University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

DBER Speaker Series | Discipline-Based Education Research Group

3-2014

Data Connections RETA: DBER, Quality Improvement in Education and Statistical Modeling

Walter Stroup
University of Nebraska - Lincoln, wstroup1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/dberspeakers

Stroup, Walter, "Data Connections RETA: DBER, Quality Improvement in Education and Statistical Modeling" (2014). *DBER Speaker Series*. 58.
http://digitalcommons.unl.edu/dberspeakers/58

This Presentation is brought to you for free and open access by the Discipline-Based Education Research Group at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in DBER Speaker Series by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Data Connections RETA

DBER, Quality Improvement in Education and Statistical Modeling

Walt Stroup

Professor, Department of Statistics, UNL

Data Connections

- $1.2 million NSF RETA (Research and Evaluation Technical Assistance), 2011-2014
- Partnership between University of Nebraska-Lincoln (UNL) and Lincoln Public Schools (LPS)
- Focused on developing, evaluating and sharing statistical models to better estimate **value-added** teacher effects on student learning

  "Coherent picture of teaching and learning"

Time Line I

- **2004-** : Math in the Middle; Nebraska Math; Statistics Department GTA Training; collaboration with Math, TLTE, English

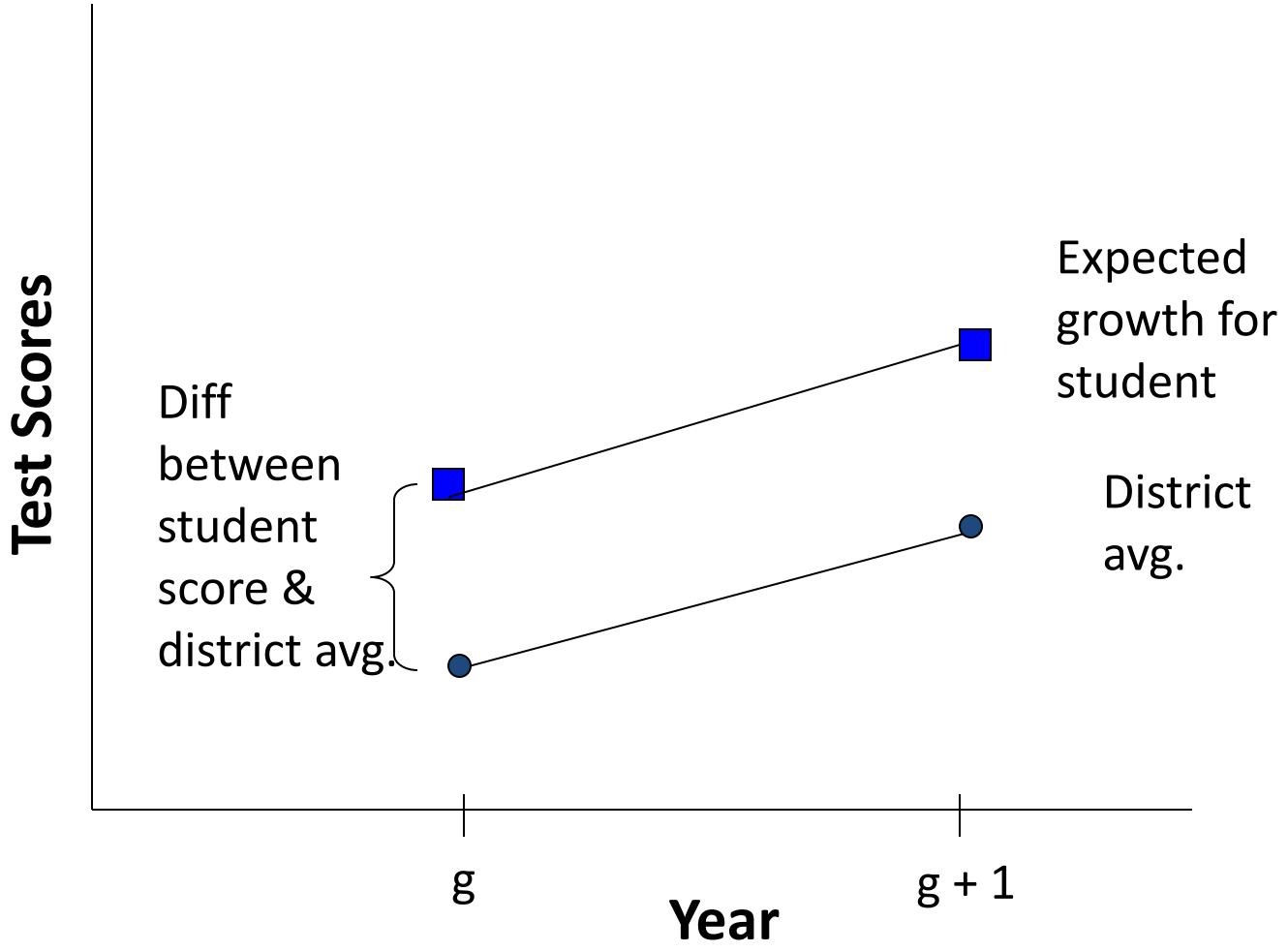
Time Line I

- **2004-** : Math in the Middle; Nebraska Math; Statistics Department GTA Training; collaboration with Math, TLTE, English
- **2009**: At NSF-MSP conference, Dept of Ed in new Obama admin speaks of using data to identify successful MSPs to scale up
- **2009**: problem – then existing statistical methods to do so were underdeveloped, controversial, poorly understood
- much **data-free ideology**

Time Line II

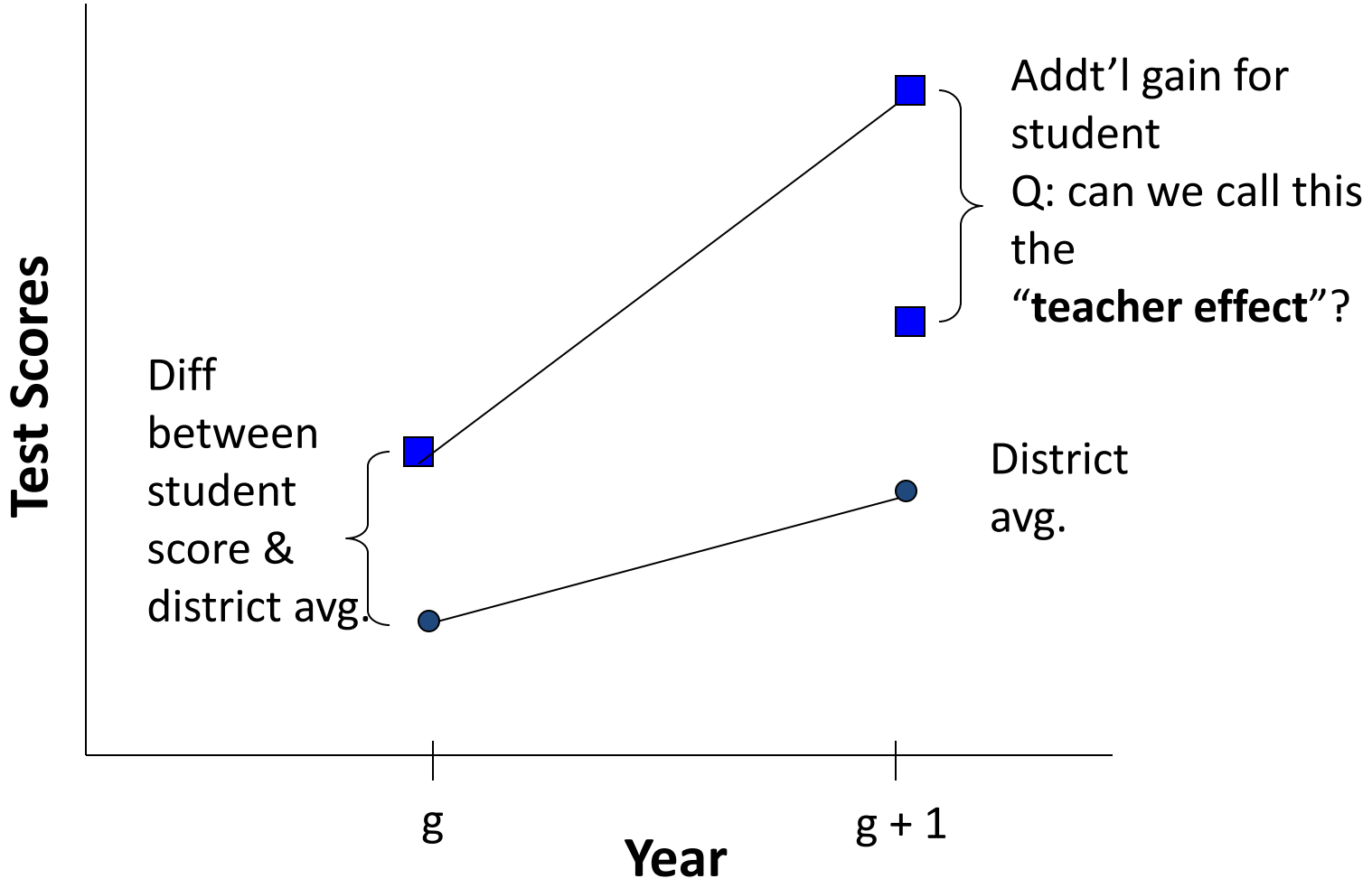
- 2011: received RETA grant
- back to 1980s
  - value added models (VAMs)
  - origins: W. L. Sanders in Knoxville, TN
  - UTK & Knox County schools
- 1990s to present
  - increased use of VAMs in education
  - many states mandate their use for evaluation
  - close VAM/No Child Left Behind/Race to the Top connection

What is Value-Added?



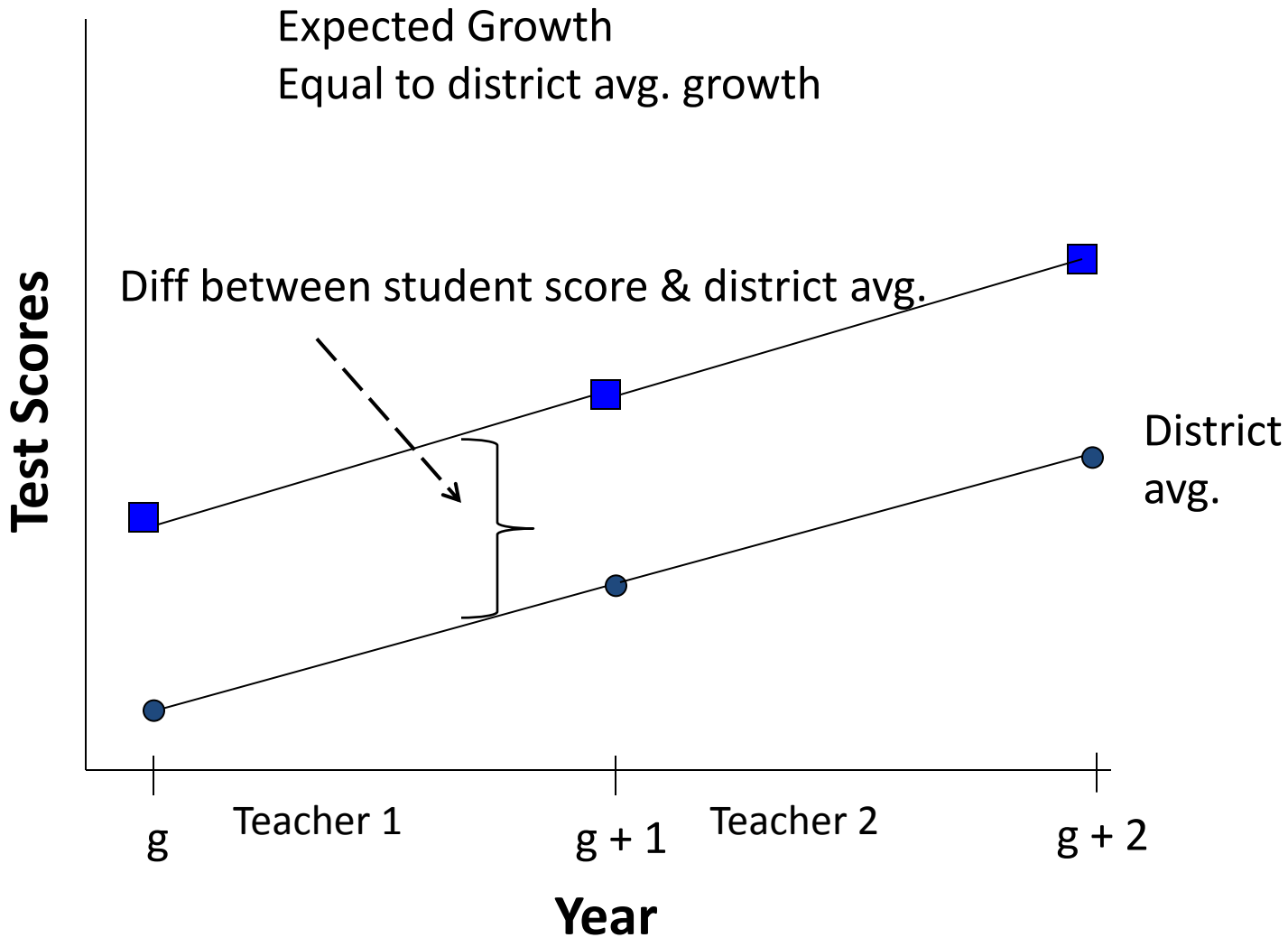
Ballou, Sanders, & Wright (2004)

What is Value-Added?

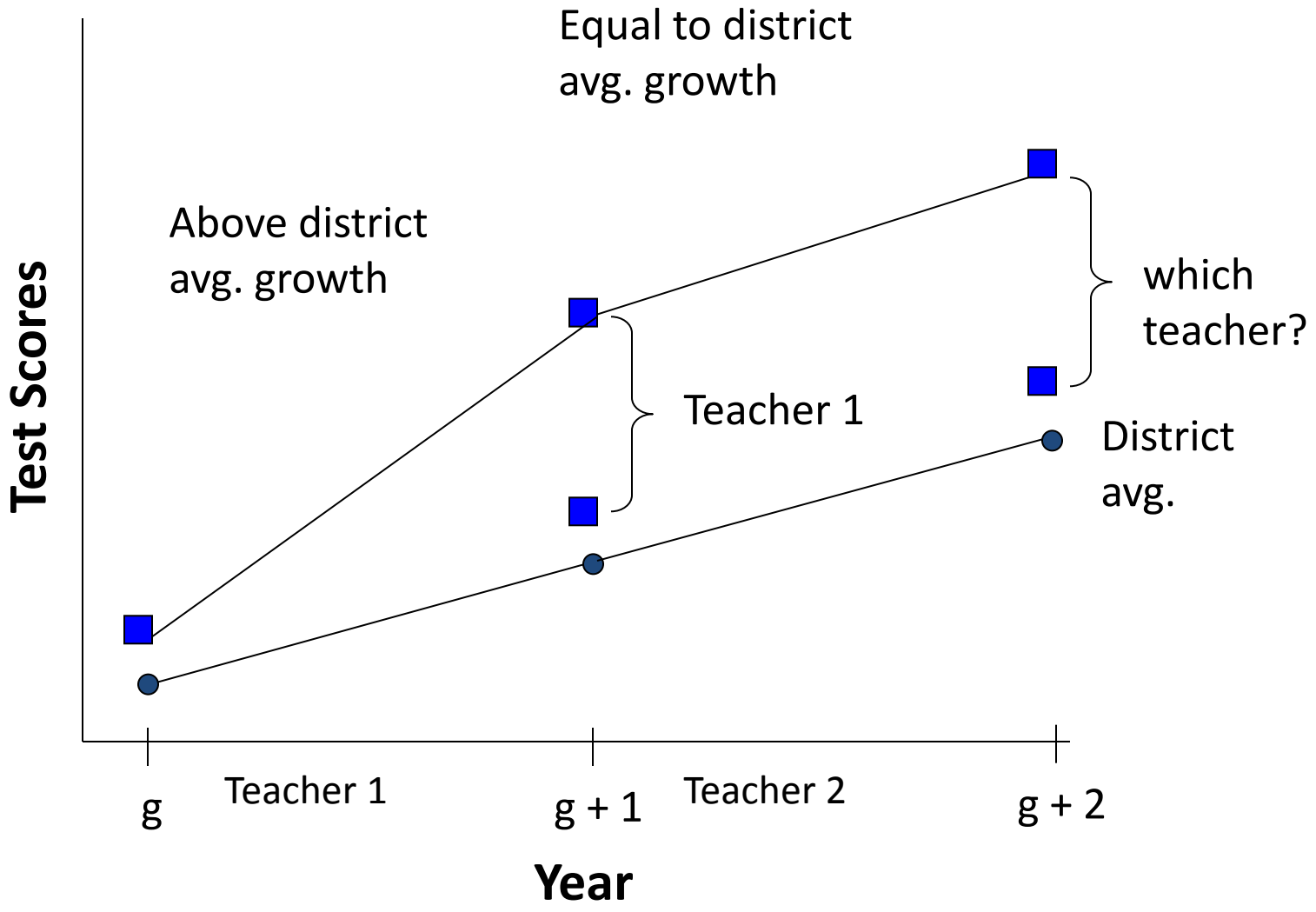


Ballou, Sanders, & Wright (2004)

What is a Layered Model?



What is a Layered Model?



What is a Layered Model?

- Usual statistical model

$$score_{g+1} = \mu + student + teacher_1$$

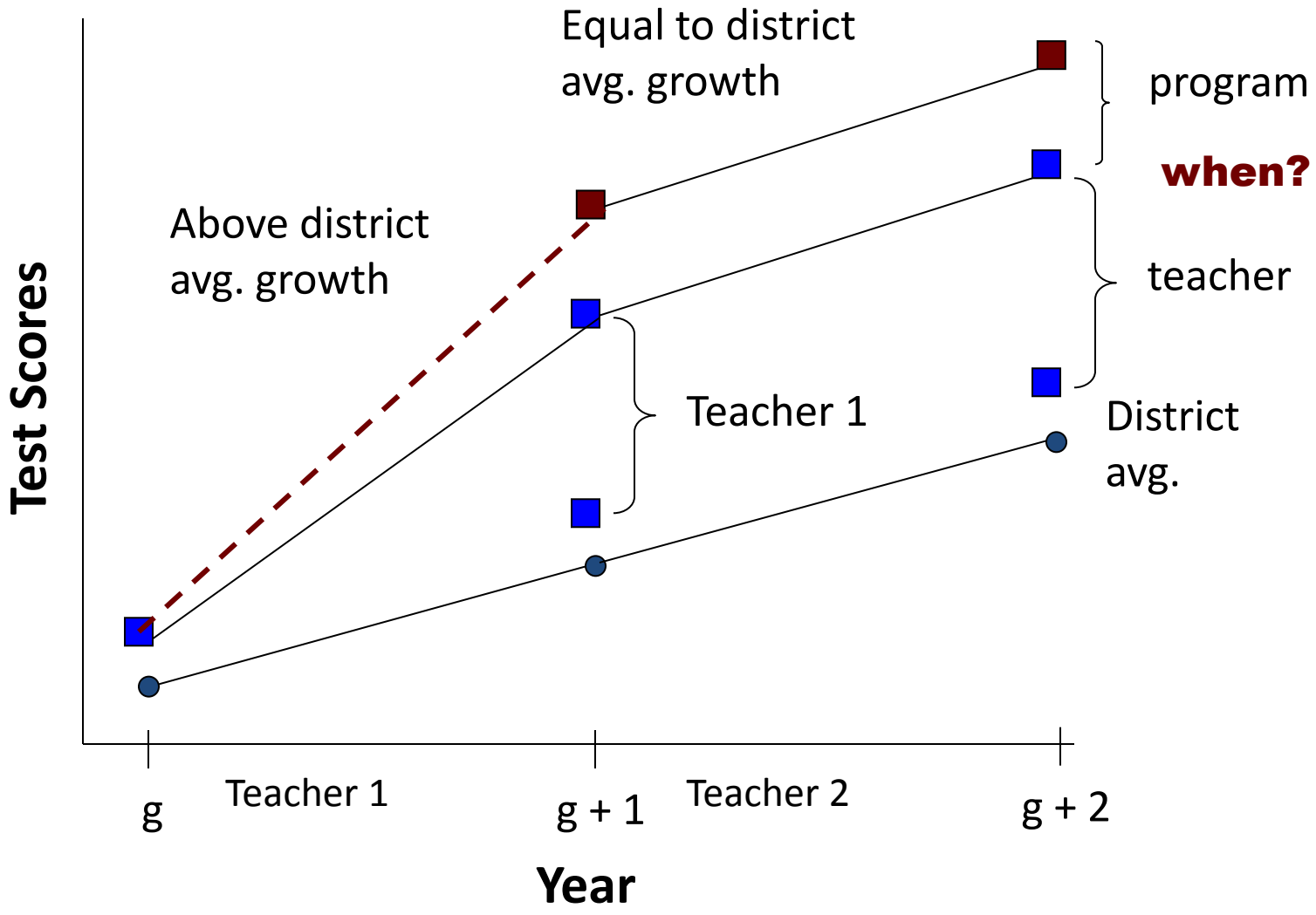
$$score_{g+2} = \mu + student + teacher_2$$

- Layered model

$$score_{g+1} = \mu + student + teacher_1$$

$$score_{g+2} = \mu + student + teacher_1 + teacher_2$$

What is a Program Effect?



What is a Program Effect?

- Layered Model with Program Effect

$$score_{g+1} = \mu + student + teacher_{1,P}$$

$$score_{g+2} = \mu + student + teacher_{1,P} + teacher_{2,N}$$

- Definition?

$$program\ effect = teacher_{1,P} - teacher_{1,N}$$

- ➔ For teachers in the program
  - you need to know their effect **before** as well as **during** the program
  - you need some assurance that their effect is stable

Two Statistical Issues

- Fixed versus Random Model Effects
- Impact of type of effect on how we estimate
  - teacher effect
  - program effect

Types of Model Effects

- If multiple studies done independently would all studies use the same ***levels*** (e.g. in pgm or not)?
- Anything special about levels in the study?
- Do the levels represent a target population?

Types of Model Effects

- If multiple studies done independently would all studies use the same ***levels*** (e.g. in pgm or not)?
- Anything special about levels in the study?
- Do the levels represent a target population?
- Fixed
  - yes
  - yes
  - no
- Random: opposite of fixed

Types of Model Effects

- If multiple studies done independently would all studies use the same ***levels*** (e.g. in pgm or not)?
- Anything special about levels in the study?
- Do the levels represent a target population?

- Fixed
  - yes
  - yes
  - no
- Random

Effects in the model

- **Program (P or N)**
- **Teachers**

How do they fit these criteria?

Estimating Model Effects

- Fixed
  - familiar to all
  - compute the mean
- Random
  - they don't teach this in intro stat
  - key to estimating teacher and program effects

Estimating a Random Effect

- Example: student “mastery”
- Let M denote mastery
- M varies among students
  - mean, denote as μ_M
  - variance, denote as σ_M^2
- Measure “mastery” by a test, denoted S
- S has measurement error
  - mean, denote as μ_S
  - variance, denote as σ_S^2

Teacher Effect on Mastery

- M varies among students
  - mean, denote as μ_M
  - variance, denote as σ_M^2
- S has measurement error
  - mean, denote as μ_S
  - variance, denote as σ_S^2
- Student mastery under teacher T
  - M+T
- Teachers in study represent target population
  - mean, denote as μ_T
  - variance, denote as σ_T^2

Estimating a Random Effect

- We want to estimate teacher effect T
- We do so via student mastery M+T
- We measure M + T by S
- Question: what is the best estimate of M+T ?
- Hint: **it is NOT the test score S**

Estimating a Random Effect

- We want to estimate teacher effect T
- We do so via student mastery M+T
- We measure M + T by S
- Question: what is the best estimate of M+T
- Hint: **it is NOT the test score S**
- What is it?
  - **E(M+T|S)**
  - depends on means and variances of M, S and T

Some Issues Addressed by RETA

- Mixed Model Methodology
  - teacher effects
  - program effects
- Requirements for valid estimates vs real world
  - models assume
    - students randomized to teachers
    - tests do not have ceiling or floor effects
  - in reality
    - student assignment not random (for good reasons)
    - tests often have ceiling / floor effects

Findings

- Randomization
  - previous studies address extreme non-randomization to “game” the VAM
  - we looked at non-random processes schools actually use
  - no impact on accuracy, some impact on precision
- Ceiling
  - sufficient impact to invalidate estimates
  - exacerbated by non-randomization
  - assessing teacher & program effect requires tests with adequate “stretch”

Implications

- VAMs can help inform quality improvement in education
- Help inform re: “how are we doing?”
- Estimates from VAMs have **Variability**

Implications

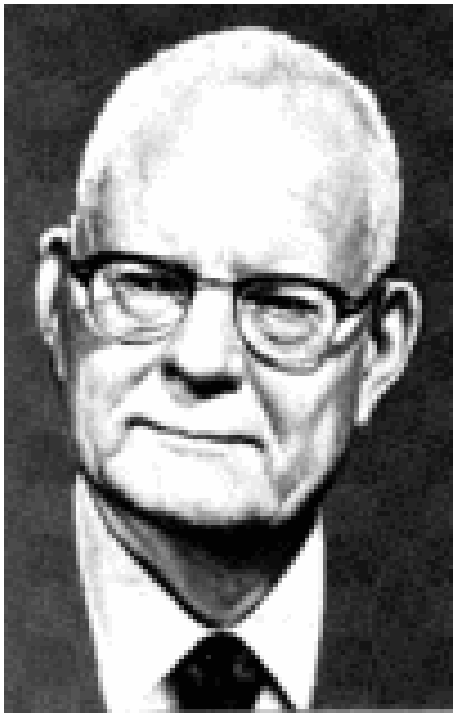
- VAMs can help inform quality improvement in education
- Help inform re: how are we doing
- Variability
  - estimates of teacher / program effects involve a mean AND a standard error
  - often reported w/o std error – **not good**
  - std errors tend to be large enough so that precise statements about individual teachers require extreme caution
  - e.g. **high likelihood of ranking teachers incorrectly**
  - **help improve: yes**; high stakes evaluation: no

Final Thought
re: statistical modeling and estimation of teacher & program effect

this is fundamentally a quality improvement enterprise

Quality Improvement

W Edwards Deming



- Preeminent figure/founding father of statistical process/quality improvement
- "Not enough to do your best. You have to know what to do, then do your best."
- "Profound Knowledge" – understanding and working with **variation**
- 14 Points
  - 3: cease dependence on inspection
  - 11: eliminate management by numbers & numeric goals
- 85/15

Deming, QI and VAM

- Deming advocated data-informed quality improvement
- Deming deplored merit evaluation in any form
- VAMs can be effectively used for QI in education IF they are used in a manner consistent with guidelines Deming articulated
  - VAMs can provide useful information when implemented appropriately
  - VAM is one tool among many