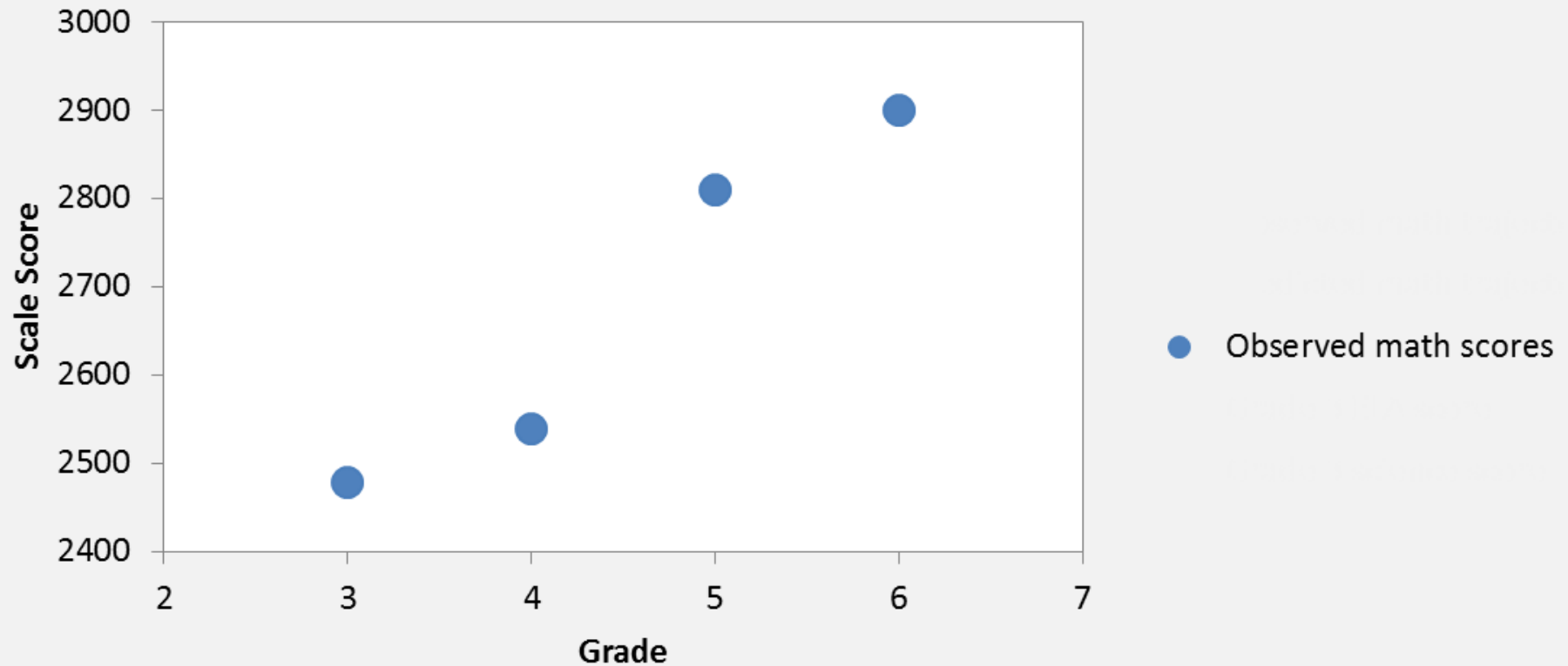
Growth Rate

- A linear regression through the points creates an observed trajectory.

- *If growth is linear,* the use of linear regression addresses measurement error.
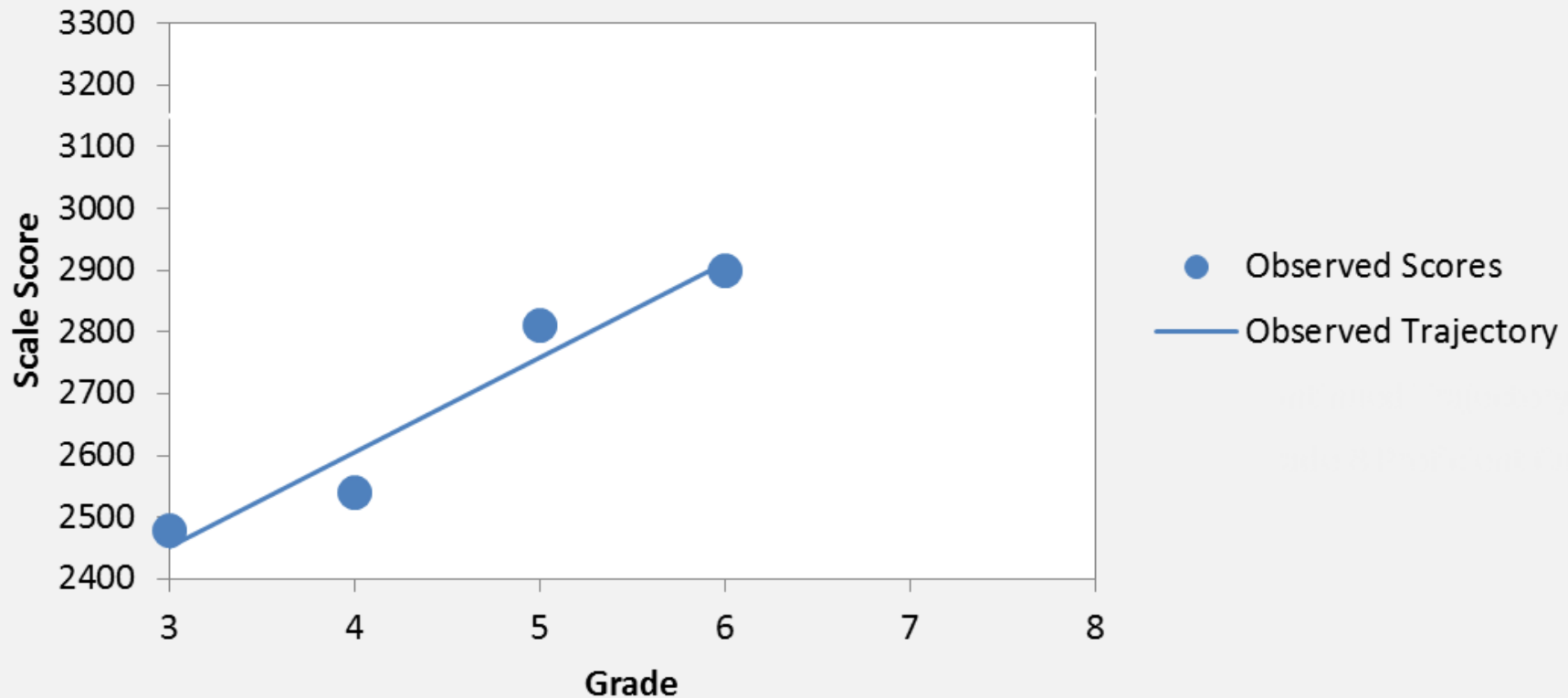
Growth Rate

- The slope of the trajectory is the estimated annual growth rate. In this case, Marc's annual growth rate is 153 points.
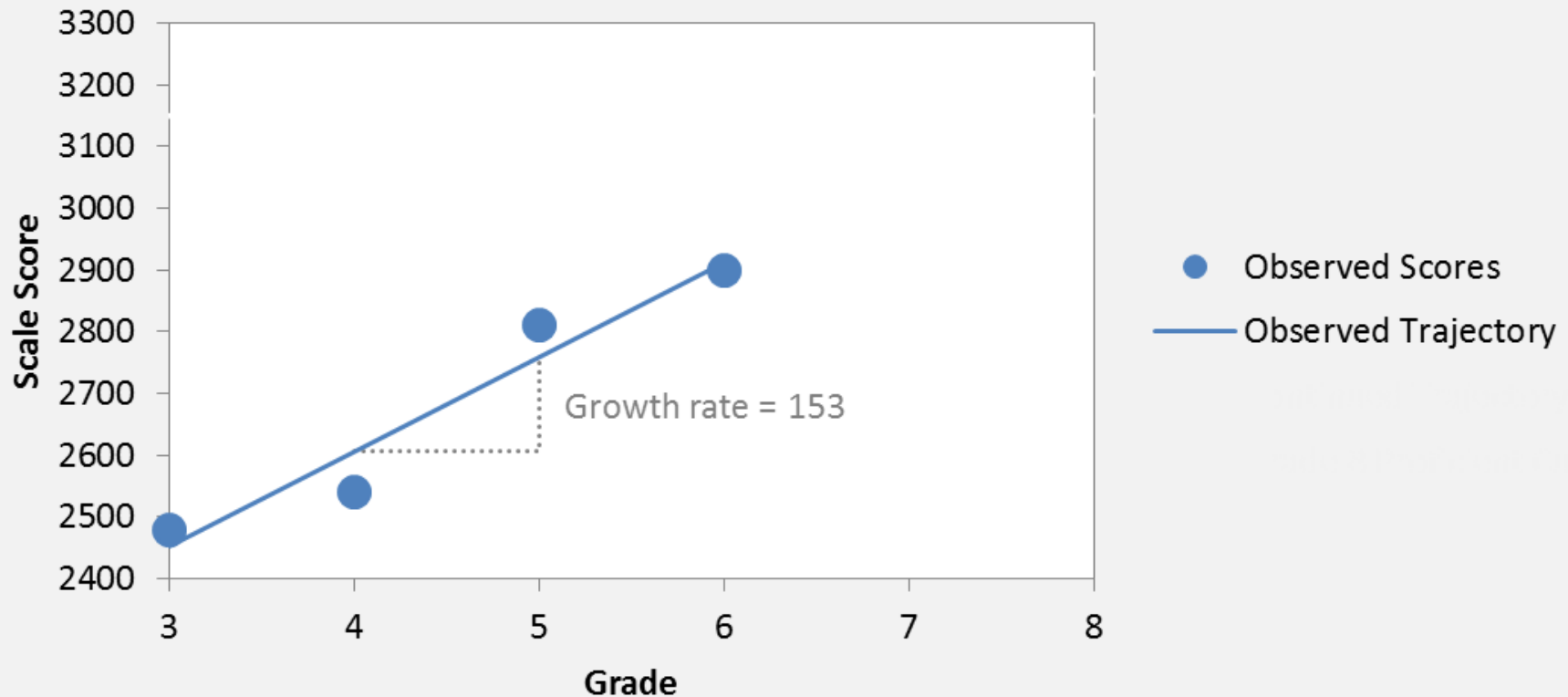
Growth Rate

- If desired, we can continue the trajectory into the future.
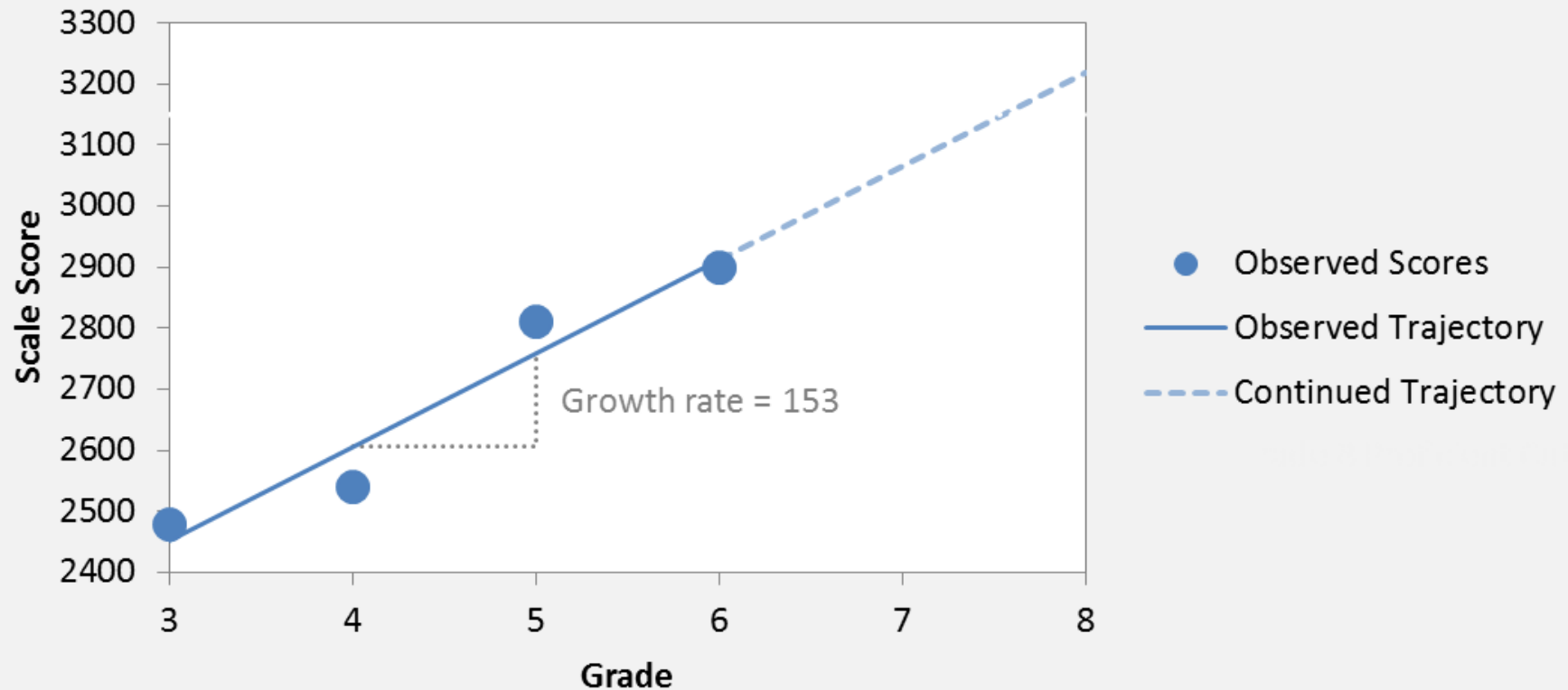
Growth Rate

- And see whether that trajectory would result in meeting a later grade's proficiency cut score.

- In this case, yes, since the two dotted lines cross.
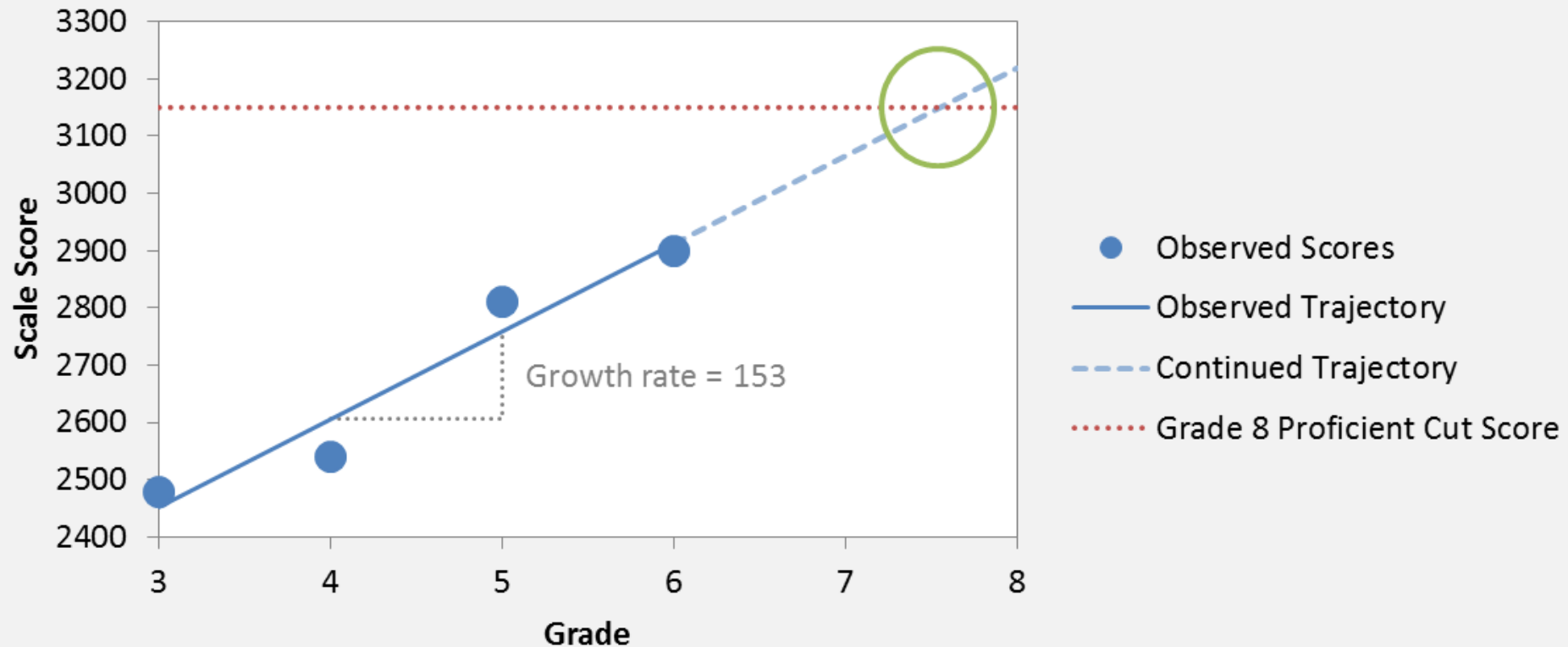
Student Growth Percentile

- We want to calculate Cara's SGP in grade 5.

- We first identify all of the students in the state that had the same scores as Cara in grades 3 and 4.
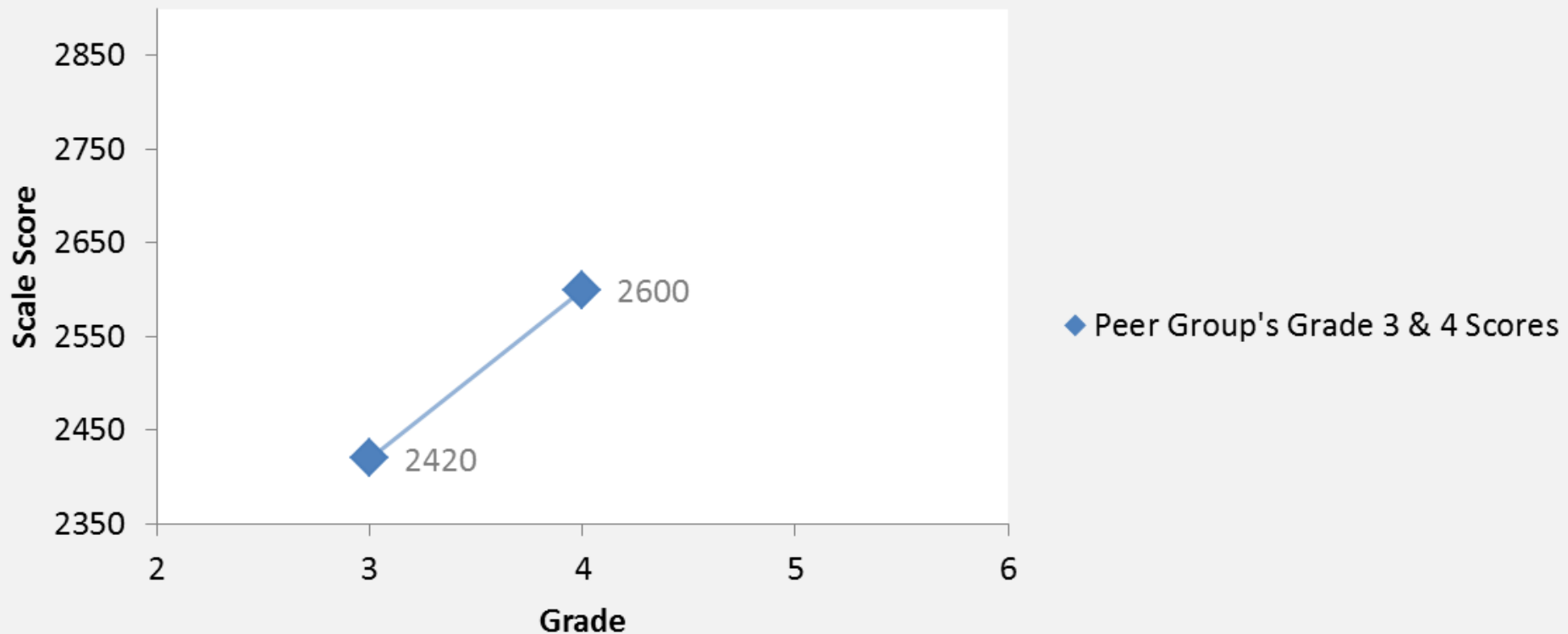
Student Growth Percentile

- There are 10 students with the same scores as Cara in grade 3 and 4 (including Cara). This is Cara's peer group.

- We then review the grade 5 scores of her peer group.
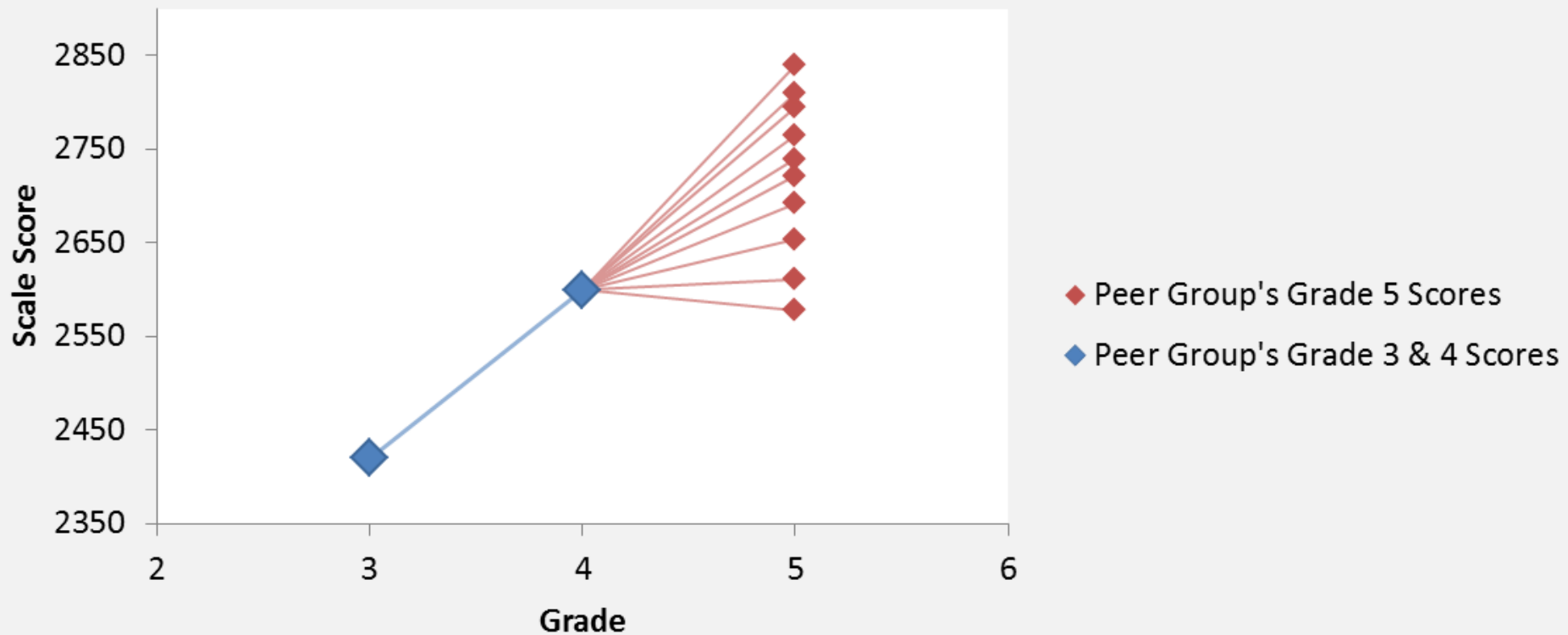
Student Growth Percentile

- We identify Cara's score in grade 5.

- Note that her grade 5 score is higher than six of the students in her peer group.
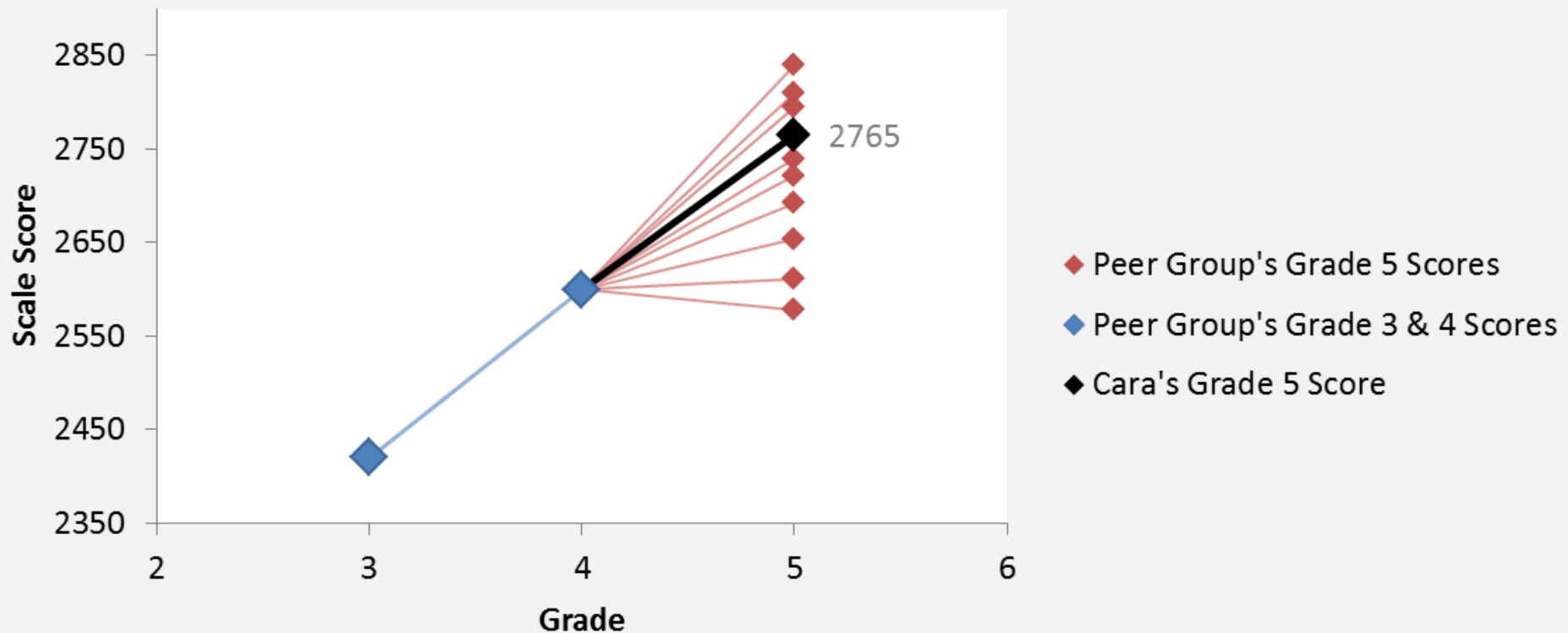
Student Growth Percentile

- Therefore, her SGP is 60.

- Computationally, it's more complicated than presented here, but this is the conceptual idea.
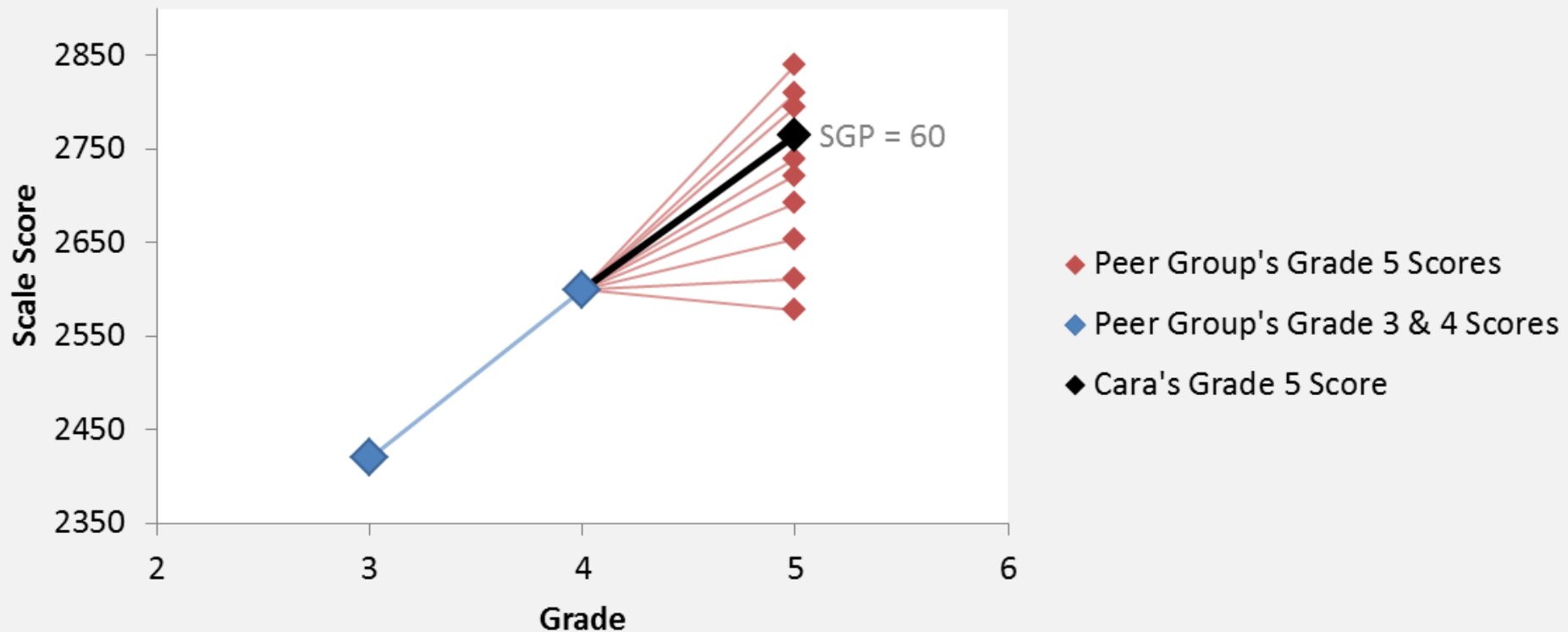
Baselined SGP

- Because SGPs always have one percent of students in each SGP from 0-99, if we set a target SGP, it is impossible for all students meet the target, even if the entire state improves (a zero-sum game).

- Baselined SGPs are similar to regular SGPs, but they are calculated *as if* the scores had occurred in the baseline year.

- If the entire state improves, baselining makes it is possible for all students to meet the target SGP.

- Baselining introduces susceptibility to drift

Transition

- Transition models track student's from one achievement level to another in the next grade.
- They may use the achievement levels (e.g., a student transitions from *Novice* to *Basic.*
- They may also use sub-categories in achievement levels (e.g., low, mid, and high).
- A table is used to track transitions.

Grade 3 Achievement Level		Grade 4 Achievement Level											
		Novice			Basic			Proficient			Advanced		
		L	M	H	L	M	H	L	M	H	L	M	H
Novice	L												
	M												
	H												
Basic	L												
	M												
	H												
Proficient	L												
	M												
	H												
Advanced	L												
	M												
	H												

Transition

- Raymond scored in the middle of the Basic achievement level in grade 3.
- He scored in the middle of the Proficient achievement level in grade 4.

Grade 3 Achievement Level		Grade 4 Achievement Level											
		Novice			Basic			Proficient			Advanced		
		L	M	H	L	M	H	L	M	H	L	M	H
Novice	L												
	M												
	H												
Basic	L												
	M								X				
	H												
Proficient	L												
	M												
	H												
Advanced	L												
	M												
	H												

Transition

- The transitions can be labeled descriptively. In this case, there are five labels:

SD   Significant Decline
D   Decline
M   Maintaining
I   Improvement
SI   Significant Improvement

Based on these labels, Raymond made a *Significant Improvement* (SI) in his achievement level from grade 3 to grade 4.

By comparison, a transition from *Low Proficient* to *Mid Basic* is a *Decline* (D).

Grade 3 Achievement Level		Grade 4 Achievement Level											
		Novice			Basic			Proficient			Advanced		
		L	M	H	L	M	H	L	M	H	L	M	H
Novice	L	M	I	I	SI	SI	SI	SI	SI	SI	SI	SI	SI
	M	D	M	I	I	SI	SI	SI	SI	SI	SI	SI	SI
	H	D	D	M	I	I	SI	SI	SI	SI	SI	SI	SI
Basic	L	SD	D	D	M	I	I	SI	SI	SI	SI	SI	SI
	M	SD	SD	D	D	M	I	I	SI	SI	SI	SI	SI
	H	SD	SD	SD	D	D	M	I	I	SI	SI	SI	SI
Proficient	L	SD	SD	SD	SD	D	D	M	I	I	SI	SI	SI
	M	SD	SD	SD	SD	SD	D	D	M	I	I	SI	SI
	H	SD	SD	SD	SD	SD	SD	D	D	M	I	I	SI
Advanced	L	SD	SD	SD	SD	SD	SD	SD	D	D	M	I	I
	M	SD	SD	SD	SD	SD	SD	SD	SD	D	D	M	I
	H	SD	SD	SD	SD	SD	SD	SD	SD	SD	D	D	M

Transition

- The transitions can be labeled with a numeric value to identify what type of transition the state values most highly.
- The values are created by a policymaking body, such as a standard setting panel, legislature, State Board, etcetera.

Based on these values, Raymond's transition from *Mid Basic* to *Mid Proficient* is worth 15 points.

By comparison, a transition from *Low Proficient* to *Mid Basic* is worth only 5 points.

Grade 3 Achievement Level		Grade 4 Achievement Level											
		Novice			Basic			Proficient			Advanced		
		L	M	H	L	M	H	L	M	H	L	M	H
Novice	L	3	5	10	15	20	26	29	30	30	31	31	32
	M	2	4	7	11	15	21	24	25	26	27	28	30
	H	1	3	5	8	12	17	20	21	22	24	26	28
Basic	L	0	3	6	7	10	13	16	18	20	22	24	26
	M	0	1	3	5	7	10	13	15	18	20	22	24
	H	0	1	2	4	6	8	11	14	16	18	20	22
Proficient	L	0	1	2	3	5	7	10	12	14	16	18	20
	M	0	1	2	3	4	6	9	10	12	14	16	18
	H	0	1	2	3	4	5	8	9	10	12	14	16
Advanced	L	0	1	2	3	4	5	7	8	9	10	12	14
	M	0	1	2	3	4	5	6	7	8	9	10	12
	H	0	1	2	3	4	5	6	6	7	8	9	12

- Start with a student's prior test scores
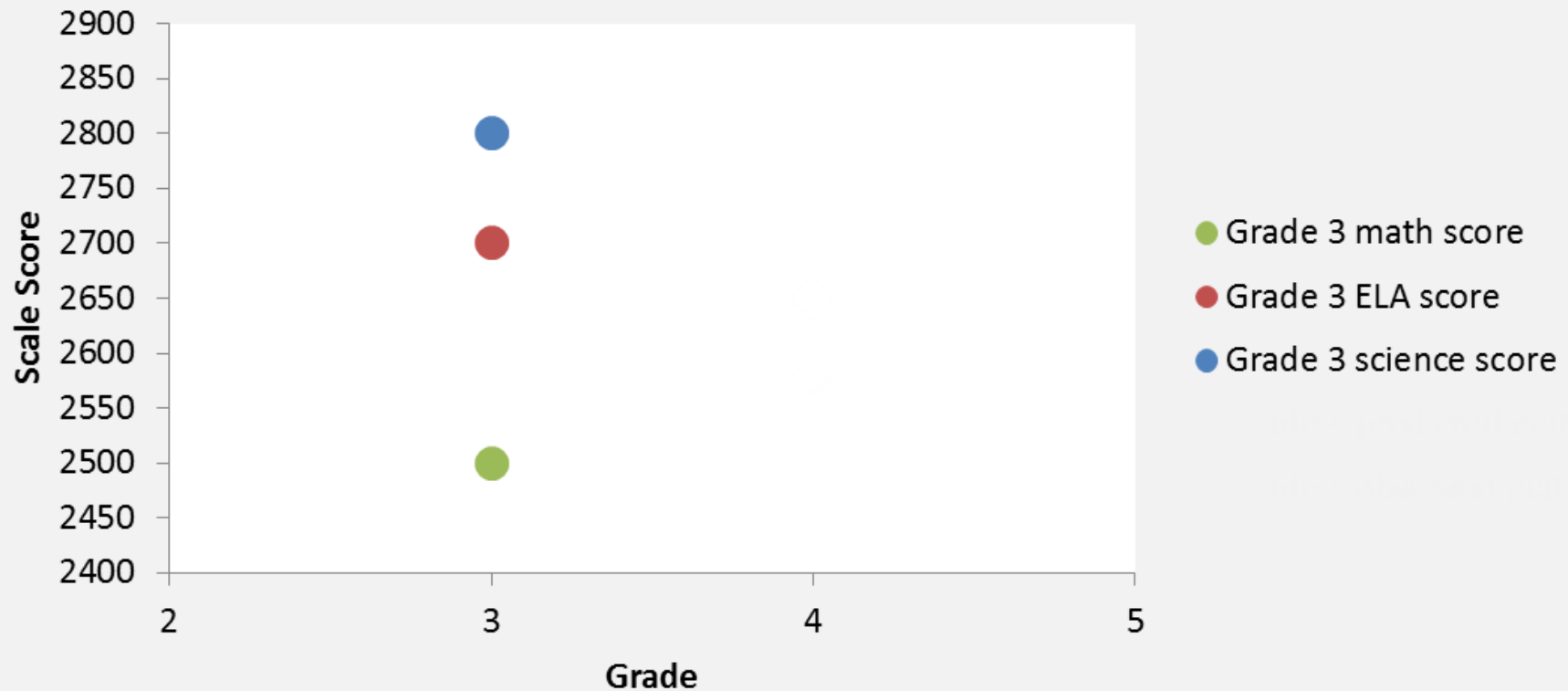
Status-based Value-added Model (VAM)

- Predict the student's current score using her prior test scores.
- The predicted score for a student is the score a student with the same set of past scores gets in a classroom/school of average "effectiveness".
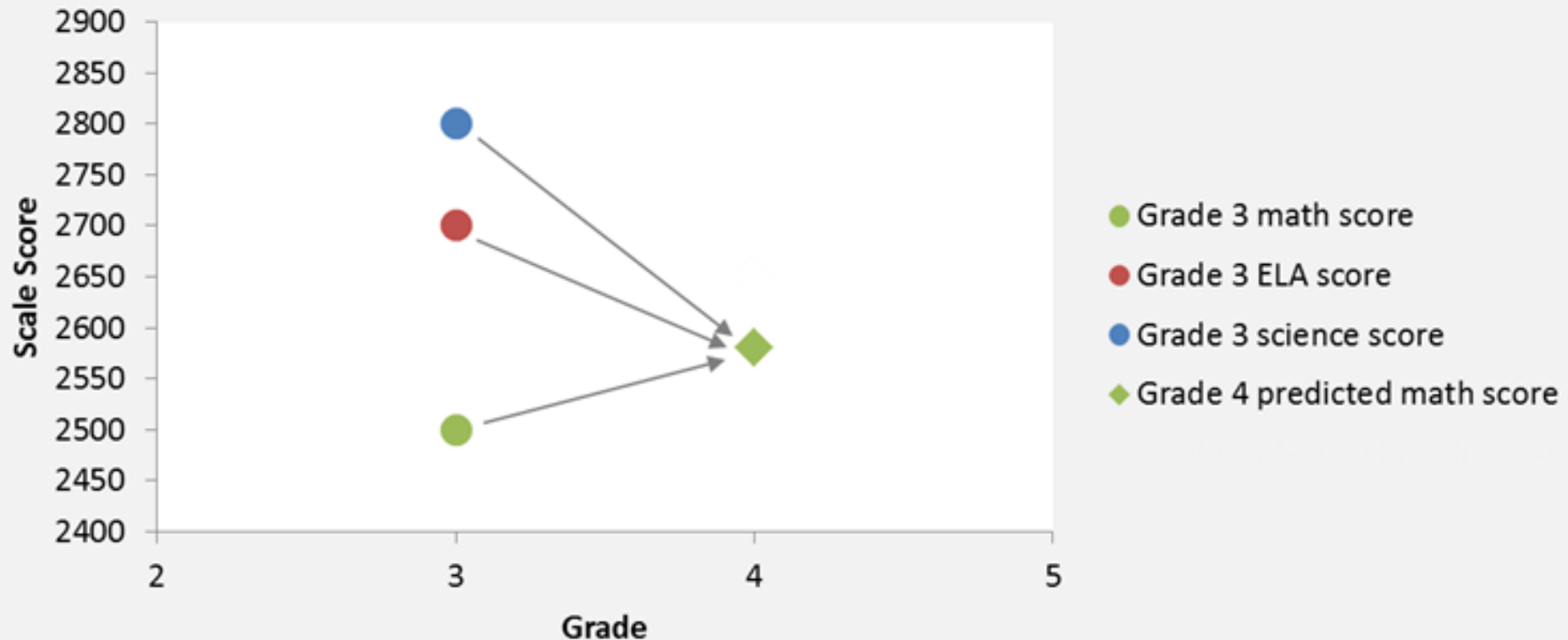
Status-based Value-added Model (VAM)

- The predicted and observed scores will be different to some degree.

- Compare the student's current observed score to his or her predicted score.
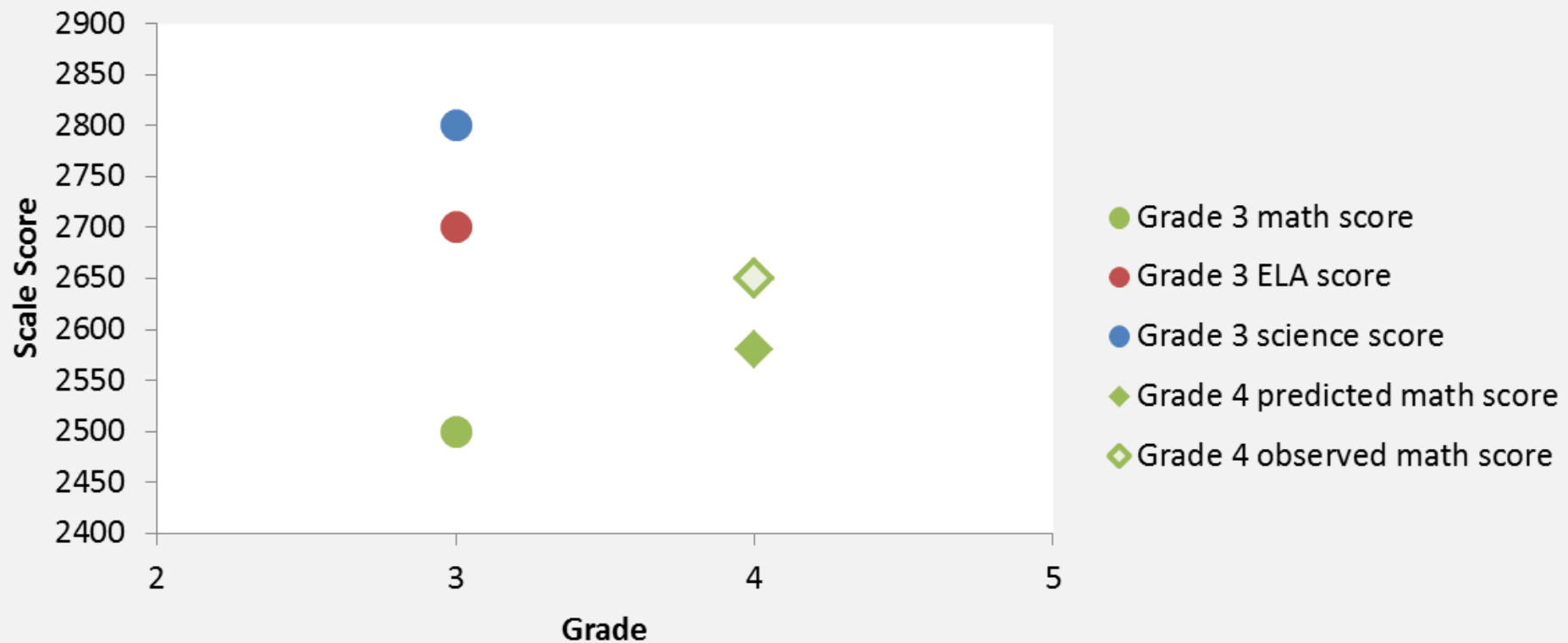
Status-based Value-added Model (VAM)

- The difference between the observed score and the predicted score is the "deviation" in the student's score.

- The average deviation in an aggregation unit (e.g., classroom, school, district) is the "value added" by that unit to its students' test scores.

Gain-based VAM

- Start with a student's prior test scores
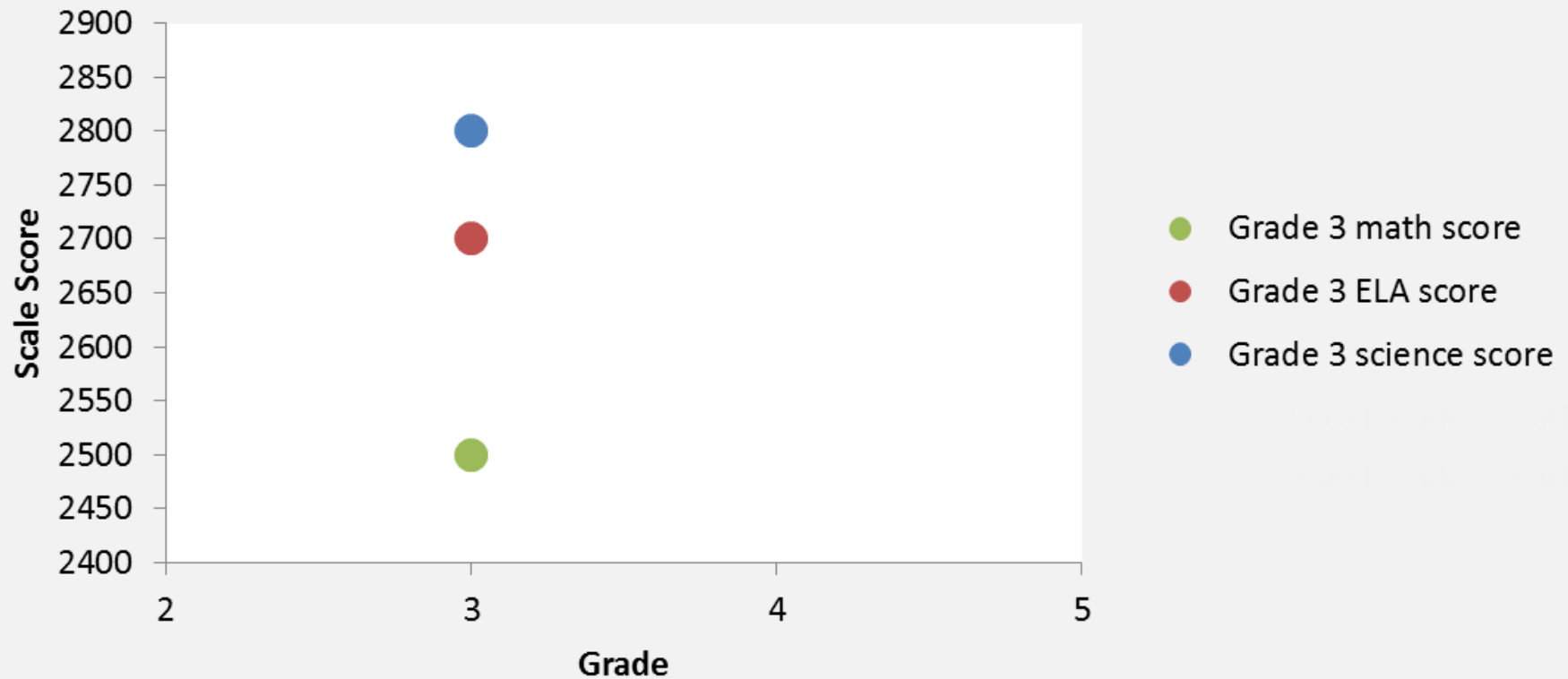
Gain-based VAM

- Predict the student's gain score using prior test scores
- The predicted gain for a student is the gain a student with the same set of past scores gets in a classroom/school of average "effectiveness".

Predicted gain = 80

Legend:
- Grade 3 math score
- Grade 3 ELA score
- Grade 3 science score
- Predicted grade 4 math gain

Gain-based VAM

- Calculate the student's observed gain, which is likely to be different than the predicted gain to some degree.
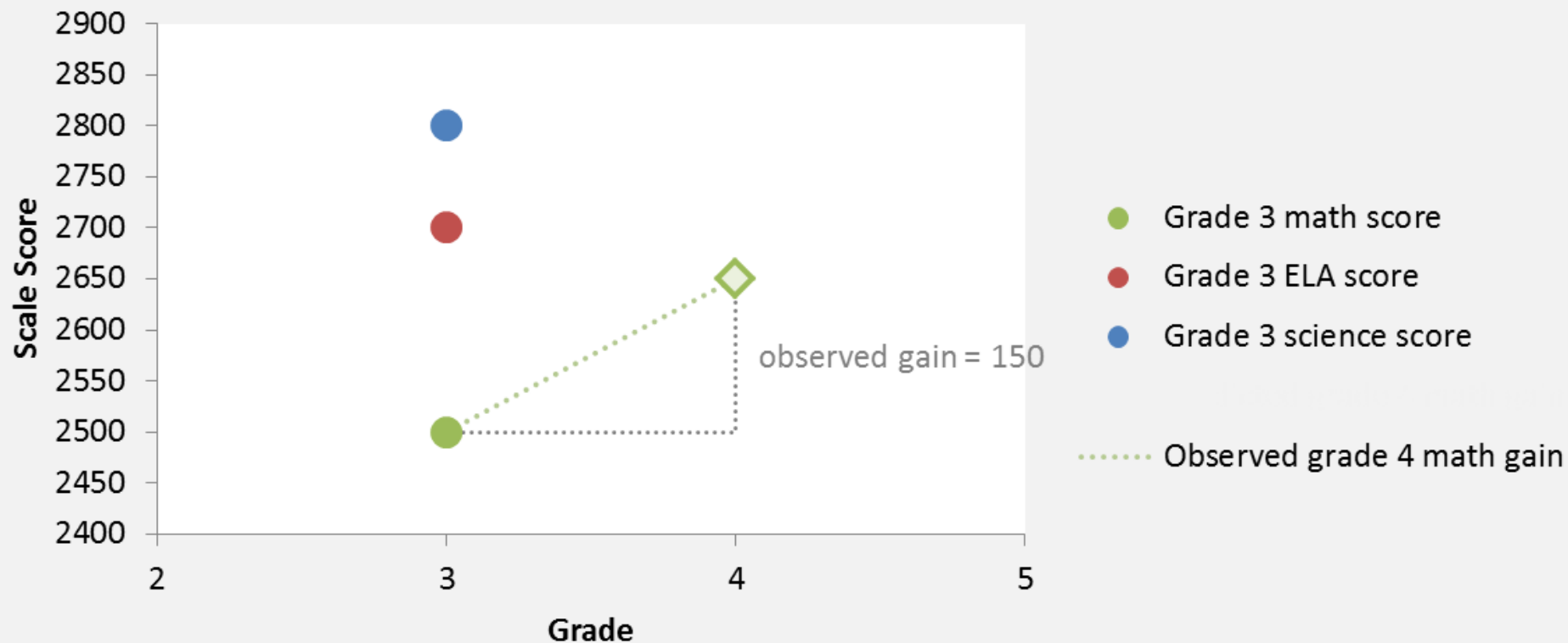
Gain-based VAM

- The difference between the observed gain and the predicted gain is the "deviation" in that student's gain score.

- The average deviation in an aggregation unit (e.g., classroom, school, district) is the "value added" by that unit to its students' gain scores.
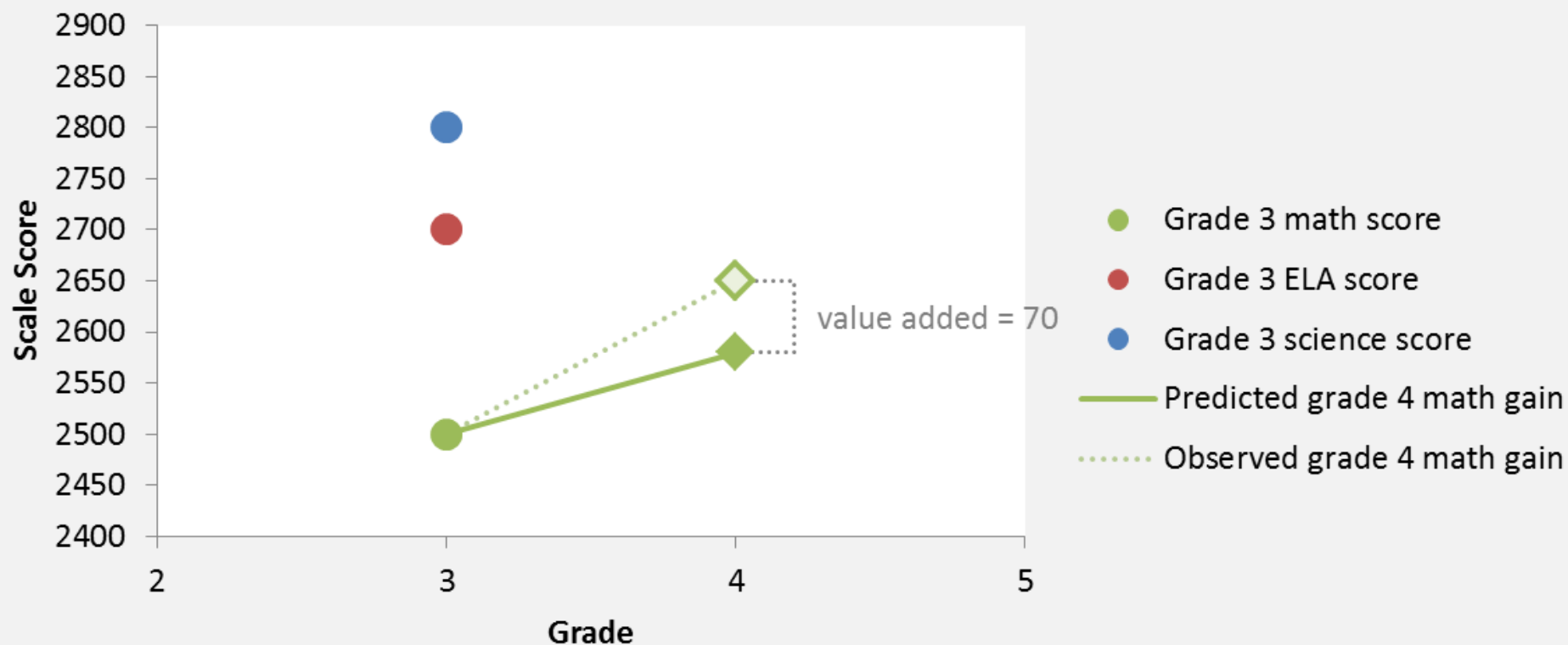
Growth-rate-based VAM

- Start with a student's observed scores in the subject area
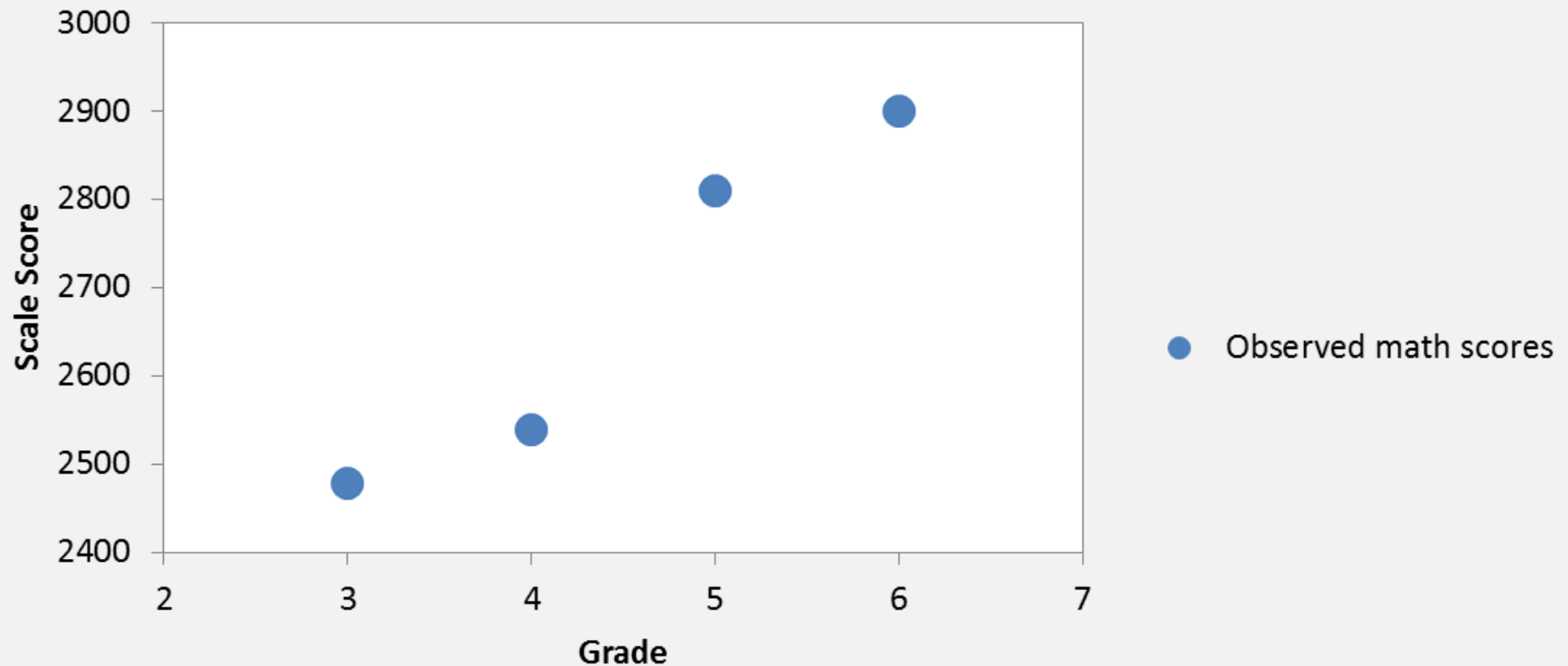
Growth-rate-based VAM

- Use linear regression to calculate a trajectory for the student.
- The use of a linear regression addresses measurement error to some degree.
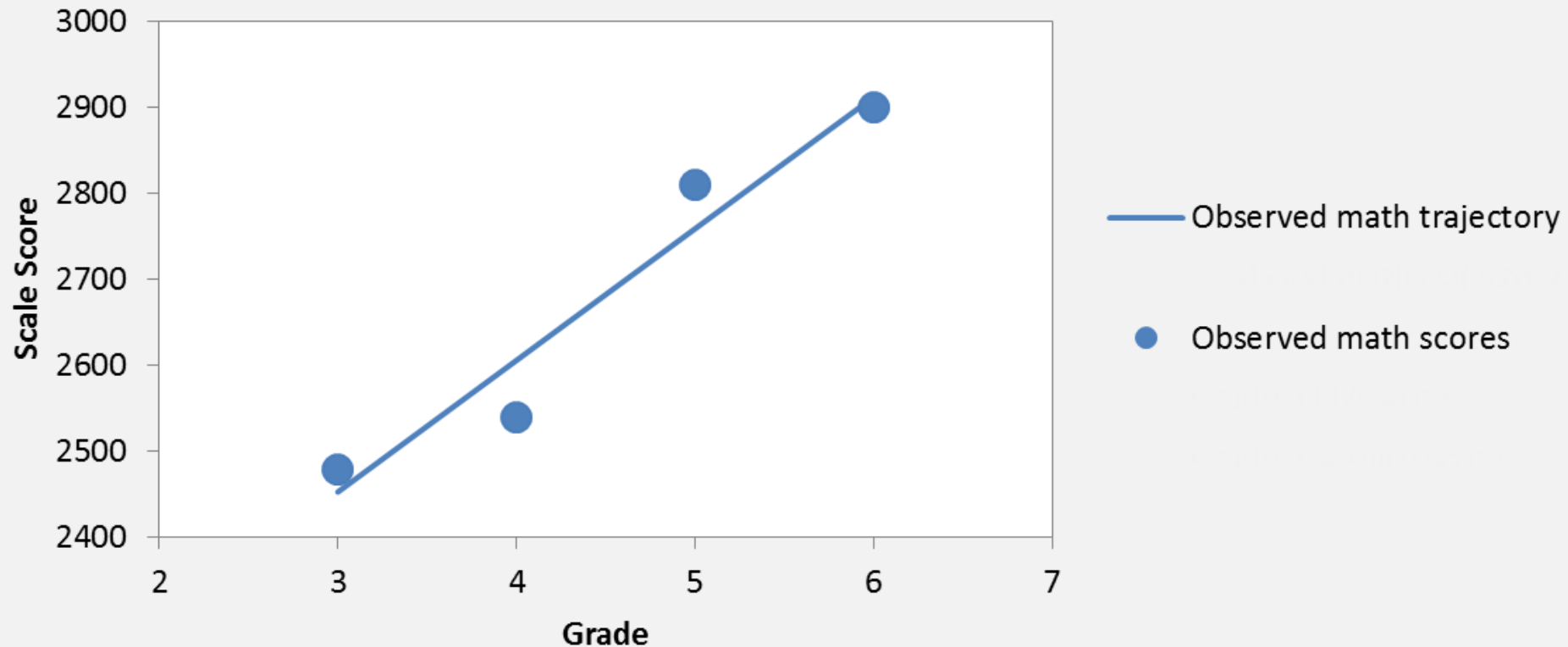
Growth-rate-based VAM

- The next step is to gather both the observed growth rate, the initial score in the subject area, and prior scores in other subject areas.
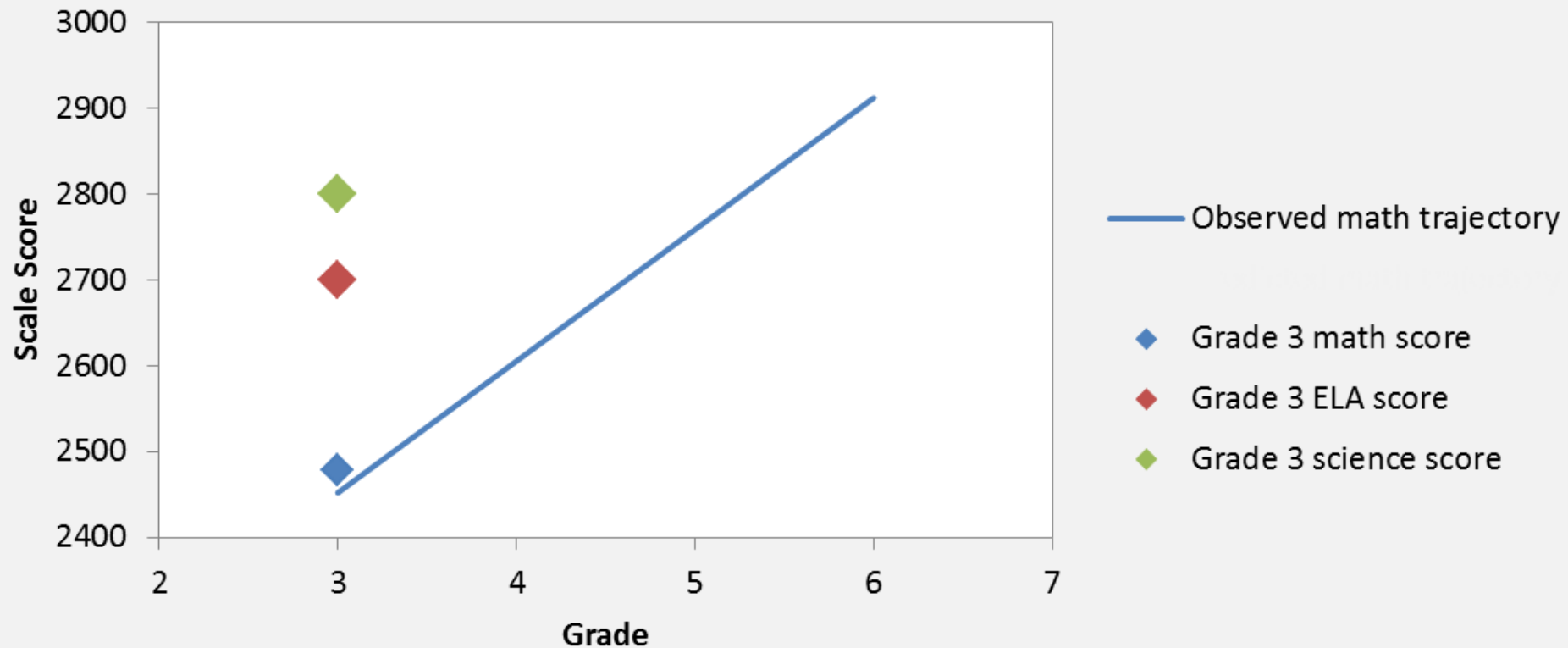
Growth-rate-based VAM

- Use the initial score in the subject area and prior scores in other subject areas to predict all students' trajectories.

- The predicted and observed trajectories are likely to be different to some degree.

- The predicted growth rate for a student is the growth rate expected for a student with the same prior scores in a school of average "effectiveness."
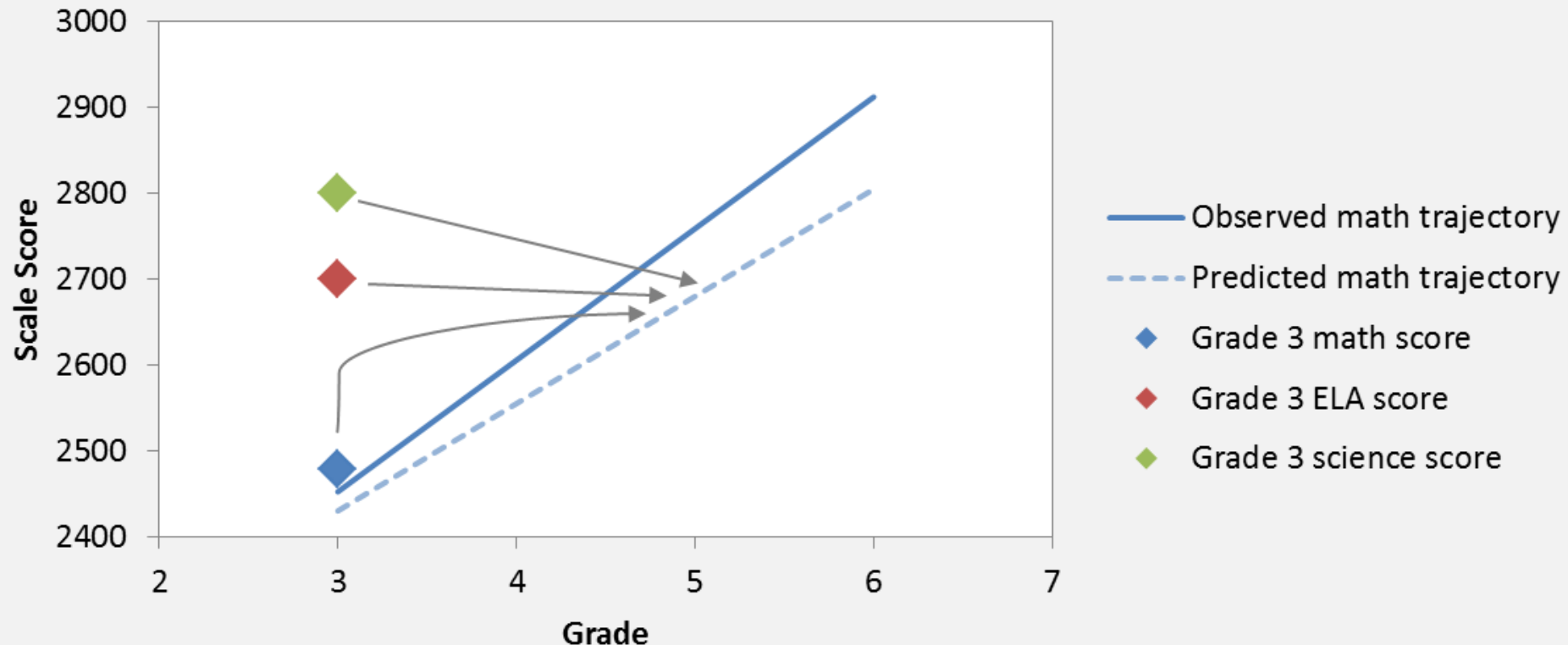
Growth-rate-based VAM

- The observed growth rate is the slope of the observed trajectory.

- The predicted growth rate is the slope of the predicted trajectory.
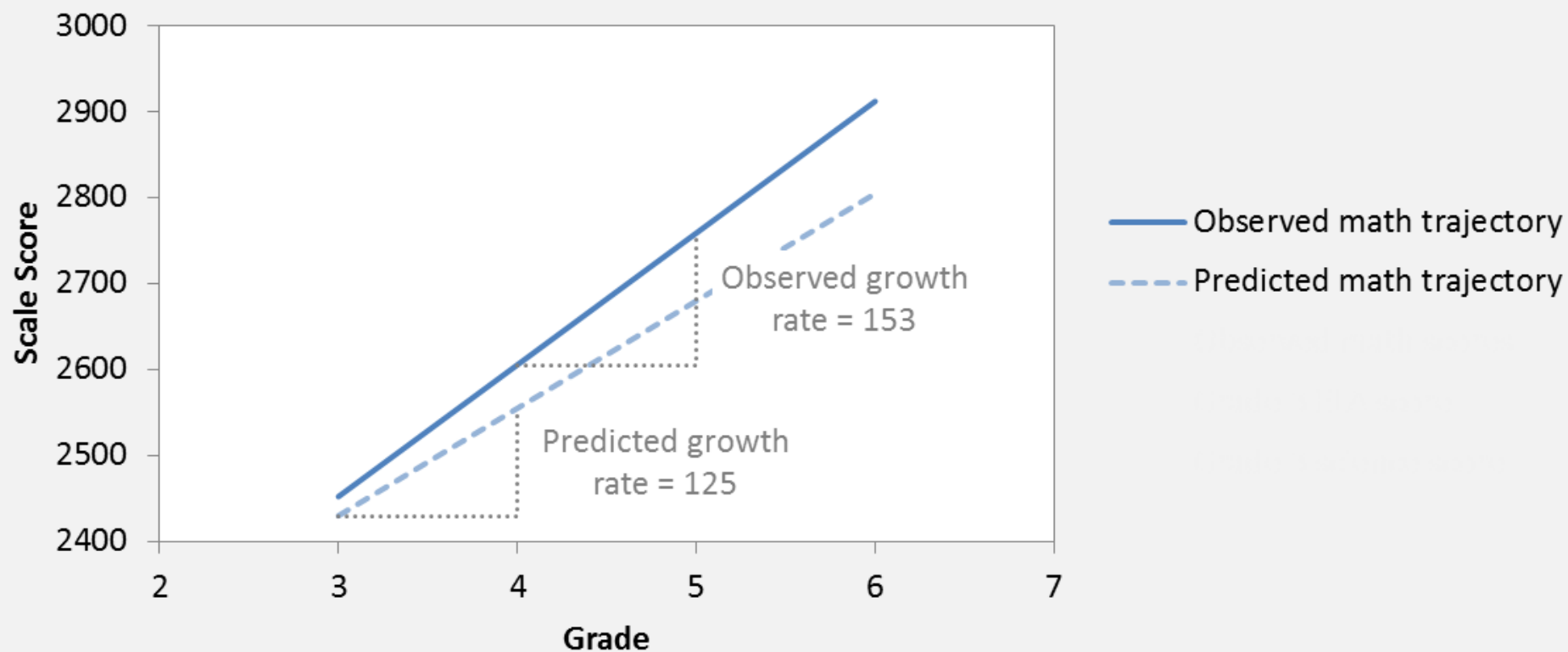
Growth-rate-based VAM

- The difference between the observed and predicted growth rates is the "deviation" in growth rate for the student.

- The average deviation in an aggregation unit (e.g., classroom, school, district) is the "value added" by the unit to its students' growth rates.
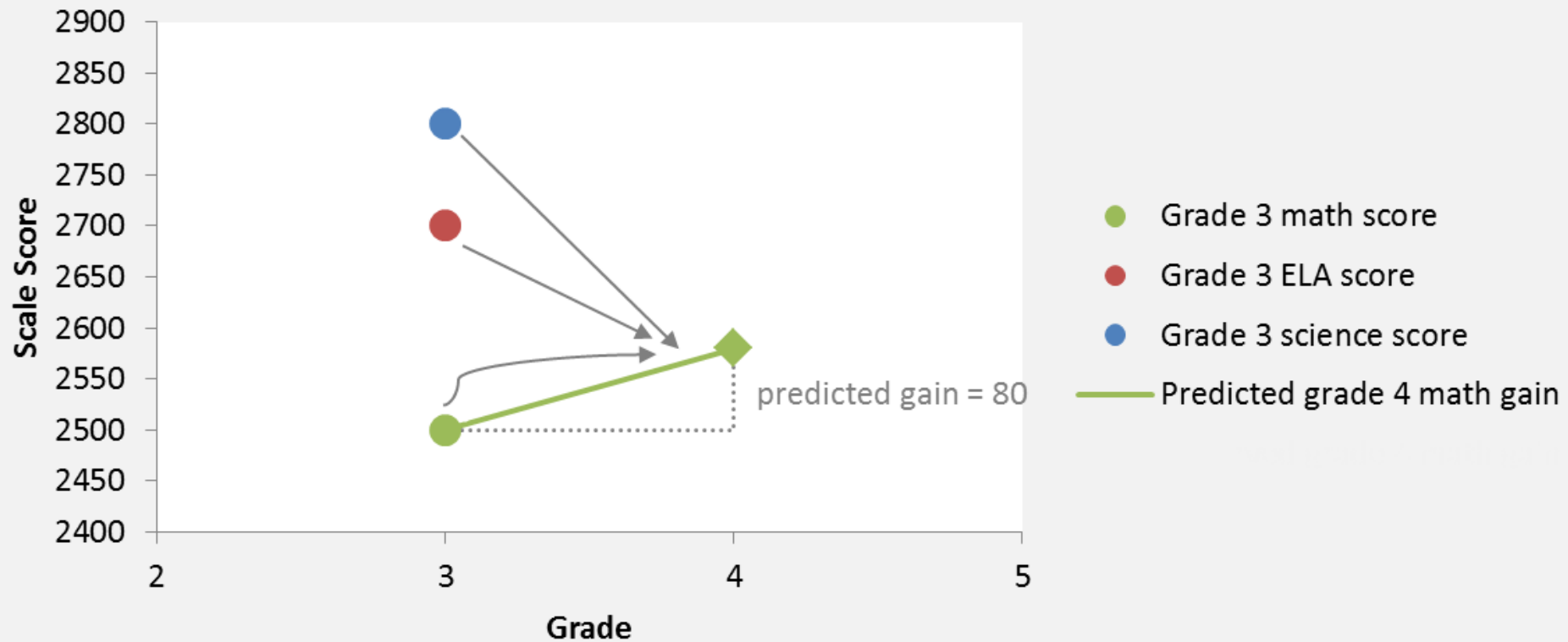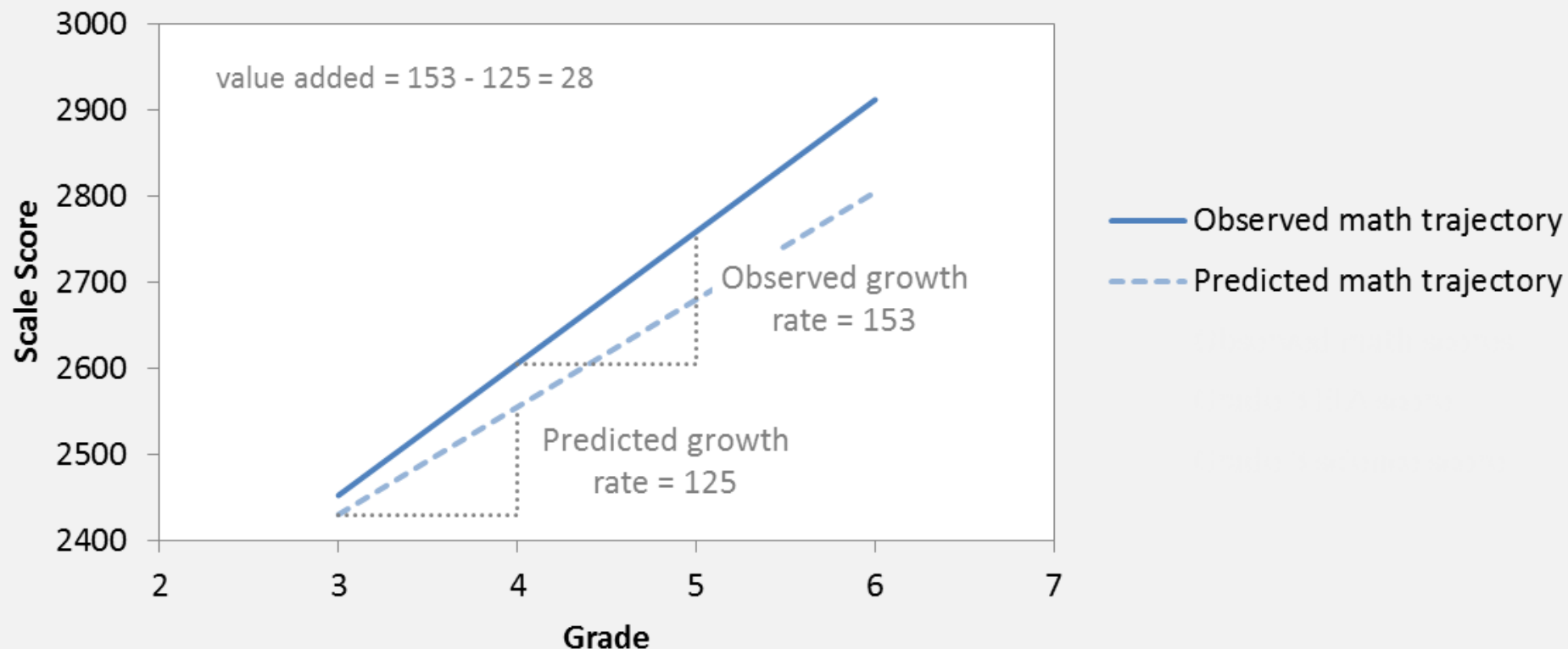
Existence of Potential Measures of Growth by Interpretation and Analytical Model

Interpretation			Analytical Model										
Focus	Description		Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced		Y	N	N	Y	N	N	Y	Y	Y	Y	N
	Norm-referenced		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
	Criterion-referenced judgment		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Future Scores	Criterion-referenced trajectory continuation		Y	N	N	Y	Y	Y	Y	N	N	N	N
	Criterion-referenced trajectory target		Y	N	N	Y	Y	Y	Y	N	N	N	N
	Criterion-referenced category projection		N	N	N	N	N	N	N	N	N	N	Y
	Criterion-referenced probability projection		N	N	N	N	N	N	N	N	N	N	Y

See white paper for explanation of why some cells do not have potential growth measures.

Important Characteristics to Consider

- **Most Important Characteristic to Consider**
  - Match of interpretation to intended use

- **Additional Important Characteristics to Consider**
  - Ability to Improve Stability by Conditioning on Other-subject Prior Test Scores
  - Degree of Match to Common Understanding of Calculating Growth
  - Whether the Measures Creates a Zero-Sum Game
  - Relative Degree of Difficulty in Aggregating Across Subjects and Scales
  - Relative Degree of Complexity of Growth Score Calculations
  - Relative Degree of Difficulty in Communication
  - Relative Degree of Correlation with Status Scores
  - Relative Degree of Unreliability/Instability of Individual Student Growth Scores
  - Relative Degree of Susceptibility to Equating/Scale Drift
  - Number of Years of Data Required from the Same Assessment to Create Growth Measures
  - Whether individual growth measures are produced (in addition to aggregate growth measures)

Ability to Improve Stability by Conditioning on Other-subject Prior Test Scores

Interpretation			Analytical Model										
Focus		Description	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores		Scale-referenced	N	-	-	N	-	-	N	Y	Y	Y	-
		Norm-referenced	N	N	N	N	Y	Y	N	Y	Y	Y	-
		Criterion-referenced judgment	N	N	N	N	Y	Y	N	Y	Y	Y	-
Future Scores		Criterion-referenced trajectory continuation	N	-	-	N	Y	Y	N	-	-	-	-
		Criterion-referenced trajectory target	N	-	-	N	Y	Y	N	-	-	-	-
		Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	Y
		Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	Y

- Improves precision and stability of growth measures, which can improve credibility.

- See white paper for explanation of ratings.

Degree of Mismatch of Growth Measure to Common Understanding of Calculating Growth

Interpretation			Analytical Model										
Focus	Description		Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced		L	-	-	L	-	-	M	H	M	M	-
Observed Scores	Norm-referenced		L	M	M	L	H	H	M	H	M	M	-
Observed Scores	Criterion-referenced judgment		L	M	M	L	H	H	M	H	M	M	-
Future Scores	Criterion-referenced trajectory continuation		L	-	-	L	H	H	M	-	-	-	-
Future Scores	Criterion-referenced trajectory target		L	-	-	L	H	H	M	-	-	-	-
Future Scores	Criterion-referenced category projection		-	-	-	-	-	-	-	-	-	-	H
Future Scores	Criterion-referenced probability projection		-	-	-	-	-	-	-	-	-	-	H

- "Traditional" growth measures may have more credibility with policymakers, though they may be less valid as measures of growth.

- See white paper for explanation of ratings.

Whether the Measures Creates a Zero-Sum Game

Interpretation			Analytical Model										
			Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Focus		Description											
Observed Scores		Scale-referenced	N	-	-	N	-	-	N	Y	Y	Y	-
		Norm-referenced	N	Y	N	N	Y	N	N	Y	Y	Y	-
		Criterion-referenced judgment	N	Y	N	N	Y	N	N	Y	Y	Y	-
Future Scores		Criterion-referenced trajectory continuation	N	-	-	N	Y	N	N	-	-	-	-
		Criterion-referenced trajectory target	N	-	-	N	Y	N	N	-	-	-	-
		Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	N
		Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	N

- A zero-sum game occurs when there will always be students with undesirable growth, even if growth improved dramatically across the entire state. This is a concern for some educators.

- See white paper for explanation of ratings.

Relative Degree of Difficulty in Aggregating Across Subjects and Scales

Interpretation			Analytical Model										
---	---	---	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Focus		Description											
Observed Scores		Scale-referenced	H	-	-	H	-	-	M	H	H	H	-
		Norm-referenced	L	L	L	L	L	L	L	L	L	L	-
		Criterion-referenced judgment	L	L	L	L	L	L	L	L	L	L	-
Future Scores		Criterion-referenced trajectory continuation	M	-	-	M	M	M	M	-	-	-	-
		Criterion-referenced trajectory target	M	-	-	M	M	M	M	-	-	-	-
		Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	M
		Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	M

- Useful for comparing growth across subject areas, for aggregating growth measures across subject areas in accountability, and for using growth from one test to another in accountability.

- See white paper for explanation of ratings.

Interpretation			Analytical Model										
Focus	Description	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced	L	-	-	M	-	-	M	H	H	H	-
Observed Scores	Norm-referenced	M	M	H	M	H	H	M	H	H	H	-
Observed Scores	Criterion-referenced judgment	L	M	H	M	H	H	M	H	H	H	-
Future Scores	Criterion-referenced trajectory continuation	M	-	-	M	H	H	M	-	-	-	-
Future Scores	Criterion-referenced trajectory target	M	-	-	M	H	H	M	-	-	-	-
Future Scores	Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	H
Future Scores	Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	H

- Simplicity of calculation can improve the acceptance of the growth measure (though it does increase the ability to ask questions that may be misinformed based on a naïve understanding).

- See white paper for explanation of ratings.

Interpretation		Analytical Model										
Focus	Description	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced	H	-	-	H	-	-	H	H	H	H	-
	Norm-referenced	L	H	H	L	L	M	H	H	H	H	-
	Criterion-referenced judgment	L	H	H	L	L	M	H	H	H	H	-
Future Scores	Criterion-referenced trajectory continuation	M	-	-	M	M	M	M	-	-	-	-
	Criterion-referenced trajectory target	M	-	-	M	M	M	M	-	-	-	-
	Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	M
	Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	H

- The ease of communicating clearly about growth measures determines the degree of effort needed to help stakeholders understand the measures.

- See white paper for explanation of ratings.

Relative Degree of Correlation with Status Scores

Interpretation			Analytical Model										
Focus	Description	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced	M	-	-	L	-	-	M	L	M	L	-
	Norm-referenced	M	M	M	L	L	L	M	L	M	L	-
	Criterion-referenced judgment	M	M	M	M	M	M	M	M	M	M	-
Future Scores	Criterion-referenced trajectory continuation	H	-	-	H	H	H	H	-	-	-	-
	Criterion-referenced trajectory target	H	-	-	H	H	H	H	-	-	-	-
	Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	H
	Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	H

- The more strongly correlated a growth measure is with status scores, the less new information is provided by measuring growth.

- See white paper for explanation of ratings.

Interpretation			Analytical Model										
Focus	Description		Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced		L	-	-	M	-	-	L	-	-	-	-
	Norm-referenced		L	L	L	M	H	M	L	-	-	-	-
	Criterion-referenced judgment		L	L	L	M	H	M	L	-	-	-	-
Future Scores	Criterion-referenced trajectory continuation		L	-	-	L	M	L	L	-	-	-	-
	Criterion-referenced trajectory target		L	-	-	L	M	L	L	-	-	-	-
	Criterion-referenced category projection		-	-	-	-	-	-	-	-	-	-	L
	Criterion-referenced probability projection		-	-	-	-	-	-	-	-	-	-	M

- The less stable growth scores are, the more difficult it will be for the growth measure to achieve credibility.

- See white paper for explanation of ratings.

Interpretation			Analytical Model										
Focus	Description		Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced		M	-	-	M	-	-	M	L	M	M	-
	Norm-referenced		M	M	H	M	L	H	M	L	M	M	-
	Criterion-referenced judgment		M	M	H	M	L	H	M	L	M	M	-
Future Scores	Criterion-referenced trajectory continuation		H	-	-	H	H	H	H	-	-	-	-
	Criterion-referenced trajectory target		H	-	-	H	H	H	H	-	-	-	-
	Criterion-referenced category projection		-	-	-	-	-	-	-	-	-	-	H
	Criterion-referenced probability projection		-	-	-	-	-	-	-	-	-	-	H

- Increasing the number of years of observed data used for calculating measures of growth can increase the stability of growth measures, but that has to be balanced with increasing susceptibility to scale drift which can decrease stability of growth measures.

- See white paper for explanation of ratings.

Interpretation			Analytical Model										
Focus	Description		Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores	Scale-referenced		2	-	-	4	-	-	2	1	2	4	-
	Norm-referenced		2	2	3	4	1	3	2	1	2	4	-
	Criterion-referenced judgment		2	2	3	4	1	3	2	1	2	4	-
Future Scores	Criterion-referenced trajectory continuation		2	-	-	4	2	3	2	-	-	-	-
	Criterion-referenced trajectory target		2	-	-	4	2	3	2	-	-	-	-
	Criterion-referenced category projection		-	-	-	-	-	-	-	-	-	-	2
	Criterion-referenced probability projection		-	-	-	-	-	-	-	-	-	-	2

- It may be necessary for a state to calculate growth scores in the early years of implementing a new assessment (sometimes in the first year, usually by the second year).

- See white paper for explanation of ratings.

Interpretation			Analytical Model										
Focus		Description	Gain Score	Z-score Gain	Baselined Z	Growth Rate	SGP	Baselined SGP	Transition	Status VAM	Gain VAM	Rate VAM	Prediction
Observed Scores		Scale-referenced	Y	-	-	Y	-	-	Y	N	N	N	-
		Norm-referenced	Y	Y	Y	Y	Y	Y	Y	N	N	N	-
		Criterion-referenced judgment	Y	Y	Y	Y	Y	Y	Y	N	N	N	-
Future Scores		Criterion-referenced trajectory continuation	Y	-	-	Y	Y	Y	Y	-	-	-	-
		Criterion-referenced trajectory target	Y	-	-	Y	Y	Y	Y	-	-	-	-
		Criterion-referenced category projection	-	-	-	-	-	-	-	-	-	-	Y
		Criterion-referenced probability projection	-	-	-	-	-	-	-	-	-	-	Y

- If a state does not produce individual student growth measures, stakeholders may question why annual assessment is performed.

- See white paper for explanation of ratings.

- All of these are addressed implicitly in the previous slides, but deserve explicit mention:
  - Fairness toward students vs. Fairness toward educators
  - Simplicity vs. Validity
  - Fairness toward students vs. New information in growth scores
  - Instability from Measurement/Sampling Error vs. Instability from Scale Drift

- Holding educators/institutions accountable for achievement scores
  - Fully privileges fairness toward students over fairness toward educators by implying that achievement gaps must be closed
- Holding educators/institutions accountable for unadjusted growth scores
  - Mostly privileges fairness toward educators over fairness toward students by implying that achievement gaps are acceptable if they do not increase.
- Holding educators/institutions accountable for demographic-adjusted growth rates
  - Fully privileges fairness toward educators over fairness toward students by implying that it is acceptable if achievement gaps increase.
- Future-focused (growth-to-standard) interpretations
  - Attempt to achieve a balanced by holding the same ultimate achievement expectations for students, but giving some time to get there.

Balancing Simplicity with Validity

- Some interpretations of growth are more complex than others, and some analytical models are more complex than others.

- Though it is not always true that a simple interpretation or analytical model is less valid for an intended use, this is often the case.

  - More complex interpretations and analytical models tend to be more capable of supporting important nuances of intended uses.

  - States will need to evaluate the degree to which the validity cost of simplicity of a growth interpretation or a growth measure is outweighed by the benefit of being simple.

Balancing Fairness toward Students with Gaining New Information from Growth Measures

- Achieving some level of fairness toward students requires the use of a future-focused interpretation of growth.

- Future-focused (growth-to-standard) interpretations of growth tend to be more correlated with status scores than other measures of growth (with a shorter number of years to get to the standard making the correlations higher).

- The higher the correlation, the less new information there is in a growth measure not already available in status measures.

  – Allowing a longer window to get to the standard may improve the amount of new information available in growth scores.

  – States will need to determine how long is reasonable to allow, and whether it is reasonable to provide a different time limit for students in elementary school versus students in high school.

Balancing Instability from Measurement/Sampling Error vs. Instability from Scale Drift.

- Possible guard against growth scores being unstable because of measurement and sampling error by
  - Using more years of data to calculate growth scores
  - Averaging across multiple years of growth scores.
- However, the more years that are involved in growth calculations, the more susceptible the measures will be to scale drift, introducing instability into growth measures.
  - Drift across a small number of years is likely to be relatively small
  - Small degrees of drift can accumulate over time to become important
- States will need to consider how much instability comes from using too few years of data vs. instability that comes from too many years of data.
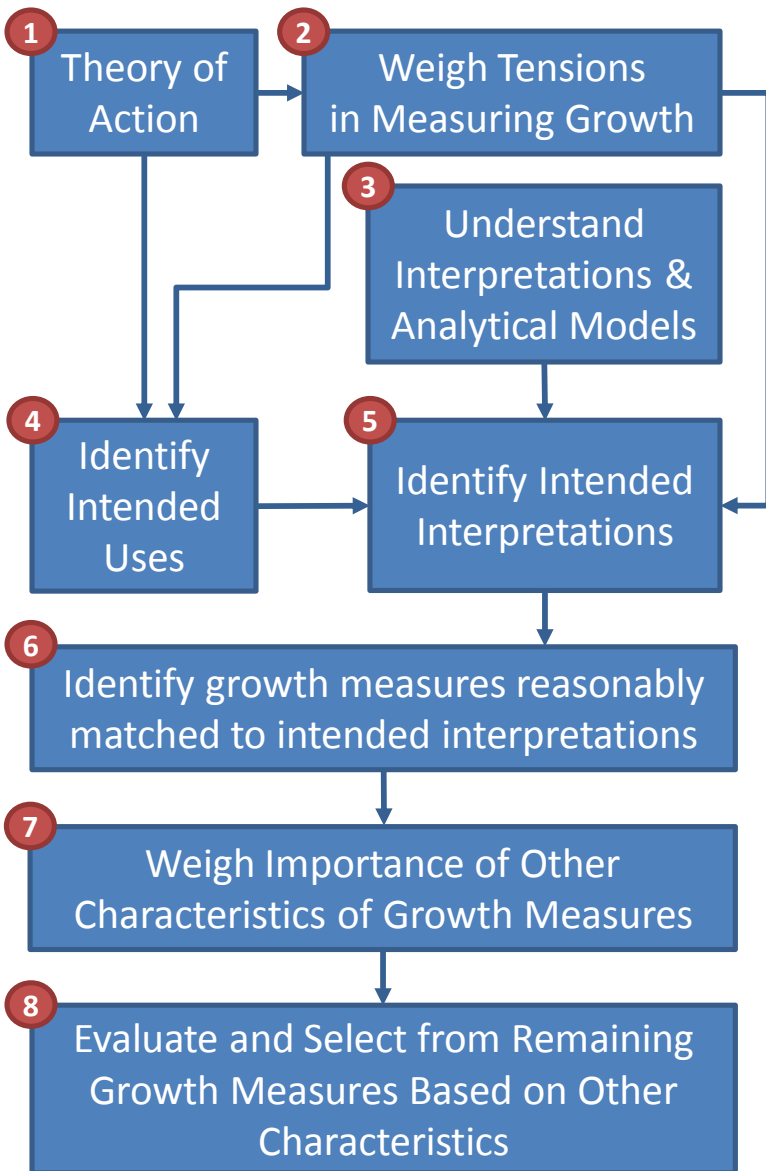
Deciding on One or More Growth Measures

- Decisions need to be made for each intended use.
- Two potential approaches defined here, though there are more.
  - Ordered decision making
    - First decision: does the interpretation match the intended use
    - Subsequent decisions: order the remaining characteristics by order of importance to the state
    - For each characteristic, identify the growth measures that meet your needs using the tables in previous slides
    - <u>A large number of potential growth measures will be eliminated once you have made two or three decisions about intended uses and desired characteristics</u>
  - A tool for rating the various growth measures
    - The ordered decision making process may result in no options. The tool provided scores each potential growth measure on the match to the interpretation desired and the importance of the remaining characteristics.
    - <u>A large number of potential growth measures will be eliminated as possibilities once you have entered your ratings into the tool.</u>

- Regardless of the approach used, no single growth model is going to meet all of your needs perfectly.
  - You will need to discuss what compromises you are willing to make.
  - You will need to identify which growth measures most closely match your intended uses and desired characteristics.

- You may need to report more than one measure of growth, for example:
  - Individual student vs. aggregate
  - Reporting vs. accountability
  - Ordered decision-making and/or the tool described in later slides may help you identify a single growth measure that is good enough for all your intended uses (or a small number of growth measures that satisfy your needs).

Process of Selecting One or More Growth Measures

1 Theory of Action

2 Weigh Tensions in Measuring Growth

3 Understand Interpretations & Analytical Models

4 Identify Intended Uses

5 Identify Intended Interpretations

6 Identify growth measures reasonably matched to intended interpretations

7 Weigh Importance of Other Characteristics of Growth Measures

8 Evaluate and Select from Remaining Growth Measures Based on Other Characteristics

1. Develop/revisit/review theory of action for what implementing a growth measure is intended to do and how it is intended to do it.
2. Discuss and weigh the tensions inherent in measuring growth to determine state priorities.
3. Understand at a basic level the types of interpretations of growth measures and the analytical models used to calculate them as background for the next steps.
4. Identify intended uses of growth measures from theory of action and priorities identified in weighing tensions.
5. Identify intended interpretations that serve the state's theory of action, intended uses, and priorities.
6. *For each intended interpretation*, identify which growth measures are reasonably matched to the states intended interpretations.
7. Discuss and weigh the importance of other characteristics of growth measures.
8. From the growth measures reasonably matched to each intended interpretations, evaluate their suitability based on their other characteristics and the importance of those characteristics.

Joseph Martineau
Senior Associate

jmartineau@nciea.org

Center for Assessment