INTRODUCTION TO VALUE-ADDED

MDE Teacher Evaluation Work Group

Dr. Robert H Meyer

Director, Value-Added Research Center

University of Wisconsin-Madison

July 30, 2012

Contents

- Brief Introduction to VARC
- Value-Added Basics
- Value-Added Conceptually: The Oak Tree Analogy
- Using Value-Added for Decision Making
  - Appropriate Model Design
  - Varied Uses of the Data

Brief Introduction to VARC

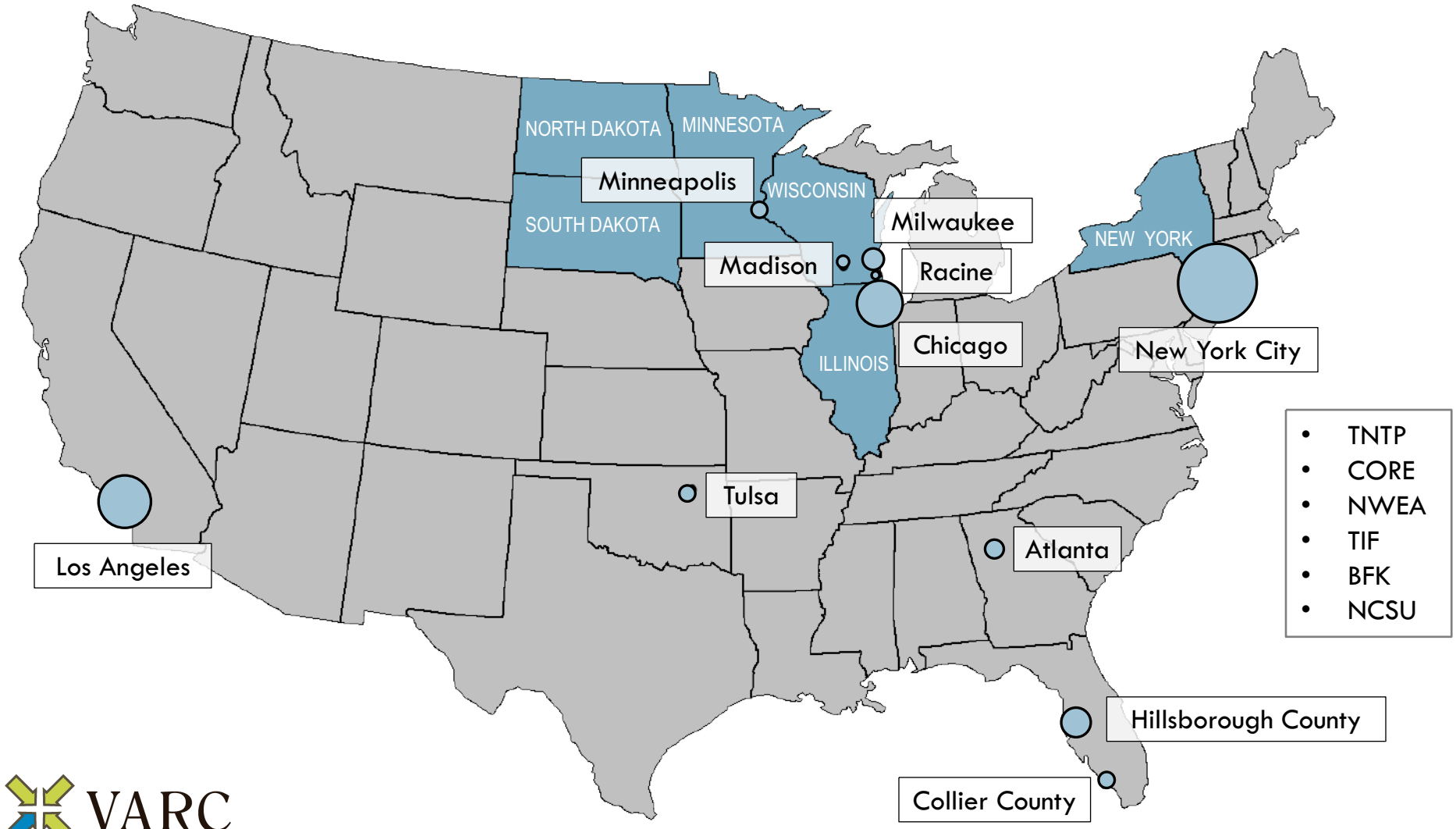
VARC

The Value-Added Research Center at the University of Wisconsin – Madison

Expertise includes:

- Student growth and value-added measures
- District, state, and IHE models
- Educator effectiveness policy
- Data systems and data quality
- Professional development and technical assistance

VARC Partners



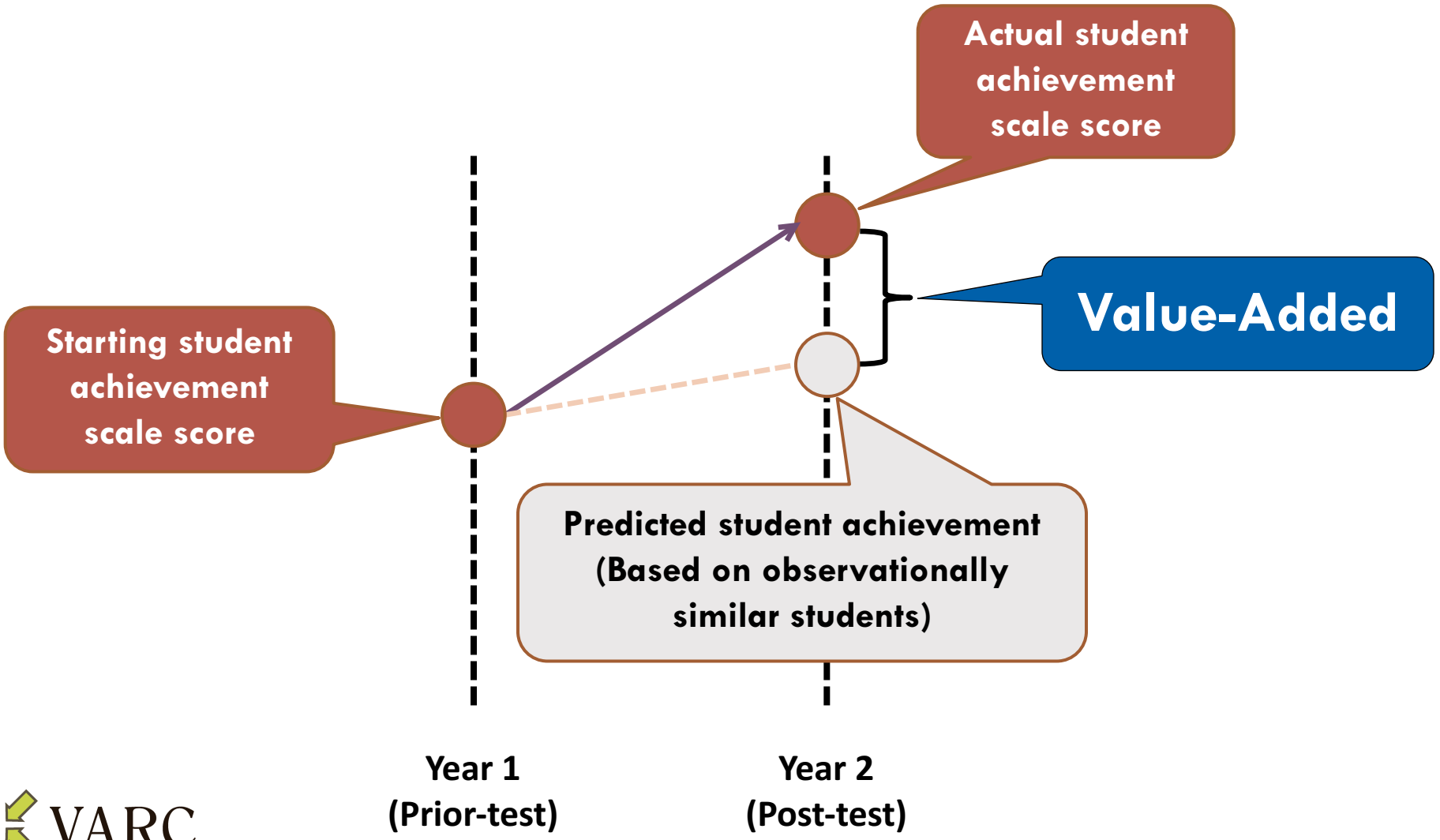
VARC

Value-Added Basics

What is Value-Added?

- It is a kind of growth model that measures the contribution of schooling to student performance on standardized assessments
- It uses statistical techniques to separate the impact of schooling from other factors that may influence growth
- It focuses on how much students improve on the assessment from one year to the next as measured in scale score points

Value-Added: A Visual Representation



VARC

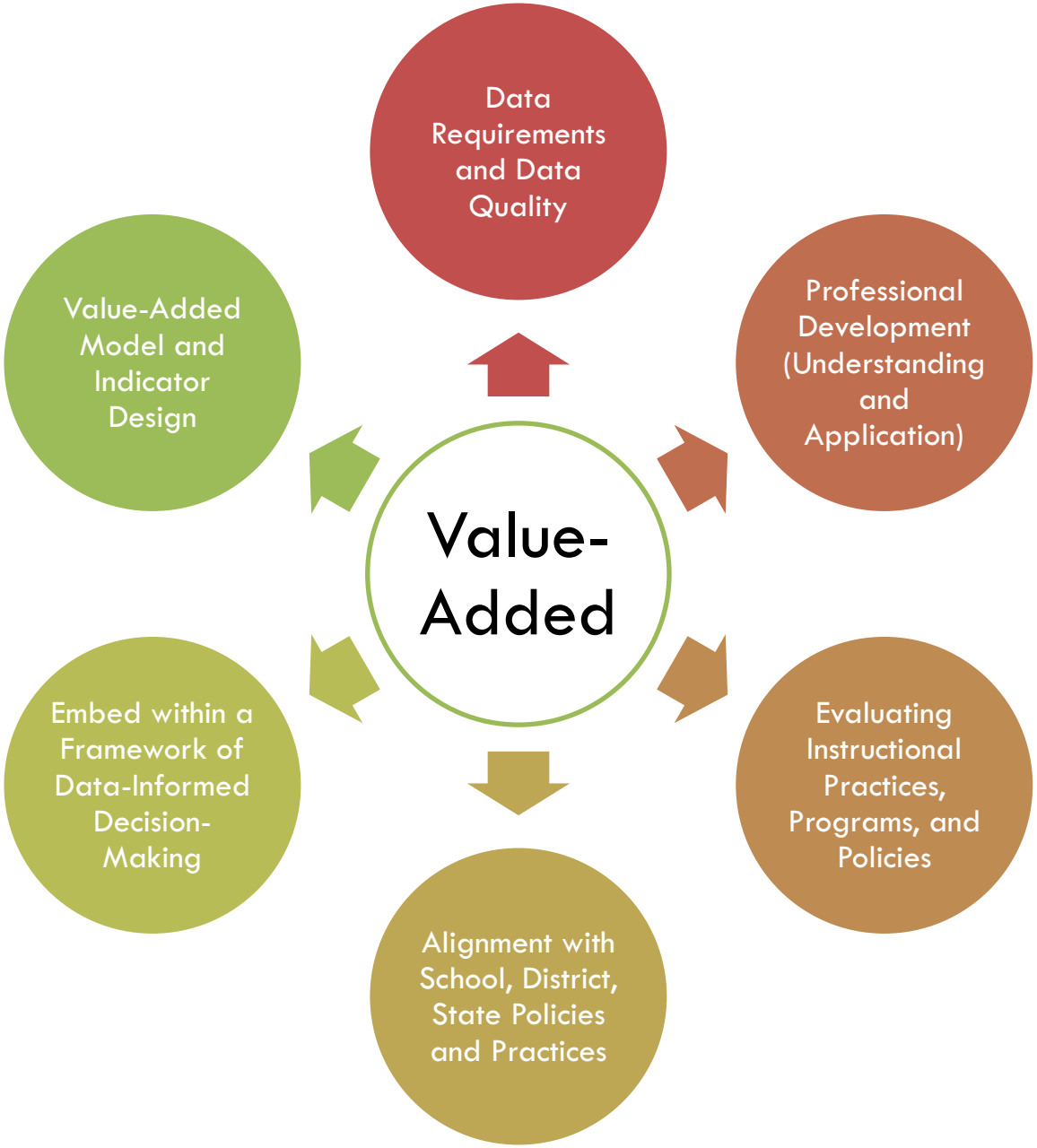
Value-Added Recipe

WSJ: Would you say your formula is the *secret sauce* ?

Rob Meyer: Maybe not a [secret] sauce, but perhaps a well-crafted recipe… made with quality ingredients…[.]

VARC

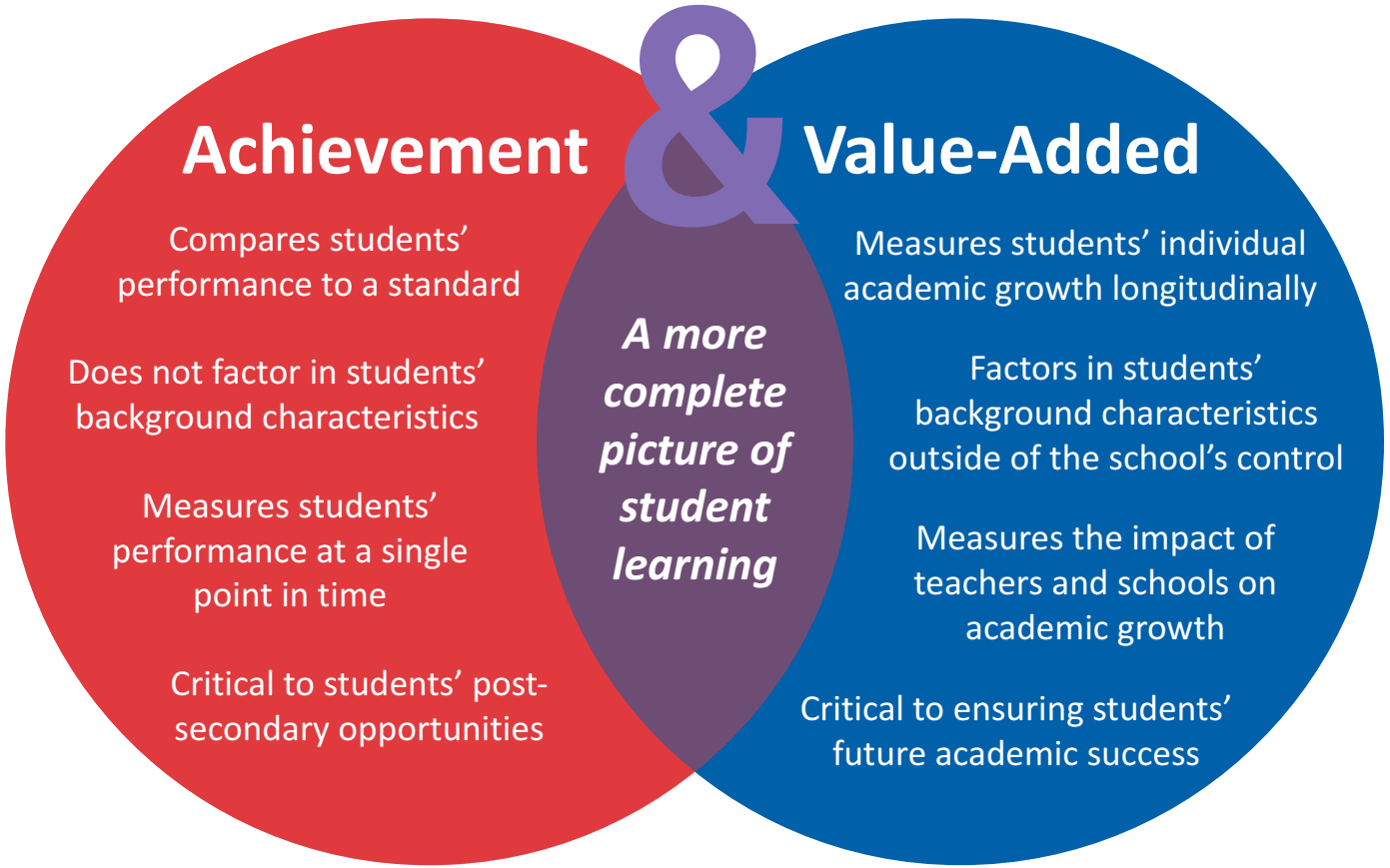
Support Required for a Value-Added System



VARC

How Can Value-Added Information be Used?

The Power of Two

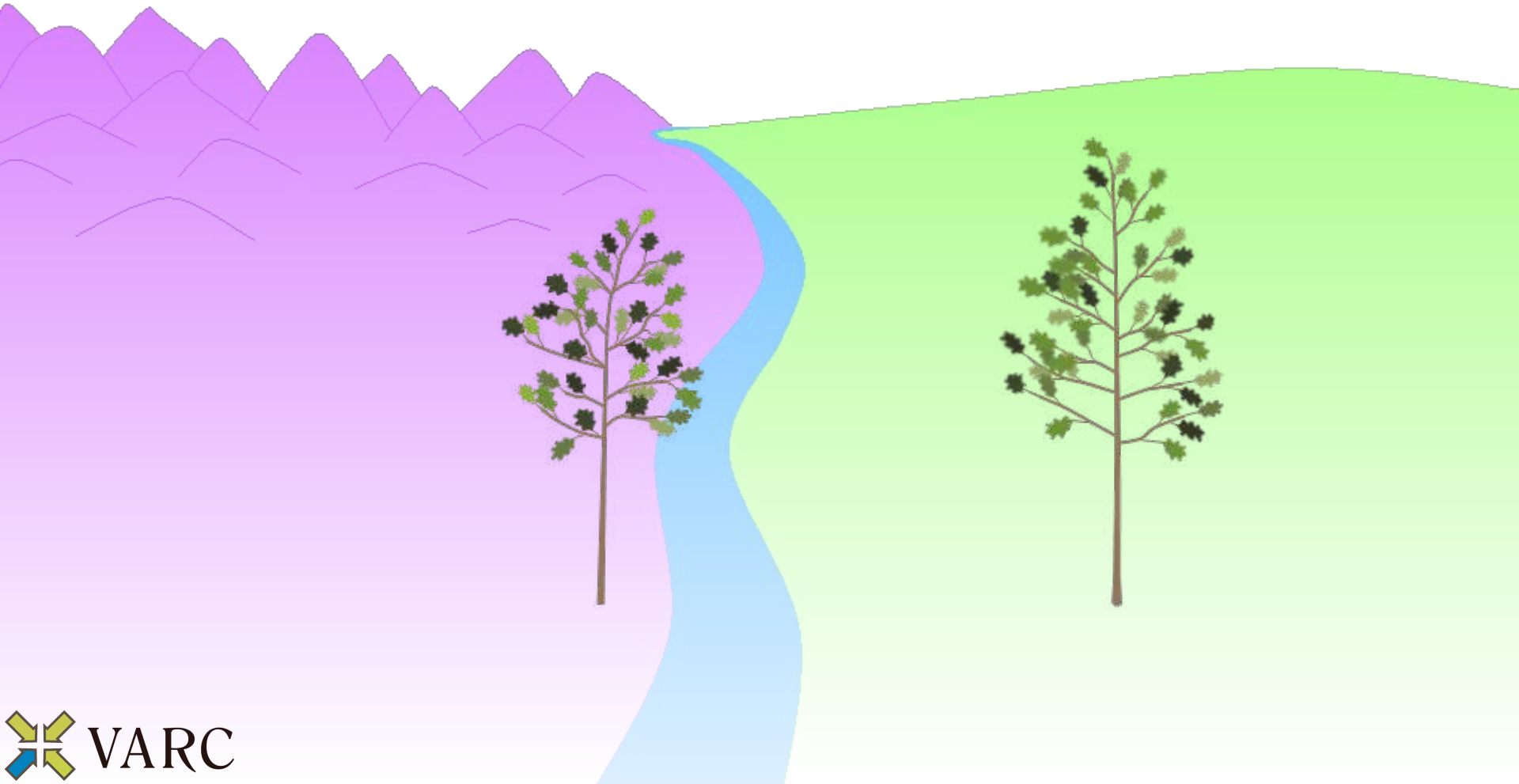


VARC

Adapted from materials created by Battelle for Kids

Value-Added Conceptually: The Oak Tree Analogy

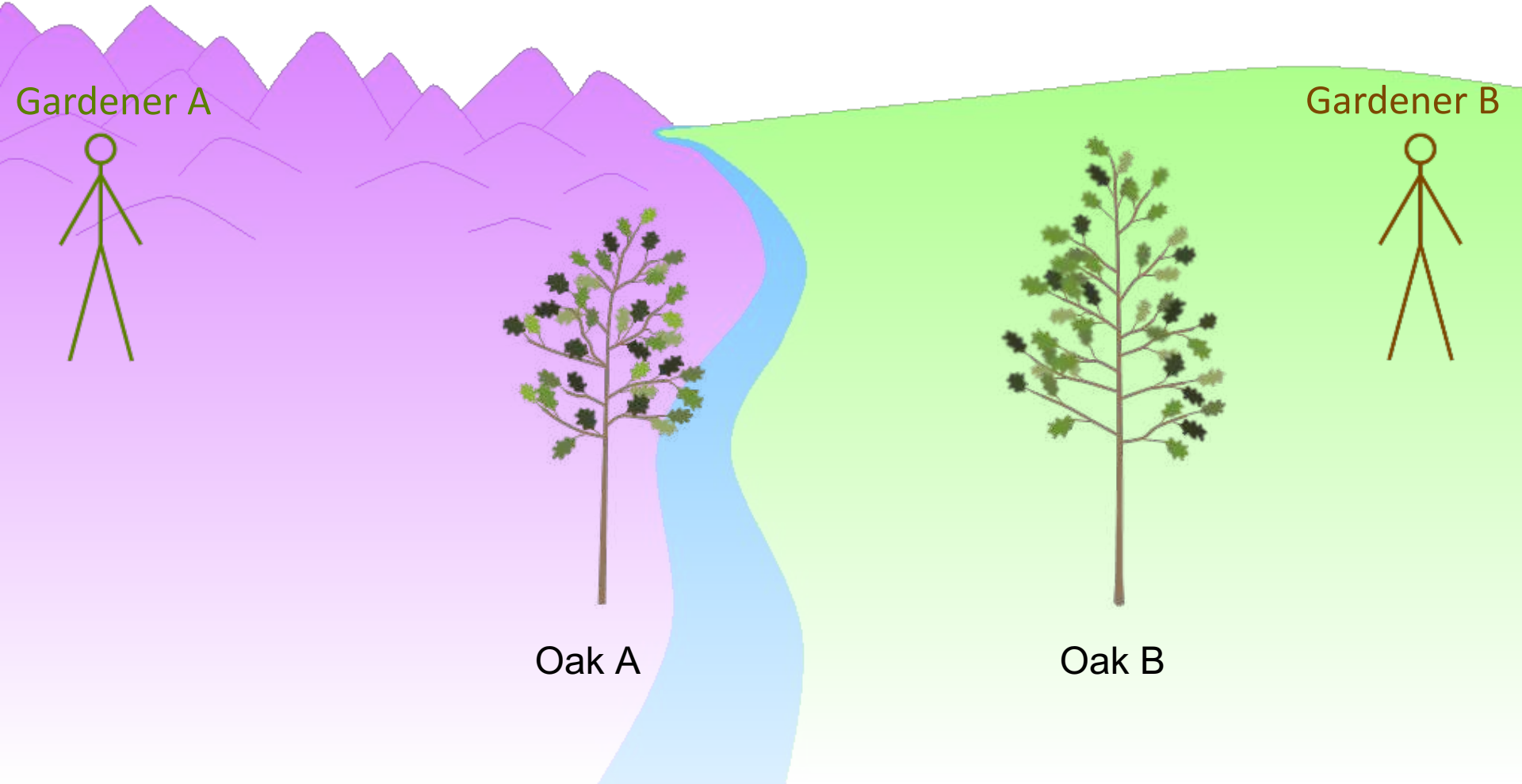
The Oak Tree Analogy



VARC

Explaining the concept of value added by evaluating the performance of two gardeners

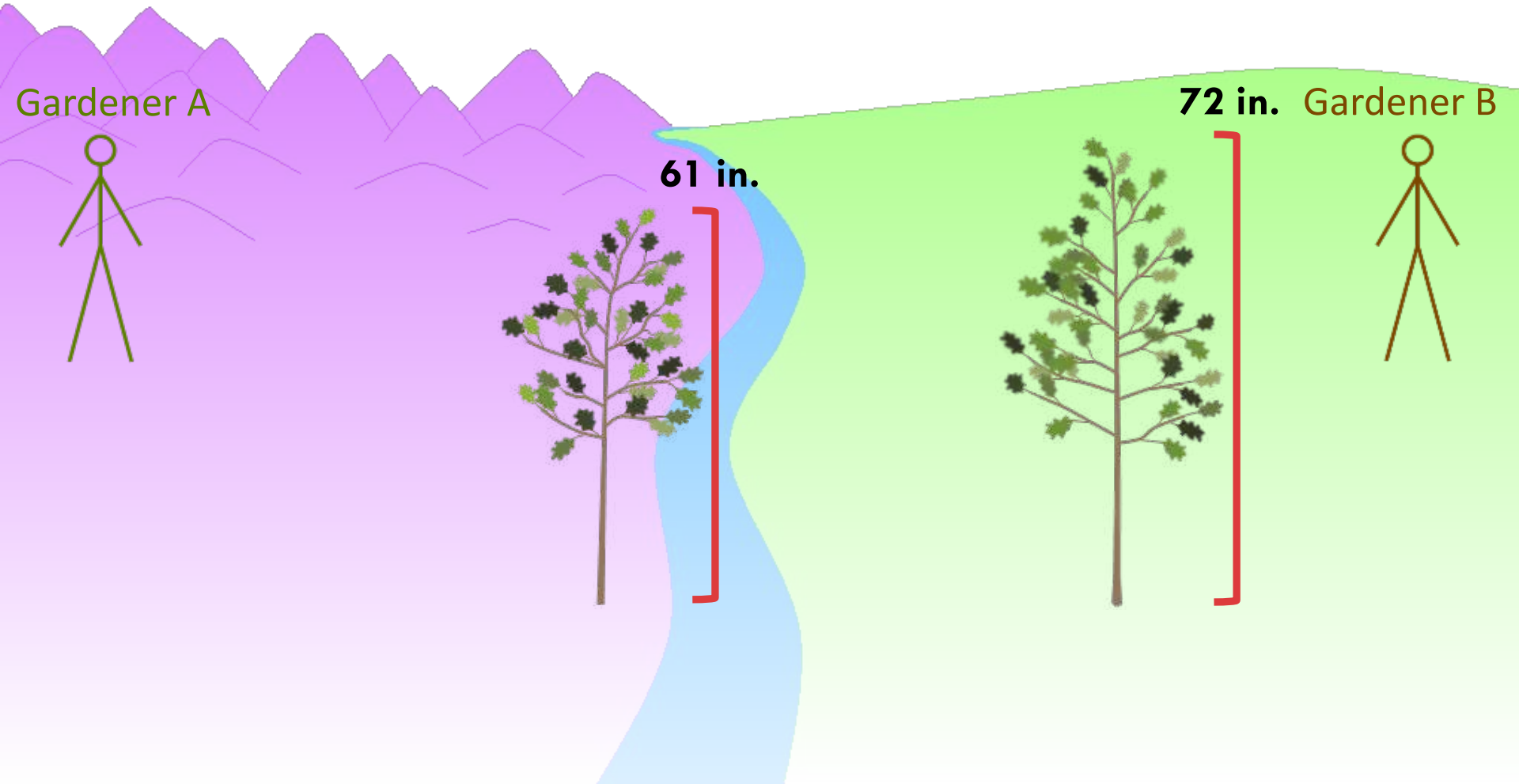
- For the past year, these gardeners have been tending to their oak trees trying to maximize the height of the trees.
- Each gardener used a variety of strategies to help their own tree grow. We want to evaluate which of these two gardeners was more successful with their strategies.



To measure the performance of the gardeners, we will measure the height of the trees today (1 year after they began tending to the trees).

- Using this method, Gardener B is the better gardener.

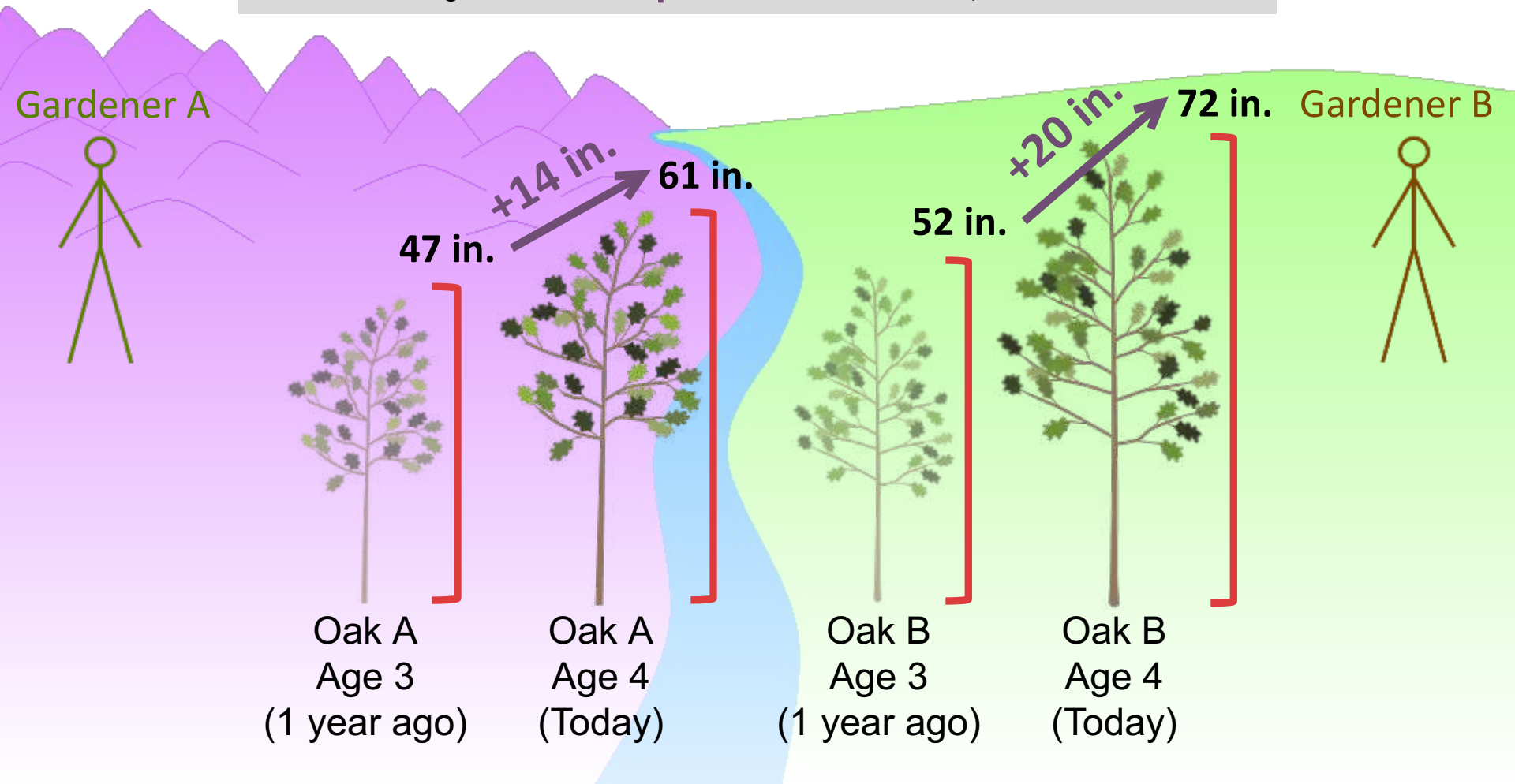
This method is analogous to using an **Achievement Model**.



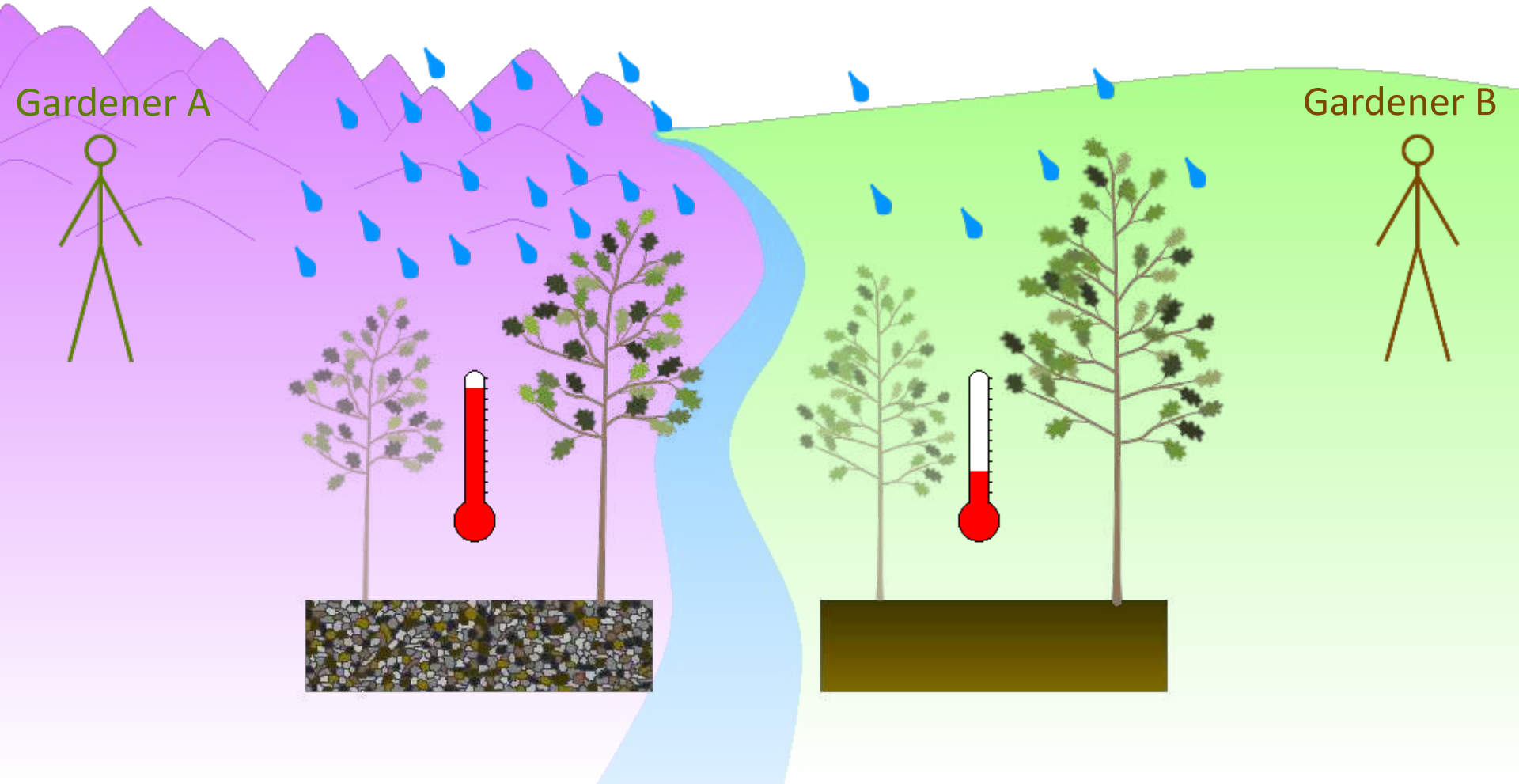
We can compare the height of the trees one year ago to the height today.

- By finding the difference between these heights, we can determine how many inches the trees grew during the year of gardener's care.
- Oak B had more growth this year, so Gardener B is the better gardener.

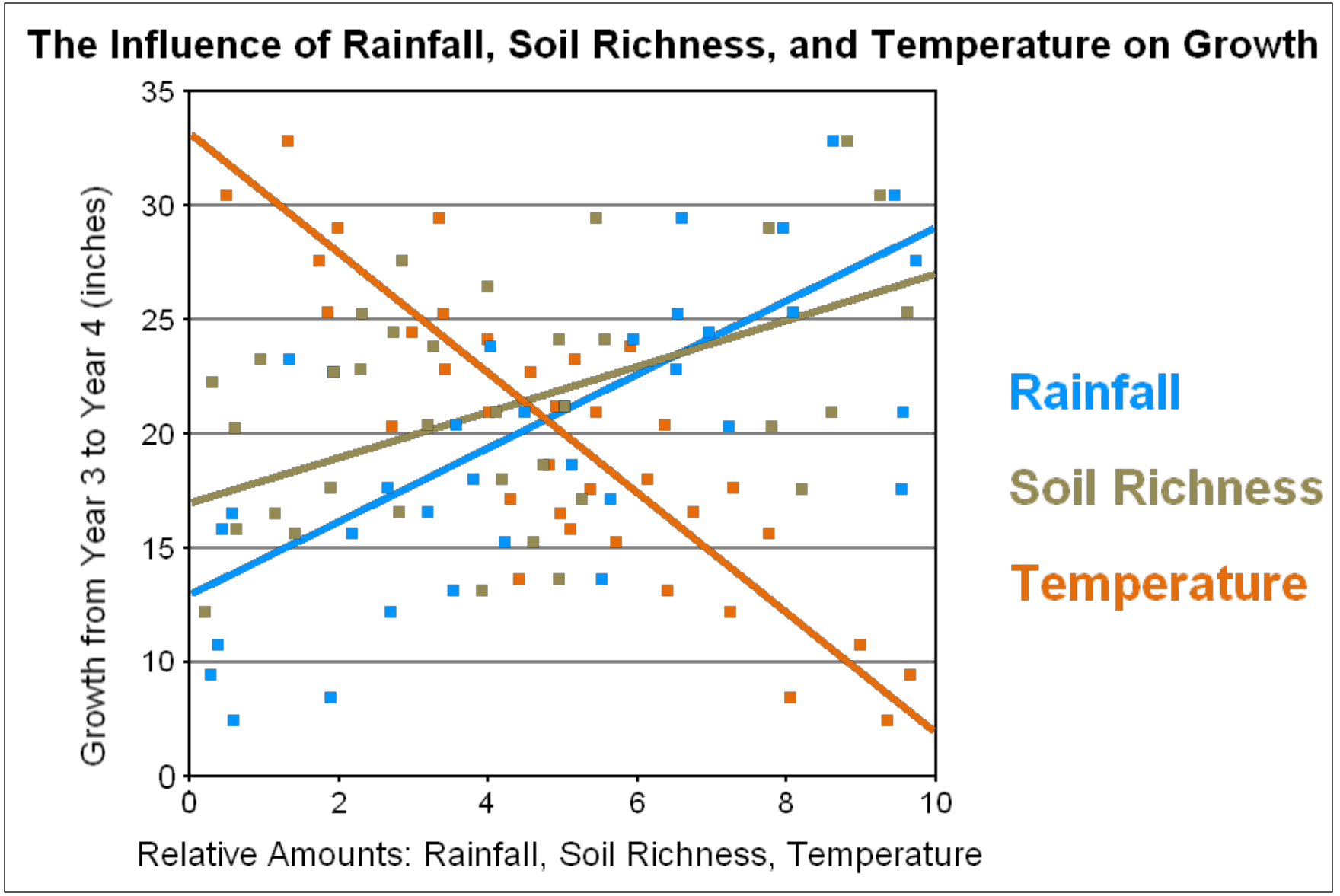
This is analogous to a **Simple Growth Model**, also called **Gain**.



External condition	Oak Tree A	Oak Tree B
Rainfall amount	High	Low
Soil richness	Low	High
Temperature	High	Low



In order to find the impact of rainfall, soil richness, and temperature, we will plot the growth of each individual oak in the region compared to its environmental conditions.



Now that we have identified growth trends for each of these environmental factors, we need to convert them into a form usable for our predictions.

Rainfall	Low	Medium	High
Growth in inches relative to the average	-5	-2	+3

Soil Richness	Low	Medium	High
Growth in inches relative to the average	-3	-1	+2

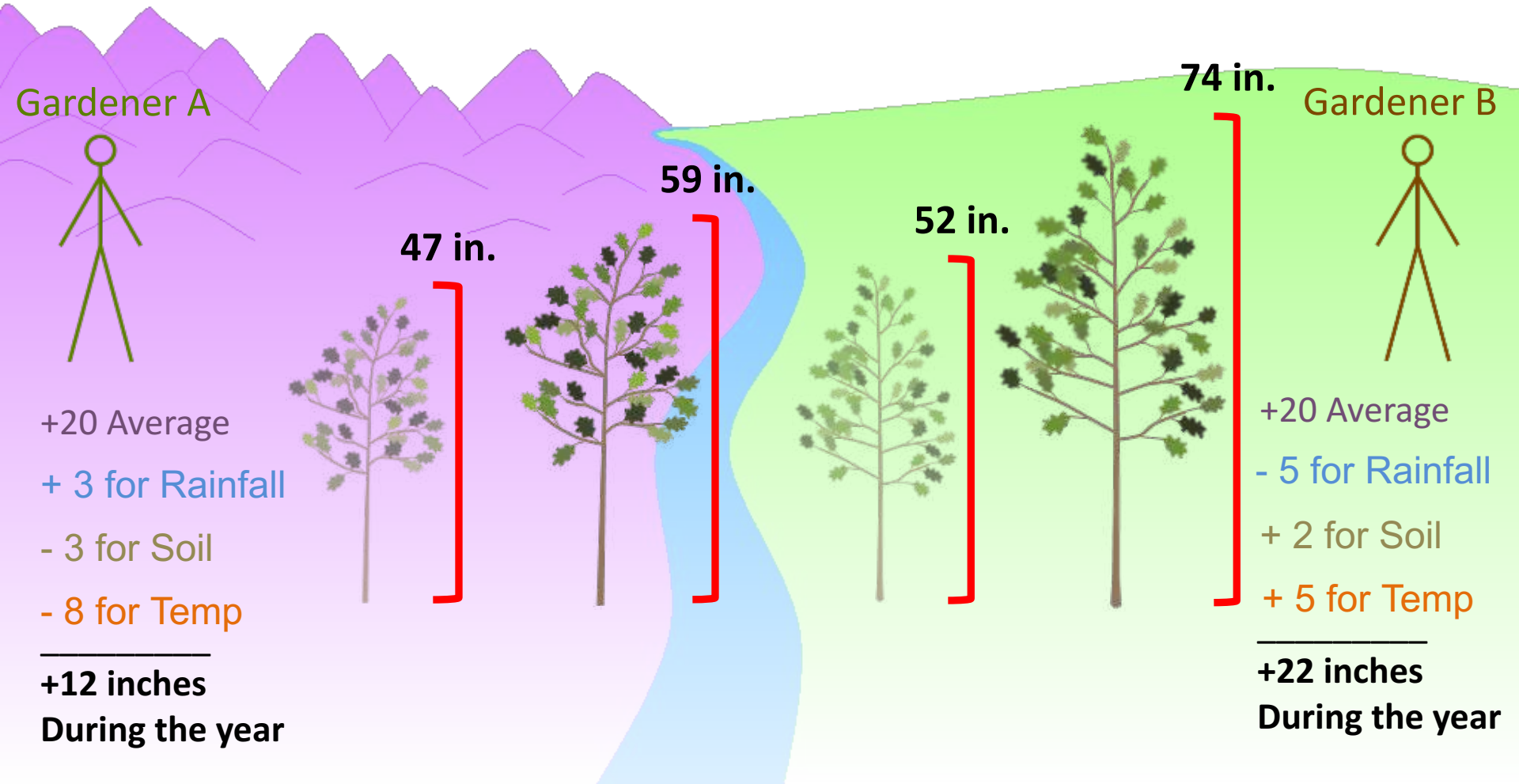
Temperature	Low	Medium	High
Growth in inches relative to the average	+5	-3	-8

Now we can go back to **Oak A** and **Oak B** to **adjust for their growing conditions**.

Now that we have refined our predictions based on the effect of environmental conditions, our gardeners are on a level playing field.

The predicted height for trees in Oak A's conditions is 59 inches.

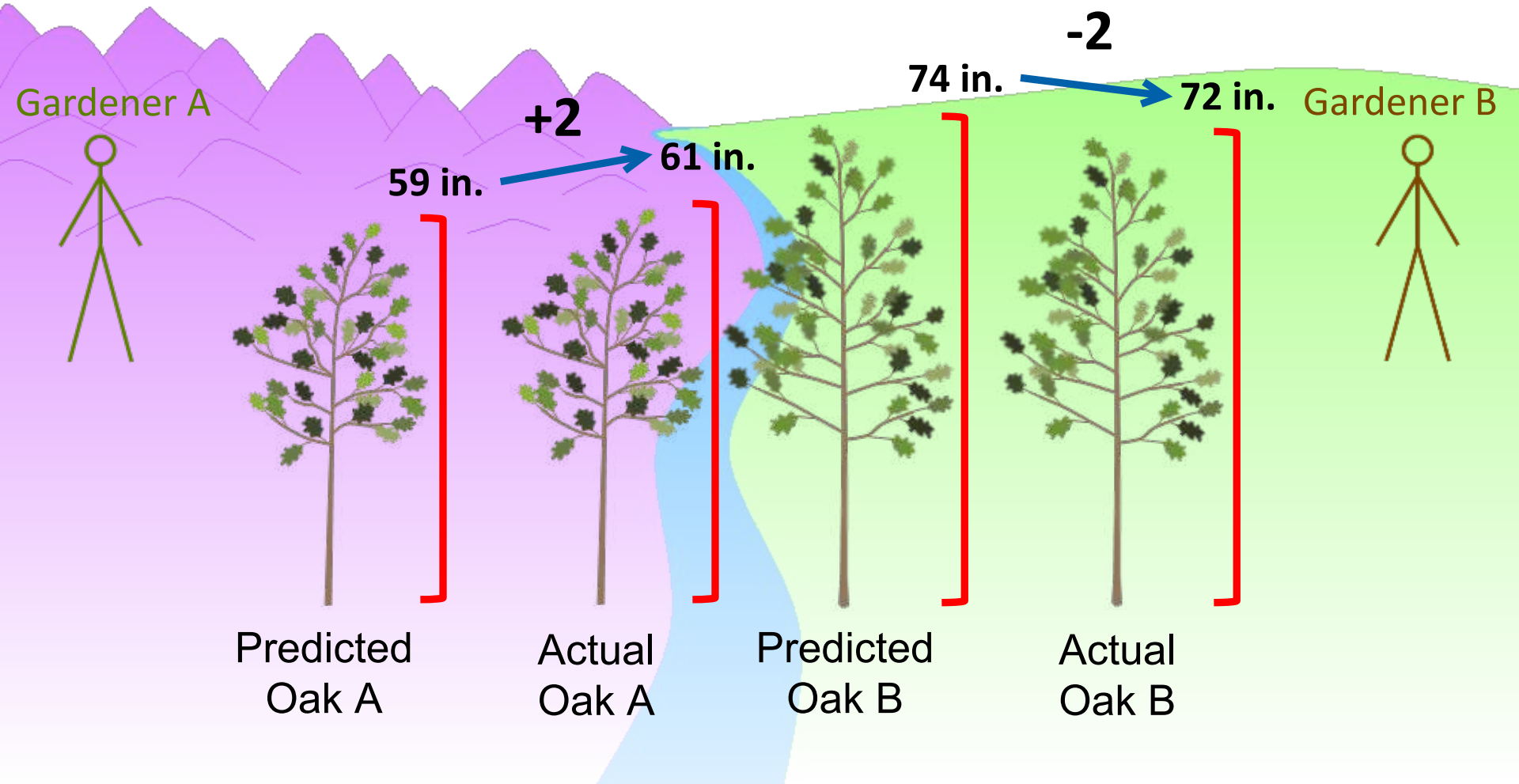
The predicted height for trees in Oak B's conditions is 74 inches.



Finally, we compare the actual height of the trees to our predictions.

Oak A's actual height of 61 inches is 2 inches **more** than we predicted.
We attribute this above-average result to the effect of Gardener A.

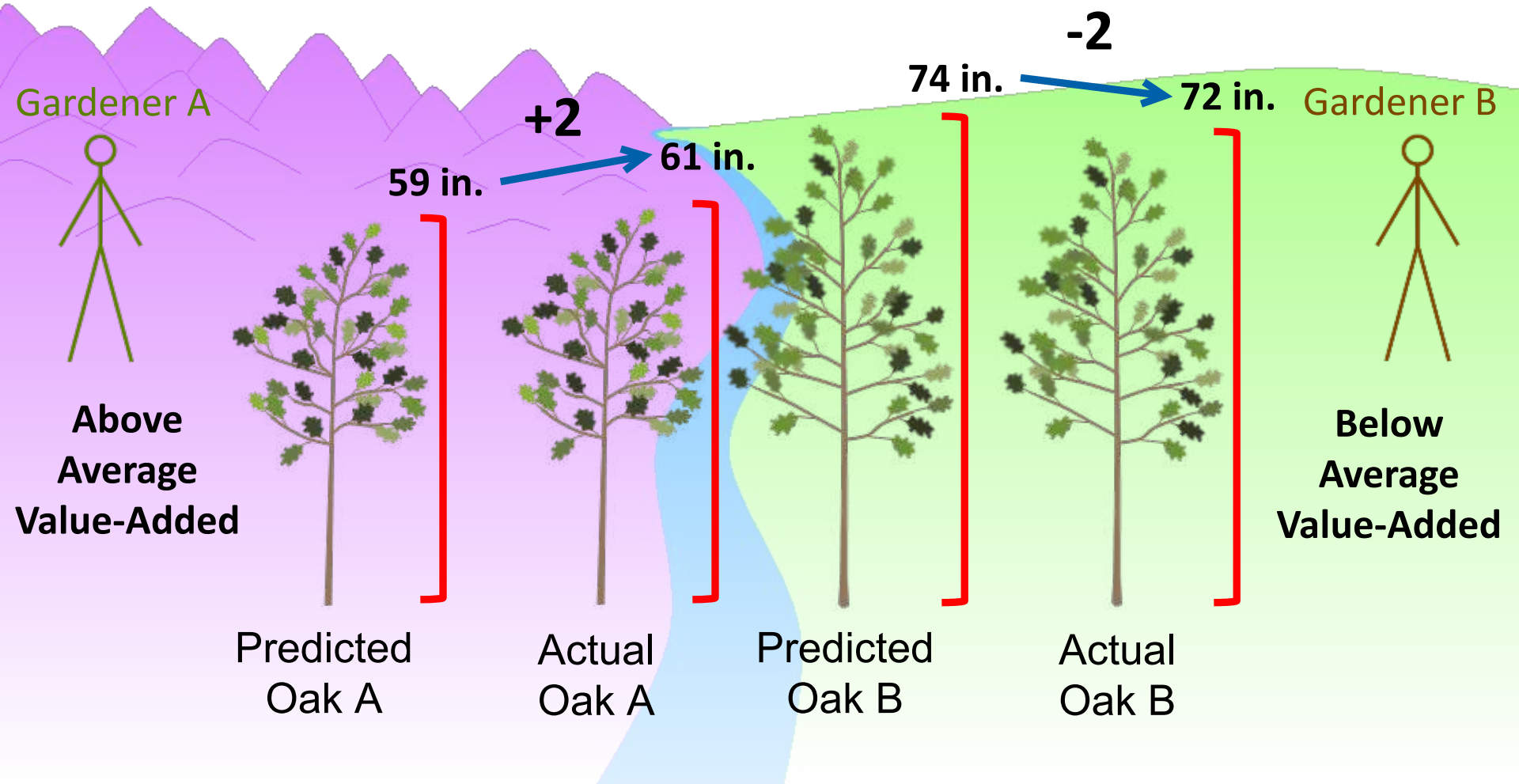
Oak B's actual height of 72 inches is 2 inches **less** than we predicted.
We attribute this below-average result to the effect of Gardener B.



Using this method, **Gardener A is the superior gardener.**

By accounting for last year's height and environmental conditions of the trees during this year, we found the "value" each gardener "added" to the growth of the tree.

This is analogous to **a Value-Added measure**



How does this analogy relate to value added in the education context?

	Oak Tree Analogy	Value-Added in Education
What are we evaluating?	• Gardeners	• Districts • Schools • Grades • Classrooms • Programs and Interventions
What are we using to measure success?	• Relative height improvement in inches	• Relative improvement on standardized test scores
Sample	• Single oak tree	• Groups of students
Control factors	• Tree's prior height • Other factors beyond the gardener's control: • Rainfall • Soil richness • Temperature	• Students' prior test performance (usually most significant predictor) • Other demographic characteristics such as: • Grade level • Gender • Race / Ethnicity • Low-Income Status • ELL Status • Disability Status

Using Value-Added for Decision Making

Appropriate Model Design

Development of a Value-Added System

How complex should a value-added model be?
Possible rule: "Simpler is better, unless it is wrong."

- Clarity: What is the objective?
- Is the model designed to inform that objective?
- Why? Achieve accuracy and fairness

VARC

Design Process: Continuous Improvement

VARC

VA Model Design Features

Inclusion of Student-Level Demographic Characteristics: Technical Validity Criterion

VARC

Inclusion of Student-Level Demographic Characteristics: Consequential Validity Criterion

VARC

Using Value-Added for Decision Making

Varied Uses of the Data

Appropriate Uses of Value-Added

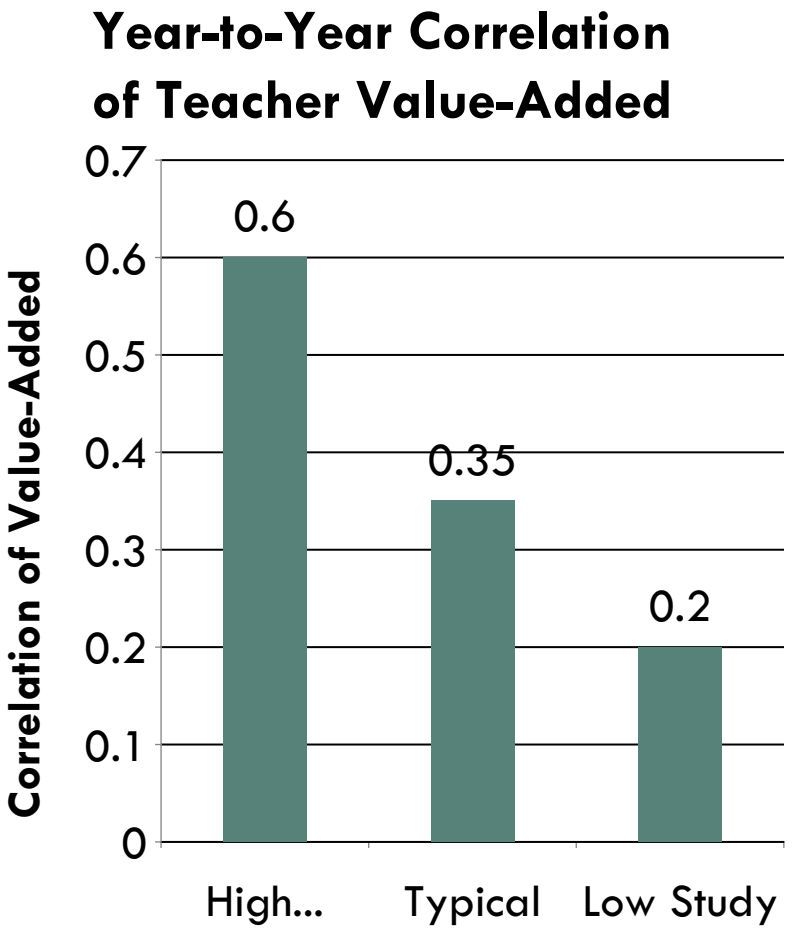
- When used for accountability, Value-Added should always be used with multiple measures
- Our preferred use is as a tool to improve student learning, but Value-Added can be an important piece of measuring educator effectiveness

Value-Added for High Stakes Decisions

- Value-Added is not a perfect measure of teacher effectiveness, but what are the alternatives?

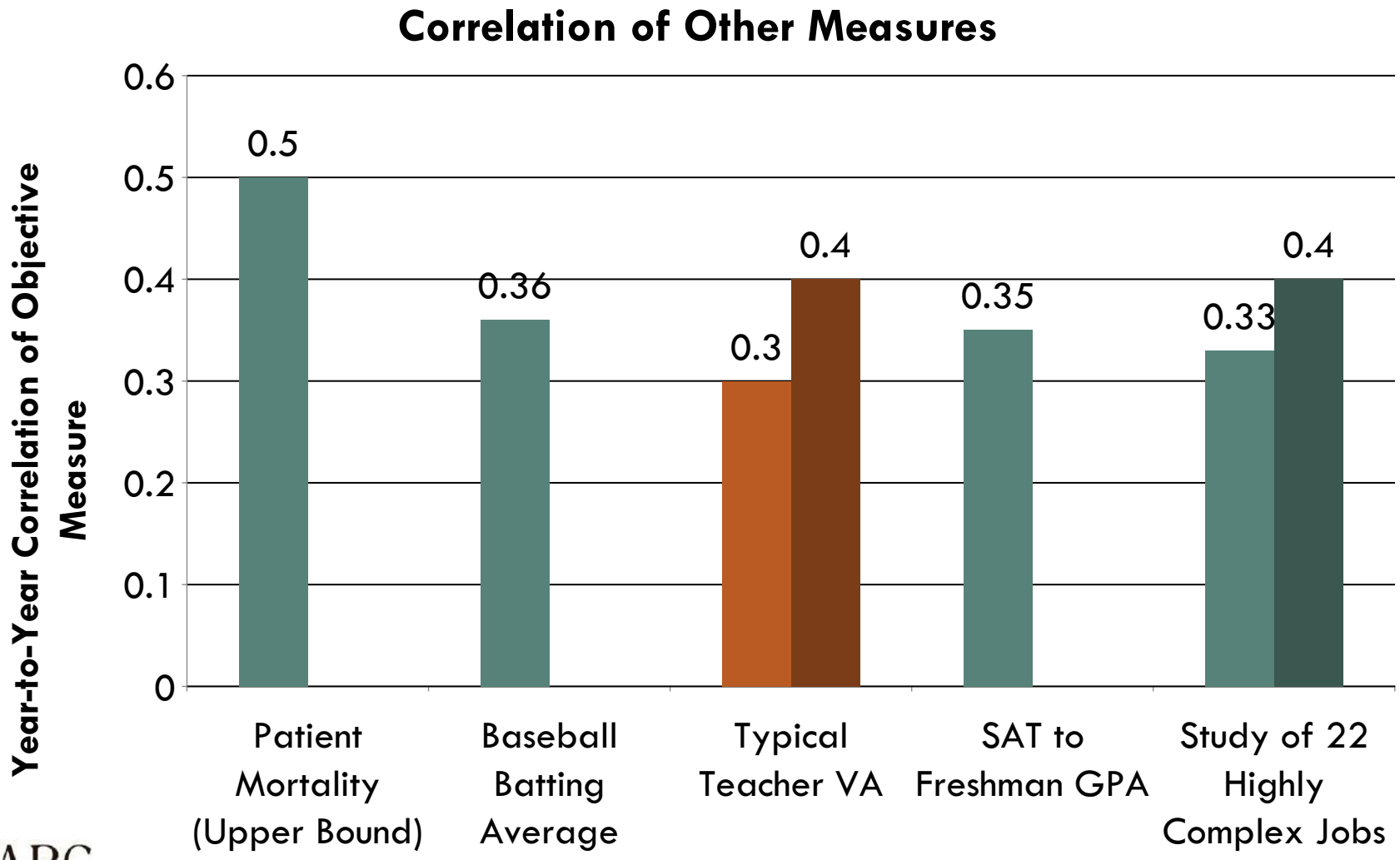
How Stable are Value-Added Measures?

- Year-to-year correlations ranged from 0.2 to 0.6 across multiple studies, with most teachers between 0.3 and 0.4
- Some of this instability is due to variation in teachers' true performance from year to year and some of it is due to error in the measure



VARC
Value-Added Research Center

How Stable are High-Stakes Measures in Other Fields?



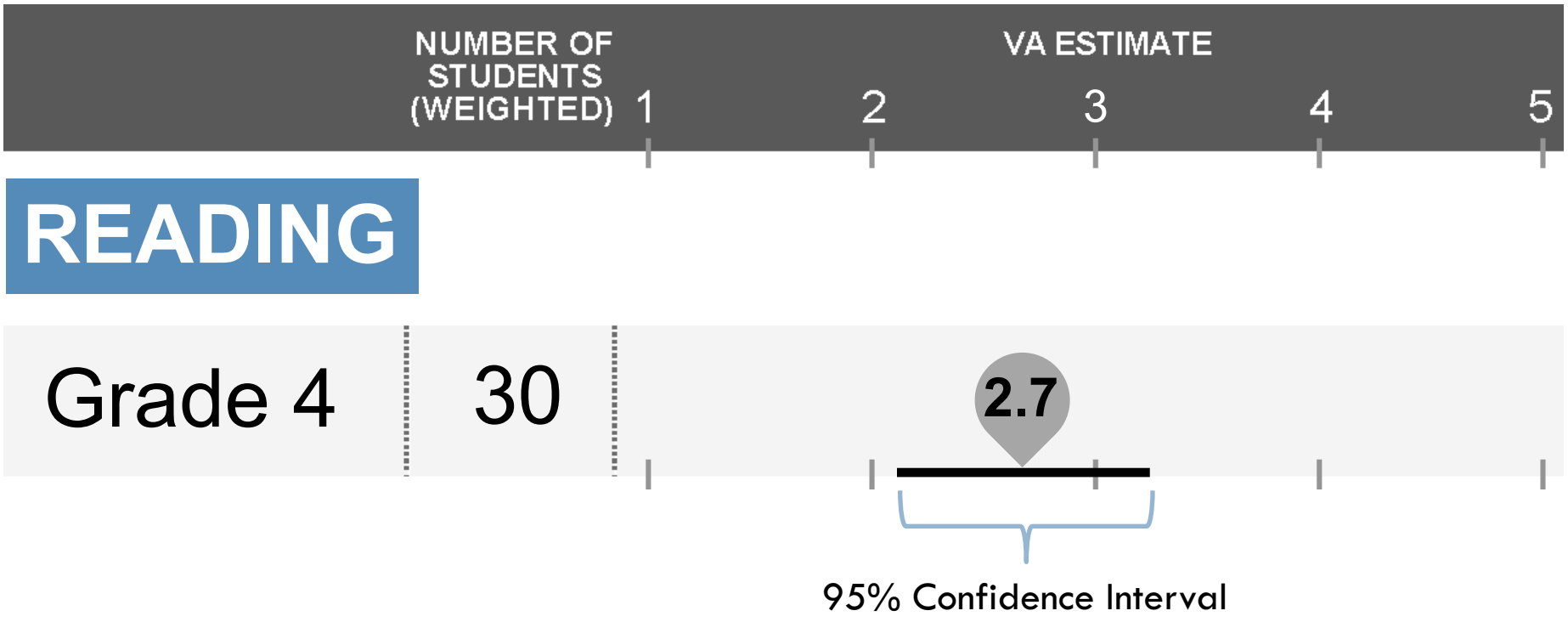
VARC
Value-Added Research Center

Managing and Improving Performance

Uses of Noisy Value-Added Estimates

- Which estimates are "good enough"?
- Good enough for which uses?
  - Trigger for additional professional development
  - Merit pay
  - Tenure decisions
  - Dismissal decisions
- Good enough compared to which alternatives?

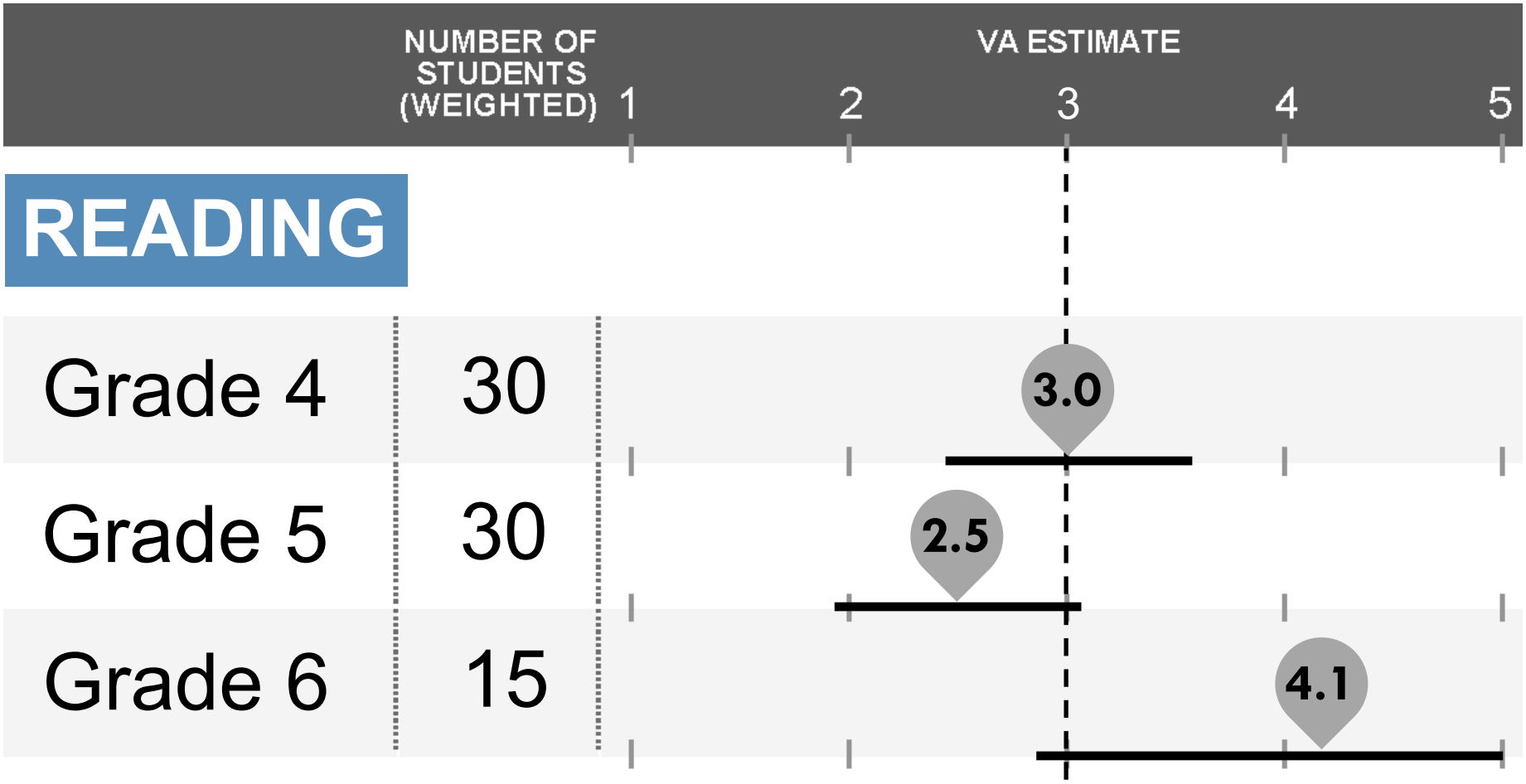
Data Interpretation



Value-Added estimates are provided with a confidence interval.

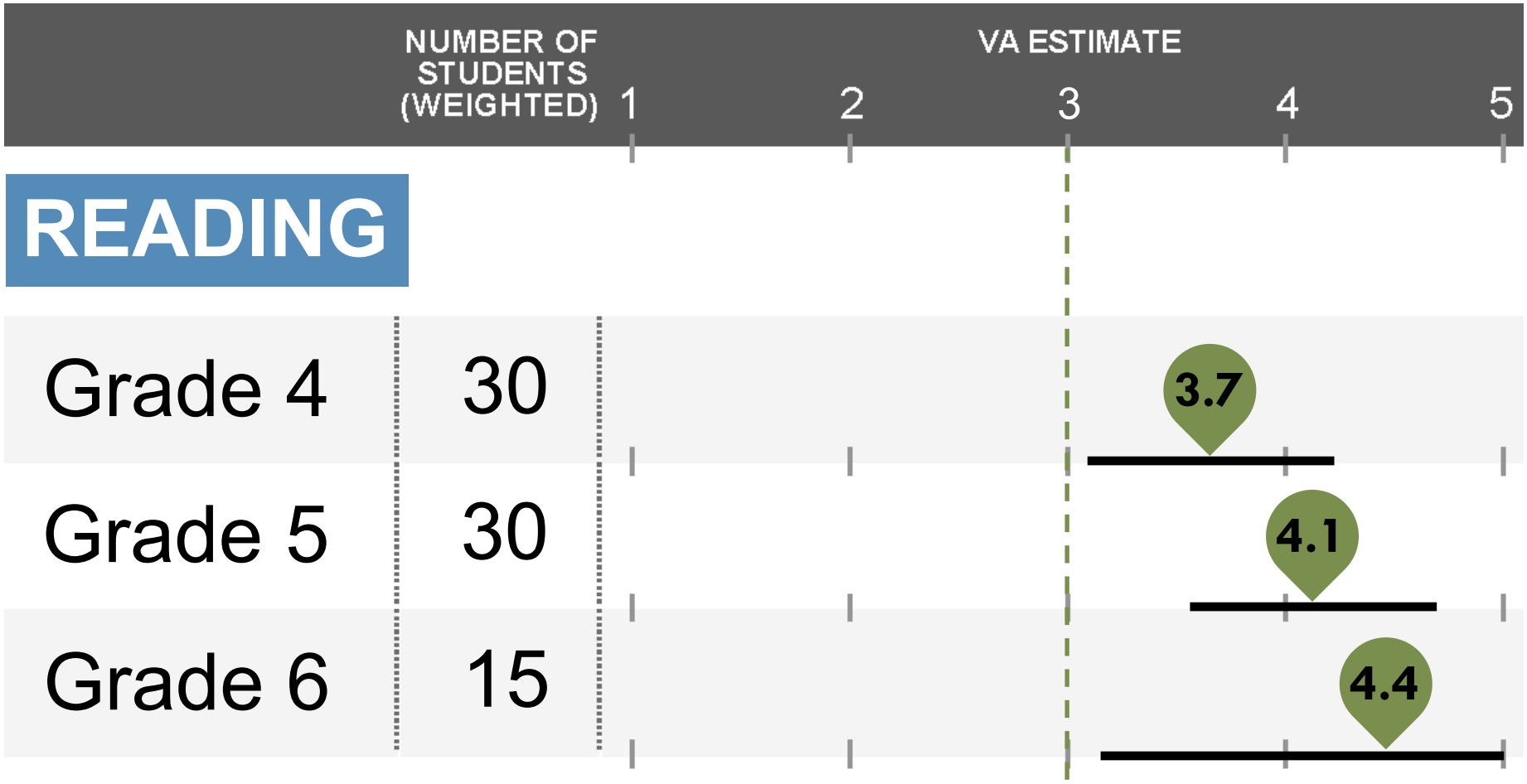
Based on the data available for these thirty 4th Grade Reading students, we are 95% confident that the true Value-Added lies between the endpoints of this confidence interval (between 2.1 and 3.3 in this example), with the most likely estimate being 2.7

Value-Added Color Coding



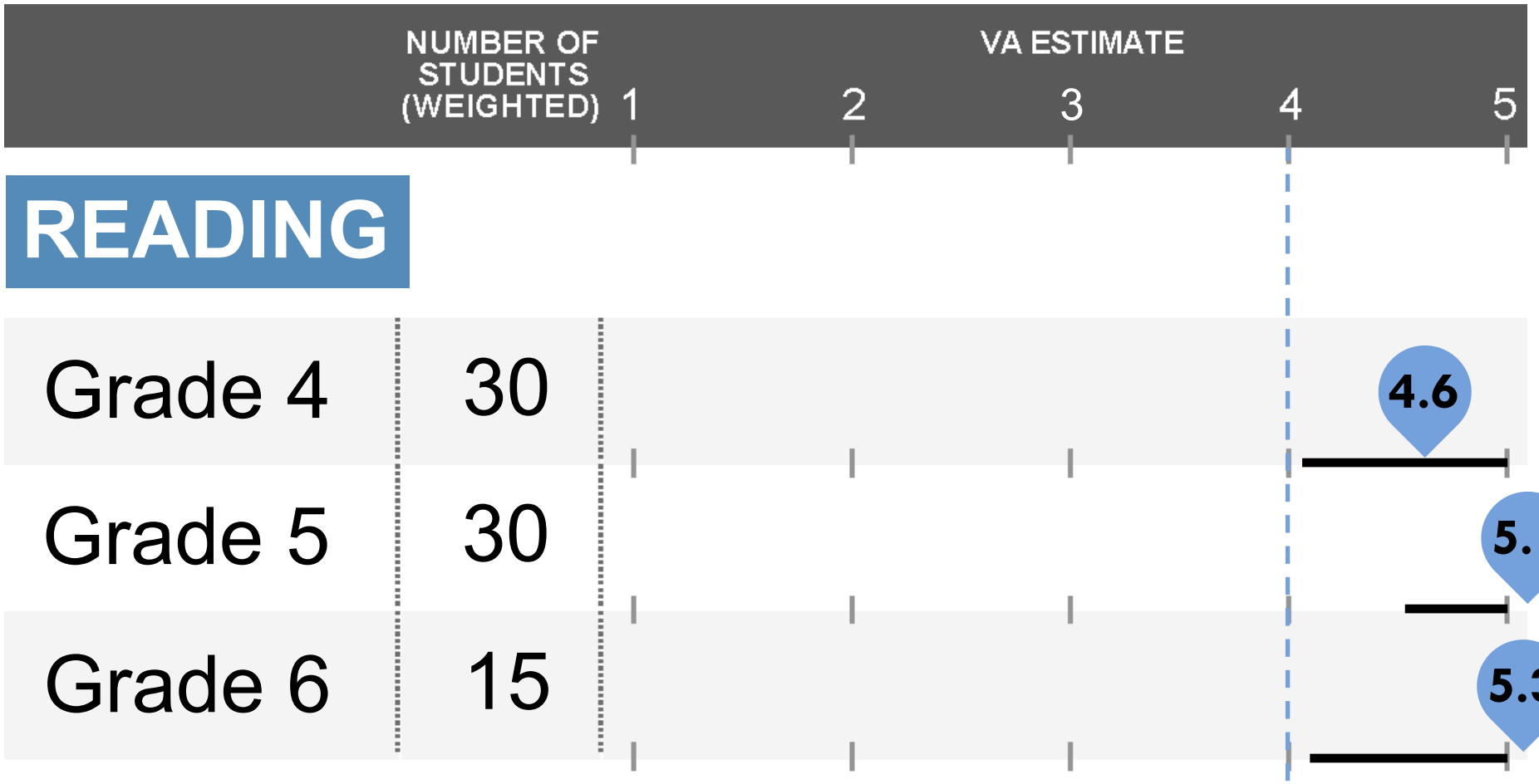
If the confidence interval crosses 3, the color is gray.

Value-Added Color Coding



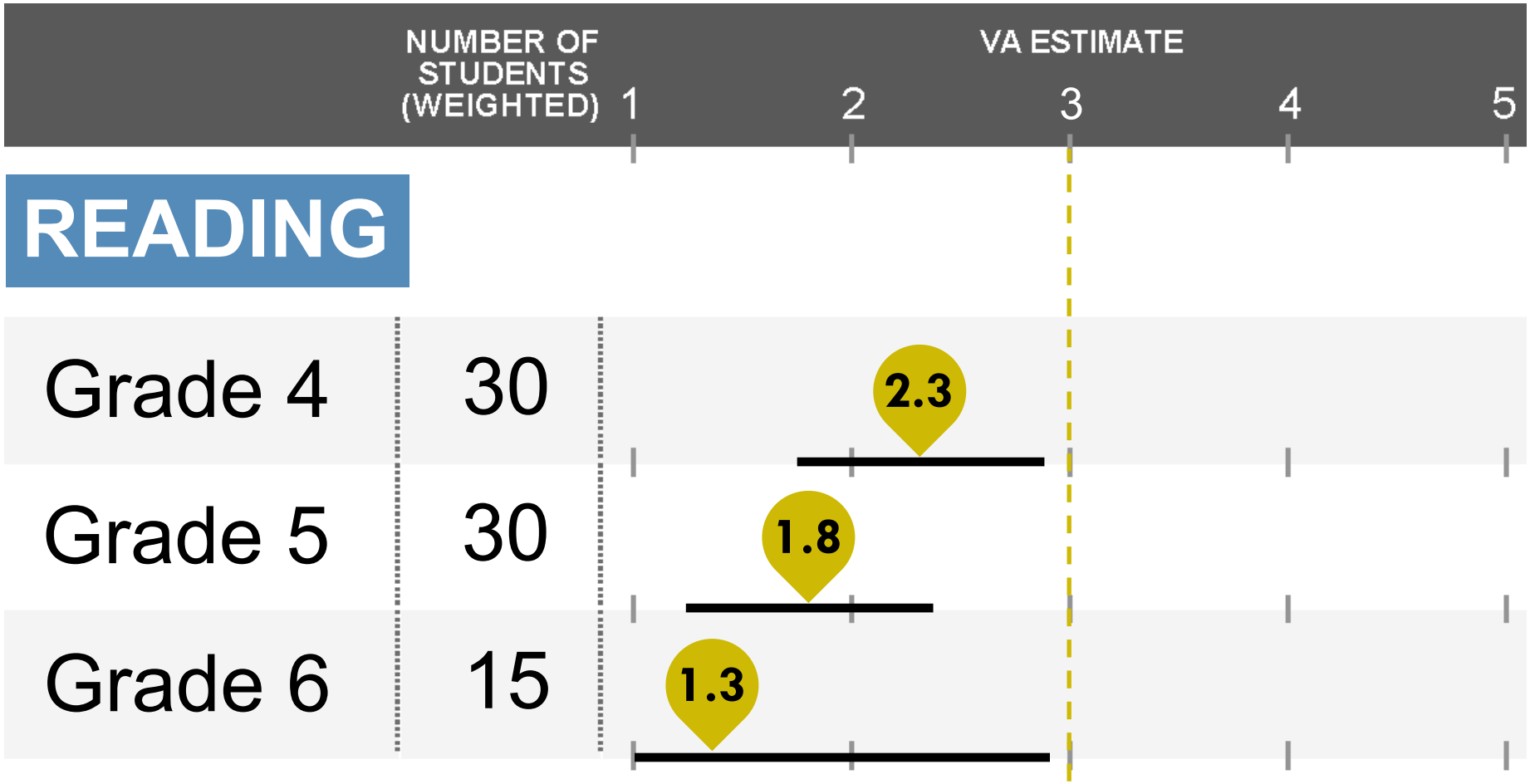
If the entire confidence interval is above 3, the color is green.

Value-Added Color Coding



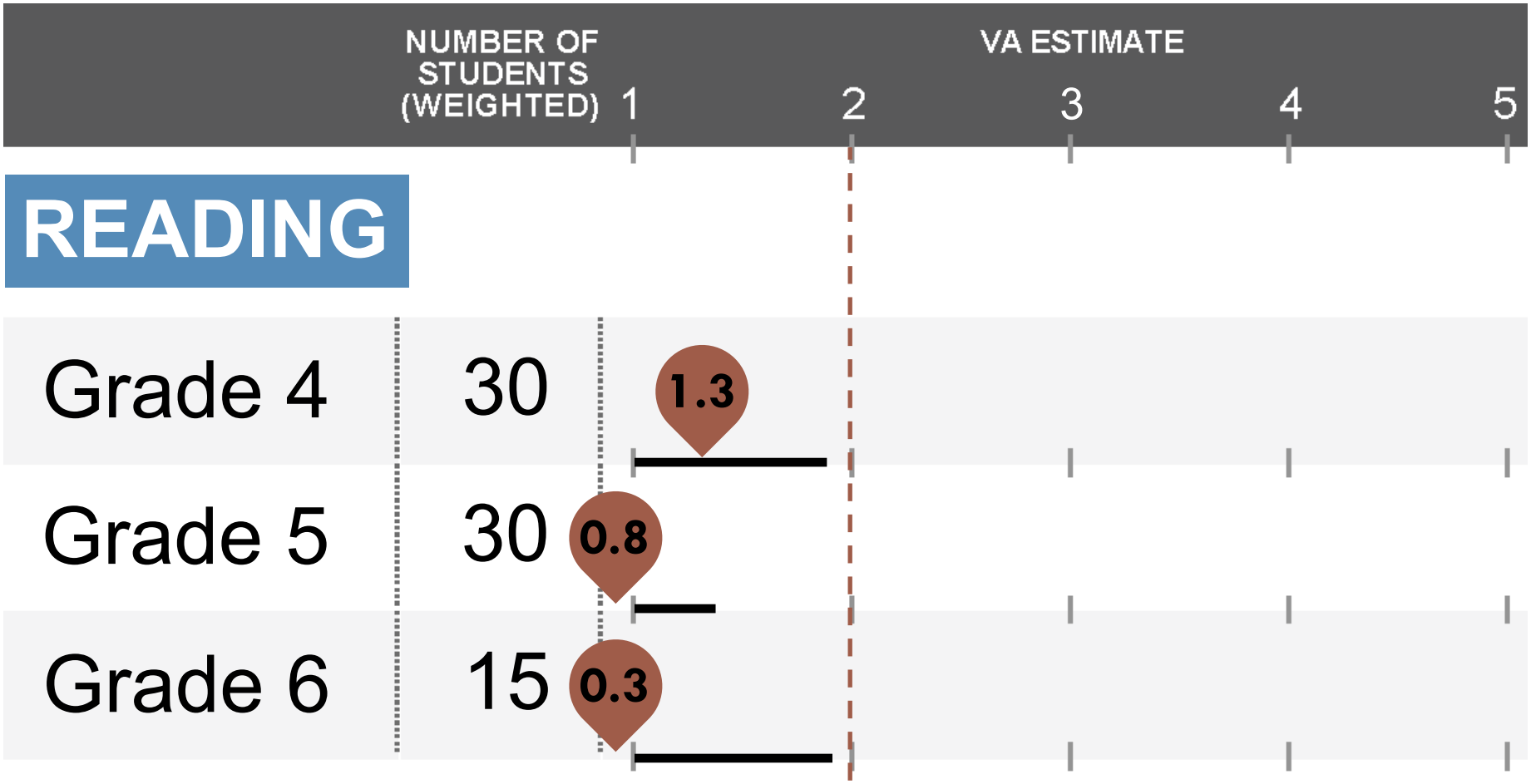
If the entire confidence interval is above 4, the color is blue.

Value-Added Color Coding



If the entire confidence interval is below 3, the color is yellow.

Value-Added Color Coding



If the entire confidence interval is below 2, the color is red.

Value-Added Color Coding

These colors are meant to categorize results at a glance, but making responsible decisions based on Value-Added estimates may require more careful use of the data.

General guidelines:

Green and Blue results are areas of relative strength. Student growth is above average.

Gray results are on track. In these areas, there was not enough data available to differentiate this result from average.

Yellow and Red results are areas of relative weakness. Student growth is below average.

Decision Making Examples

To put this into context, let's go through a few examples of decisions that might be made based on Value-Added results.

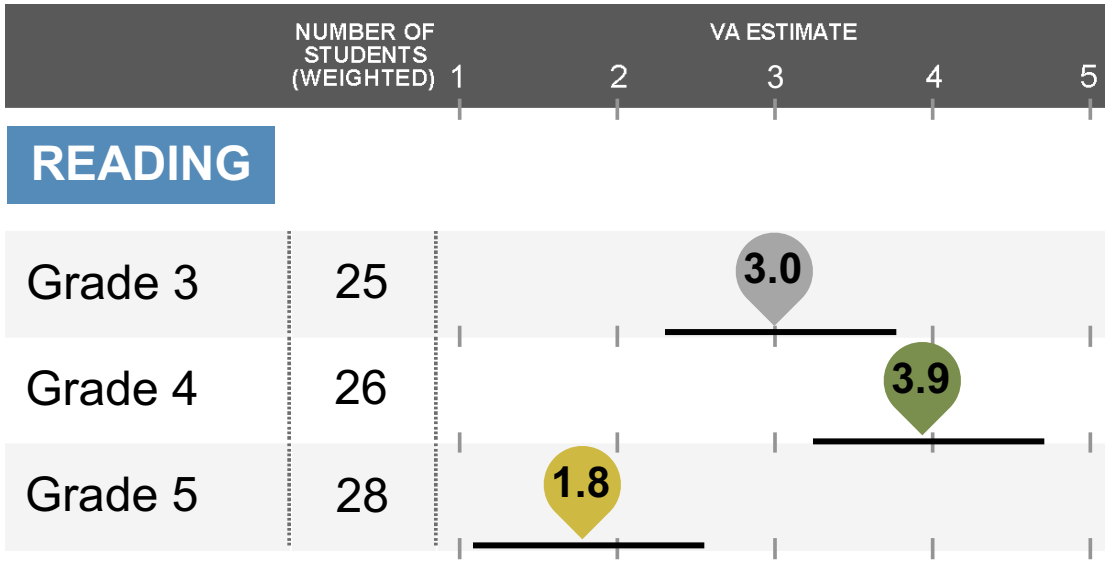
School-Level

- Which grade-level teams should get additional help from a literacy coach?
- How do I interpret gray results, and what can I learn from them?

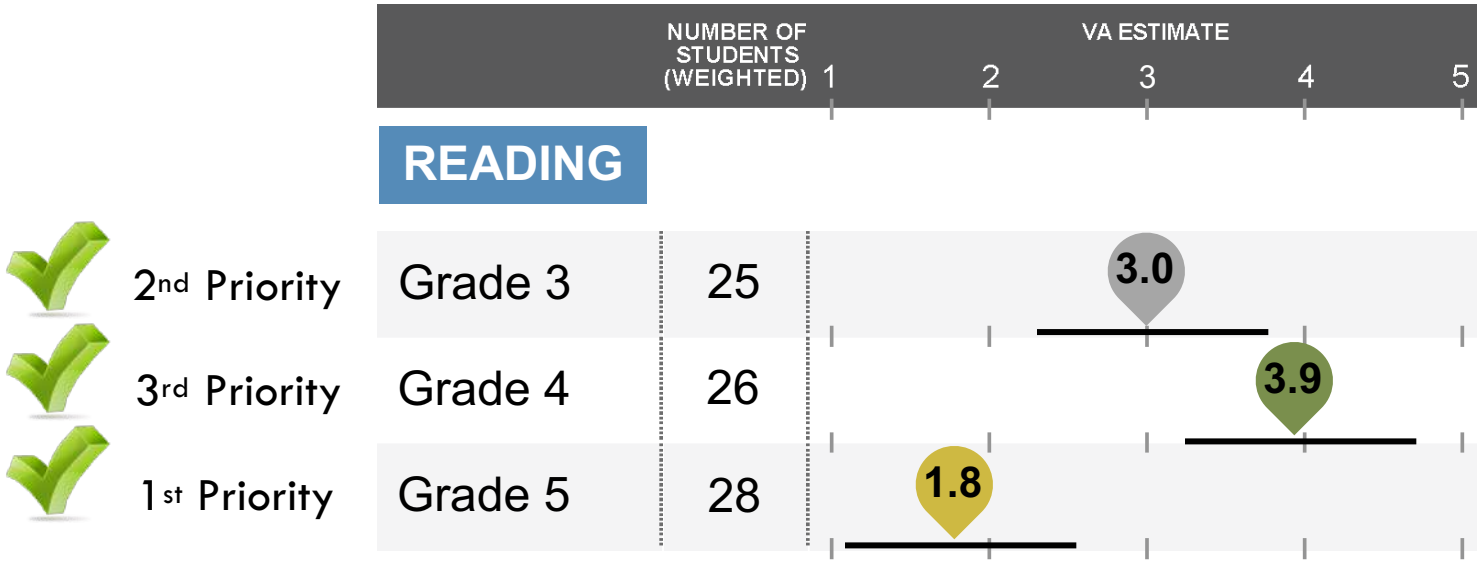
District-Level

- Are there particular schools or groups of schools that require more support?

Which grade-level teams should get additional help from a literacy coach?



Which grade-level teams should get additional help from a literacy coach?



This is a relatively low-stakes decision.

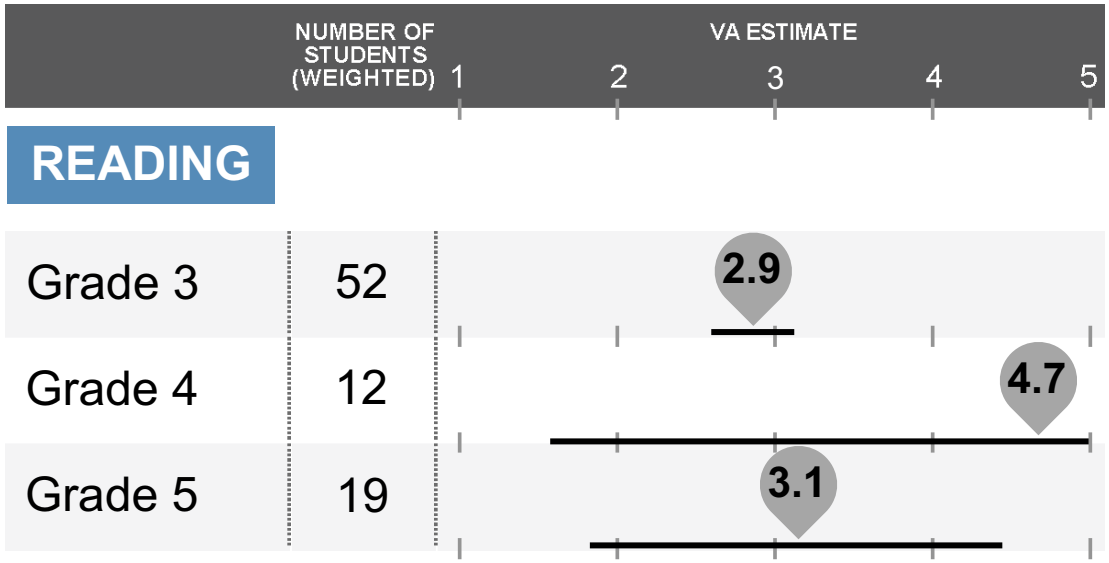
A literacy coach may be beneficial to any of these teams. There is little risk in providing this resource to all the teachers.

The limiting factor is likely to be availability of this resource. If possible, provide it to all teachers, but limited allocation may be based on area of most need.

How do I interpret gray results, and what can I learn from them?

These three teams each have gray estimates.

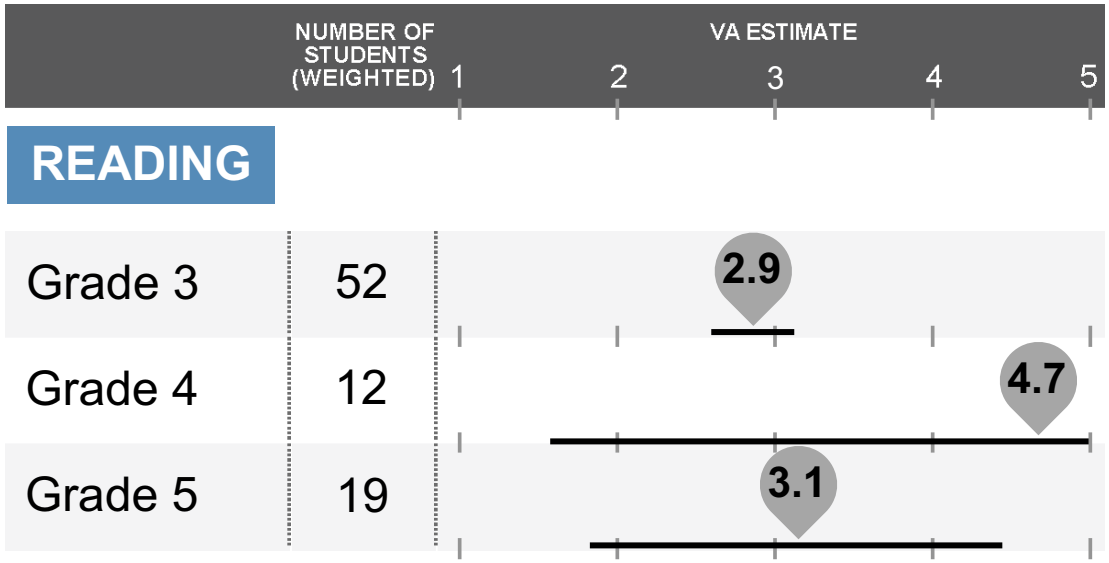
Would you interpret them the same way?



How do I interpret gray results, and what can I learn from them?

These three teams each have gray estimates.

Would you interpret them the same way?



Grade 3 – The tight confidence interval around the gray estimate indicates we can be confident that this team's Value-Added was close to average.

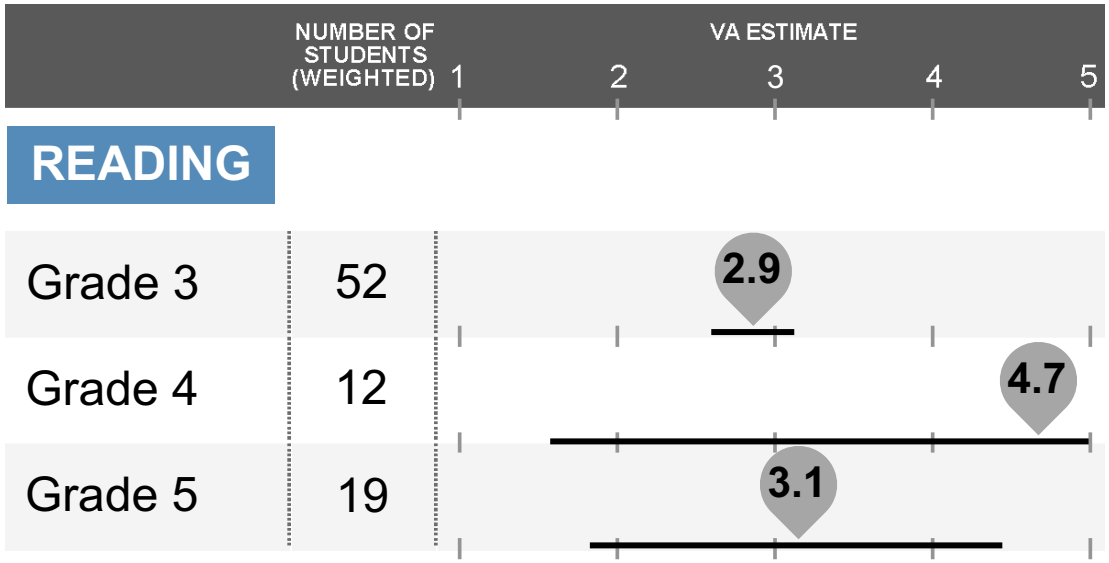
Grade 4 – The best estimate of Value-Added is above average (4.7). However, since it was based on a small amount of data (12 students), we cannot say with confidence that it was above average. This team may actually have below average Value-Added.

Grade 5 – The best estimate is average Value-Added (3.1). However, the wide confidence interval indicates that there was not enough data to rule out above or below average Value-Added.

How do I interpret gray results, and what can I learn from them?

These three teams each have gray estimates.

Would you interpret them the same way?

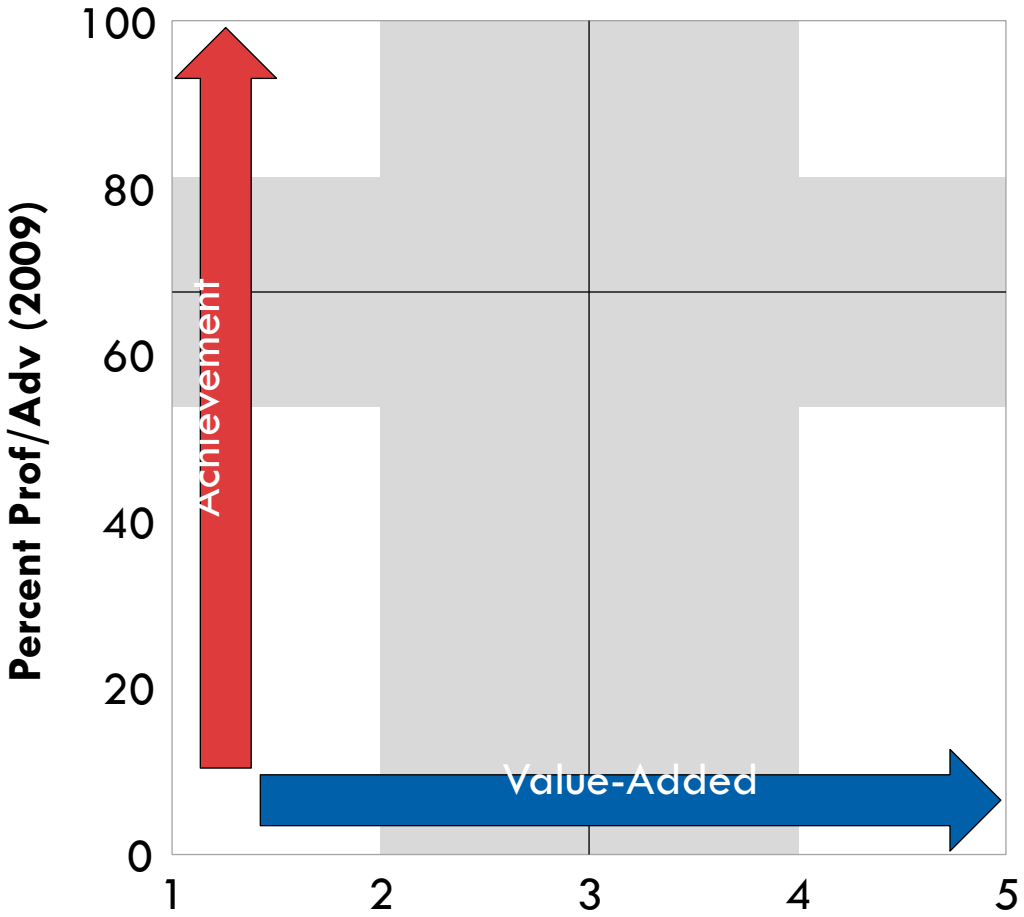


As always, consider multiple data sources when making decisions.

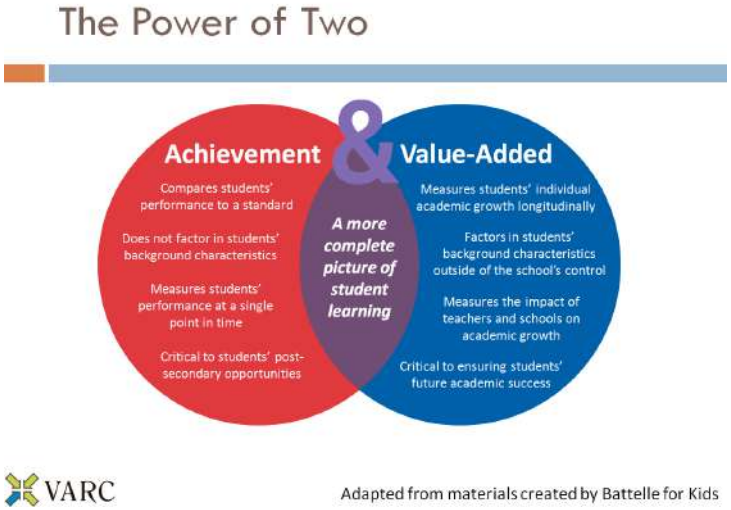
The 3rd grade team has the most certain Value-Added estimate can be treated as one of the average teaching teams in the state.

The 4th and 5th grade teams have less certain estimates and it is particularly important that additional sources of information are considered before making decisions about professional development, resource allocation, staffing assignments, etc.

Quadrant Analysis with Scatter Plots

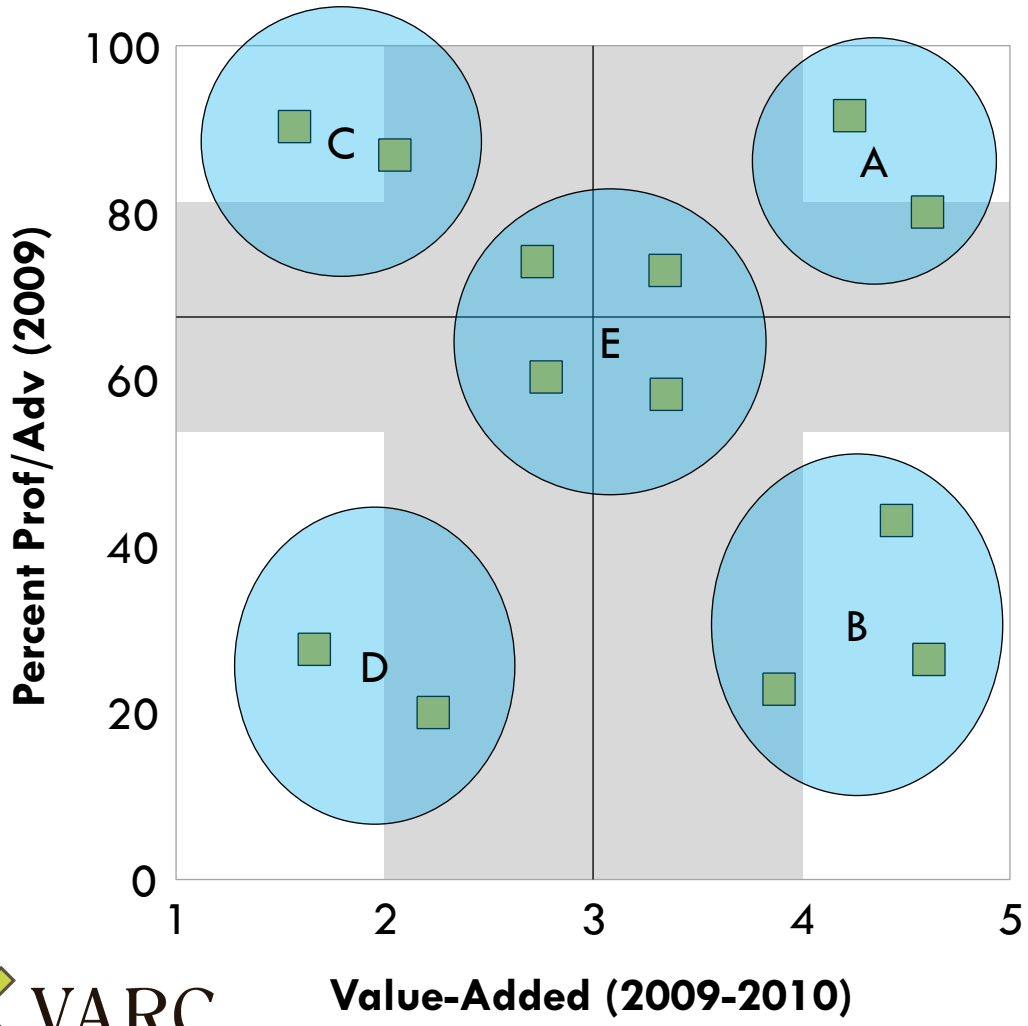


These scatter plots are a way to represent **Achievement** and **Value-Added** together



VARC

How to Read the Scatter Plots



A. Students know a lot and are growing faster than predicted

B. Students are behind, but are growing faster than predicted

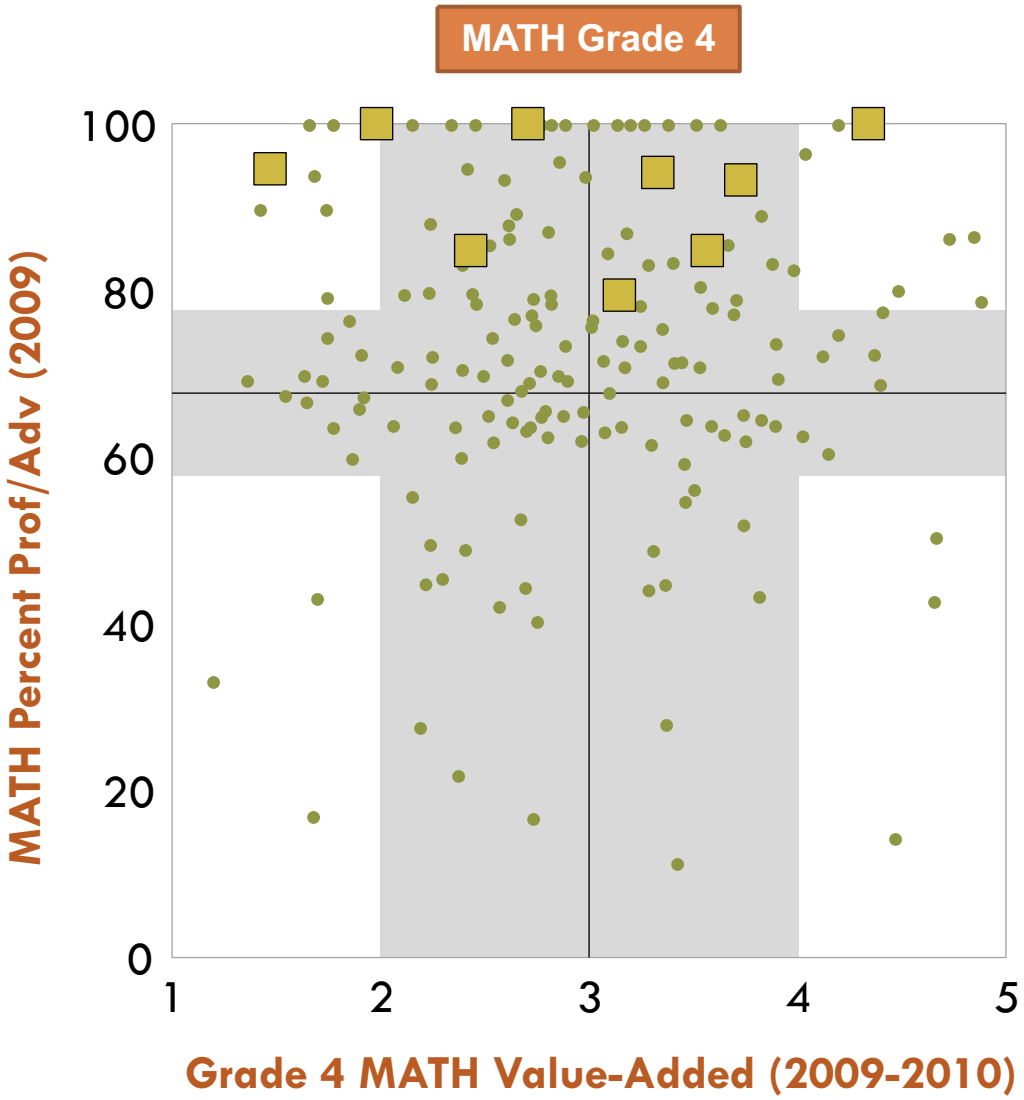
C. Students know a lot, but are growing slower than predicted

D. Students are behind, and are growing slower than predicted

E. Students are about average in how much they know and how fast they are growing

VARC

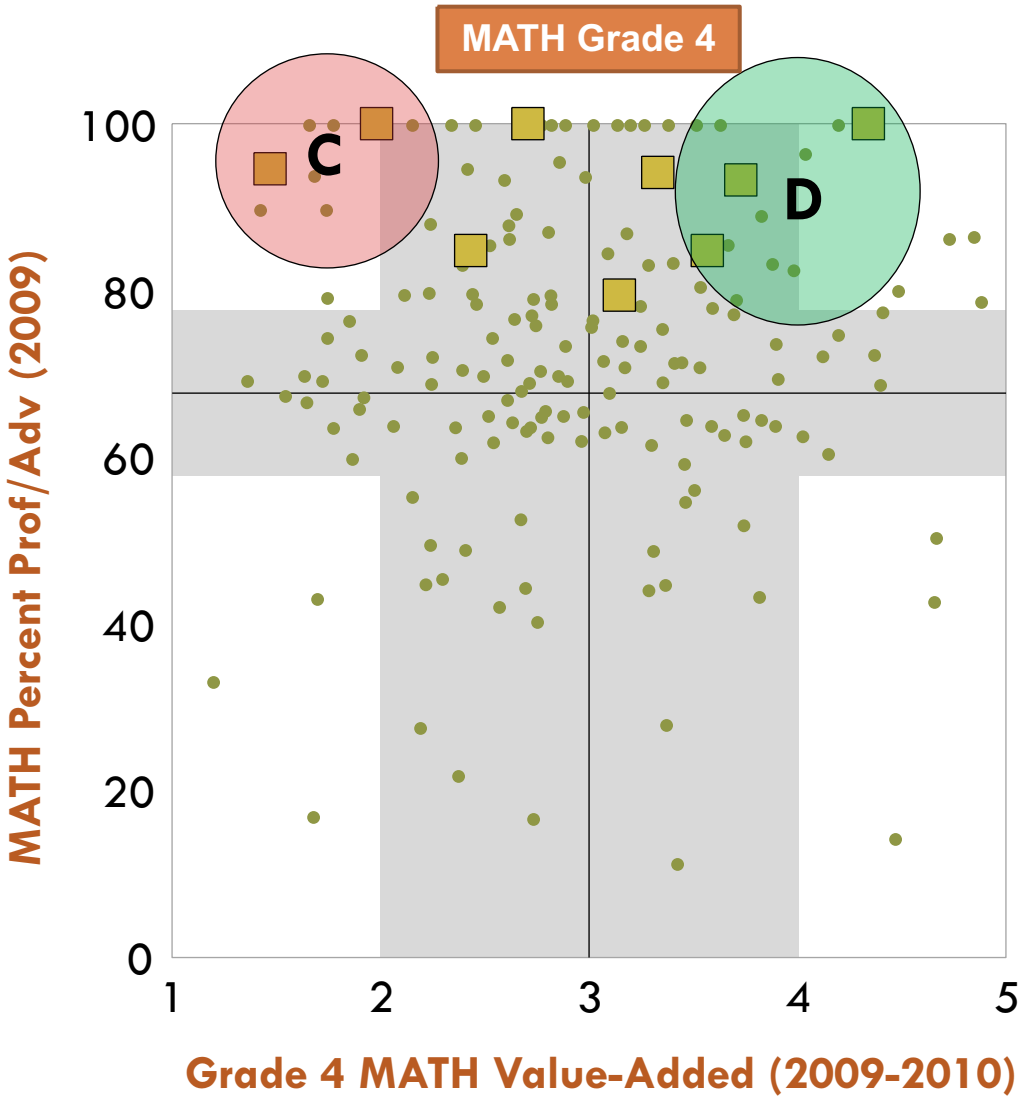
Are there particular schools or groups of schools that require more support?



This scatter plot shows a fairly high performing district.

Schools in your district
Schools in the state

1. Are there particular schools or groups of schools that require more support?



- What would you tell a principal in group C who said their Value-Added was low because their students had no room to grow on the test?
- How can we learn from the success of group D and bring that knowledge to group C?
- Are there programs or resources that group D is receiving that we could also provide to group C?

Attainment or Value-Added

Based on these examples, which type of data do you think districts should focus on providing to educators:

A) Attainment Data

B) Value-Added Data

VARC

Combining Multiple Measures of Teacher Effectiveness

Combining Performance Measures

Schemes for Combining Multiple Measures of Effectiveness

VARC

Thank You