Models for School Accountability and Program Evaluation

UCLA Graduate School of Education & Information Studies
Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, and Student Testing

CRESST

Purpose:

- **State educational system performance**

  District performance

  School performance

  School is facilitating adequate progress towards objectives.

  Specific aspect of schooling process is having a significant effect on student outcomes.

Data:

- Universal student ID

  Assessment results

  - Metrics
    - Does the scale matter – yes
    - Continuous vs Categorical
    - Does the metric matter – yes

Data

- Universal student ID
  Linked to teachers (need teacher ID)
  Linked to schools

Data

- Student characteristic information
- **Gender**
- **Title1**
- **Lunch status**
- **Language status**
- **Race/Ethnicity**
- **Gifted/Talented**
- **SWD**
- **Program**
- **School entry Date**

- Assessment results
-
- **NRT**
- **CRT**
- **Performance assessments**
- **OTL (minimally attendance)**

Data

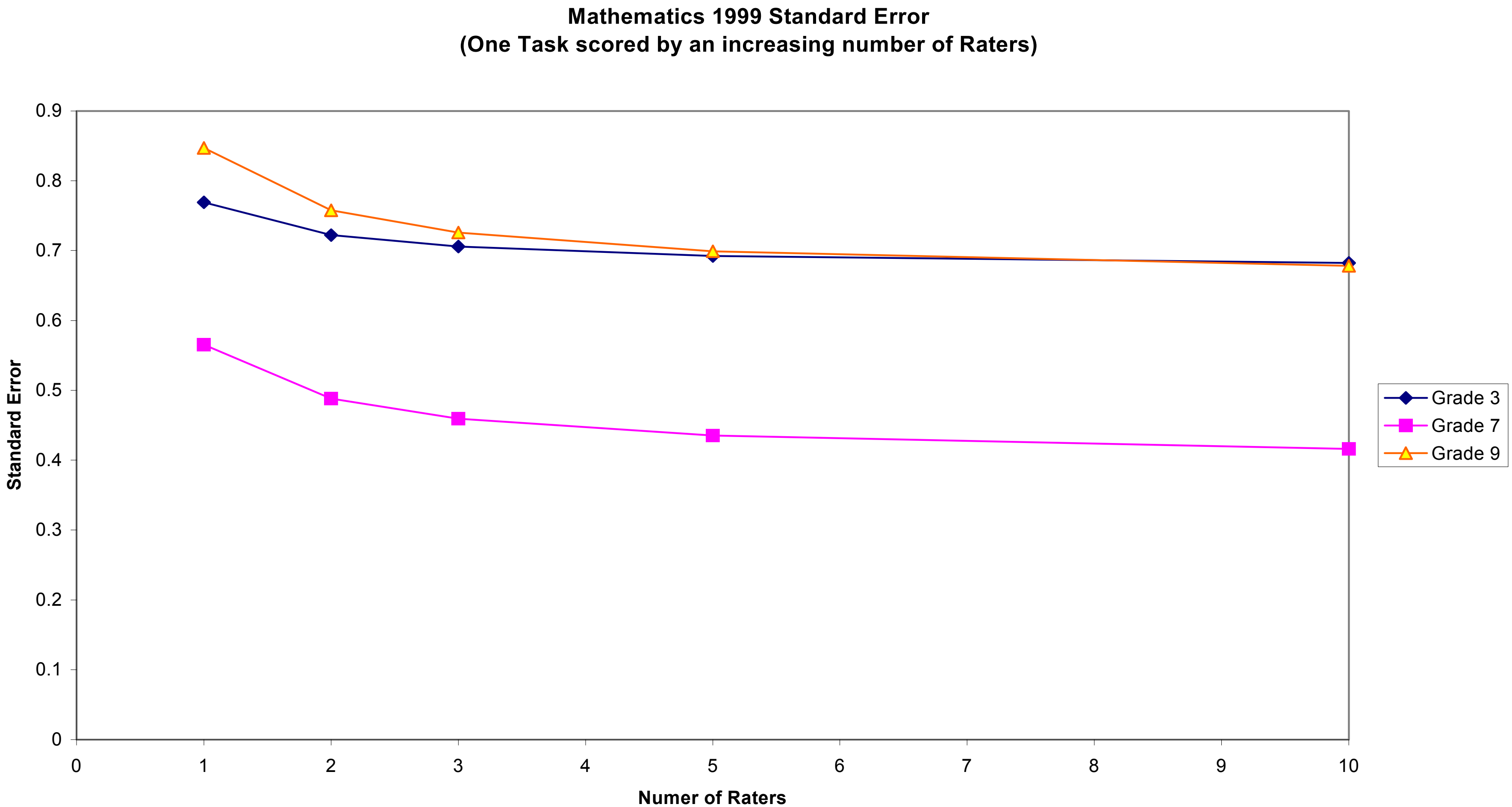
- Assessment Results
- Choice of metric
- **Use metric with equal interval scale**
- **Continuous vs. categorical (proficiency levels) data**
- ***Categories can be based on underlying continuous score or rater judgment***

Data: Computational Requirements

- Standard “high End” PC can accommodate most analyses.
- Longitudinal analyses produce occasion files that are years*students in size.
- Standard time for data cleaning (as likely currently required).
- Time to create datasets for analysis – 3 to 5 days.
- Time to run analysis 30 minutes to 3 hours.

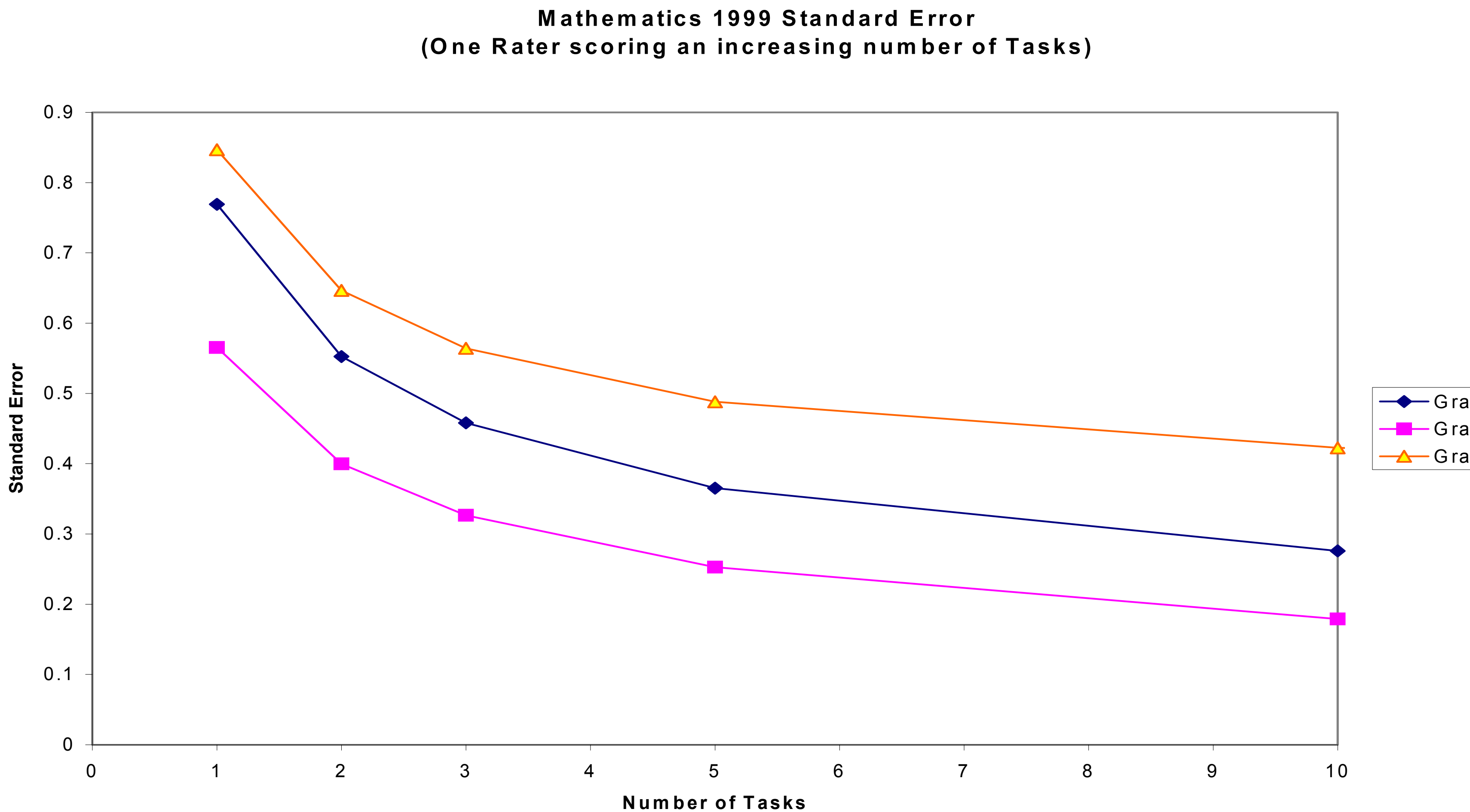
Generalizability Studies (1999 Math)

Standard Errors with increasing number of raters



Generalizability Studies (1999 Math)

Standard Errors with increasing number of tasks

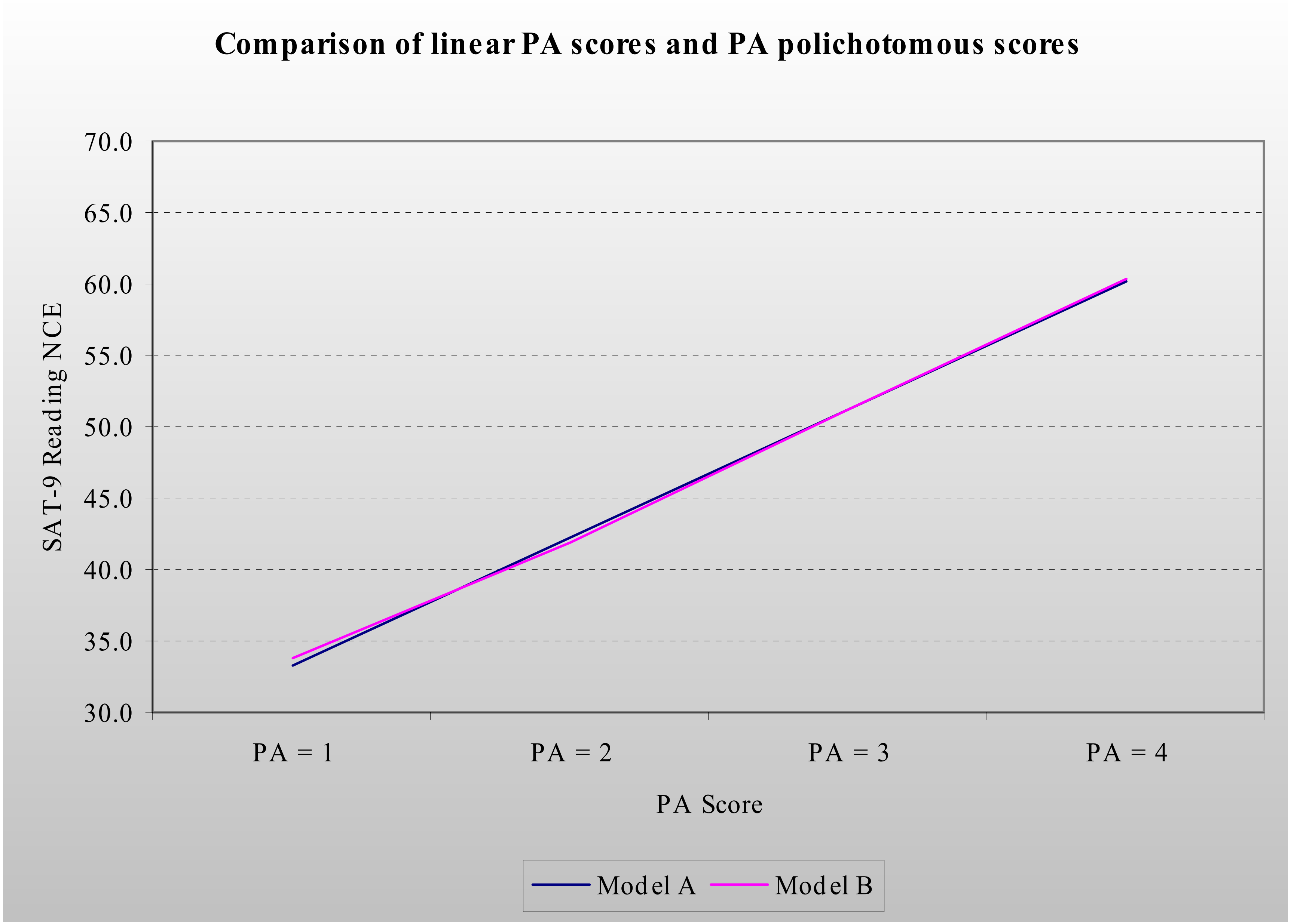


Generalizability Studies

Percentage of misclassifications under different measurement conditions (varying numbers of tasks or raters)

1 Rater	**Observed Score (True Score = 3)**			
	1	**2**	**3**	**4**
1 Task (SE=.77)	3%	27%	40%	30%
2 Tasks (SE=.55)	0%	20%	58%	22%

1 Task	**Observed Score (True Score = 3)**			
	1	**2**	**3**	**4**
1 Rater (SE=.77)	3%	27%	40%	30%
2 Raters (SE=.72)	1%	25%	44%	28%


Comparison of linear PA scores and PA polichotomous scores
SAT-9 Reading NCE
70.0
65.0
60.0
55.0
50.0
45.0
40.0
35.0
30.0
PA = 1
PA = 2
PA = 3
PA = 4
PA Score
Model A
Model B

Correlations	MATH NCE	Read Prof	Math Prof	Read Pass	Math Pass
READING NCE	0.66	0.98	0.64	0.80	0.55
MATH NCE		0.64	0.98	0.50	0.85
Read Prof			0.63	0.83	0.54
Math Prof				0.48	0.87
Read Pass					0.45

Data

- Equal interval metric
- Otherwise means, differences, and confidence intervals will have no meaning)
  - either Scale Scores or NCEs
  - but not Percentile ranks (although these can be converted to NCEs)
  - not GEs as these are based on extrapolation

Choice of Metric:

- **IRT-based scale scores**
- **Vertically equated scores across grades and years**
- **Theoretically represent growth on a continuum that can measure academic progress over time**
- **Change from year to**

- Change represents a relative position from year to year not absolute growth in achievement
- Relative standing compared to a norming population

Data

- The metric matters depending on which question we are interested in addressing.
  - **Questions concerning absolute growth require a vertically equated scale score.**
  - **Relative questions are less sensitive to the metric.**
    - **ranking schools**
    - **comparison of various performance measures**
    - **comparison of schools based on growth**
    - **comparison of achievement gaps (static or longitudinal)**
    - **value added comparisons**

Sampling Conditions for Monte Carlo Study

Total Number of Schools	Students Sampled (%)	mean n
60	25%	31.3
60	50%	65.6
60	75%	98.5
60	100%	130.9

Focus is on metric – however, also addresses effect of school sample size.

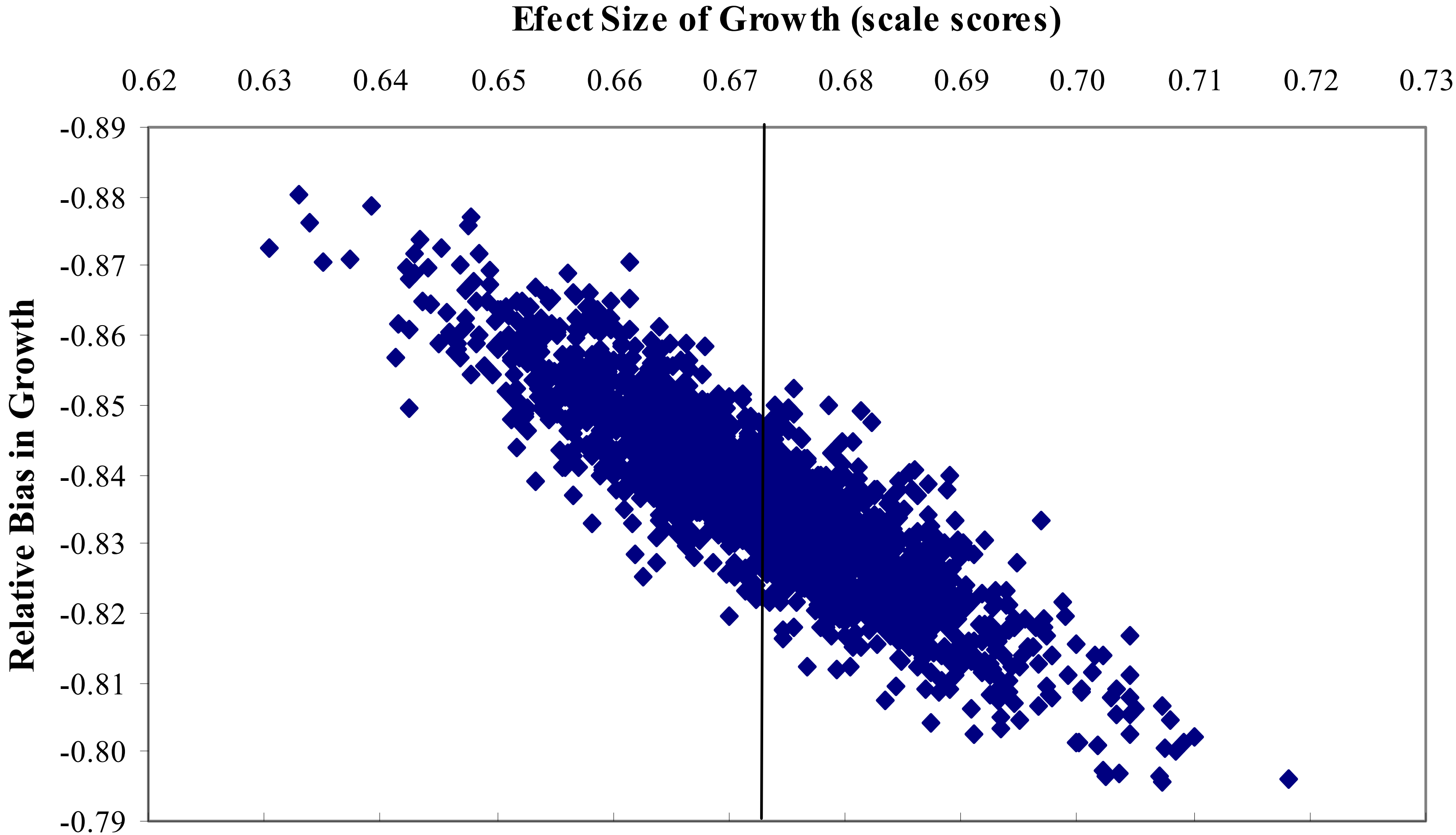
Summary of Results Describing SAT-9 Reading Achievement

	25%		50%		75%	
SAT-9 Reading Achievement	NCE	SS	NCE	SS	NCE	SS
Mean Initial status (g_{000})						
Student Predictors						
Special Education (γ_{010})	-0.47	-0.44	-0.47	-0.44	-0.47	-0.44
Low SES (γ_{020})	-0.36	-0.4	-0.35	-0.4	-0.35	-0.39
LEP (γ_{030})	-0.34	-0.35	-0.33	-0.34	-0.32	-0.33
Minority (γ_{040})	-0.48	-0.54	-0.48	-0.54	-0.48	-0.53
Girl (γ_{050})	0.1	0.1	0.1	0.1	0.1	0.1
School Predictors						
LAAMP Effect (γ_{001})	0.03	0.04	0.02	0.03	0.02	0.02
Minority (γ_{002})	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Low (γ_{003})	0.13	0.1	0.17	0.15	0.2	0.17
Mean Growth (g_{100})	0.07	0.64	0.07	0.63	0.07	0.63
Student Predictors						
Special Education (γ_{110})	0	-0.03	0	-0.03	0	-0.03
Low SES (γ_{120})	0.05	0.06	0.05	0.06	0.05	0.06
LEP (γ_{130})	0.07	0.07	0.07	0.07	0.07	0.07
Minority (γ_{140})	-0.03	-0.02	-0.03	-0.02	-0.03	-0.02
Girl (γ_{150})	0.01	0.01	0.01	0.01	0.01	0.01
School Predictors						
LAAMP Effect (γ_{101})	0.01	0.01	0.01	0.01	0.01	0.01
Minority (γ_{102})	0.11	0.14	0.12	0.15	0.12	0.16
Low (γ_{103})	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08

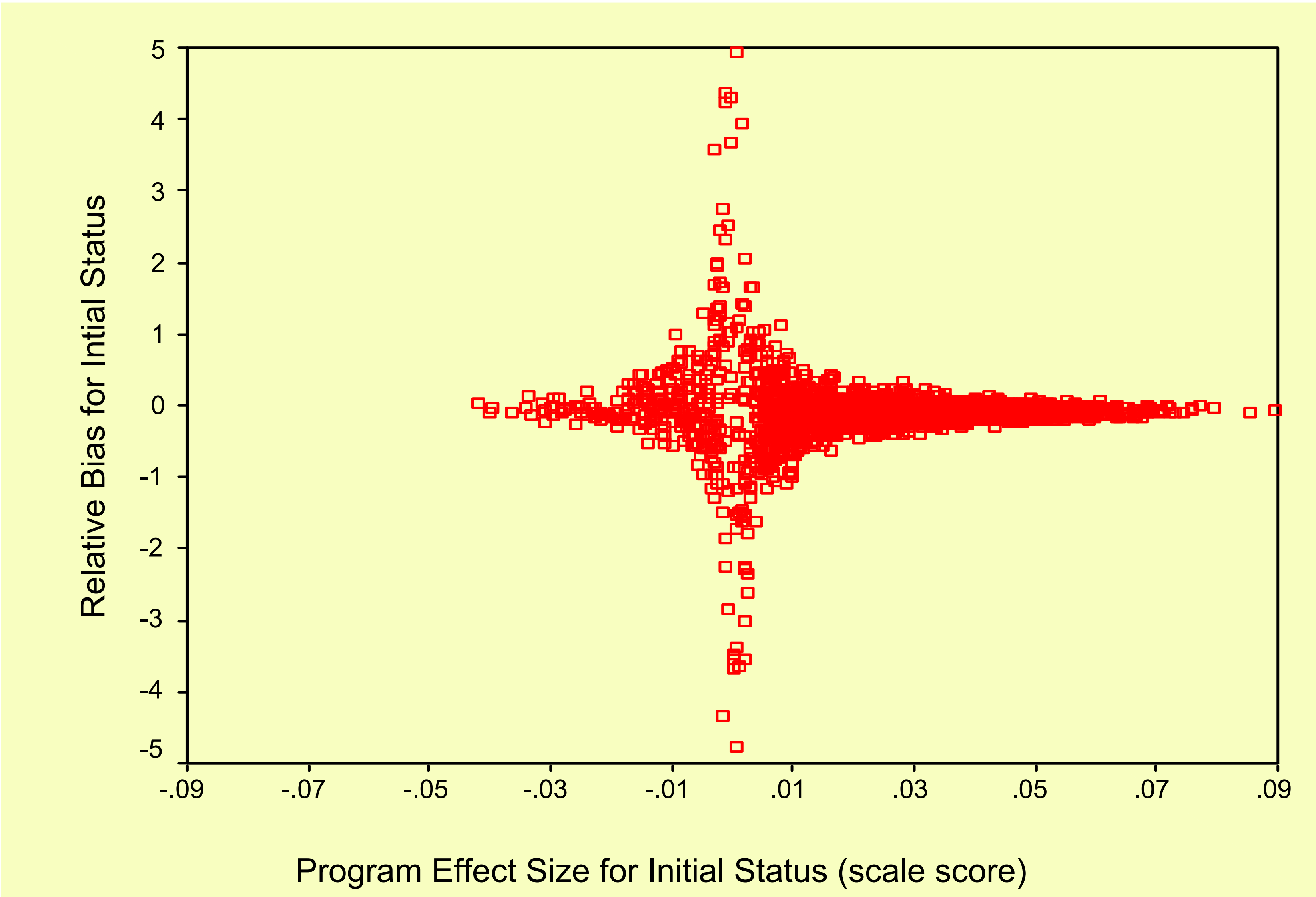
Percent Reduction in Between School Variation in Growth

Sampling	Reading		Math	
Condition	NCE	SS	NCE	SS
Model 2 to 4				
25%	24.5	23.7	9.3	9.2
50%	24.4	25.5	9.6	9.7
75%	24.5	26.4	9.2	9.3
Model 1 to 4				
25%	43.8	52.2	16.8	16.8
50%	42.7	51.9	16.4	16.5
75%	42.9	52.3	16.1	16.1

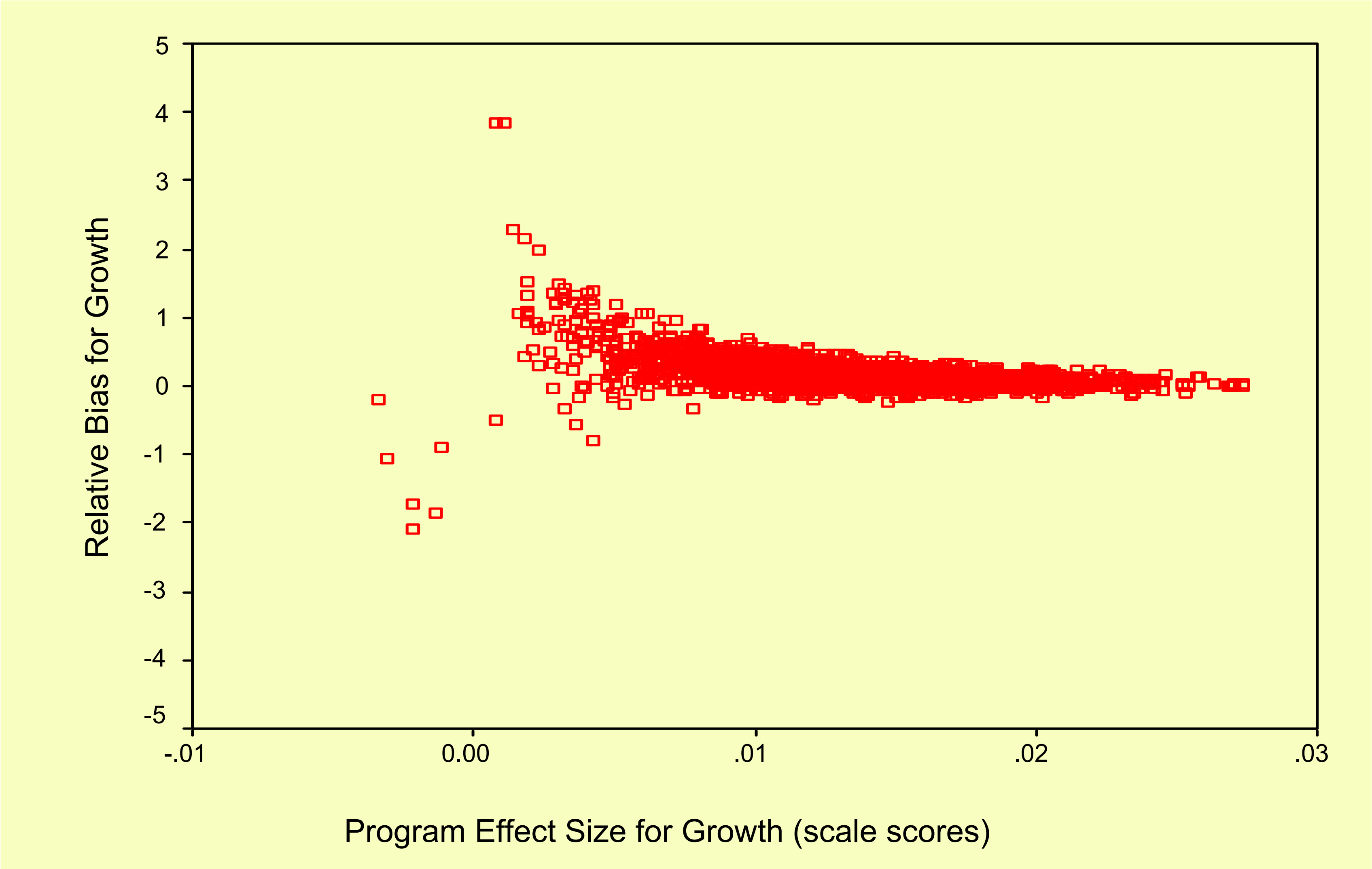
Comparison of Relative Bias to the Effect Size of Growth



Relationship between Relative Bias in NCEs for Initial Status



Relationship between Relative Bias in NCEs for Growth



Models of School Performance

-
-
-
-
-

Cross Sectional

Longitudinal

Cohort

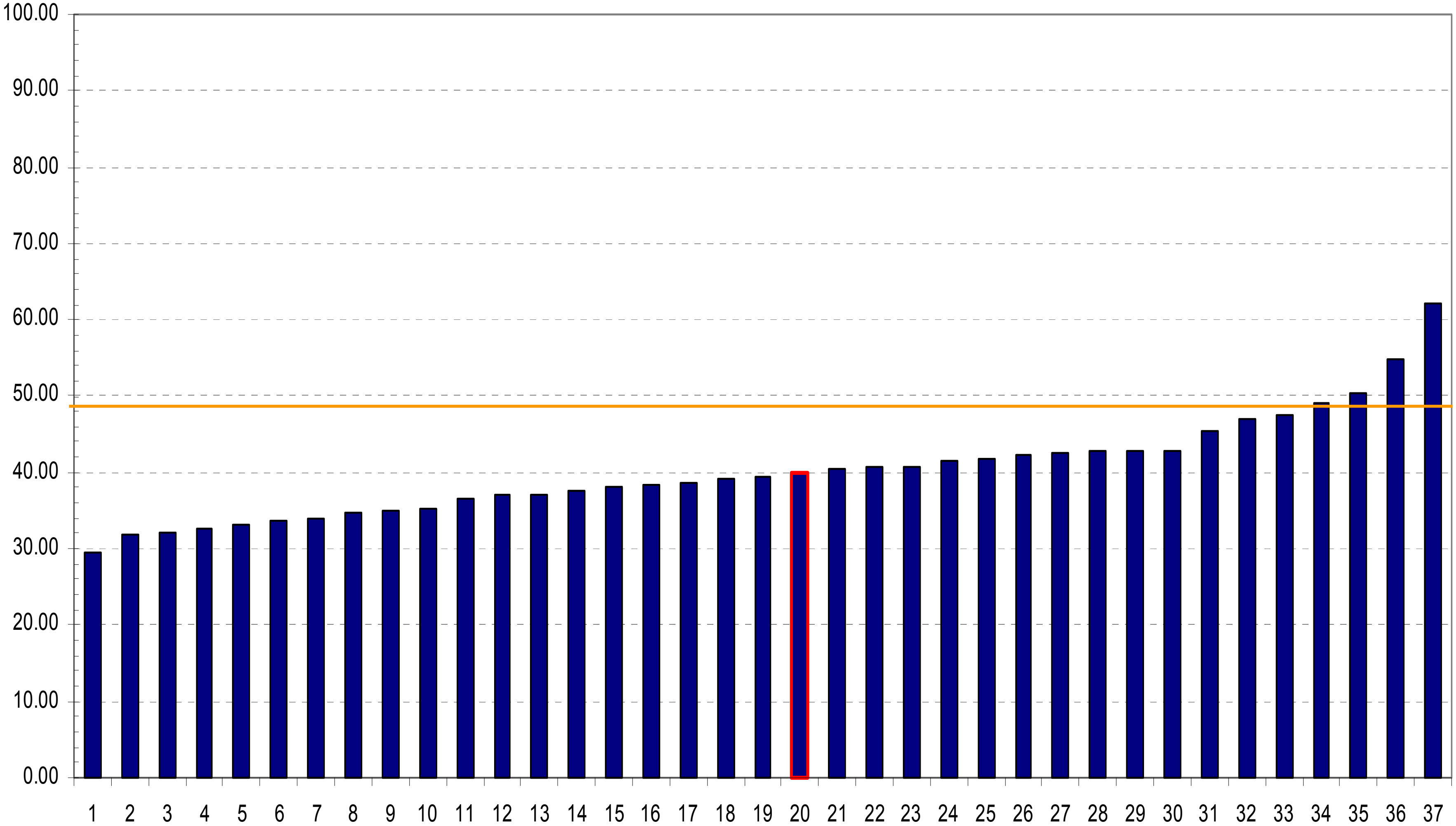
Panel

School Means as a Measure of Performance

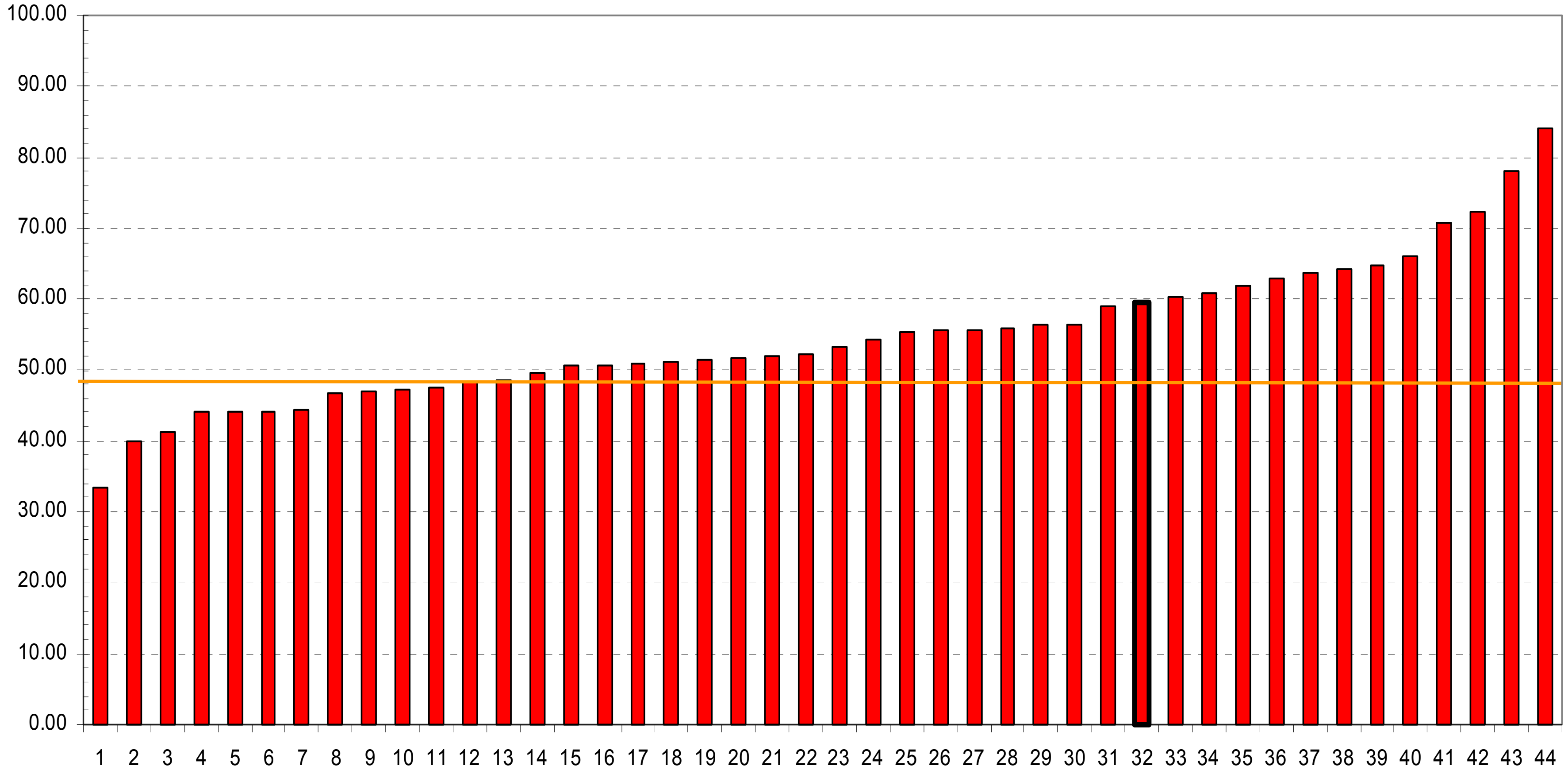
- Minimal data requirements.
- Simple calculations.
- Straightforward explanation to consumers.

- Assume equally precise measures of school performance.
- Do not adjust for inputs.
- Potential confounding factors artificially attribute all of the variation in student scores to schools.
- Potential ecological fallacy.

School Mean Performance - Mathematics CAT/6 NCEs
District 8

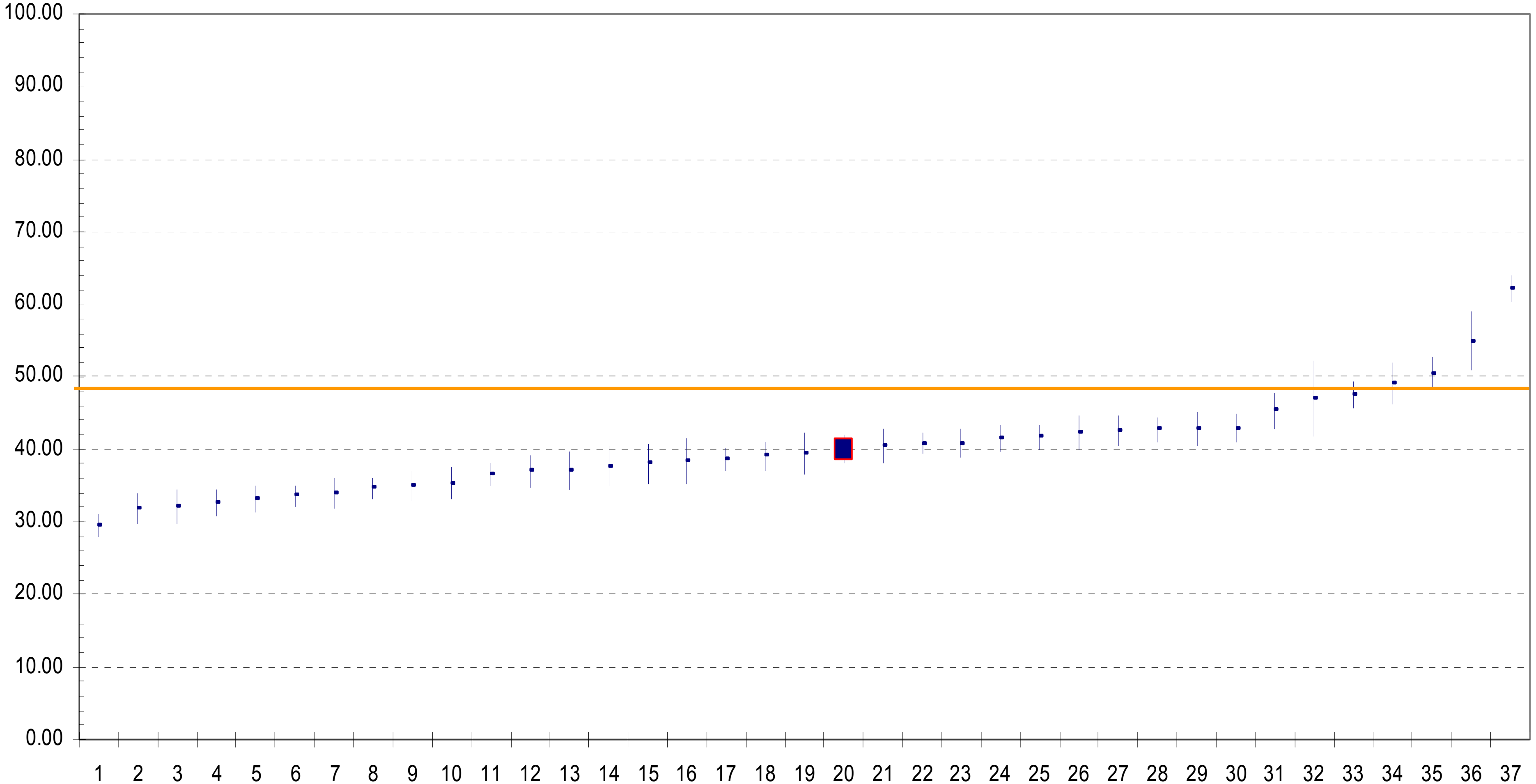


School Mean Performance - Mathematics CAT/6 NCEs
District 3

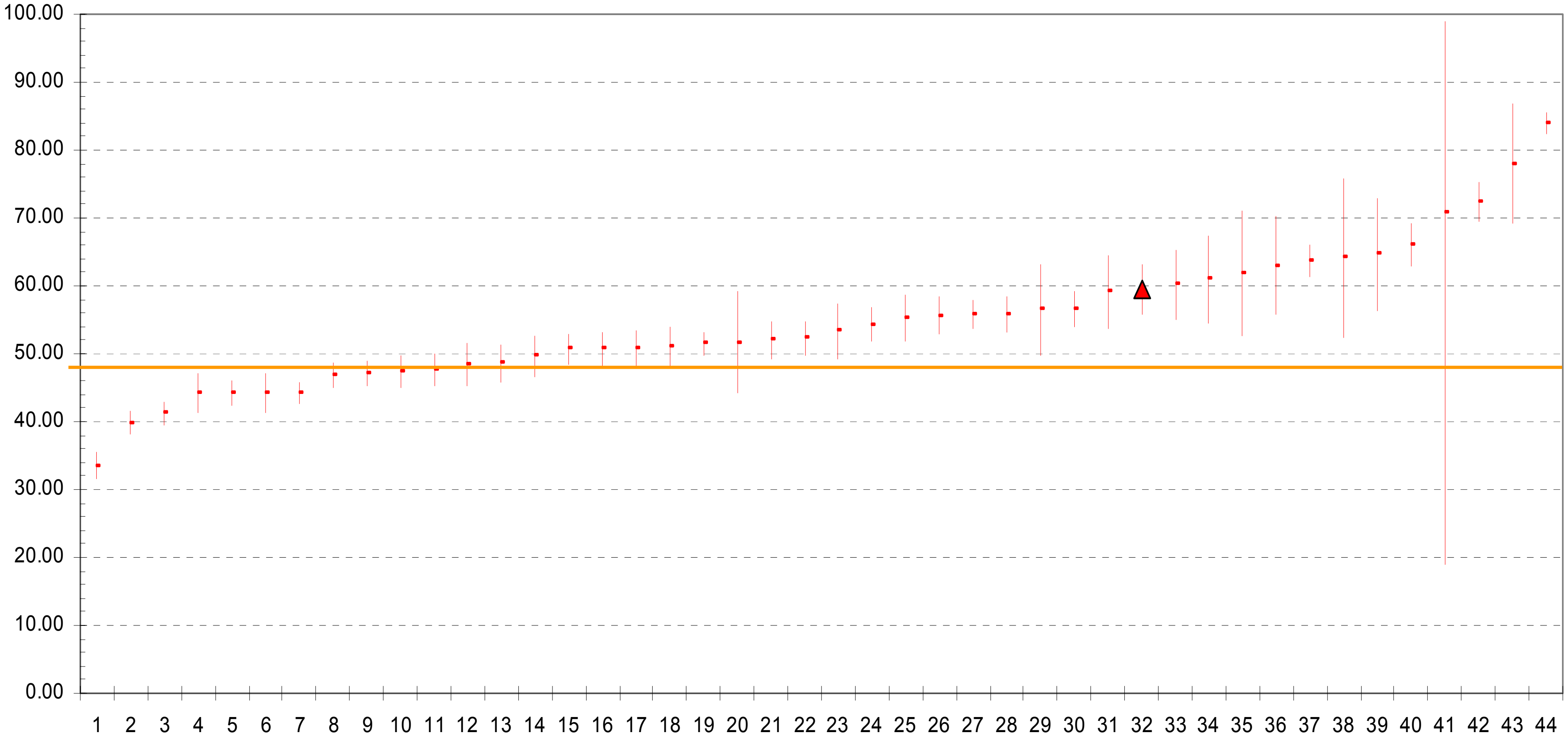


- Means can rank order schools efficiently and can be used as both a relative and absolute measure.
- But there is no indication of precision.

School Mean Performance with 95% Confidence Intervals
District 8



School Mean Performance with 95% Confidence Intervals
District 3



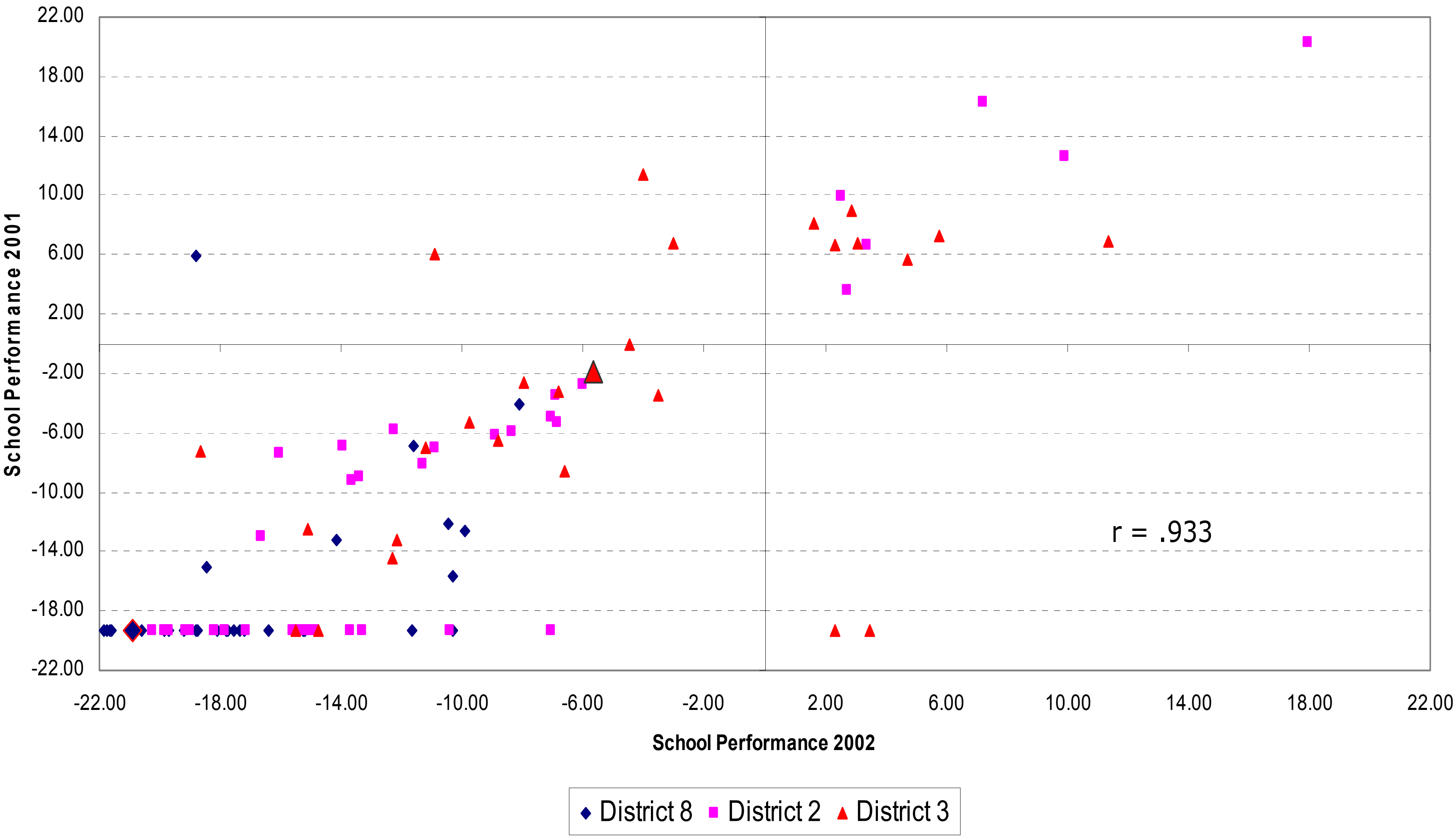
Adding Confidence intervals demonstrate precision and also allows one to determine whether mean differences are due to random variation or true differences among schools.

Correlations of Mean School Performance by Year

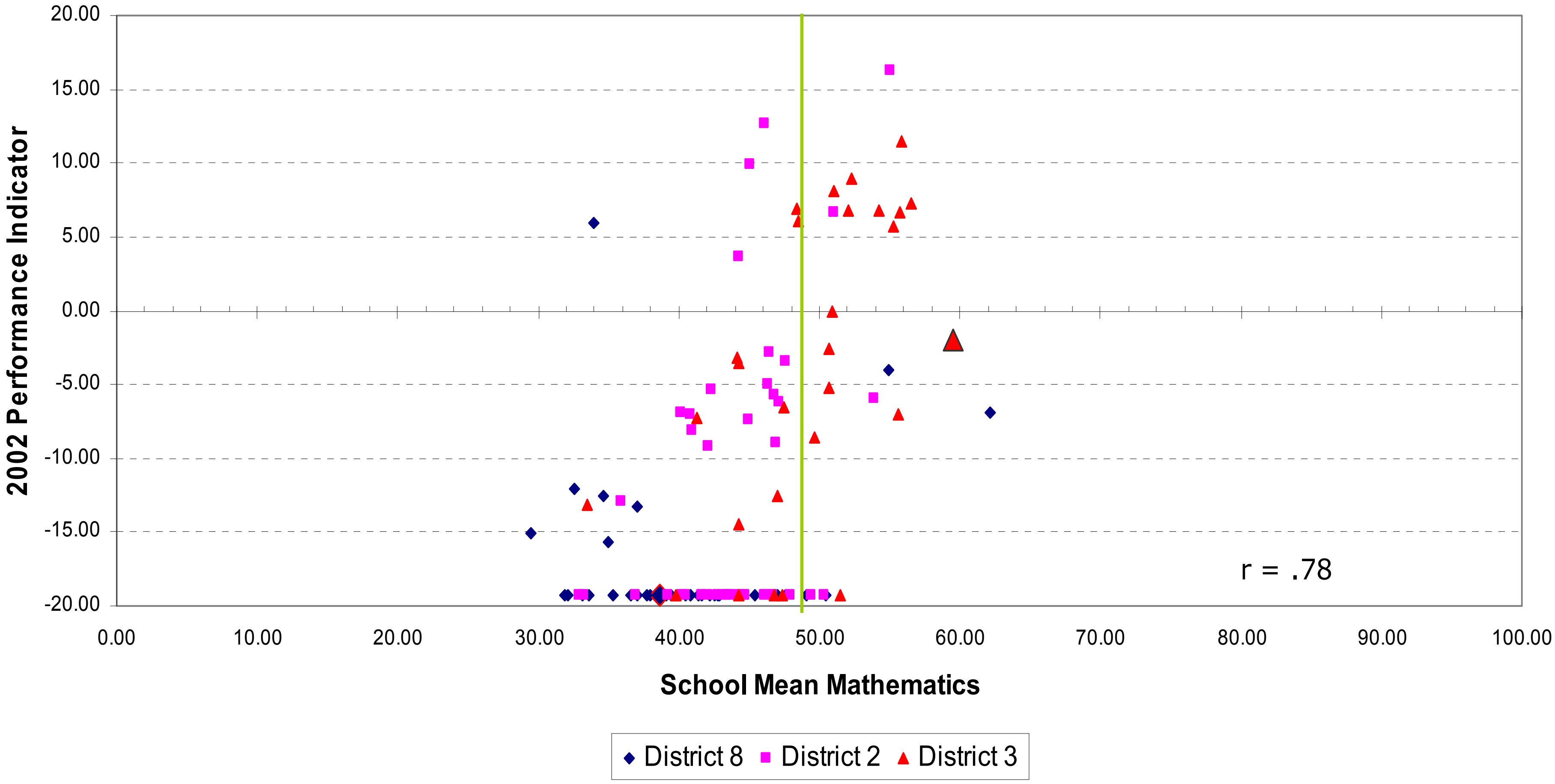
	MATH1994	MATH1995	MATH1996	MATH1997	MATH1998	MATH1999	MATH2000	MATH2001	MATH2002
MATH1993	0.96	0.91	0.89	0.87	0.90	0.90	0.88	0.88	0.87
MATH1994		0.94	0.92	0.90	0.92	0.91	0.90	0.89	0.89
MATH1995			0.93	0.90	0.91	0.90	0.88	0.87	0.88
MATH1996				0.96	0.93	0.90	0.90	0.87	0.89
MATH1997					0.94	0.92	0.91	0.89	0.89
MATH1998						0.97	0.95	0.94	0.93
MATH1999							0.97	0.95	0.94
MATH2000								0.96	0.95
MATH2001									0.96

If consistency is comfort – aka validating, then means over time should provide some ease of mind as they are highly correlated over time*

Consistency of School Performance Indicator

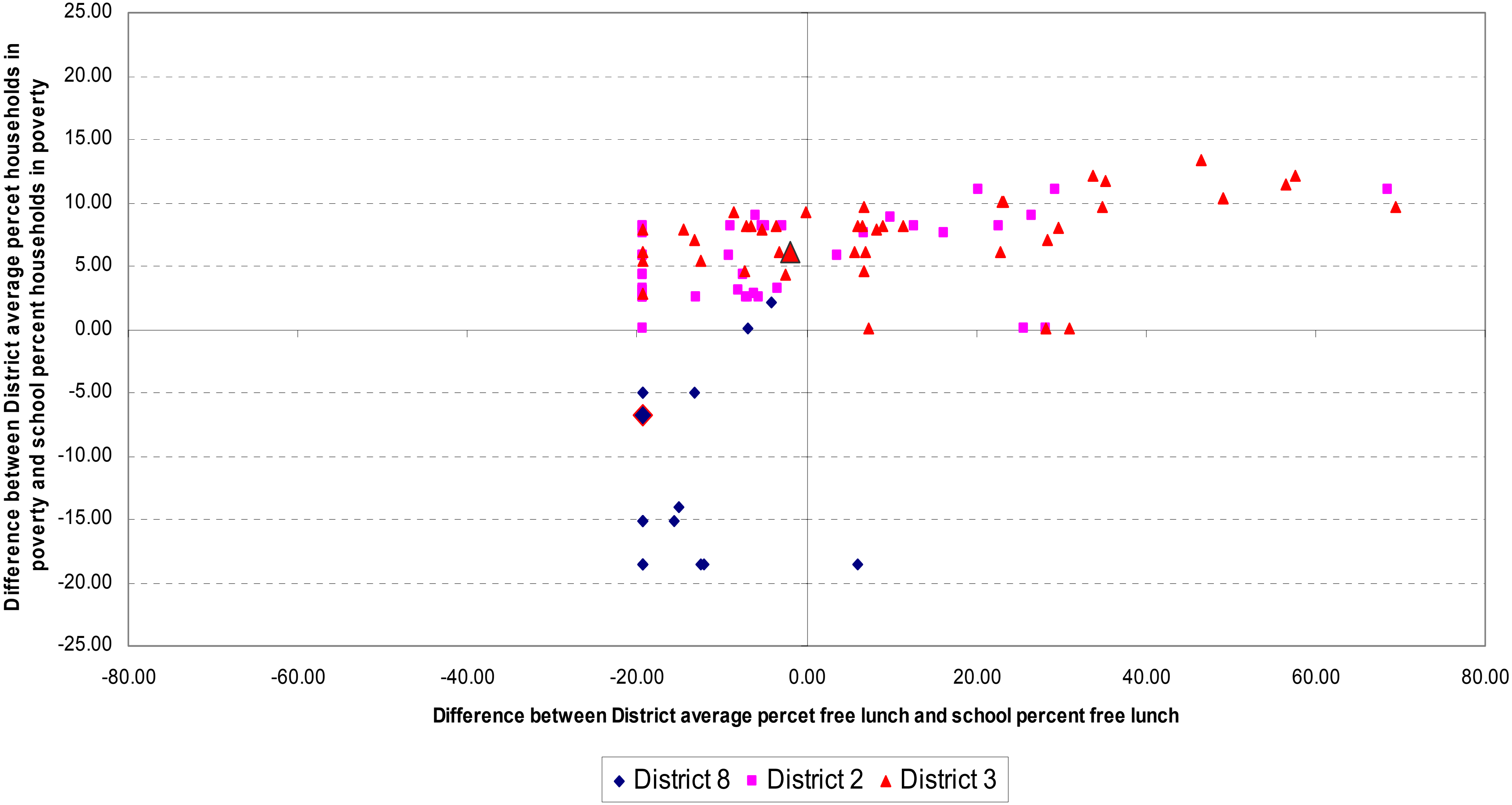


Cmoparison of 2002 Performance Indicator with 2002 school mean Mathematics perfromance



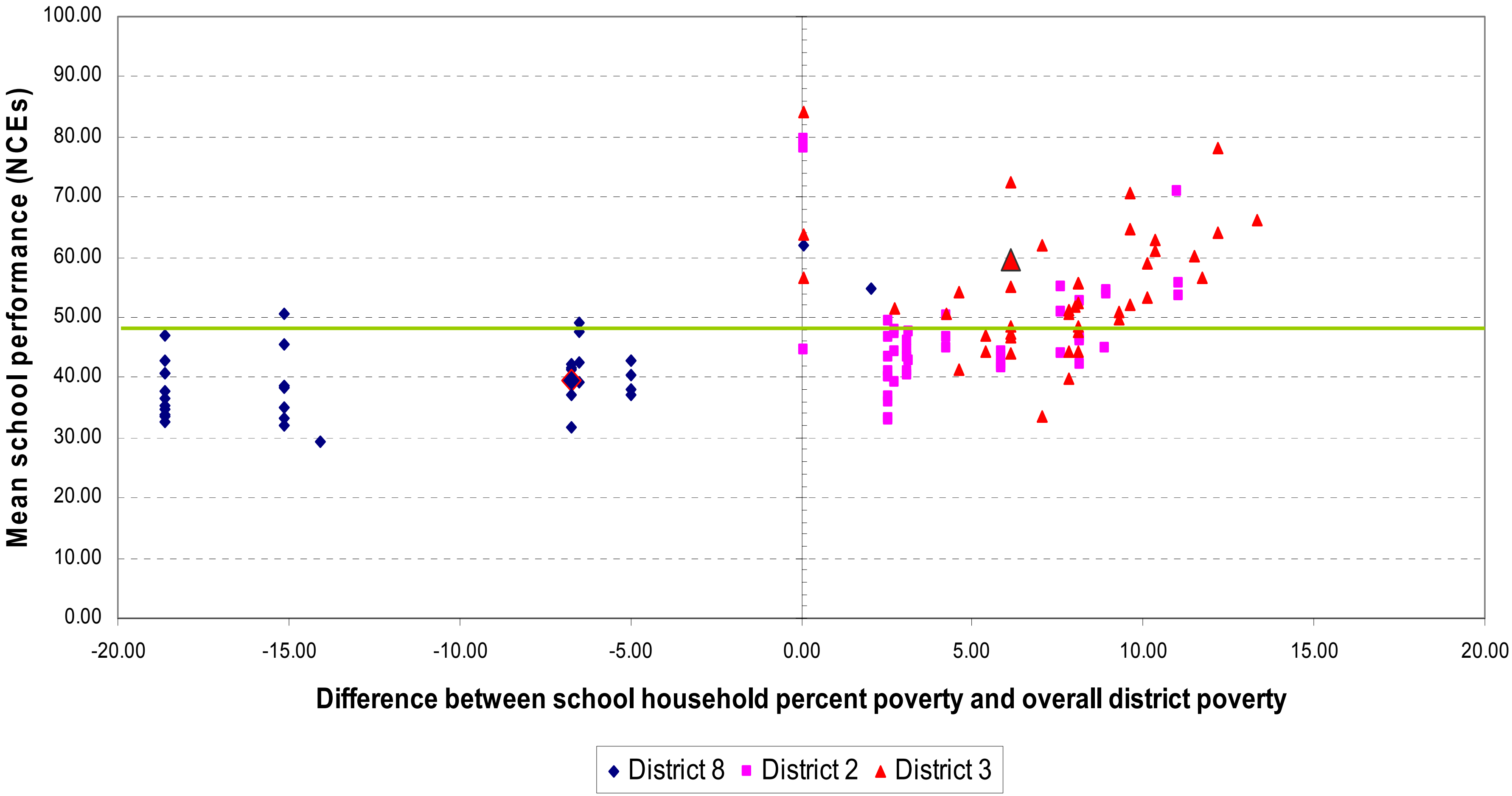
Relationship truncated due to several schools at the minimum.

Percent Free Lunch and Percent deviation in HH Poverty



Serious truncation and non-linearity issues

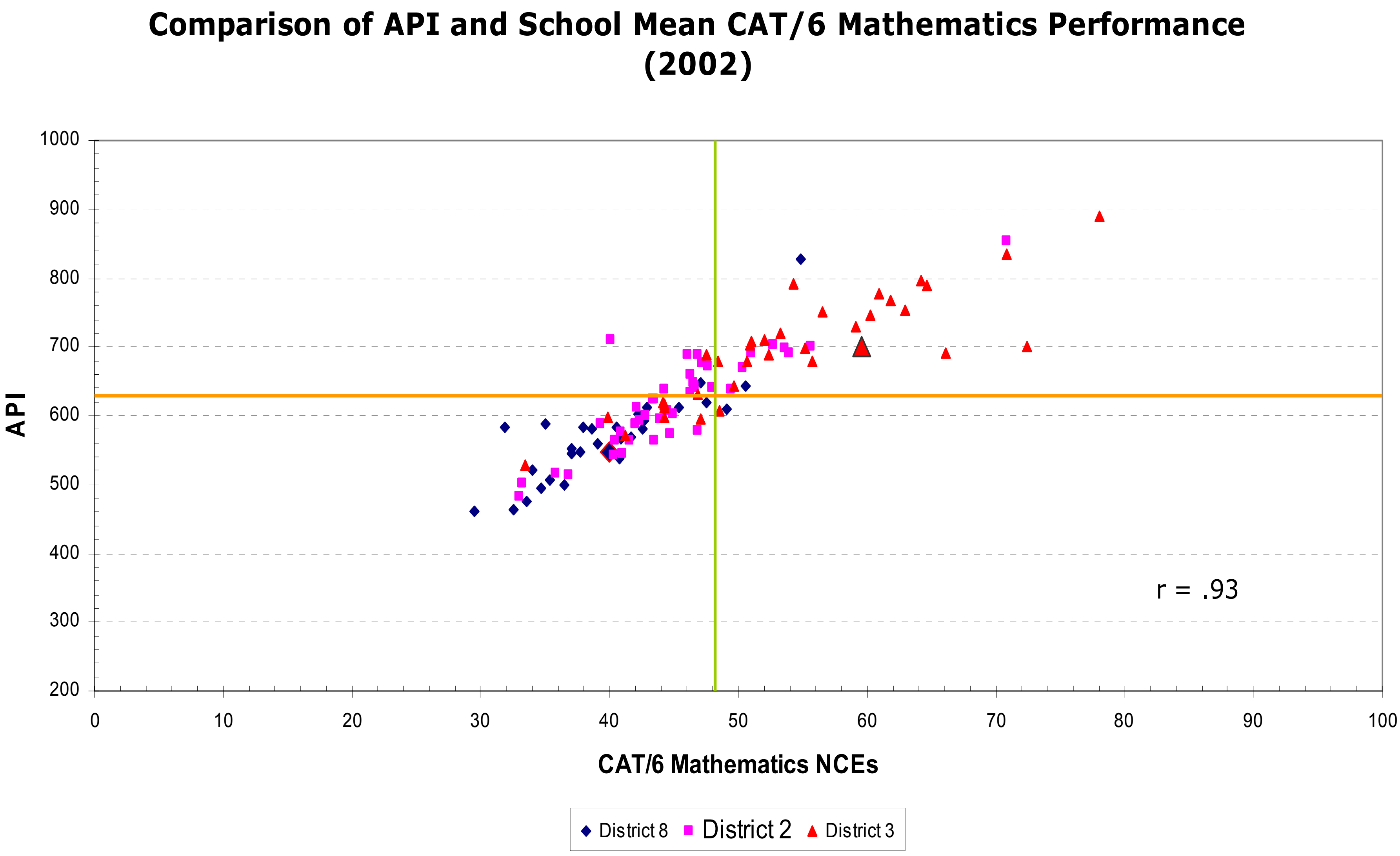
Comparison of mean school performance and percent of households in poverty



All but two school in District 8 are below average in poverty….

Can simply use percent poverty as measure of school performance.

Move to a weighted compiled performance index - API



Models for School Accountability and Program Evaluation

Part II

Pete Goldschmidt

UCLA Graduate School of Education & Information Studies
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

6th Program, Edward F. Reidy, Jr. Interactive Lecture Series: Incorporating Measures of Student Growth into State Accountability Systems
October 7-8, 2004
Nashua, NH

Means by no means

- Tracking unadjusted school means as a measure of school performance places a significant emphasis on school enrollment characteristics.

- Performance indices not taking individual student characteristics into account will also over-emphasize school enrollment characteristics.

Do you think I mean the means mean what you think means mean…..

- Using means also over-emphasizes the effect of student background characteristics.
- Assumes that all of the variation in student outcomes is between schools.
- Under-estimates standard errors—that is, over-estimates precision of mean estimates— making it appear as if schools are statistically different when they are not.
- Simply comparing mean performance and aggregated school characteristics leads an ecological fallacy.

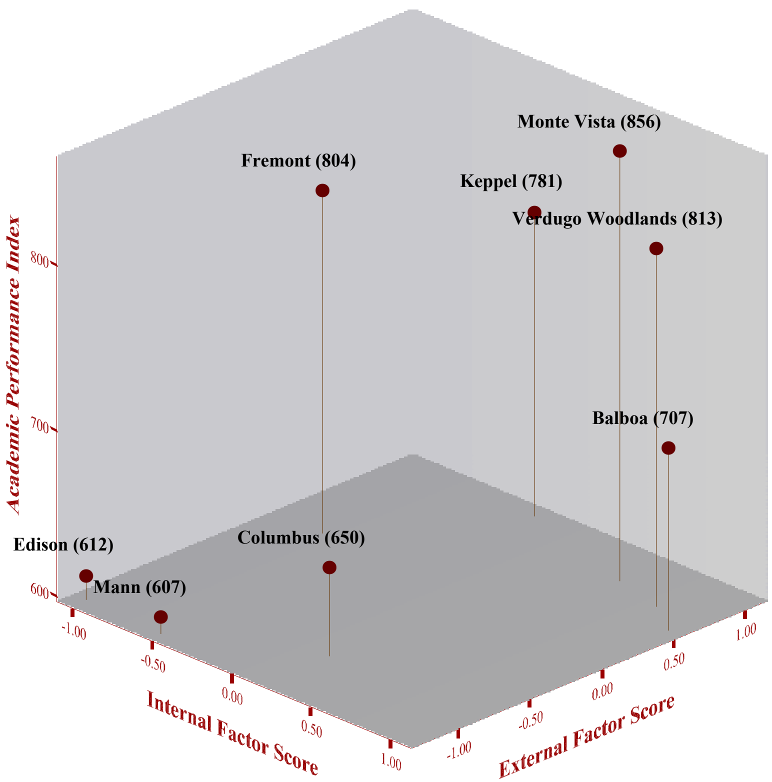
- Using means also over-emphasizes the effect of student background characteristics.
  - Within-school variability is lost upon aggregation, which will increase relationships between aggregate variables.
- Assumes that all of the variation in student outcomes is between schools
  - In fact, only about 30% (elementary schools) of the variability in student outcomes is between schools. The majority of the variability is within schools.

Evidence suggests that simple means or performance indices that aim to capture variation in school performance capture exactly that portion of school performance *not* controllable by the schools.

- Example:
- Raters completed a school quality survey and these survey results were compared with existing measures of school quality.
  - The survey results indicated that school quality can generally be summarized into two factors:
    - The survey results
- External factors – factors outside of school control;
- Internal factors – factors controllable by schools.

Correlations Among Various Measures of School Quality

	API 2000	API 2001	Overall Quality	External Factor	Internal Factor
API 2000	1	.995	.757	.889	.405
API 2000		1	.754	.884	.419
Overall Quality			1	.900	.857
External Factor				1	.584
Internal Factor					1



- External Factors:
  - Students
  - Parents
  - Facilities
- Internal Factors:
  - Leadership
  - School improve.
  - School org.
  - Resource mgmt.
  - Teamwork
  - Curriculum
  - Instruction
  - Professional Dev.

- Under-estimates standard errors—that is, over-estimates precision of mean estimates— making it appear as if schools are statistically different when they are not.
- Simply comparing mean performance and aggregated school characteristics leads an ecological fallacy (inferences about individuals based on group data).
  - Variables have different meanings at different levels of aggregation.
  - When we use aggregate measures to proxy for individual characteristics, we don't know whether the estimate is significant because the aggregate measure is actually significant or whether it is significant due to the underlying omitted individual characteristic.

Moving beyond unconditional means

- Can use conditional mean school performance.
- Three ways to estimate:
  - OLS regression ignoring school membership, then aggregate residuals to the school.
  - OLS regression of school averages.
  - Multilevel models.

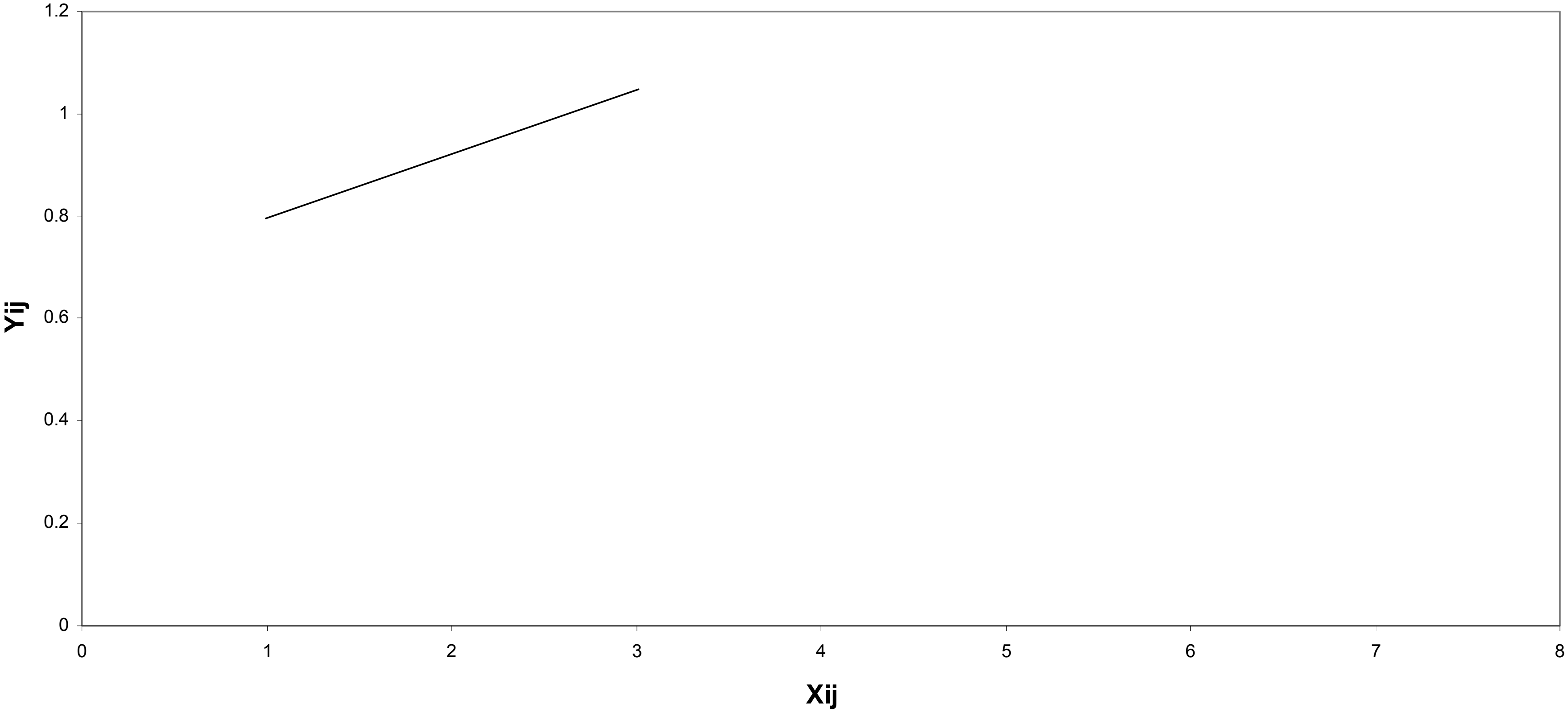
Example: Total, between, and within school relationships

Data

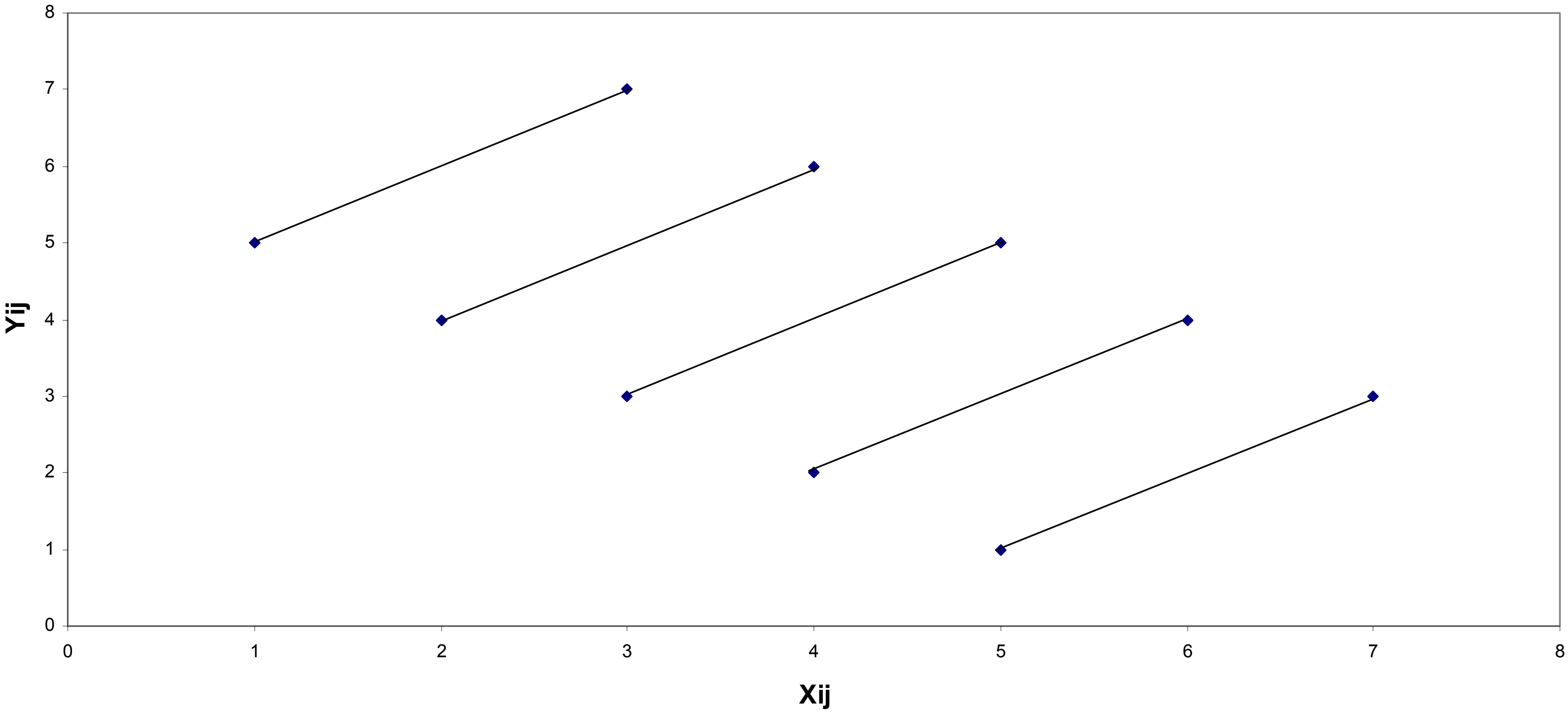
j	I	xij	Xbar	yij	ybar
1	1	1	2	5	6
1	2	3	2	7	6
2	1	2	3	4	5
2	2	4	3	6	5
3	1	3	4	3	4
3	2	5	4	5	4
4	1	4	5	2	3
4	2	6	5	4	3
5	1	5	6	1	2
5	2	7	6	3	2

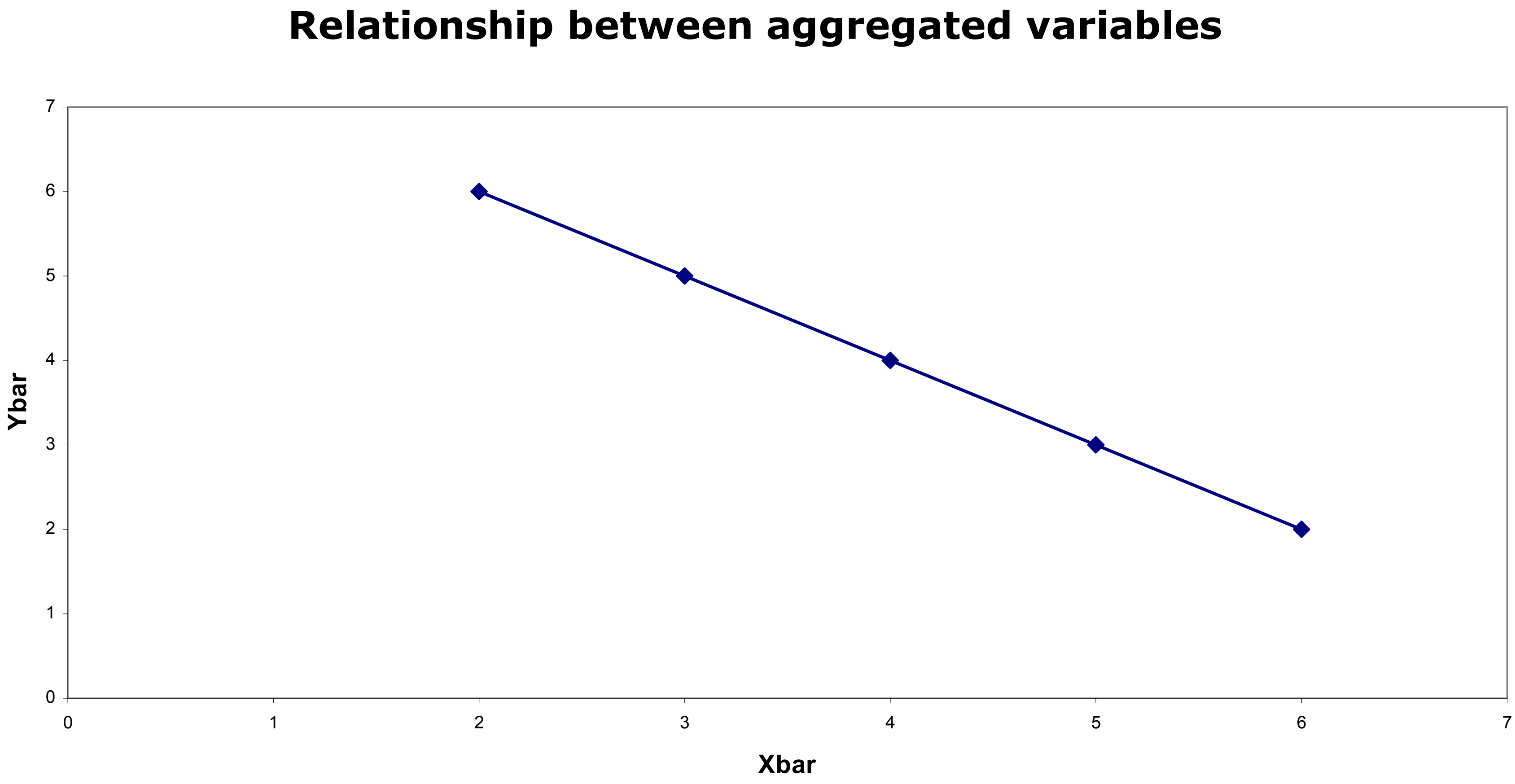
Where j = school i = student
y= test score x = hrs studying

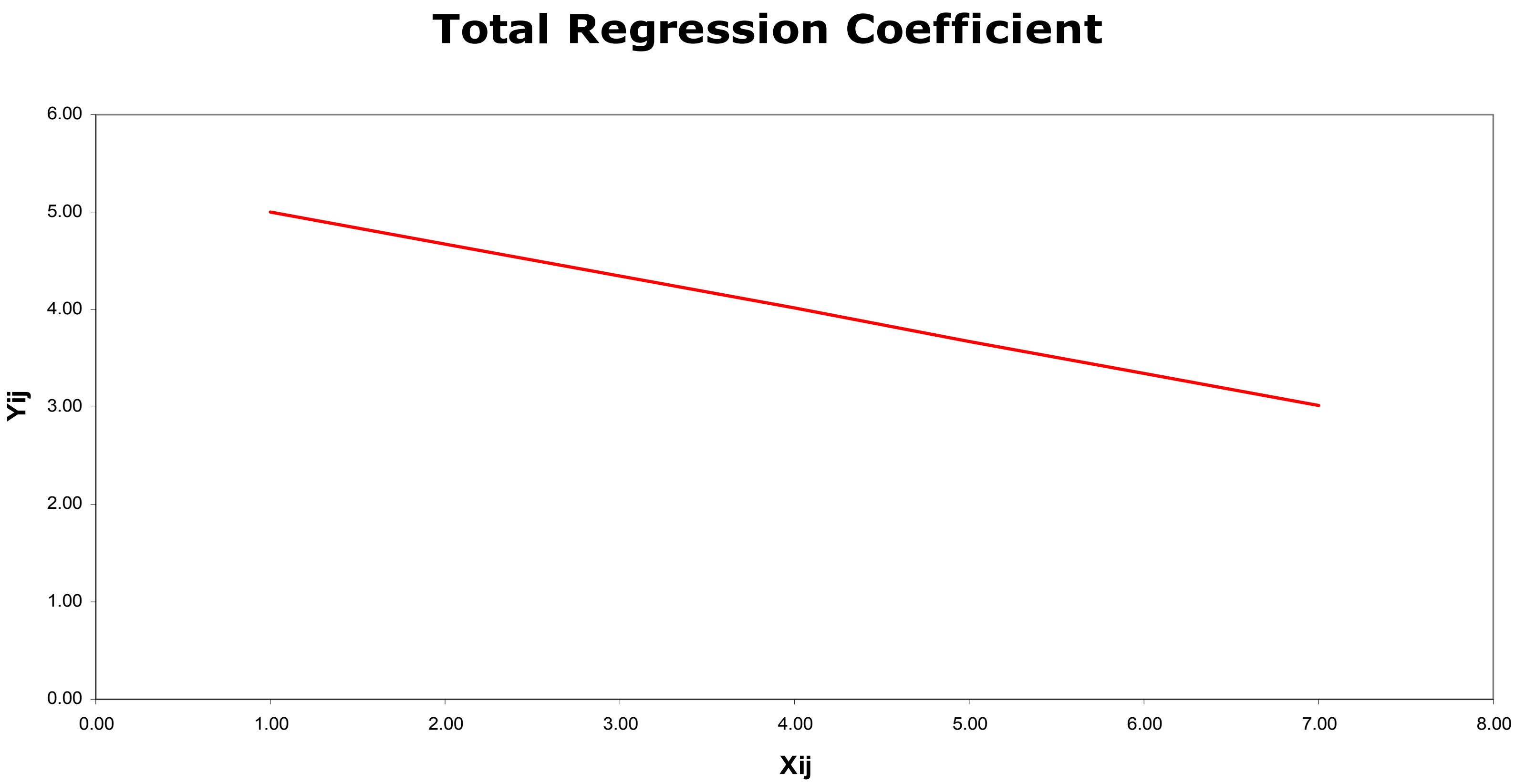
Regression within a single unit



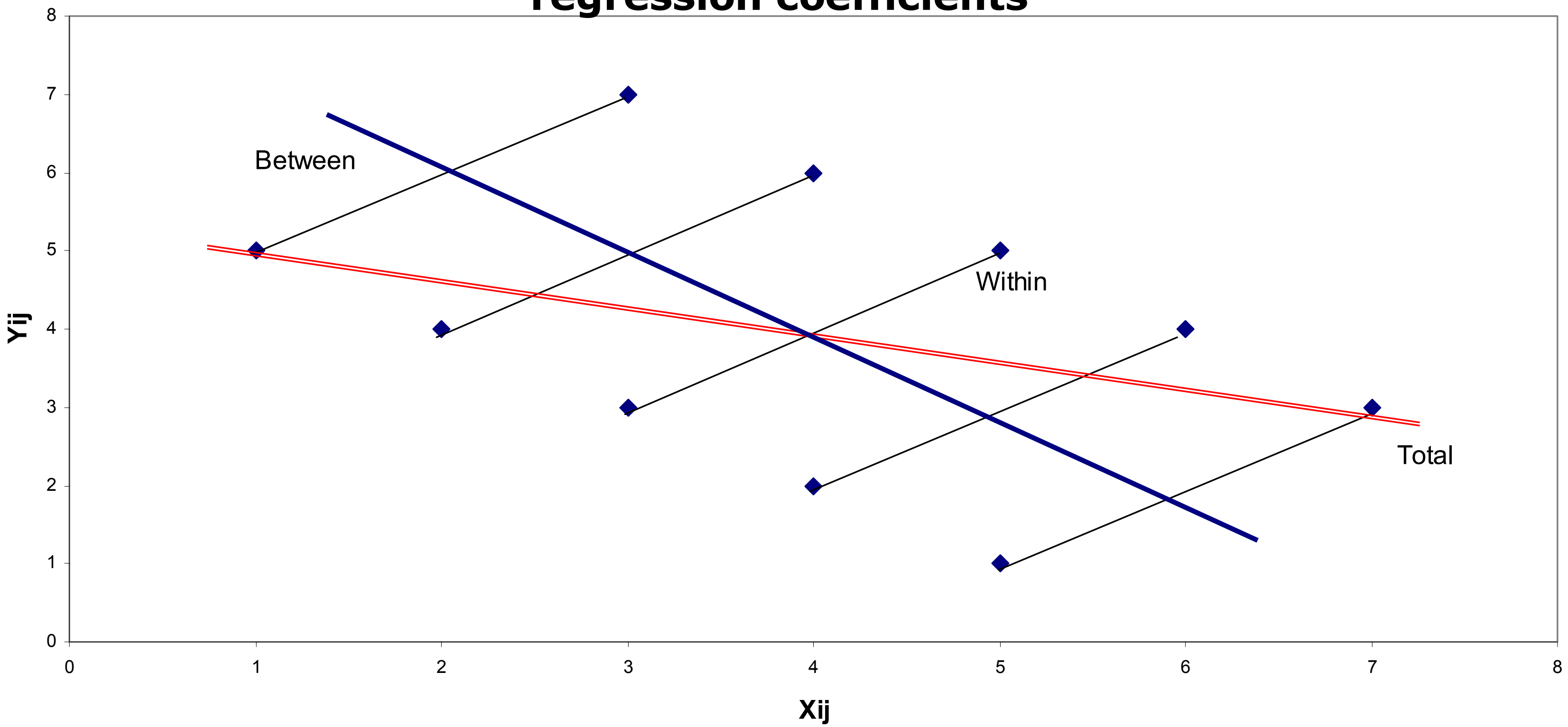
Regressions within units




Relationship between aggregated variables
Ybar
Xbar
0
1
2
3
4
5
6
7


Total Regression Coefficient
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Yij
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
Xij

Relationship among the total, within, and between regression coefficients



Quasi Value Added – the cross sectional case

$A_j = \beta_1 PctB_1 + \dots \beta_n PctB_n + E_j$

$E_j = A_j - b_1 PctB_1 + \dots b_n PctB_n$,

Where b is estimated β.

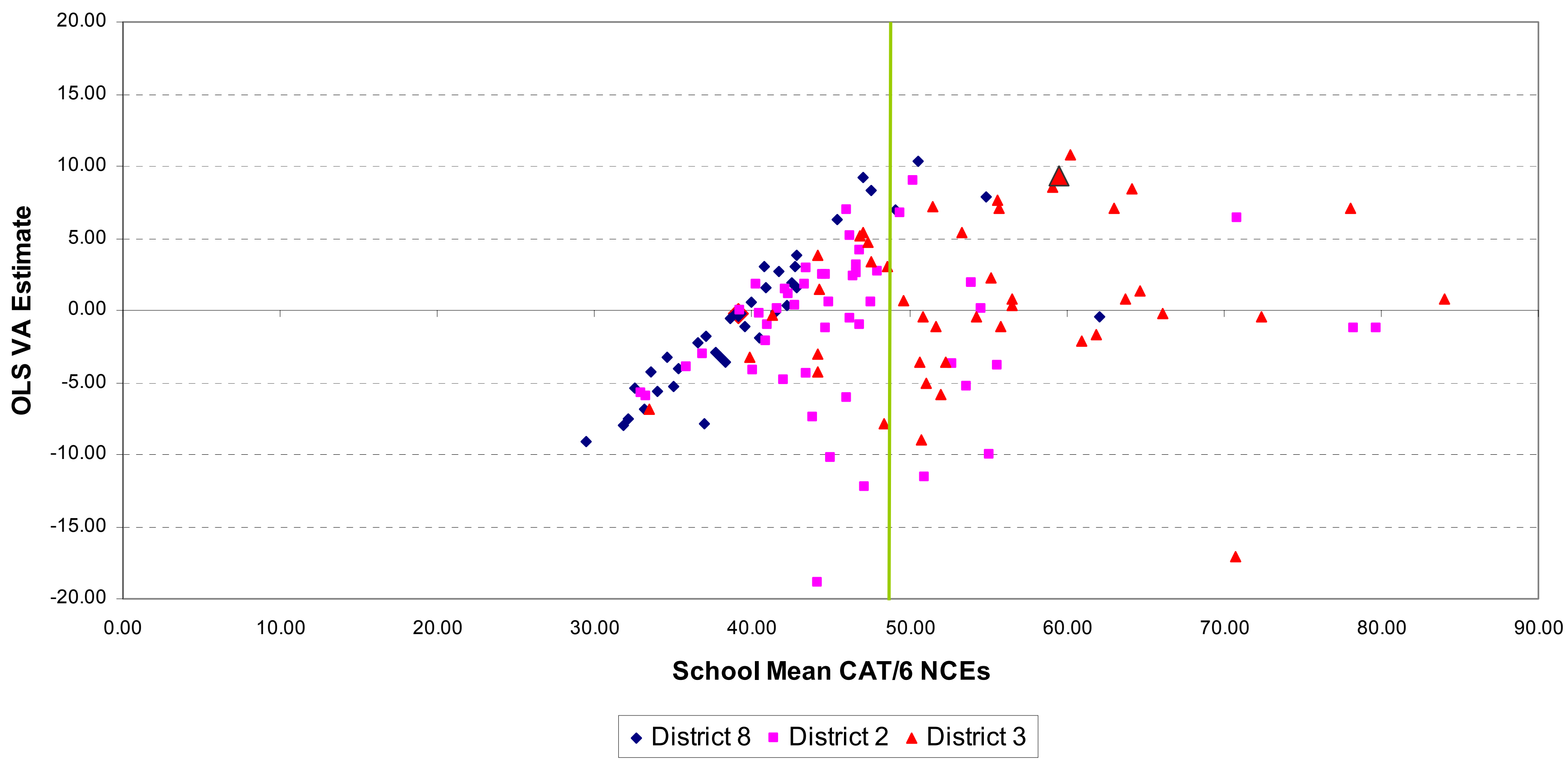
Value Added for school j = $VA_j = E_j$

That is the value added for a school is the observed mean minus what the predicted mean would be, given the existing percentages of student background characteristics.

The value added estimates are uncorrelated with anything on the right hand side (RHS) of the equation.

Can only be calculated where there is complete data for each school.

Comparison of OLS Aggregate VA Estimates and School Means



Quasi Value Added – the cross sectional case

Correlations of cross-sectional Value Added estimates									
	VA 94	VA 95	VA 96	VA 97	VA 98	VA 99	VA 00	VA 01	VA 02
VA 94	0.84	0.58	0.58	0.52	0.54	0.37	0.53	0.47	0.36
VA 95		0.63	0.61	0.52	0.54	0.40	0.53	0.44	0.38
VA 96			0.79	0.63	0.61	0.42	0.49	0.47	0.41
VA 97				0.83	0.73	0.48	0.59	0.50	0.44
VA 98					0.79	0.55	0.63	0.53	0.42
VA 99						0.69	0.76	0.67	0.55
VA 00							0.68	0.58	0.48
VA 01								0.83	0.69
VA 02									0.81

This creates an adjusted mean, but does not control for sources of internal invalidity.

Need to account for potential sources of invalidity—i.e. alternative explanations for hypotheses (the school is responsible for the observed performance). Need to reduce as much as possible rival hypotheses.

Either: Random assignment of students and teachers, or students act as their own controls, through longitudinal models.

Value Added Basics

The underlying assumption for value added models is:

$$A_{it} = f(B_{it}, P_{it}, S_{it}, I_{it}, E_{it}), \qquad (1)$$

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.

If we assume that (1) holds for any time t, then we can consider change in achievement from t to t`.

$A_{it`} - A_{it} = f(.)$

Then by simply adding A_{it} to both sides, we get a familiar model:

$$A_{it`} = f(B_{it`-t}, P_{it`-t}, S_{it`-t}, I_{it}, A_{it}, E_{it}) \qquad (2)$$

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 A_{it}.

Also, remaining variables measured contemporaneously, but this is generally not too problematic since only going back from t` to t.

Another strike against OLS in studying school quality

$$A_{it`} = \beta_0 + B_i\beta_1 + A_{it}\beta_2 + SI + E_{it`} \qquad (3)$$

This model assumes that the marginal effect of S is independent of A or B, and that they are substitutes.

This model also assumes that the effect of A and B are constant.

More likely that:

$$A_{it`} = B_i\beta_{1j} + A_{it}\beta_{2j} + E_{it`}$$

and that

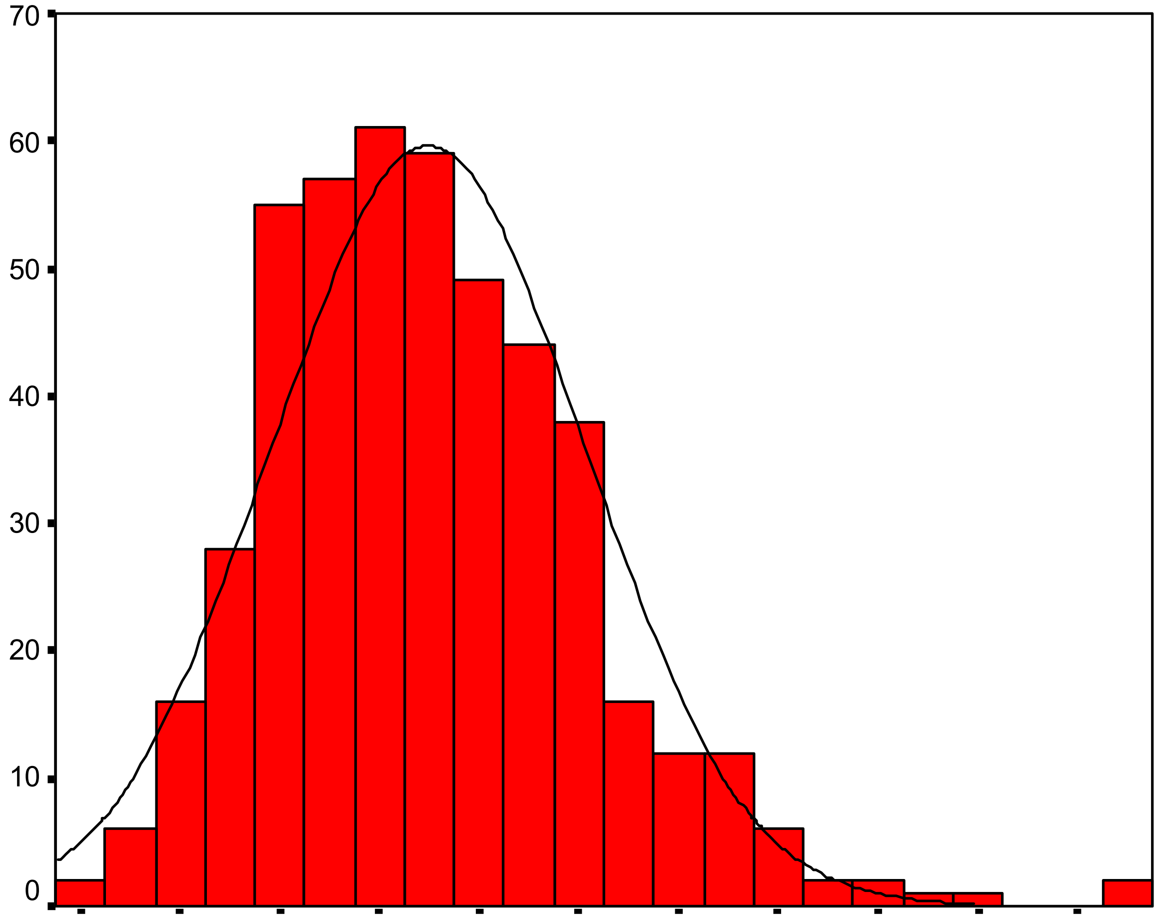
$$\beta_0 = \gamma_{00} + \gamma_{01}S + u_{oo}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}S + u_{1o}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}S + u_{20}$$

Which implies that:

Distribution of β_{2j}



Mean = γ_{20}, standard deviation var $(U_{20}) = \tau_{20}^{1/2}$

Models for School Accountability and Program Evaluation

Part III

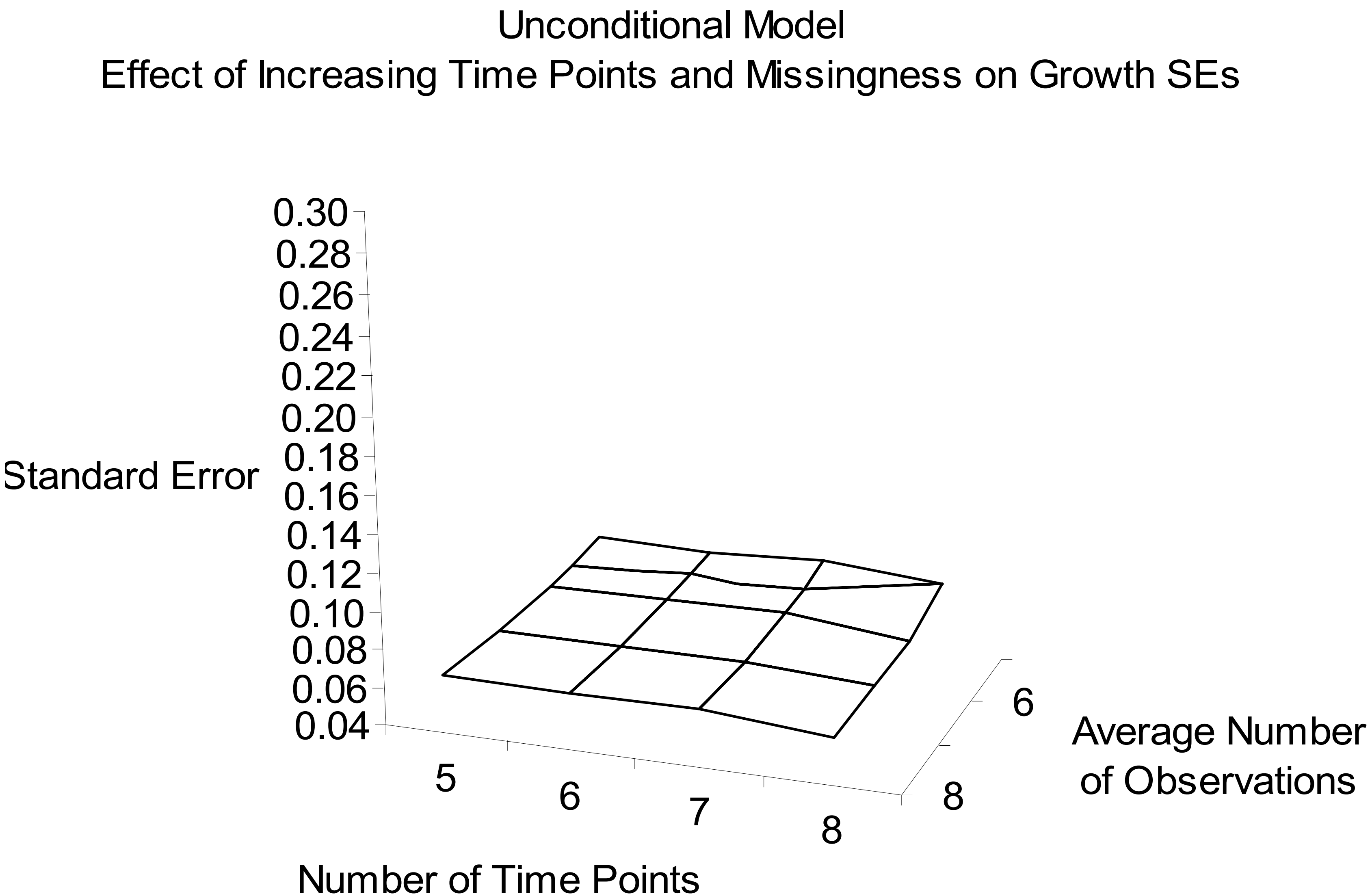
UCLA Graduate School of Education & Information Studies
Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, and Student Testing

Advantages of multiple time points:

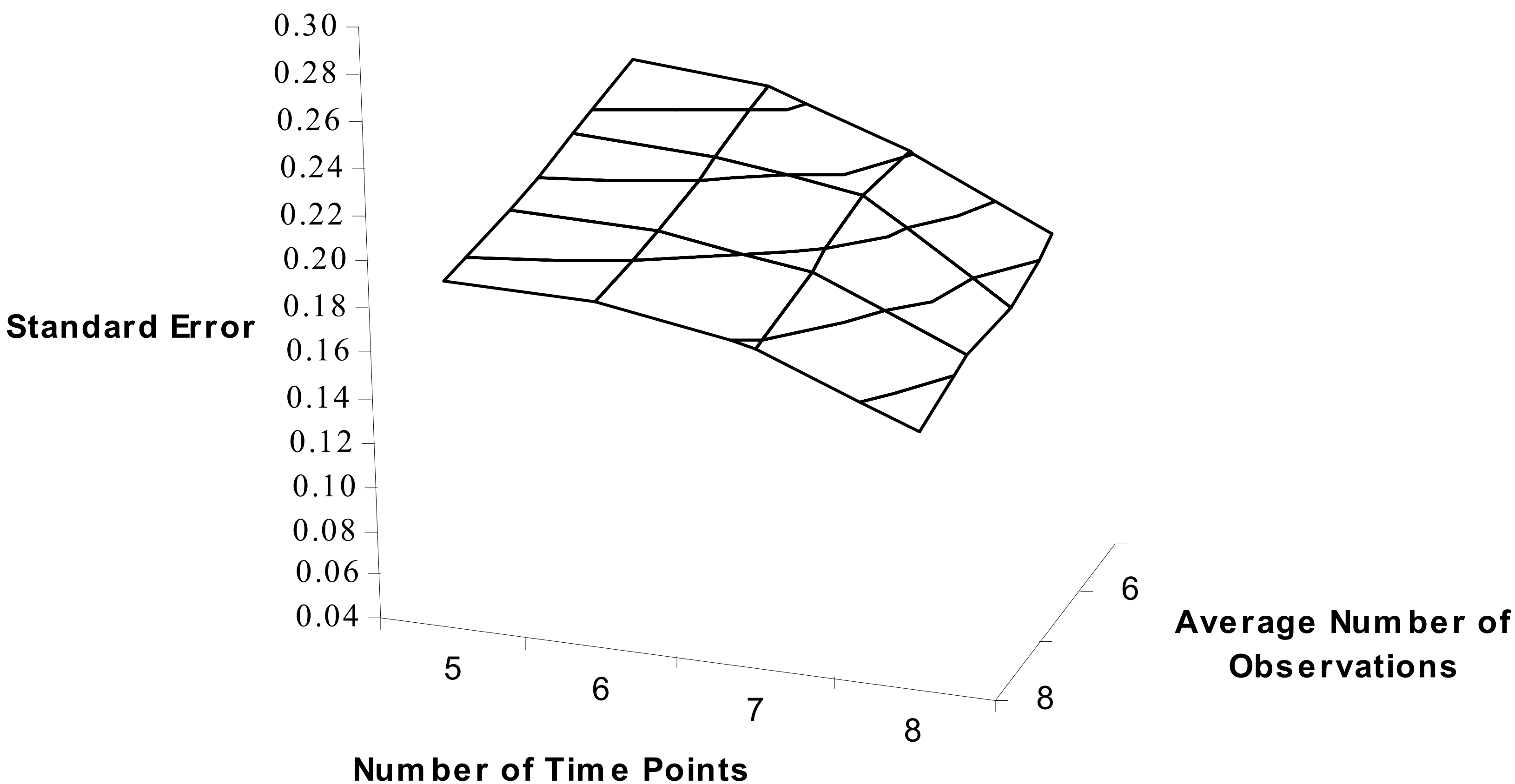
Avoid spurious negative correlation between pre-test and gains

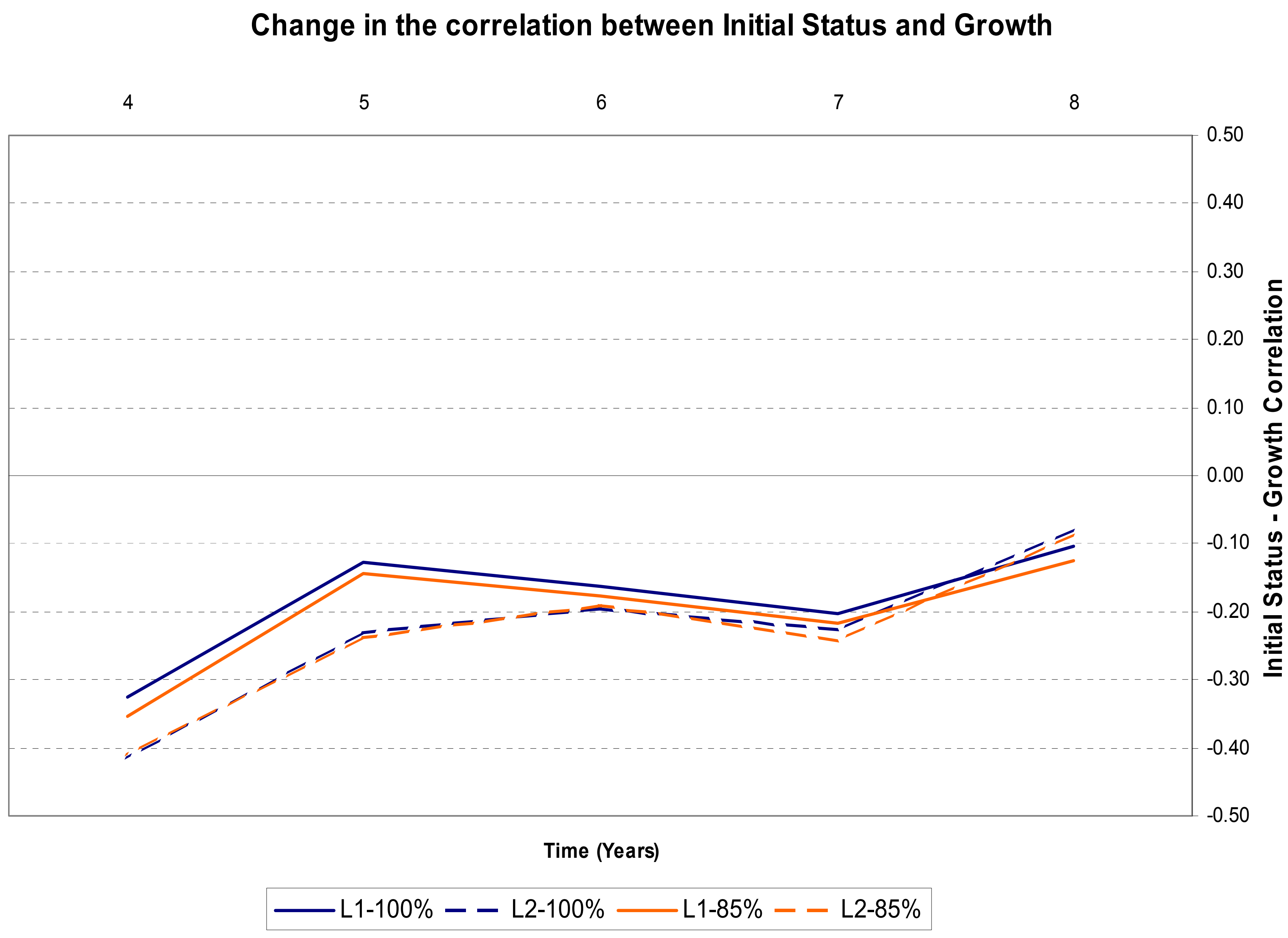
In fact evidence suggest that as occasions are added to the mode; the correlation between initial status and growth (in absolute value) decreases.

Generate more precise estimates of change as add occasions.


Unconditional Model
Effect of Increasing Time Points and Missingness on Growth SEs
Standard Error
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
5
6
7
8
Number of Time Points
6
8
Average Number
of Observations

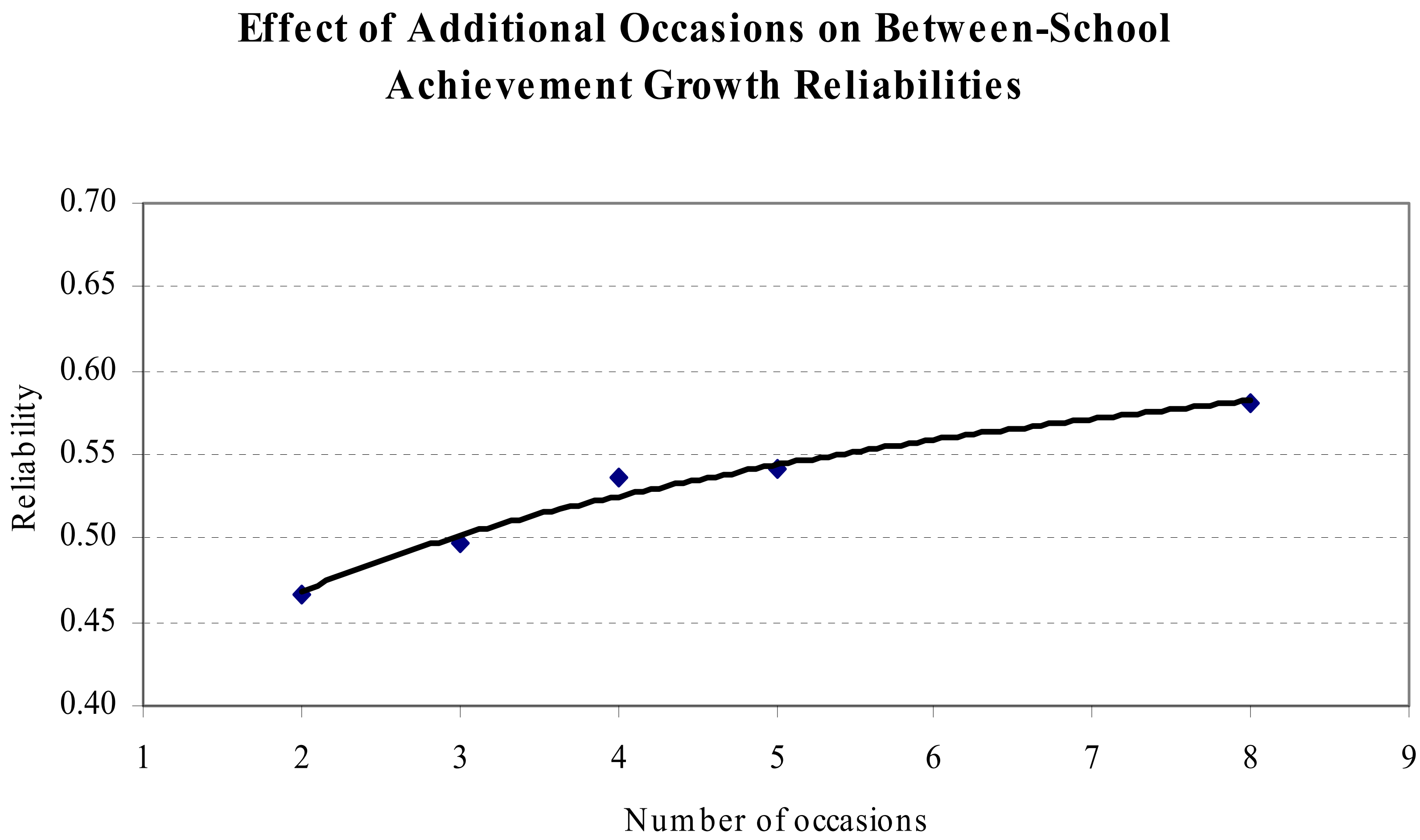
Model 3
Effect of Increasing Time Points and Missingness on Growth SEs





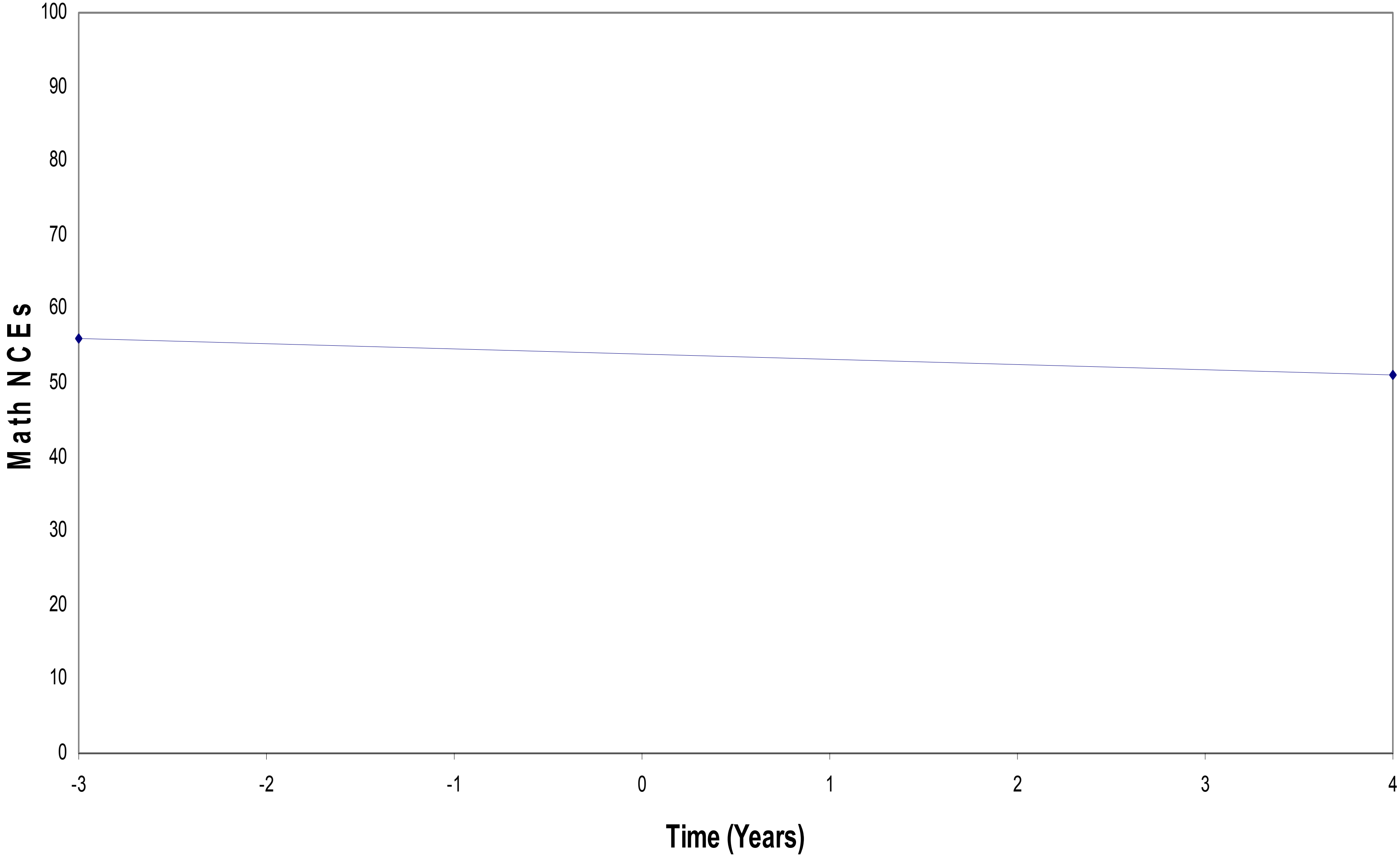
Change in the correlation between Initial Status and Growth
4
5
6
7
8
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
Initial Status - Growth Correlation
Time (Years)
L1-100%
L2-100%
L1-85%
L2-85%

Generate more reliable estimates with additional occasions

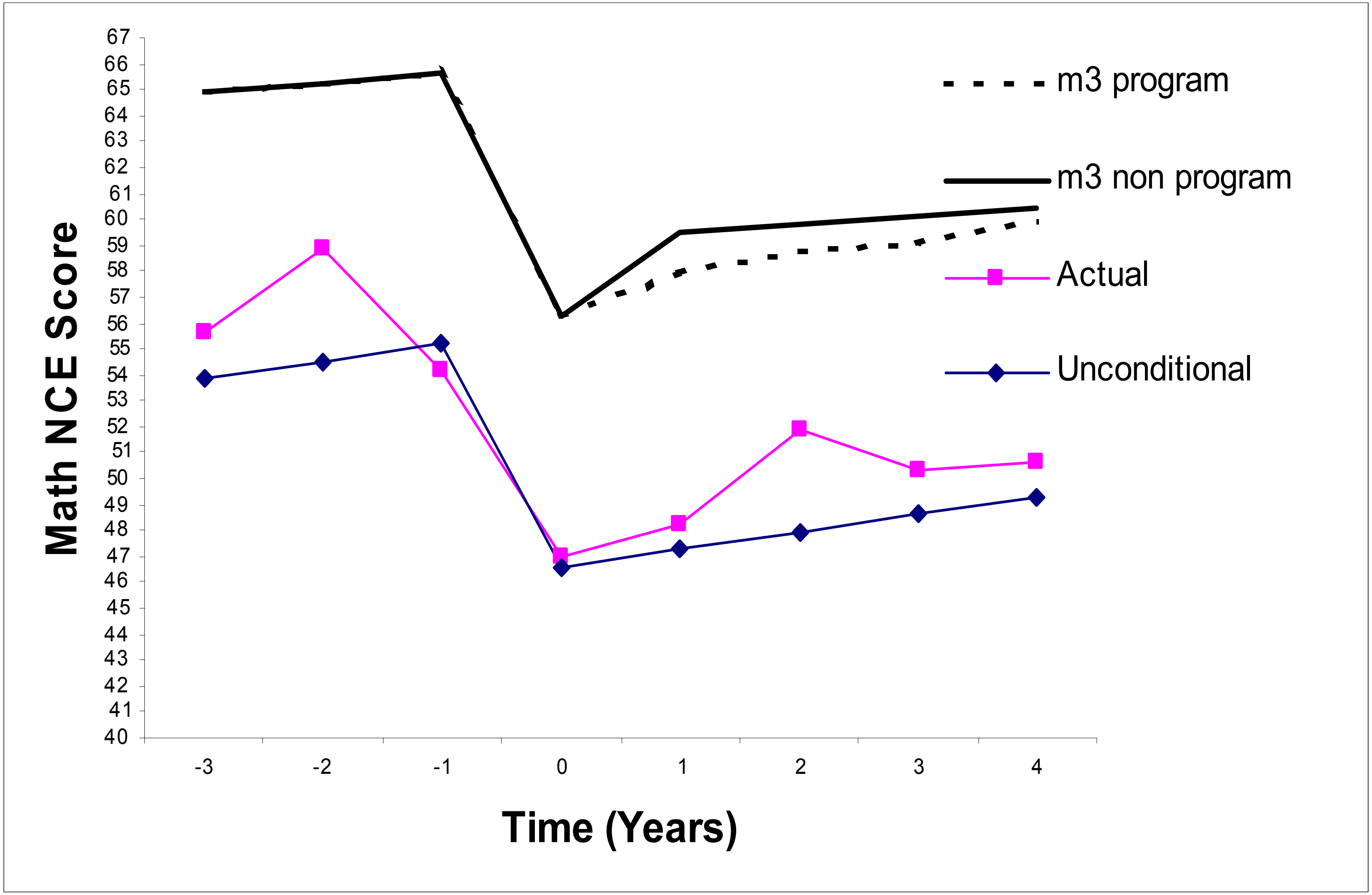


The higher the reliability, the greater the ability to detect true differences among schools.

Two Time Point Estimated Trajectory



Multiple occasions allow for a more accurate portrayal of change over time.



Types of Longitudinal Models

Longitudinal Growth Panel Models (LGPM)

-
-
-
-

LGPM:

where

$\boldsymbol{Y_{tij}}$ is the outcome at time ***t* for student *i* in school *j***

(TIME)$_{tij}$ is 0 at Grade 5 which is the initial status year for the Program, 1 at Grade 6, etc.

We can add grades backwards (Grade 4 = -1, etc.) This add precision to the t=o estimates.

π_{0ij} *is the initial status of student* ***ij* or the expected outcome at start of program or**

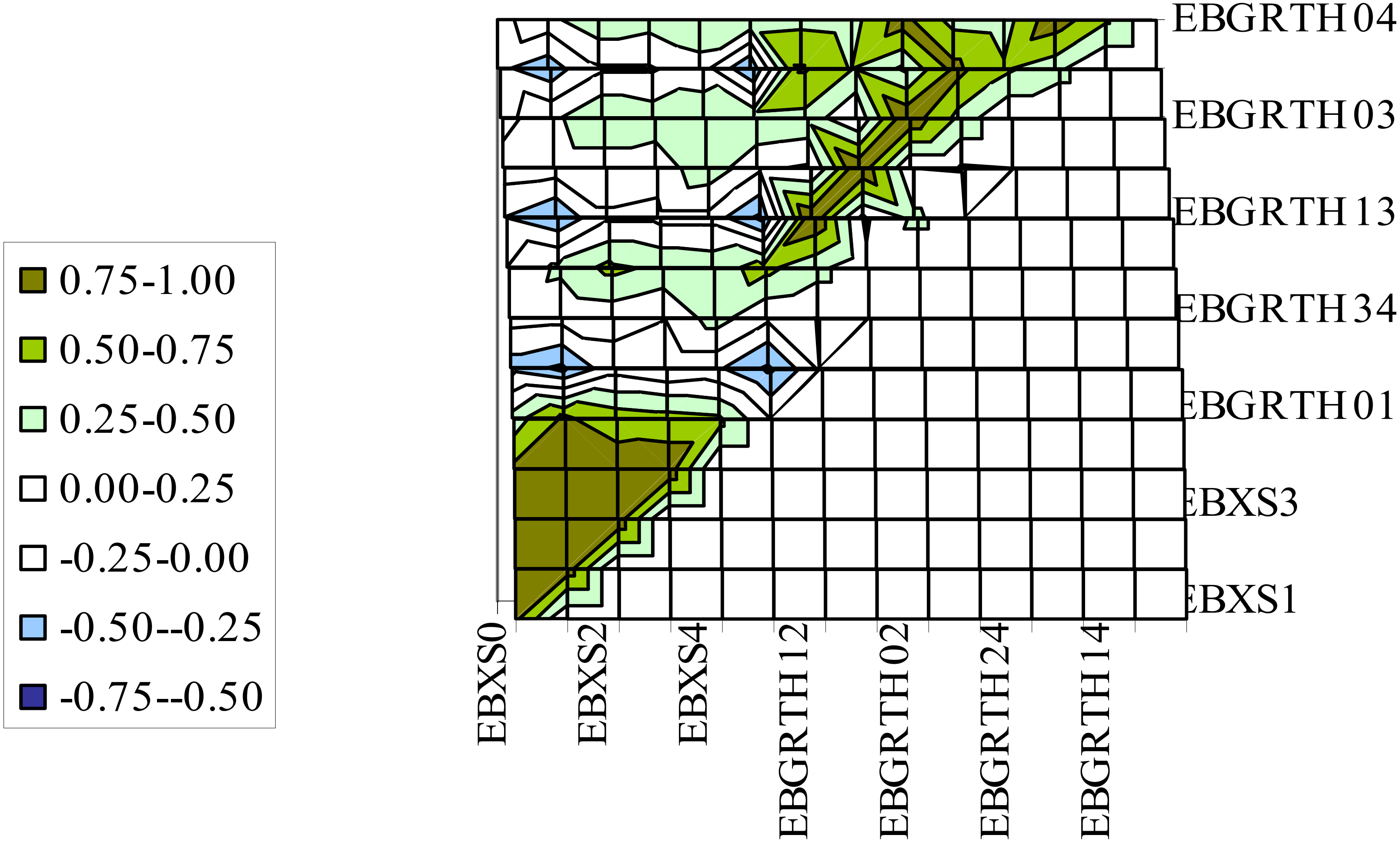
Grade 5;

π_{1ij} is the learning rate for student ***ij* during the school year; and**

LGPM:

Advantages of LGPM:

- Direct measure of growth.
- Do not need complete outcome data for each student.
  - Previously demonstrated that growth model estimates are robust to sample sizes down to 30 per school (grade).
  - Growth model estimates are robust to missingness (given missingness is not systematic – unless can model systematic part).



Correlations Among EB Estimates

Fairly inconsistent results, unless extending data

LGPM:

Is school mean status related to school mean growth?

	Not met AYP				Met AYP			
	Estimate	s.e.	t	approx p	Estimate	s.e.	t	approx p
Mean annual growth	0.45	0.46	0.97	0.331	1.85	0.74	2.51	0.012
Mean effect of "disadvantagedness" (school)	0.56	0.26	2.17	0.030	1.12	0.53	2.11	0.035
Effect of mean status on growth	0.002	0.001	2.00	0.046	-0.001	0.002	-0.80	0.426

	Total Effect	Direct Effect	Indirect Effect	approx p	Total Effect	Direct Effect	Indirect Effect	approx p
Mean annual growth	1.35	0.45	0.90	0.023	1.27	1.85	-0.58	0.213
Mean effect of "disadvantagedness" (schooll)	0.39	0.56	-0.16	0.031	1.26	1.12	0.14	0.215

Note, latent model is:

B10 = G100* + G101*(PCTLOW) + G102*(B00) + U10*

6) **Is school mean status related to growth (does where a school starts impact growth – i.e. chances of meeting AYP)?**

Question and Model

1) What difference does the school a child attends make in a child
   achievement growth?
   a. Student variables grand mean centered – means adjuste
      for differing school enrollments.
   b. How would the average student do in the school.
   c. Common value added model.

2) Do the effects of student background vary among schools?
   a. Group mean center variables – means are school means.
3) Do schools play a mediating role in terms of student factors tha
   affect achievement growth?

Group mean center student variables and add mean student lev
variables into school level model

Do schools Matter?

Fixed effects	Coefficient	SE	*p*
Average 19998-99 Intial Status	578.8	3.20	0.00
Average change (growth) in Reading	22.7	0.36	0.00

Random Effect		Variance Component		*p*
Level 1				
	Within-student (temporal) variation (residua	276.7		
Level 2				
	Within-school variation – initial status	2356.6	31,368	0.00
	Within-school variation – growth	40.0	31,368	0.00
Level 3				
	Between-school variation – initial status	656.0	62	0.00
	Between-school variation – growth	7.4	62	0.00
Variation between schools				
	In 1998-99 status	21.8%		
	In growth	15.6%		

Effect of School variability in Iniitial Status and Achievement Growth

	mean + 95%CI		Effect Size
I.S.	528.6	629.0	2.09
Growth	17.4	28.0	0.22

Models to address questions 1 through 3

	(1)	(2)	(3)	Differ +/- 2
Fixed Effects				
Model for Intial Status				
Model for school mean IS				
Intercept	577.8 *	579 *	614 *	
Pct ELL			-0.848	-
Model for within school between ELL and IS Intercept	-32.2 *	-33 *	-32 *	
Model for Achievment Growth				
Model for school mean growth Intercept	22.7 *	22.5 *	19.3 *	
Pct ELL			0.077 *	
Model for within school between ELL and gro	4.5 *	4.3 *	4.4 *	
Variance components				
Initial Status		655.5		
Growth		6.94		
Initial Status - ELL		147.8 *		

Notes:

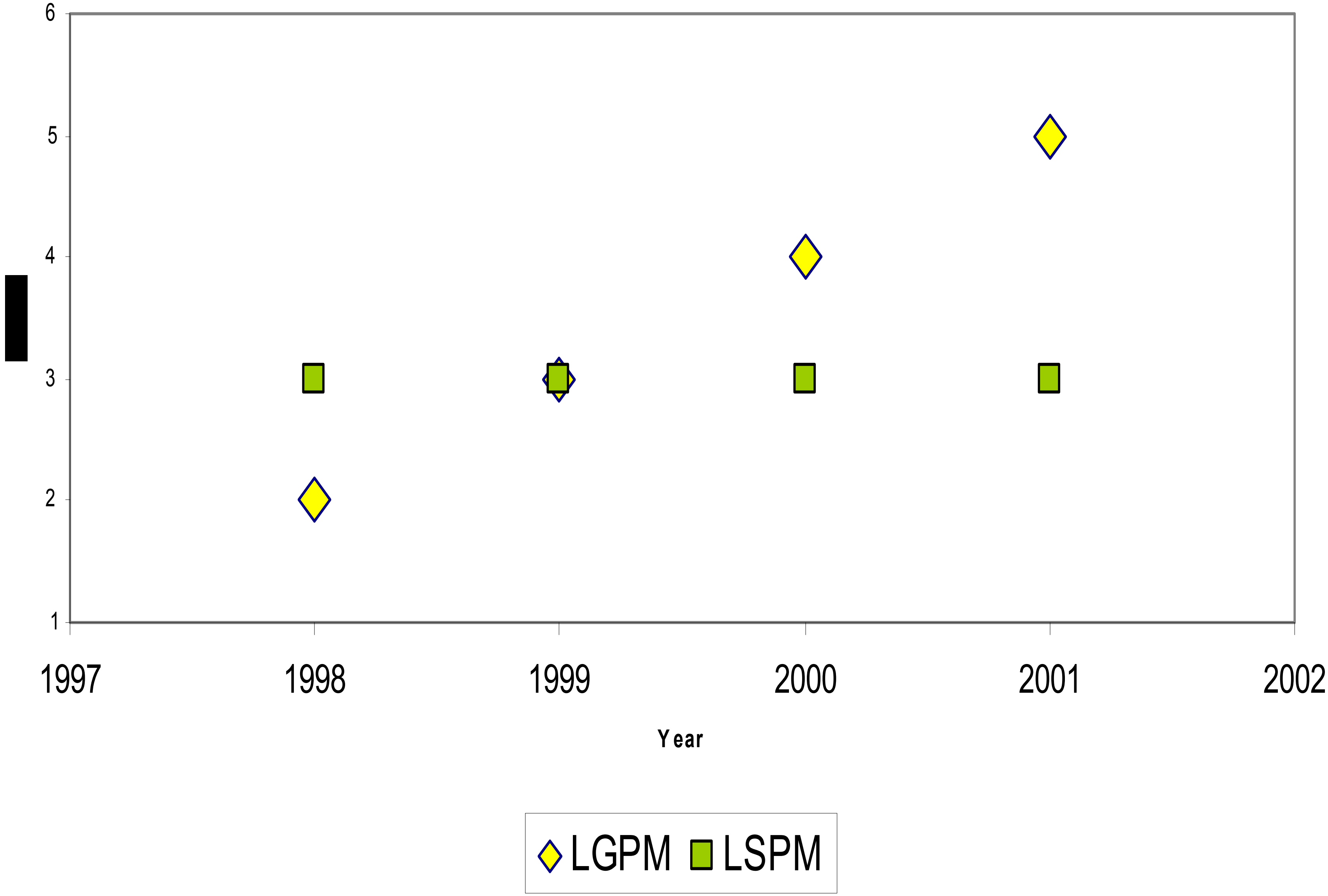
Longitudinal School Productivity Model (LSPM)

where Y_{ijt} is the outcome for student i (i = 1,..,n_j) in school j (j = 1,…,J) at occasion (or occasion (or cohort) t (t = 1,…T). β_{jt0} are estimates of performance for each school j school j and occasion t.

The level-2 (between-cohort; within-school) model, where we include a time metric such metric such that we estimate initial status and growth rate for school j:

where $Time_{tj}$ takes on values of 0, 1, 2, 3, 4 such that θ_{j0} represents the status at the first year (i.e., $Time_t$ = 0) or initial status of school j. In addition θ_{j1} represents yearly improvement / growth rate during the span of time for school j. As such, the above level-2 model specifies a school-level linear growth modeling in the sense that the school mean at 5 different time points (β_{jt0}) is regressed on the time metric ($Time_{tj}$). The residual u_{jt} represents random year-to-year fluctuations in school's performance.

Comparison of Longitudinal growth models: Panel design vs. school productivity design



LSPM

Simpler data requirements because no need to match students year to year.

Can sample different students each years (rather than track a sample of students over time.

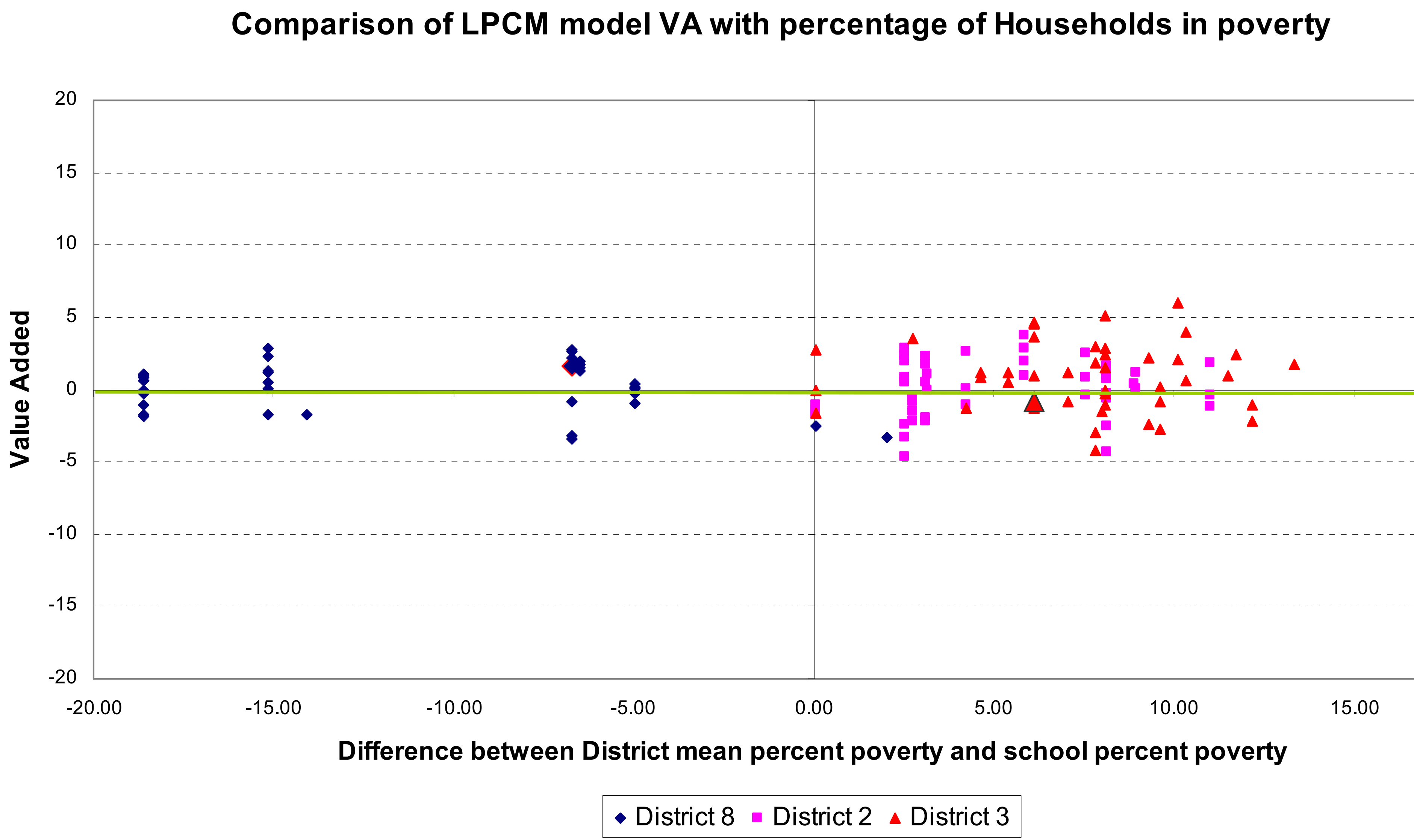
Results between LGPM and LSPM may match (this warrants further study)

Table 3: Correlations between indicators of school performance

SAT9 Reading	LGPM SS	LSPM NCE	LSPM SS
LGPM NCE	0.95	0.69	0.64
LGPM SS		0.68	0.66
LSPM NCE			0.95

SAT9 Mathematics	LGPM SS	LSPM NCE	LSPM SS
LGPM NCE	0.97	0.25	0.29
LGPM SS		0.31	0.36

Longitudinal Panel Cohort Model (LPCM)


Comparison of LPCM model VA with percentage of Households in poverty
Value Added
20
15
10
5
0
-5
-10
-15
-20
-20.00
-15.00
-10.00
-5.00
0.00
5.00
10.00
15.00
Difference between District mean percent poverty and school percent poverty
District 8
District 2
District 3

Longitudinal Program Evaluation Models (LPEM)

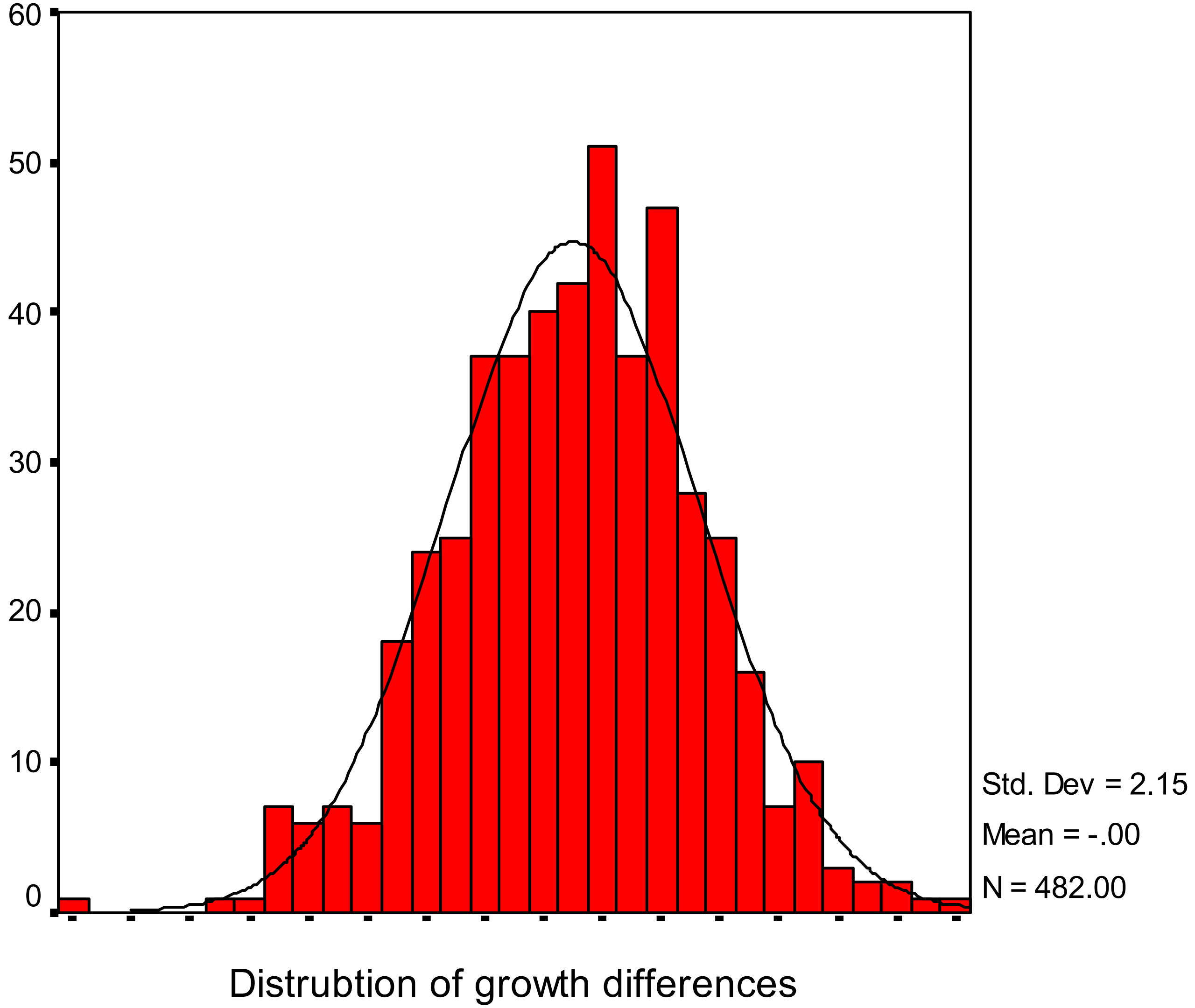
Examples:

- Comparison of growth dynamic between schools meeting AYP and those not meeting AYP.
- Examining specific dimensions of school quality.
- School reform evaluation.

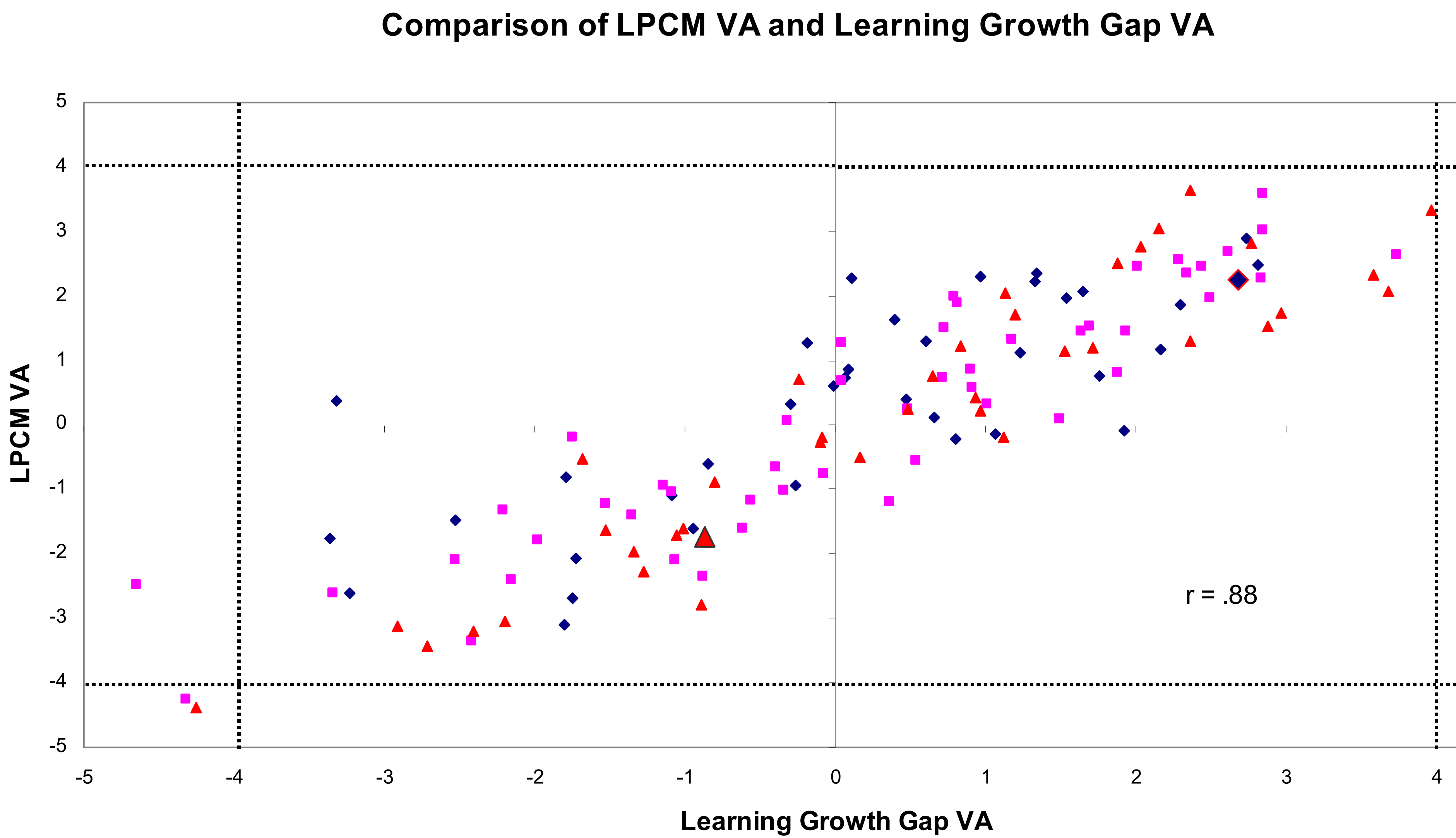
Comparison of Subgroup Performance in Schools Not Meeting and Schools Meeting

	Not Met AYP			Met AYP		
			Vary Among			Vary Among
	Estimate	Effect Size	Schools	Estimate	Effect Size	Schools
School mean status in 2001-2002	37.4		yes	41.4		yes
Achievement gap for Low SES students	-11.7 *	-0.56	yes	-11.8 *	-0.56	yes
Achievement gap for ELL students	-12.7 *	-0.61	yes	-11.2 *	-0.54	yes
Achievement gap for Minority students	-5.1 *	-0.24	yes	-4.6 *	-0.22	no
Achievement gap for Spec. Educ. students	-7.4 *	-0.35	yes	-6.4 *	-0.30	yes
Achievement gap for GATE students	16.9 *	0.80	-	15.8 *	0.75	-
School mean annual growth	1.4 *		yes	1.4 *		yes
Achvmnt growth diff for Low SES students	0.2	0.01	no	0.9 *	0.04	no
Achvmnt growth diff for ELL students	0.3 *	0.01	yes	0.8 *	0.04	yes
Achvmnt growth diff for Minority students	-0.2	-0.01	no	0.1	0.00	no
Achvmnt growth diff for Spec. Educ. students	-0.4	-0.02	no	0.6 *	0.03	no
Achvmnt growth diff for GATE students	1.1 *	0.05	no	0.8 *	0.04	no

Examining specific dimensions of school quality



Mean = γ_{20}, standard deviation var(U_{20}) = $\tau_{20}^{1/2}$


Comparison of LPCM VA and Learning Growth Gap VA
LPCM VA
Learning Growth Gap VA
r = .88
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4

Change-from-baseline with posttest effect for program students

$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{TIME})_{tij} + \pi_{2ij}(\text{TEST})_{tij} + \pi_{3ij}(\text{PREVLATE})_{tij} + \pi_{4ij}(\text{POSTVOTH})_{tij} + e_{tij},$
(level-1)

$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{PROGRAM})_{ij} + r_{0ij},$ (level-2)
$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{PROGRAM})_{ij} + r_{1ij},$ (level-2)
$\pi_{2ij} = \beta_{20j},$ (level-2)
$\pi_{3ij} = \beta_{30j} + \beta_{31j}(\text{PROGRAM})_{ij} + r_{3ij},$ (level-2)
$\pi_{4ij} = \beta_{40j} + r_{4ij},$ (level-2)
$\beta_{00j} = \gamma_{000} + \gamma_{010} + u_{00j},$ (level-3)
$\beta_{10j} = \gamma_{100} + \gamma_{110} + u_{10j} + u_{11j}$ (level-3)
$\beta_{20j} = \gamma_{200},$ (level-3)
$\beta_{40j} = \gamma_{400} + u_{40j}.$ (level-3)

Conclusions

Choice of models depends on:

- Research questions
- Policy goals

Conclusions