Measuring Student Growth: So Much More than Subtracting Two Numbers!

Assessment Learning Network

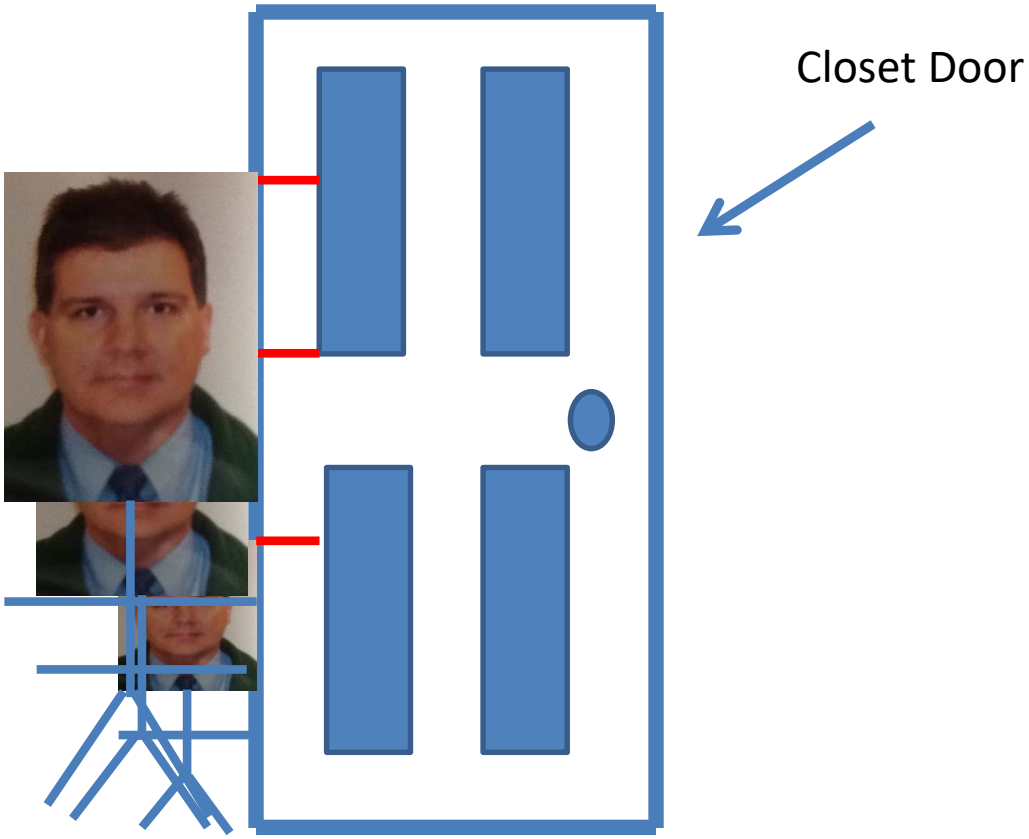
Friday, March 2, 2018

What Makes Measuring Growth So Tricky?

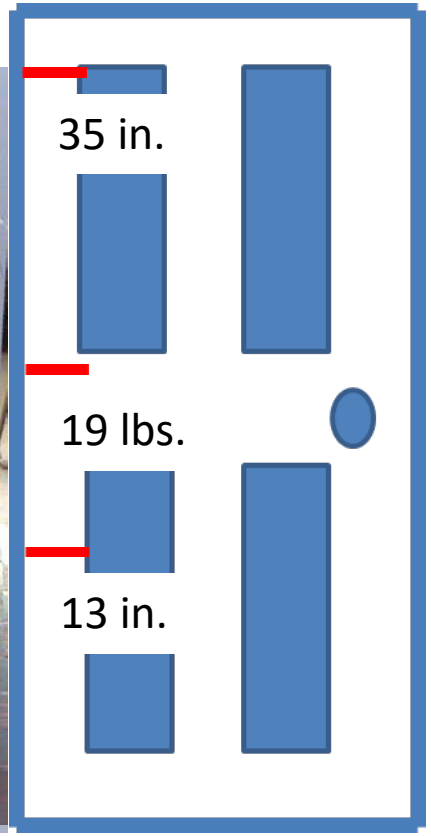
We use it all the time!

We All Understand Growth!

We All Understand Growth!



We All Understand Growth?



Closet Door

Our common understanding of growth is a *unidimensional* concept.

If our measurements span different dimensions, growth is difficult to compute and interpret

When we measure and track height, we are tracking a single dimension.

When we measured width, weight, and height, those are different dimensions, although each is an aspect of *size*….a *multidimensional* measurement

Is Math, as measured on the MEAP (Fall 2011), More Like Height or Size?

Grade 3 MEAP Math N.ME…

- 02.01 – Count to 1,000 by 1's, 10's, 100's
- 02.02- Read, Write #'s to 1000, relate to quantities
- 02.03 – Compare and order numbers to 1,000 using < and >

Grade 4 MEAP Math N.ME…

- 03.16 – Understand that fractions may represent a part…
- 03.17 – Recognize, name, and use equivalent fractions
- 03.18 – Place fractions…
- 03.19 – Understand any fraction…
- 03.20- Numberline rep of + and – of fractions
- 03.21- Relate decimal fractions to parts of a dollar

If different tests are measuring different stuff, how do we interpret rising test scores from one grade to the next?

Do increasing scores mean that students getting “taller” or are they getting “bigger”?

A lack of unidimensionality can be problematic in the measurement of growth.

Perhaps we could move forward by asking a different question. Stay tuned, we'll come back to this in the afternoon.

If we do have unidimensionality, are their other concerns?

WHAT EXACTLY ARE WE SUBTRACTING WHEN WE CALCULATE GROWTH?

I need three volunteers…

- We have basically a unidimensional scale
  - Right?
- Each of the three volunteers started at the same point on our scale
- Each volunteer “grew” on our scale from the first test to the second
  - Who grew the most?
- What might this look like in education?

A Quick Test of Addition

- 1 + 1 = _____
- 9 + 5 = _____
- 8.2 + 3.3 = _____
- ½ + 1/3 = _____
- $6\frac{2}{3} + 7\frac{3}{4}$ = _____
- $\sum_{n=1}^{100}\left(n-(n-1)\right)^{n}$ = _____

Student A:

Test 1 = 1

Test 2 = 5

Student B:

Test 1 = 5

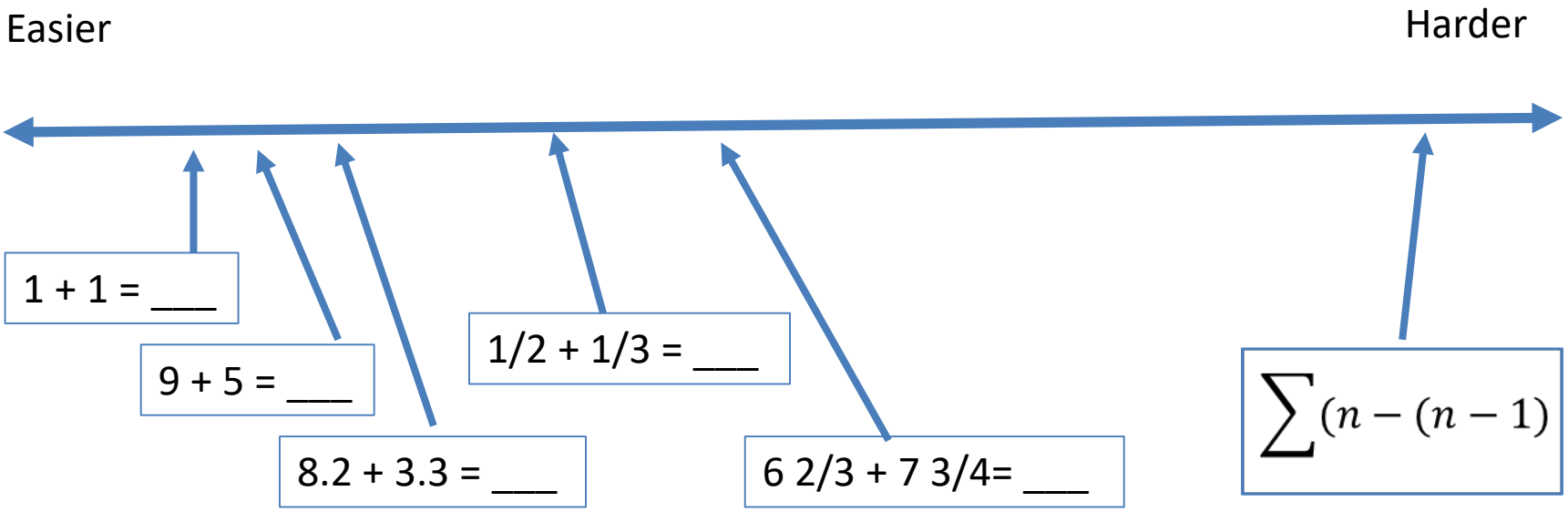
Test 2 = 6

Which student grew more?

Raw scores don’t represent an equal interval scale…

THIS IS WHY SUMMATIVE TESTS ARE SCALED

If our quick math test were scaled



In addition to the growth model chosen, we need to think about

- The nature of the content that we are measuring.
  - If we're using unidimensional models, is our content unidimensional?
  - Are we forcing multi-dimensional to be unidimensional by our scaling?
- The distribution of the data that we are feeding the growth model
  - Does the model make assumptions about the data?

THE MODEL IS ONLY ONE CONSIDERATION WHEN MEASURING GROWTH

In your experience

WHAT ISSUES HAVE YOU RUN INTO IN MEASURING GROWTH?

That’s probably enough for now. Let’s break for

LUNCH!

Measuring Student Growth: So Much More than Subtracting Two Numbers!

Assessment Learning Network

Friday, March 2, 2018

A Highly Recommended Resource

- Castellano, K. & Ho, A. (2013). A practitioner's guide to growth models
- Downloadable from: http://scholar.harvard.edu/files/andrewho/files/a_pracitioners_guide_to_growth_models.pdf
- This portion of the presentation will be loosely organized around this book.

Before we get started, let’s keep some things in mind

LET’S START WITH THIS

"…a growth model may seem like a concise, perhaps even single step procedure capable of achieving many desired goals and outcomes. Such a definition overlooks the multiple components of operational growth models and the complexity and judgement that are required as they increasingly attempt to serve multiple purposes." - p. 17

"All models are wrong, but some models are useful."

George E.P. Box

Seven “Flavors” of Growth Models

- Gain Score
- Trajectory
- Categorical
- Residual Gain
- Projection
- Student Growth Percentile
- Multivariate

Growth as we know it…

THE GAIN SCORE MODEL

The Gain Score Model

- The simplest and most intuitive model
  - The closet door model is a gain score model
- Growth is expressed in absolute terms
  - Your growth score doesn't depend on others
- Data at two points is required
- Growth is given by the equation:

Growth = ScoreNow – ScoreBefore

The quality of the scale is very important

The Gain Score Model

- The scale is very important
  - Both tests the same (scale) (pre-post test)
  - Both tests share a vertical scale (grade to grade)
- Gain scores can be aggregated by calculating a mean
- If comparisons are going to be made, the scale needs to be equal-interval

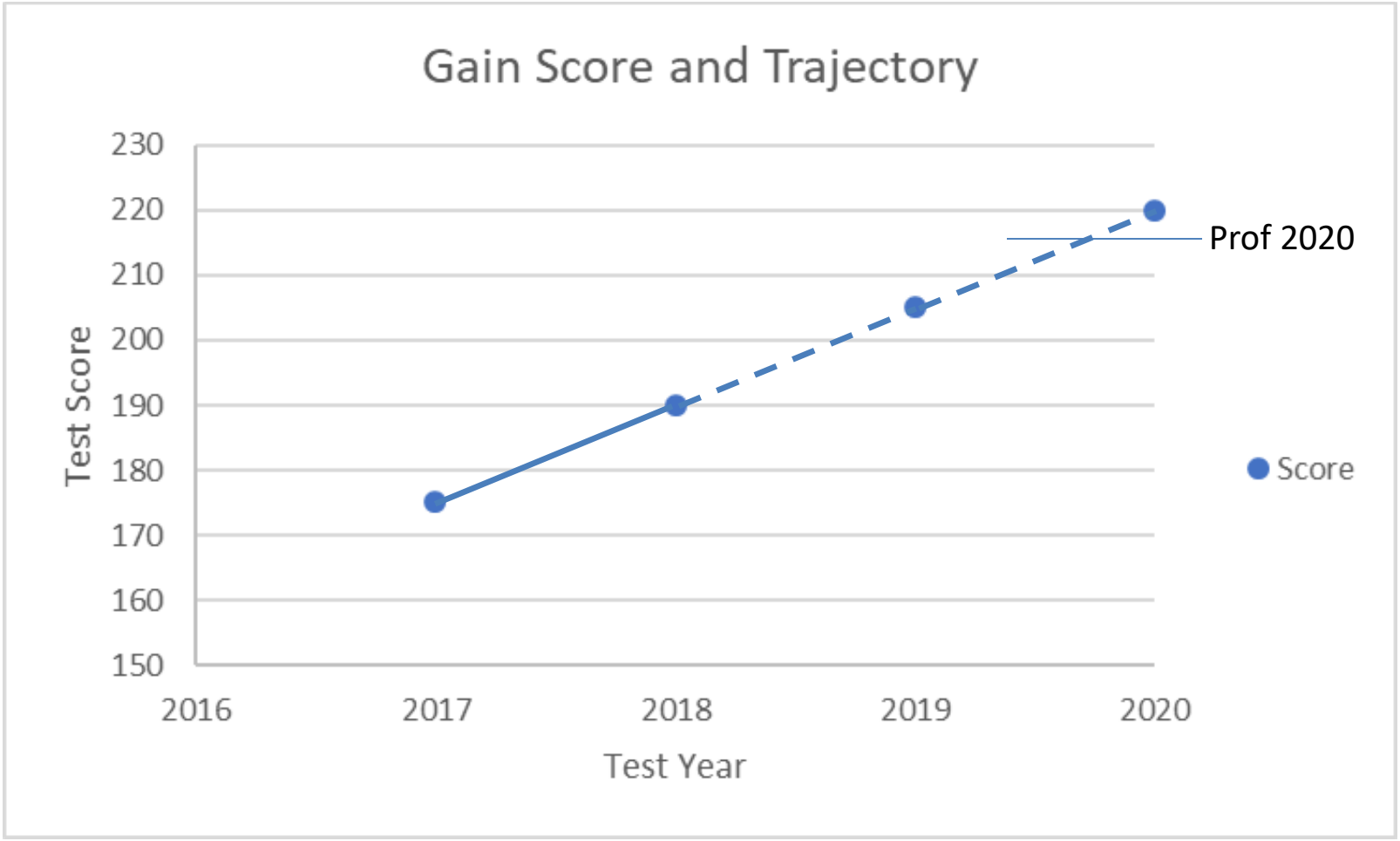
What might the future hold?

THE TRAJECTORY MODEL

The Trajectory Model

- An extension of the gain score model
- Projects the observed gain score (slope) out a number of years to predict future performance
- Often the future projections are compared to future cut scores to see if the student is “on track”

The Trajectory Model



The Trajectory Model

- Usually, the trajectory is based on consistent, linear growth
  - Is this reasonable?
- Has similar scale requirements as the gain score model
  - Vertical scale
  - Equal interval

The Trajectory Model

- There are a couple of options for aggregating the trajectory model
  - Calculate a mean (average) gain for the group over the next x years
  - Calculate the percent of test takers who are “on track” to be proficient at some point in the future
- What are the implications for use in an accountability model?

Michigan gets some pub!

THE CATEGORICAL MODEL

The Categorical Model

Test 2

Test 1

Performance Level	Novice	Partially Proficient	Proficient	Advanced
Novice				
Partially Proficient				
Proficient				
Advanced				

The Categorical Model

Test 2

Test 1

Performance Level	Novice	Partially Proficient	Proficient	Advanced
Novice	Maintained	Improved	Really Improved	Super Improved
Partially Proficient	Declined	Maintained	Improved	Really Improved
Proficient	Really Declined	Declined	Maintained	Improved
Advanced	Super Declined	Really Declined	Declined	Maintained

The Categorical Model

- Communicates growth as a progression from one performance level to another
- Michigan piloted this model and some call it the "Michigan Model"
  - Some at MDE really liked that! ☺
- Doesn't explicitly have the same strict scale requirements as other methods
  - Are assumptions made about the underlying scales, however?

The Categorical Model

- A little more about scale and the categorical model
  - The actual scales of the tests aren't used, the performance levels (categories) that sit on top of the scales are used
  - If we have four categories sitting on top of 80 scale score points, what have we done?
  - Implicit assumptions that the categories represent distinct levels of achievement

The Michigan Model (Performance Level Change)

Year X Grade Y MEAP Performance Level		Year X+1 Grade Y+1 MEAP Performance Level								
		Not Proficient			Partially Proficient		Proficient			Advanced
		Low	Mid	High	Low	High	Low	Mid	High	Mid
Not Proficient	Low	M	I	I	SI	SI	SI	SI	SI	SI
	Mid	D	M	I	I	SI	SI	SI	SI	SI
	High	D	D	M	I	I	SI	SI	SI	SI
Partially Proficient	Low	SD	D	D	M	I	I	SI	SI	SI
	High	SD	SD	D	D	M	I	I	SI	SI
Proficient	Low	SD	SD	SD	D	D	M	I	I	SI
	Mid	SD	SD	SD	SD	D	D	M	I	I
	High	SD	SD	SD	SD	SD	D	D	M	I
Advanced	Mid	SD	SD	SD	SD	SD	SD	D	D	M

The Categorical Model

- Typically summarized by listing the percentages of students in each category
  - X% Improved
  - Y% Declined, etc.
- Comparisons between different grades require strong assumptions about the underlying scales of the tests used
  - This may be very difficult to establish

Making use of more advanced statistical techniques

THE RESIDUAL GAIN MODEL

The Residual Gain Model

Test 1	Test 2
120	218
125	228
130	230
135	238
140	231
145	237
150	246
155	258
160	260

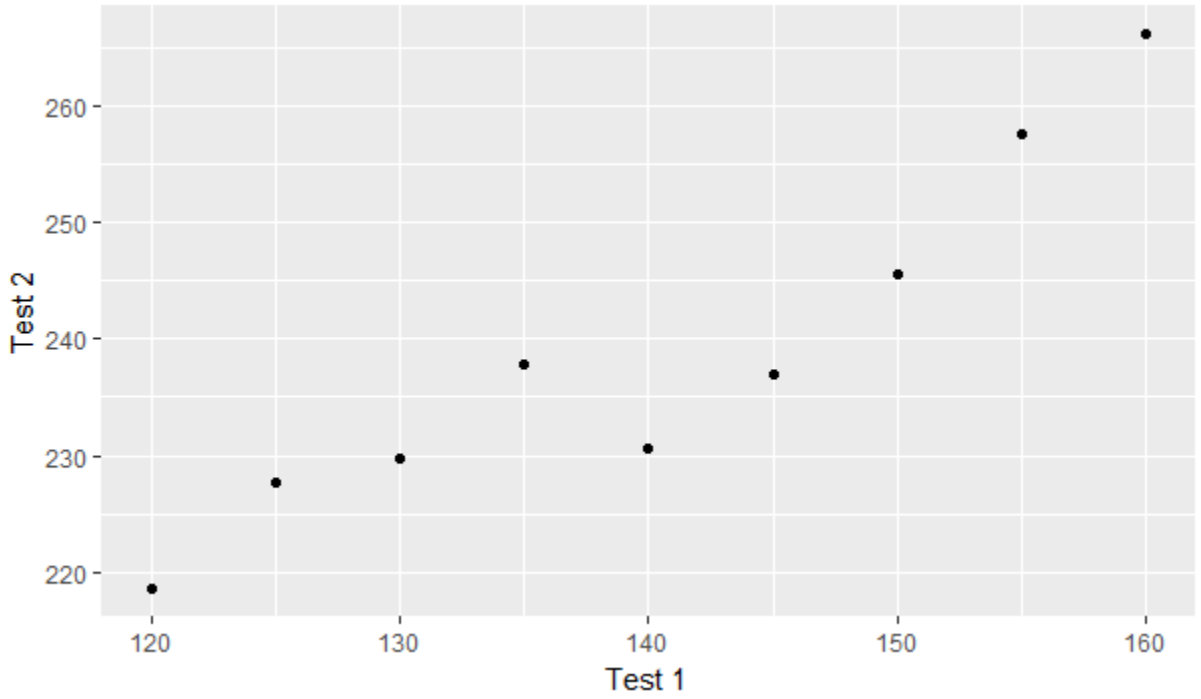
But first….

"A reasonable first step, it seems, would be to graph the cr@p out of it."

George E.P. Box

The Residual Gain Model

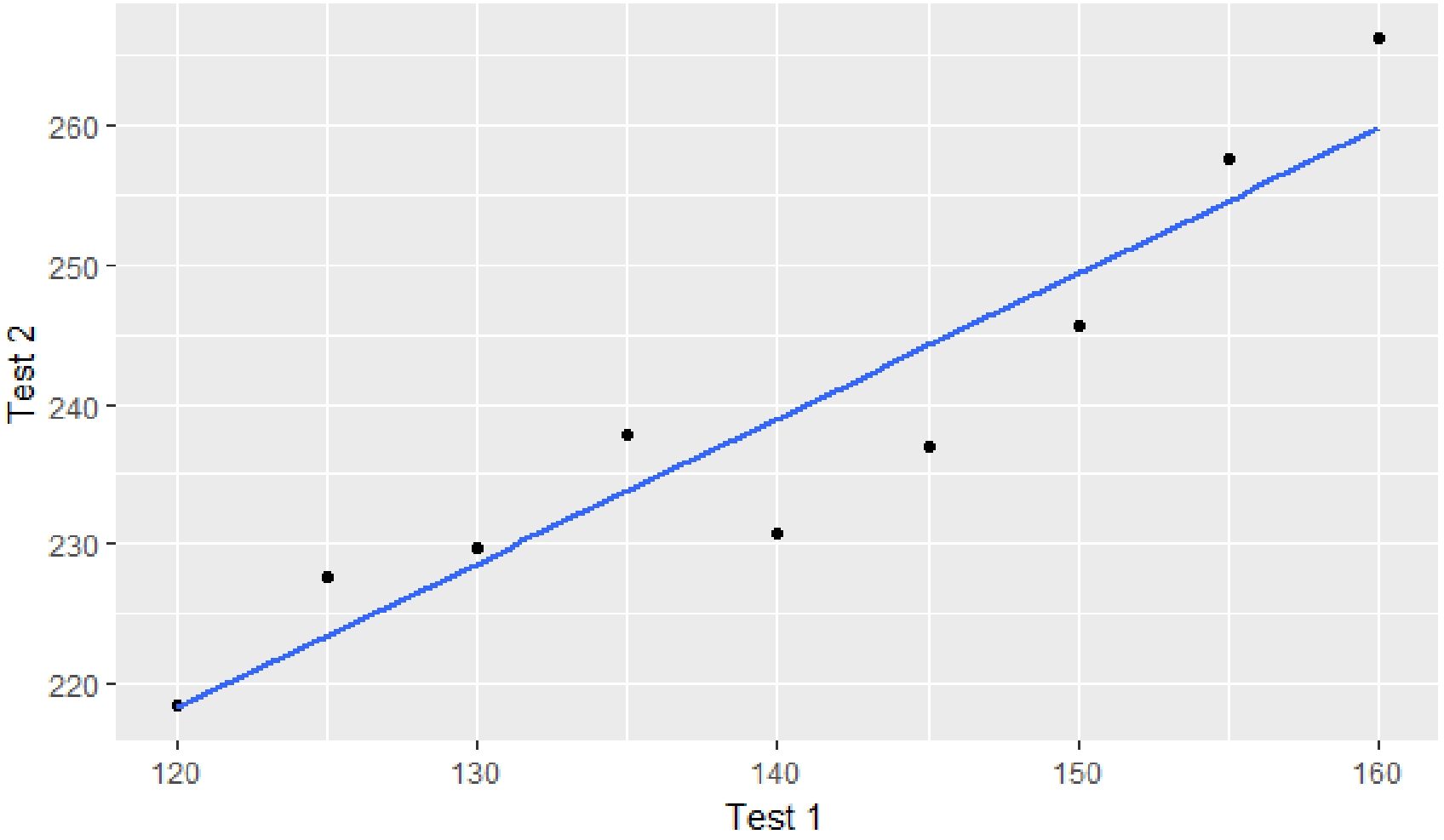
Test 1	Test 2
120	218
125	228
130	230
135	238
140	231
145	237
150	246
155	258
160	260



The Residual Gain Model

- If the plot of the two variables shows a linear relationship, we can fit a linear model to capture the relationship between the two variables.
  - This is linear regression
- The residual is difference between the observed Test 2 score and the predicted Test 2 score from the model

The Residual Gain Model



The Residual Gain Model

- For these data, the regression equation is given by PredTest2 = 93.4 + 1.04*Test1
  - Remember y = mx+ b from algebra? Same thing!
  - Find your algebra teacher and thank them! ☺
- We can use this model to get predicted scores for each of the test takers and then calculate residuals
  - Remember: Residual = Actual - Predicted

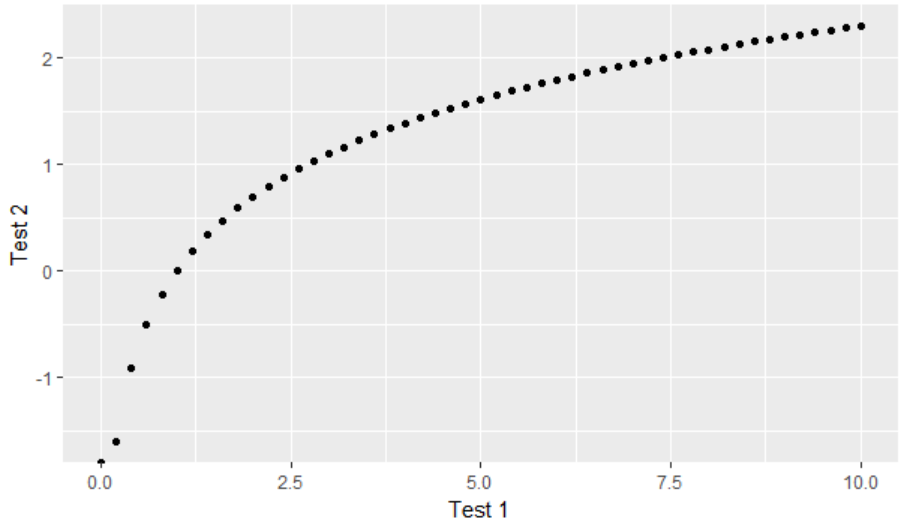
The Residual Gain Model

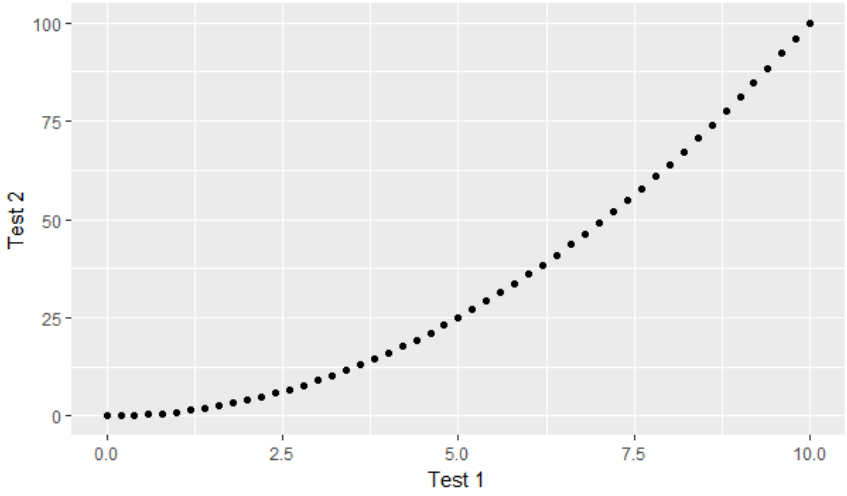
Test 1	Test 2	Model (Predicted)	Residual	Growth more/less than predicted
120	218	218	0	Same
125	228	223	5	More
130	230	229	1	More
135	238	234	4	More
140	231	239	-8	Less
145	237	244	-7	Less
150	246	249	-3	Less
155	258	255	3	More
160	266	260	6	More

More About Residual Gains

- Typically linear models are fit
  - Are all relationships linear?

Linear Relationships?





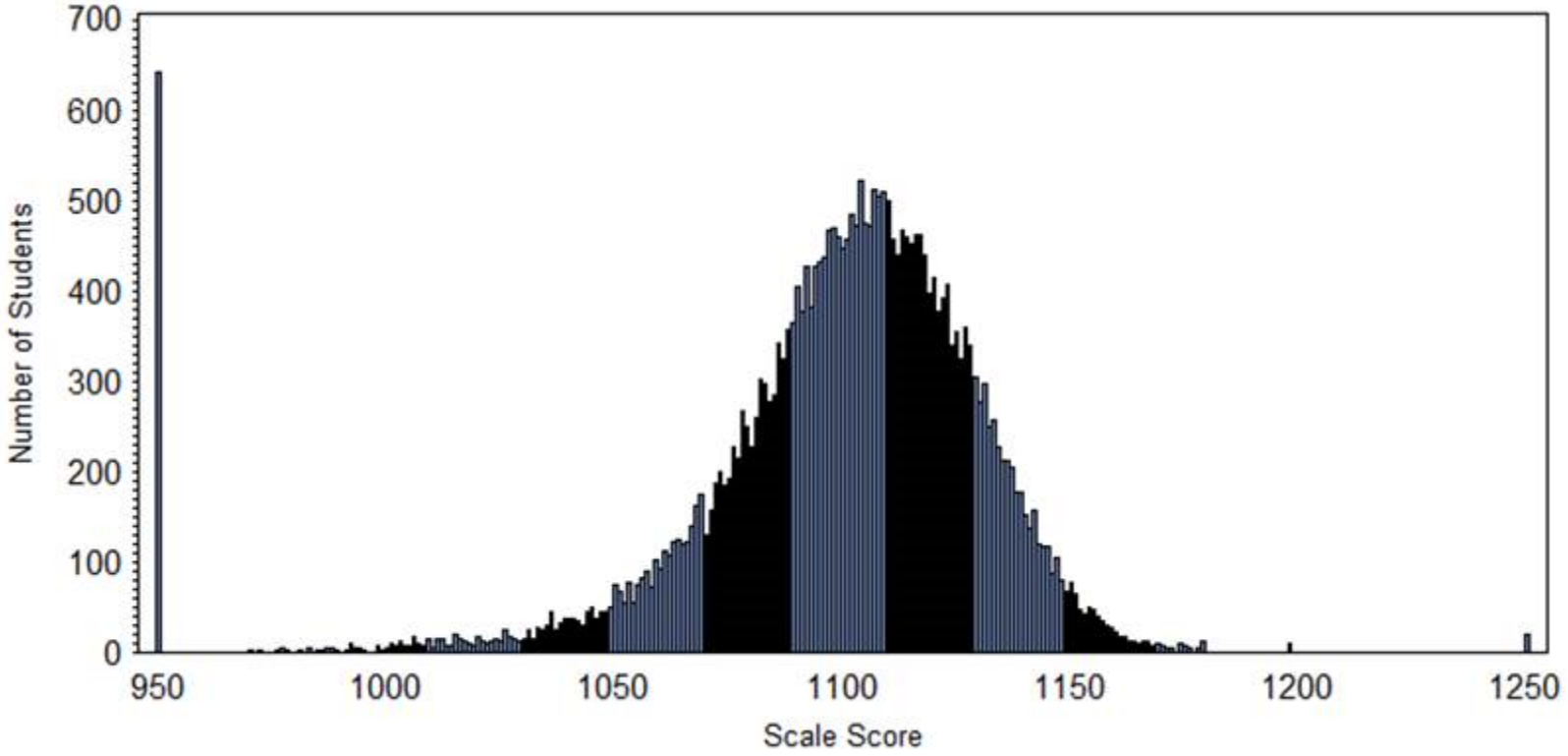
More About Residual Gains

- Typically linear models are fit
  - Are all relationships linear?
- Note there are residuals above the line and residuals below the line
  - Always the case with least squares regression
  - Does this make the model criterion or norm referenced?
  - What are the implications for aggregation?

More About Residual Gains

- In our simple example, we were only using one prior test score. In practice, multiple test scores can be used.
- Linear regression relies on some assumptions
  - Linear relationship between the variables
  - Assumptions about the distribution of data

Normally Distributed?



More About Residual Gains

- In our simple example, we were only using one prior test score. In practice, multiple test scores can be used.
- Linear regression relies on some assumptions
  - Linear relationship between the variables
  - Assumptions about the distribution of data
  - Assumptions about the relationships between aspects of the model
    - This gives us one of the great words: heteroscedasticity

Whew!.....

WHAT QUESTIONS MIGHT YOU HAVE?

Modeling into the future

THE PROJECTION MODEL

The Projection Model

- Fitting and diagnosing a linear regression model takes quite a bit of effort.
  - It would be nice to be able to use it for more than one year.
  - The linear model summarizes the linear relationship between two sets of scores
  - Since the model remains after the cohort of students moves on, it can be re-used.

The Projection Model

- Step 1: Fit an appropriate linear model for the relationship between test 1 and test 2 based on a group of students who have taken both tests. (Residual Gain Model)
- Step 2: Give test 1 to a new group of students
- Step 3: Insert the test 1 scores from the new group of students into the regression equation to get the new students' predicted test 2 scores
- Step 4: Compare the predicted scores to standards to see who is "on track" and who isn't

The Projection Model

Test 1	Test 2	Model (Predicted)	Residual	Growth more/less than predicted
120	218	218	0	Same
125	228	223	5	More
130	230	229	1	More
135	238	234	4	More
140	231	239	-8	Less
145	237	244	-7	Less
150	246	249	-3	Less
155	258	255	3	More
160	266	260	6	More

$$Test2 = 93.4 + 1.04 * Test1$$

Test 1	Model (Predicted)	On Track? (Cut=250)
122	220	No
137	236	No
142	241	No
152	251	Yes
157	257	Yes
159	259	Yes

I know what you're thinking!

- Since we have predicted test 2 scores, could we wait until these students take test 2 and compute residuals for how they actually did?
- Yes, but…
  - Since the predicted scores came (long) before the test 2 scores, there was time for instructional intervention
  - Remember the prediction paradox from Dr. Perie?
    - Predicted to do well -> Does well
    - Predicted to not do well -> gets intervention -> Does well

The Projection Model

- The strength of the projection model is to identify at-risk students while there is still time to take instruction action to improve learning as measured by test 2
- What are the implications for use in an accountability system?

The Projection Model

- Given these issues, some prefer to work with the predictions and the residuals
  - Percentage of students that scored higher than their prediction (positive residuals)
  - Average Residual
    - Can be tricky- remember there are always positive and negative residuals
- Require large models
  - Scores from multiple groups

Getting some context

THE STUDENT GROWTH PERCENTILE MODEL

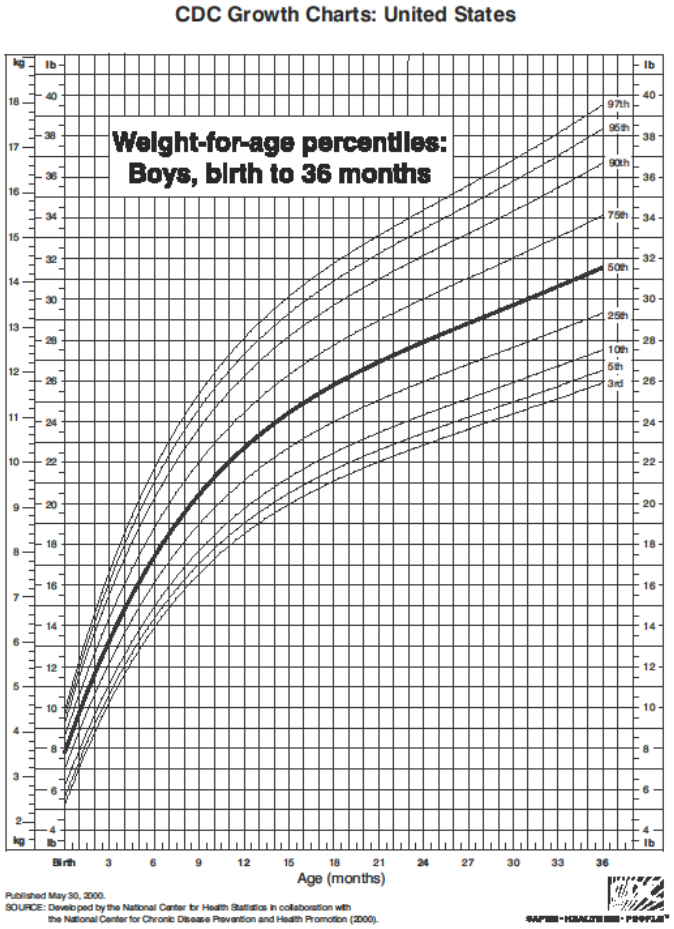
Mason Grew 2 Inches Last Year.

- Measuring height the “right way” on the door! ☺
- How do we interpret that statement?
  - It depends, right?
- If Mason is my 2 year old nephew, we’re concerned because that growth is low
- If Mason is my 9 year old nephew, we think it sounds about right.
- If Mason is my 39 year old brother, we think this is quite strange.

Sometimes, evaluating growth is helped by having a frame of reference.

Perhaps it matters where you started.

If you're a parent, you probably recognize this…



Audience Participation!

LET’S SEE HOW THIS MIGHT WORK

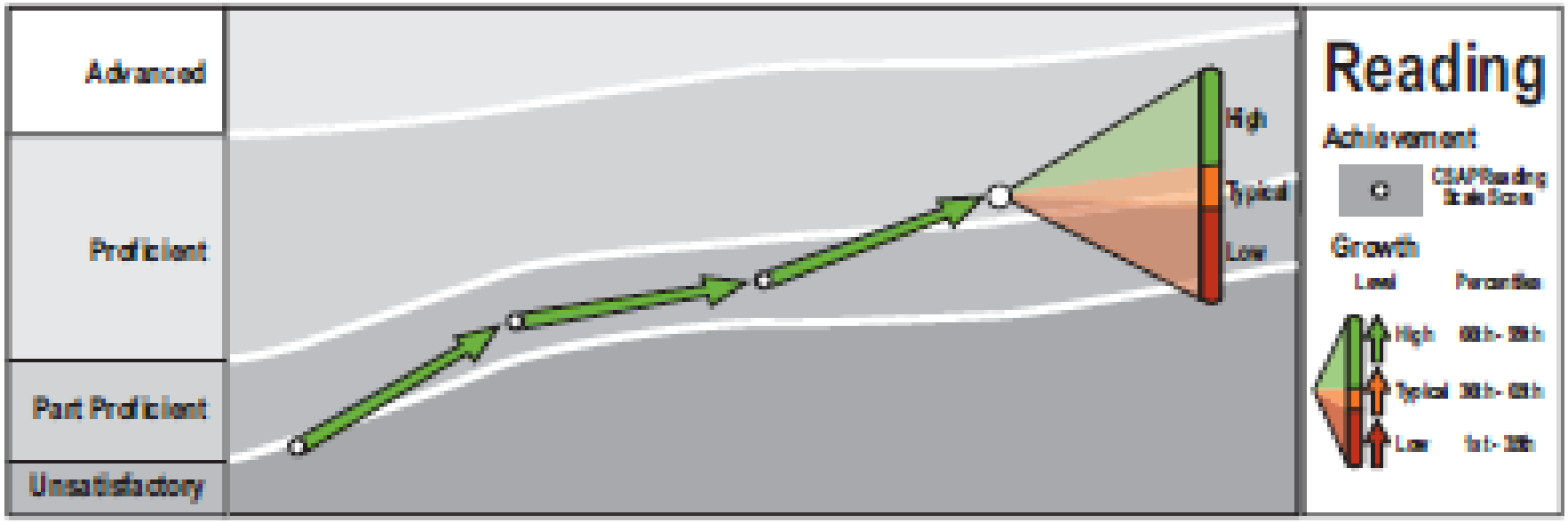
Student Growth Percentiles

- Presents student growth compared to students with the same prior test score(s).
  - Test score history
- Conceptually simple
  - Analagous to the height/weight tables
- Quite complicated in practice
  - Large sample sizes
  - Quantile regression
  - Sparse N techniques

Student Growth Percentiles

- Doesn't have the strict scale requirements as some of the others
  - Simplifies calculations, complicates interpretation
- Norm referenced
  - Implications for accountability
  - Standards-based?
- Two methods of aggregation
  - Traditionally: Median
  - Emerging Research supports using the mean

SGPs and Projection



The wild west

THE MULTIVARIATE MODEL

The Multivariate Model

- A bit of a misnomer, there is no default multivariate model
  - Statistical models are built that include the variables you believe are relevant to student growth
  - A wide variety of variables are often included
    - Prior test scores for a student
    - Teacher(s) of record for a student
    - Participation in various programs
    - Demographic variables

The Multivariate Model

- A wide variety of statistical methods can be employed depending on the nature of your variables and theory of what growth looks like
  - Structural Equation Modeling
  - Hierarchical Linear Models
  - Mixed-Methods Modeling
- Keep in mind the models need data to work with
  - We never get away from thinking about our scales!

The Multivariate Model

- Multivariate models are often the models used in “value-added” modelling
- Caution is needed in interpreting value-added
- Causality is determined more by the experimental design than the statistical model employed
- Even more scrutiny is needed for the scales when there is a desire to establish causality

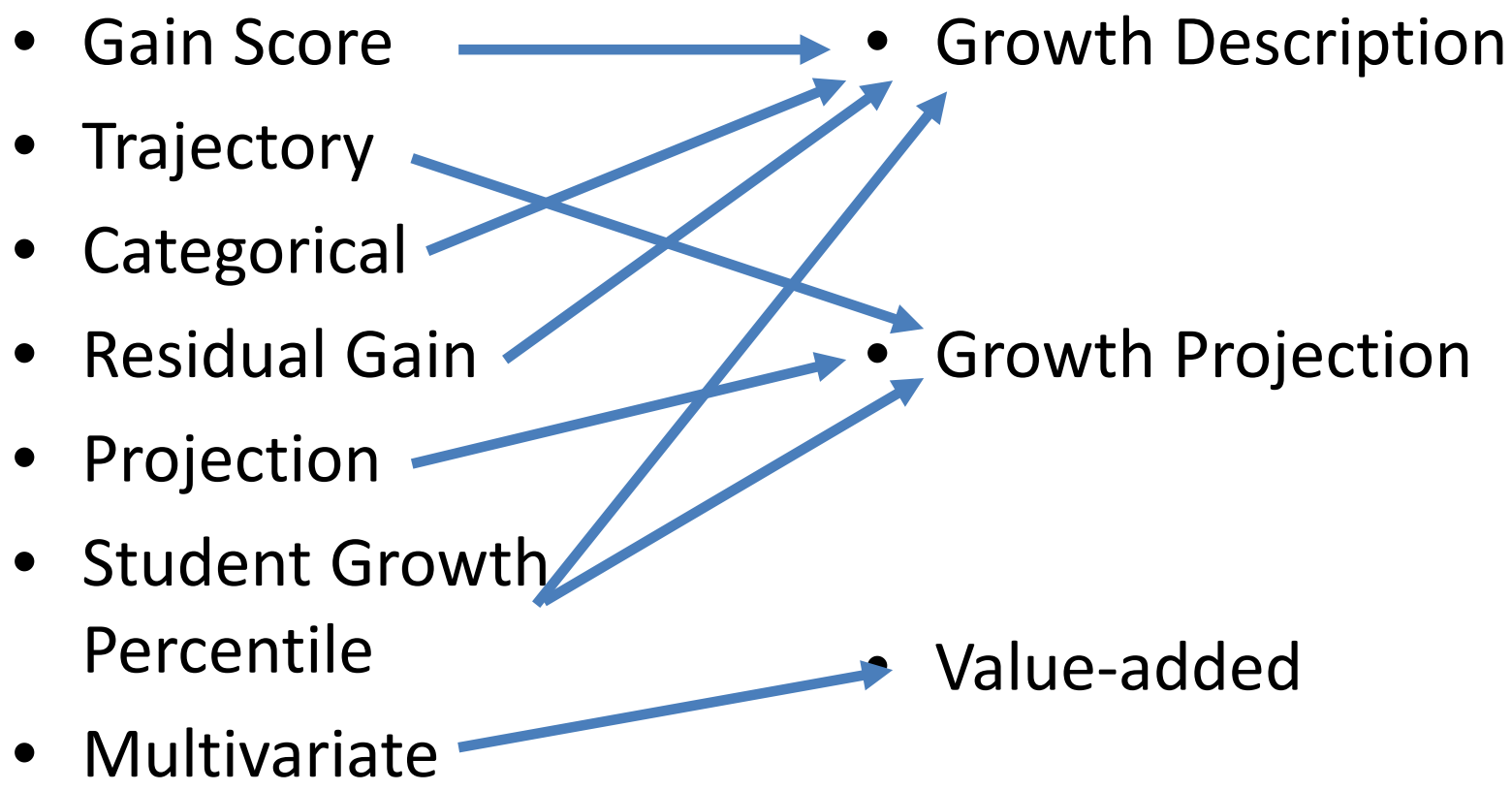
We’ve looked at seven growth models

CLASSIFICATIONS OF THE MODELS

Classifying Growth Models

- Three over-arching uses/interpretations
  - Growth Description
  - Growth Projection
  - Value Added
- Can we classify our models into these categories?

Classifying Growth Models



Many thanks!

Jim Gullen (and friends)

Michigan Assessment Consortium

jgullen@michiganassessmentconsortium.org