# Practical Considerations for Using Value Added Models for Monitoring Teacher Effectiveness

**Pete Goldschmidt, Ph.D.**

*Value Added Modeling Meeting*

***Sacramento, CA, October, 2008***

California State University Northridge
MICHAEL D EISNER COLLEGE OF EDUCATION
National Center for Research on Evaluation, Standards, & Student Testing
CRESST

# Introduction

Moving from status to growth and Value Added

Considerations regarding
- Status to growth
- Measurement issues
- Data issues
- modeling issues

Practical implications

# Status Accountability Model – based on unconditional mean performance

Irrespective of everything else going on – how is this teacher performing right now?

Assumes that:

- All student success is attributable to the current school (in the current year).
- By extension all student success is attributable to current teacher.
    - Also assumes that students do not bring any "human capital" inputs with them to the school.
    - There are no selection effects – the students in this school/class are like any other students in any other school/class in the district/state. One could bring in any other students from any other school and they would perform equally well.
    - There are no compositional effects.

# Moving Beyond Status

## Considerations

- Begin with questions:
  - ✓ What do we consider a “good” teacher to look like?
  - ✓ Is there an appropriate assessment system in place?
  - ✓ What additional data requirements are there?
  - ✓ Is there capacity to utilize various model choices?

# Value Added Basics

- The underlying assumption for value added models is:
  - ✓ Ait = f(Bit, Pit, Sit, Iit, Eit), (1)

  where for student i at time t Achievement A, is some function of:
  - ✓ Student background (B)
  - ✓ Peer and other influences (P)
  - ✓ School inputs (S)
  - ✓ Innate ability (I)
  - ✓ And luck (E).
- Model is cumulative and past inputs may affect current Achievement.
- Also would need independent measure of innate ability, gathered before any S has occurred.
- These are tremendous data requirements, and generally infeasible.

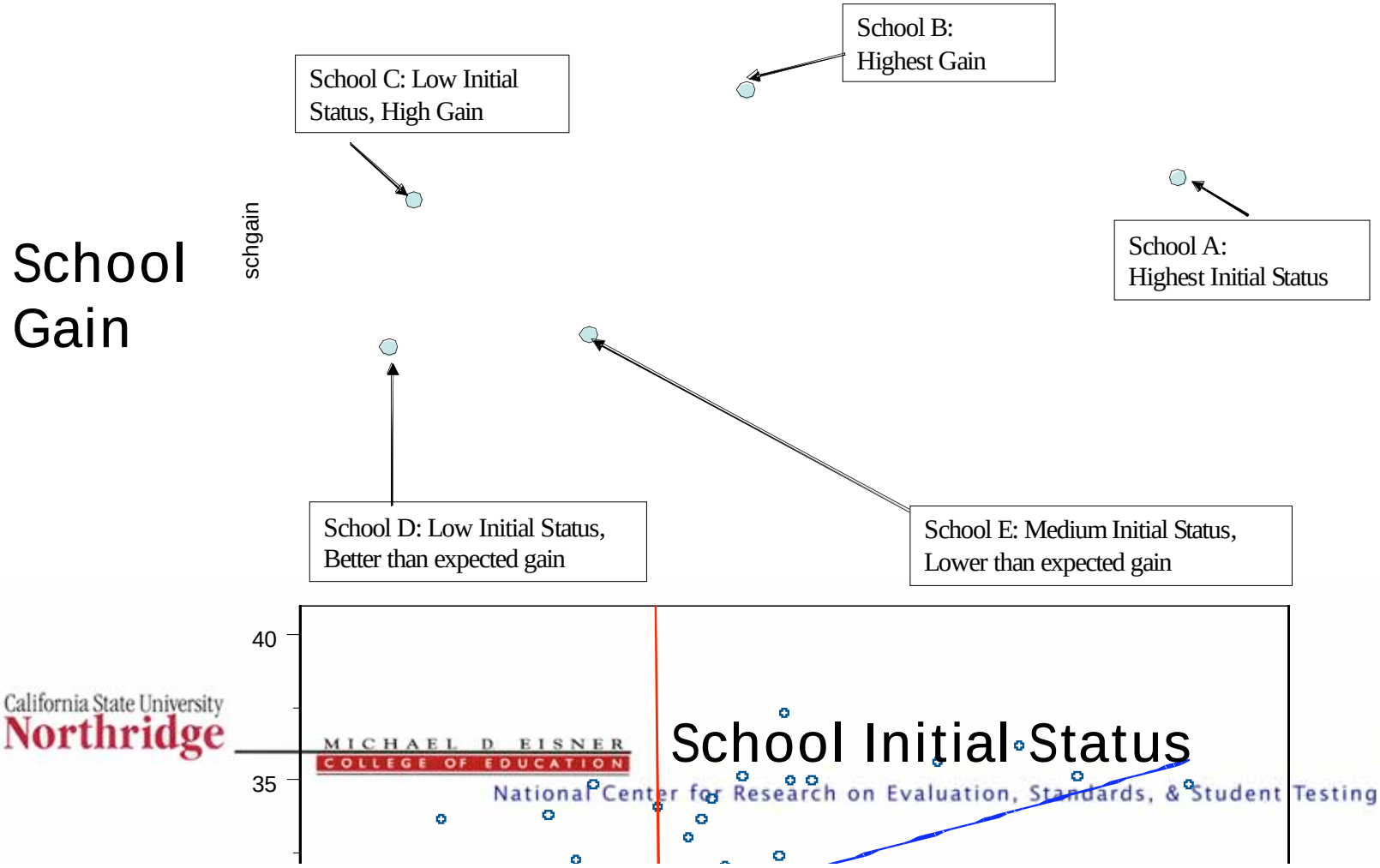
# Value Added Basics

- If we assume that (1) holds for any time t, then we can consider change in achievement from t to t`.
  - ✓ Ait` - Ait = f(.)
- Then by simply adding Ait to both sides, we get a familiar model:
  - ✓ Ait` = f(Bit`-t, Pit`-t, Sit`-t, Iit, Ait, Eit ) (2)
- Still lack measure of I, and omitting variables will increase the effect of included variables if there is a correlation between the omitted variable and the included variables.

  However:
  - ✓ *Once student B is included in the model the effect of omitting I is small; and, effect lessoned because include Ait.*
  - ✓ *Also, remaining variables measured contemporaneously, but this is generally not too problematic since only going back from t` to t.*

# Value Added example (schools)

- Based on LGPM but incorporates measurement error and uses latent initial status to predict growth.




California State University Northridge

MICHAEL D EISNER COLLEGE OF EDUCATION

National Center for Research on Evaluation, Standards, & Student Testing

CRESST

# Potential effects of change

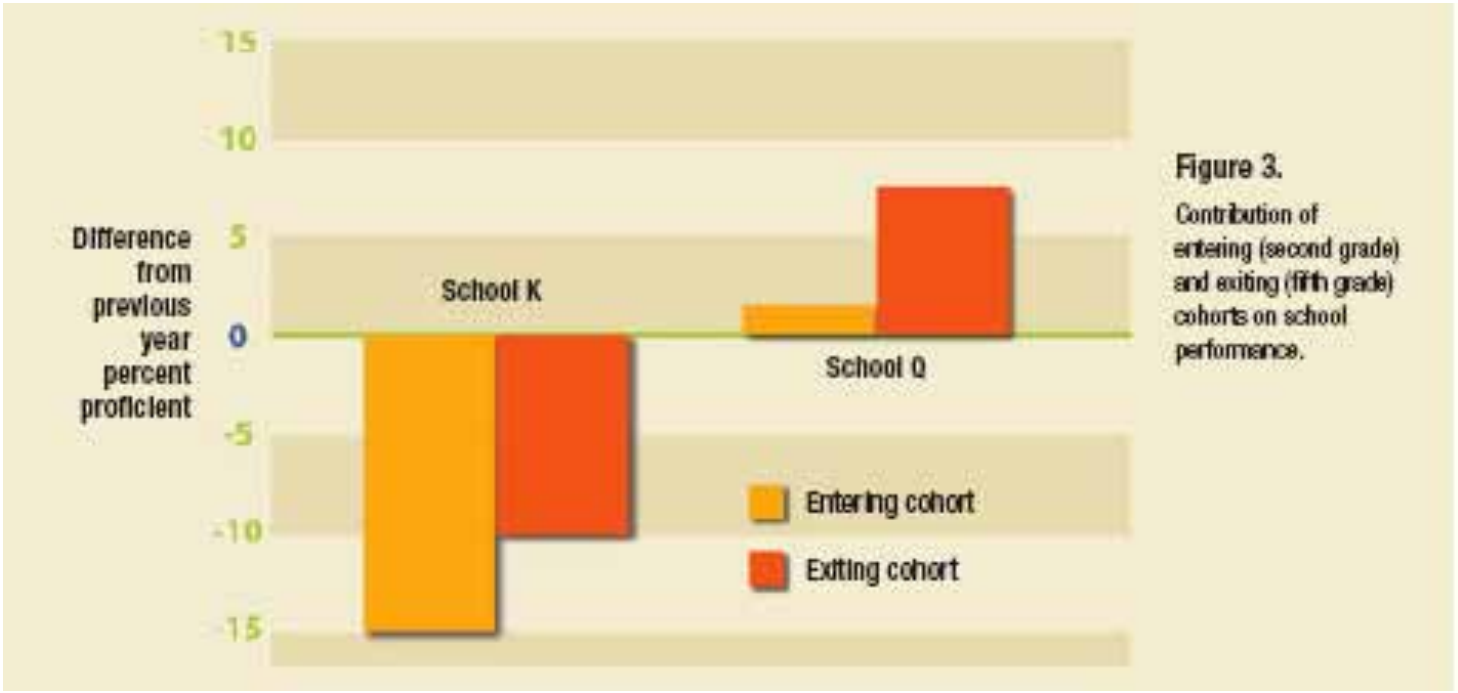
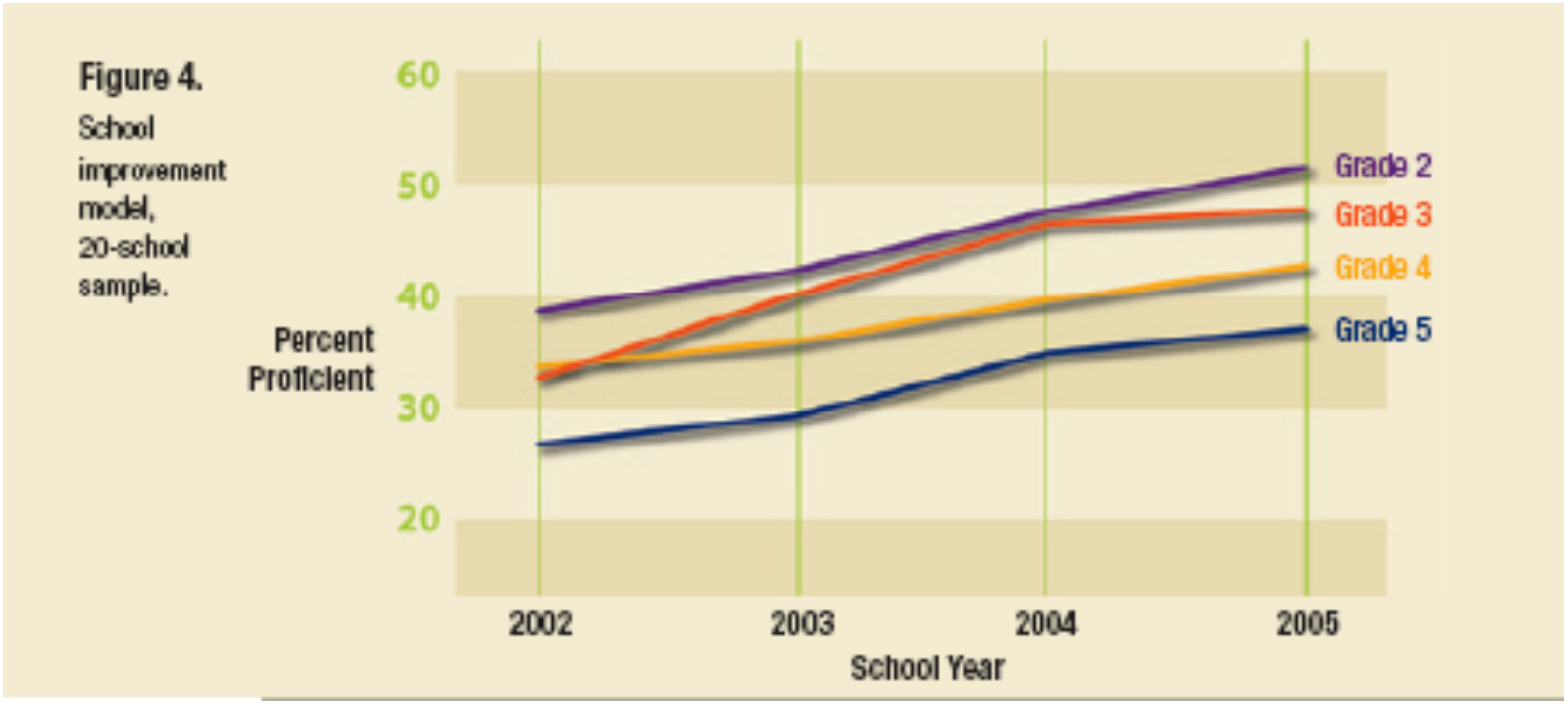



**Figure 3.**

Contribution of entering (second grade) and exiting (fifth grade) cohorts on school performance.

# Growth as a basis for Value Added I



# Growth as a basis for Value Added II

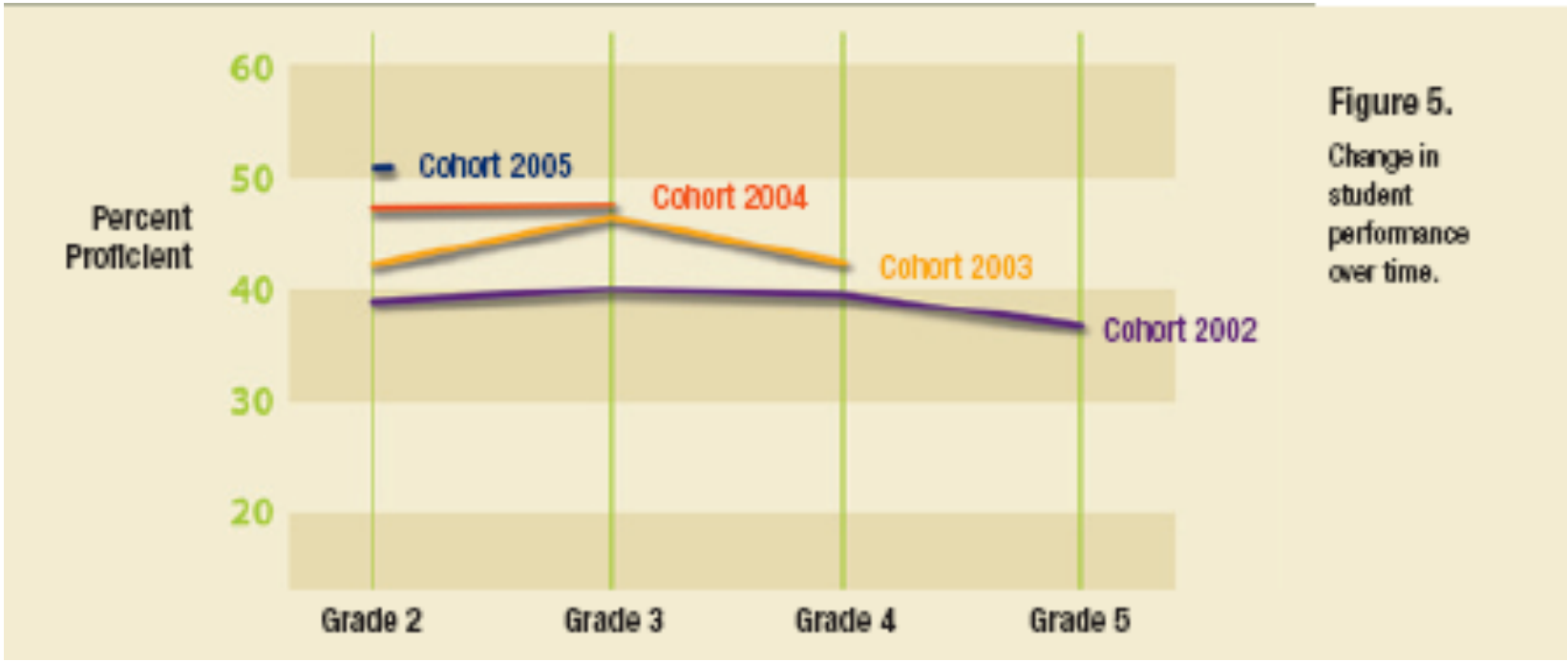



**Figure 5.**
Change in student performance over time.

# Comparing LPGM and LSPM

- Considering both types of growth
- Correlation between school ranks based on two approaches ranges from 0.25 to 0.65.
- A majority of the variation in individual student growth is within schools (as much as 90%).
- Important to consider whether individual growth affected by cohort student is in.

# Longitudinal Cohort Panel Growth Model

$$\text{math}_{ijkl} \sim \text{N}(XB,\ \Omega)$$

$$\text{math}_{ijkl} = \beta_{0ijkl}\text{intercept} + \beta_1 \text{grade2}_{ijkl} + \beta_2 \text{cyear}_{kl}$$

$$\beta_{0ijkl} = \beta_0 + f_{0l} + v_{0kl} + u_{0jkl} + e_{0ijkl}$$

$$\left[f_{0l}\right] \sim \text{N}(0,\ \Omega_f) \ :\ \Omega_f = \left[\sigma^2_{f0}\right]$$

$$\left[v_{0kl}\right] \sim \text{N}(0,\ \Omega_v) \ :\ \Omega_v = \left[\sigma^2_{v0}\right]$$

$$\left[u_{0jkl}\right] \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[\sigma^2_{u0}\right]$$

$$\left[e_{0ijkl}\right] \sim \text{N}(0,\ \Omega_e) \ :\ \Omega_e = \left[\sigma^2_{e0}\right]$$

- where $\text{math}_{ijkl}$ is the math score at time *i*, for student *j* in cohort *k*, in school *l*.

| Random effects | Variability Breakdown |
|---|---|
| Level 1 | |
| error | |
| Level 2 | |
| Between students within cohorts, schools | |
| Initial Status | 84.9% |
| Individual growth | 42.7% |
| Level 3 | |
| Between cohorts, within schools | |
| Initial Status | 6.7% |
| Individual growth | 42.2% |
| Cohort growth | 45.2% |
| Level 4 | |
| Between schools | |
| Initial Status | 8.4% |
| Individual growth | 15.1% |
| Cohort growth | 54.8% |

# Measurement Issues

Using growth for monitoring performance is preferable to static indicators of performance.

- Important to link assessments to their uses:
  - ✓ Need to consider assessments
  - ✓ Need to consider standards
  - ✓ Need to consider what growth is measuring and how it is represented
  - ✓ Need to consider score representation
    - ❖ i.e., metrics and valid inferences from growth model results

# Assessments as Indicators of Student Academic Performance

- In terms of schools or teachers, we are interested in:
  - Achievement tests as they are intended to measure knowledge and skills.

  As opposed to:

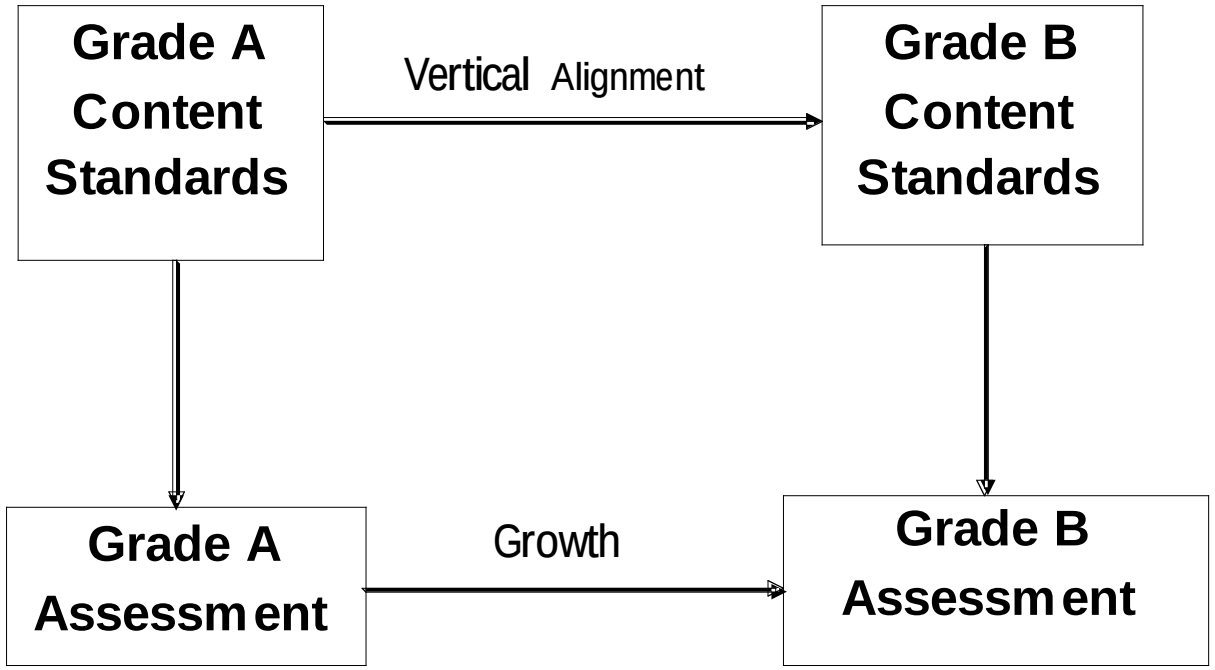
  - Aptitude tests that are generally used to predict future performance.

# Aspects Affecting Individual Assessments and Assessments Across Time

For all assessments, surveys, or instruments, need to consider measurement issues related to:

- Validity Issues
- Validity Over Time (alignment issues)
- Precision
- Reliability

# Growth Must be Based on Vertically Aligned Content Standards



# Vertical Alignment

Issues

- How are content standards/objectives related from one grade to the next?
- Knowledge or skills extended to wider range of content
- Deeper understanding (cognitive processes) for the same content
- New or different content and/or skills
- Need to consider the nature of alignment
  - ✓ *Depth of knowledge*
  - ✓ *Range of Content, etc.*
- Need to consider the quality of alignment.
  - ✓ *Are content standards clearly articulated across grades?*

# Precision, Reliability, and Growth

- Reliability
  - ✓ Describes how much of the between-person variability in observed scores is attributable to variability in true scores.
  - ✓ Better reliability in rates of change estimates creates a better ability to detect true differences in trajectories among individuals (or groups).
  - ✓ Better precision generally leads to higher reliability.
  - ✓ If there is little variation in true growth, then despite good precision reliability will be low (making it difficult to detect between person differences in growth).

# Reliability and Growth

- Gain scores are not inherently unreliable.
  - ✓ Gain scores also benefit from not normalizing performance as residual gains from covariance adjustment models do.

# Test Metrics – Uses and Misuses

## Issues

- ✓ The appropriateness of the metric depends upon the uses of the results
- ✓ Generally, scale scores are best for analyses – although more difficult to interpret
- ✓ NCEs are easily interpretable and can readily be used in accountability models
- ✓ Proficiency categories result in a loss of information due grouping data into categories
  - ❖ No information regarding within category changes in performance

# Data Issues

- Value Added requires individual student data
- Linked over time
- Linked to each teacher

# Modeling:
# Value Added and Teacher Effects

- TVASS explicitly attempts to model teacher effects
  - ✓ Assumes linear and additive teacher effects
- Rand Model – extension and generalization of TVASS
- In general models attempting to model specific teacher effects require:
  - ✓ A substantial amount of data
  - ✓ Extensive computing capacity,
  - ✓ And if modeling entire system only (approx) 5% of teacher "statistically" differ from average effectiveness
- Other models base teacher effects on school effectiveness (e.g., NC)

# Modeling Issues

- TVASS uses five years of teacher data
  - ✓ Median time in profession in CA?
  - ✓ Student school changes
  - ✓ Teacher school changes
- How treat teacher effects?
  - ✓ Cumulative
  - ✓ Additive
  - ✓ Decaying
    - ❖ Two year effect

# Conclusions

- Valued Added better way to examine teacher effects than status
- Need to consider assessments underlying analyses
- Intensive data requirements
- Value Added models can be quite complex

# Conclusions

- A single number summary of teacher, school, or student performance is less desirable than multiple indicators.
- Studies indicate that both sanctions and rewards tend to produce similar stresses on teachers
- Could use school results to monitor teachers as a group, but significant within school variability in student performance (at least as much, if not more, than between schools.

## Contact

Pete Goldschmidt, Ph.D

Pete.Goldschmidt@CSUN.edu

(818) 677-4601