

**Dr. Dale Ballou
Vanderbilt University**

**Lessons from Value-Added
Assessment in Tennessee**

PRESENTED AT THE CONFERENCE

**VALUE-ADDED MODELING:
ISSUES WITH THEORY AND APPLICATION
October 21 and 22, 2004**

ORGANIZED BY

**Dr. Robert W. Lissitz, Director
MARYLAND ASSESSMENT RESEARCH CENTER
FOR EDUCATION SUCCESS (MARCES)
Department of Measurement, Statistics and Evaluation
University of Maryland
College Park, MD 20742**

SPONSORED BY

Maryland State Department of Education

Tennessee Value-Added Assessment

Model (for vertically sequenced tests)

Outcome measures are test scores taken in a given subject and grade level.

Consider the sequence of reading test scores for a student who is first tested in 1994 in 2nd grade.

Notation (subscripts for students and teachers have been suppressed):

Y_t^k = test score in year t , grade k

b_t^k = mean test score in year t , grade k in the district

u_t^k = contribution of teacher in grade k to test score in year t .

e_t^k = student-level error in year t , grade k

$$Y_{94}^2 = b_{94}^2 + u_{94}^2 + e_{94}^2$$

$$Y_{95}^3 = b_{95}^3 + u_{94}^2 + u_{95}^3 + e_{95}^3$$

$$Y_{96}^4 = b_{96}^4 + u_{94}^2 + u_{95}^3 + u_{96}^4 + e_{96}^4$$

...

etc.,

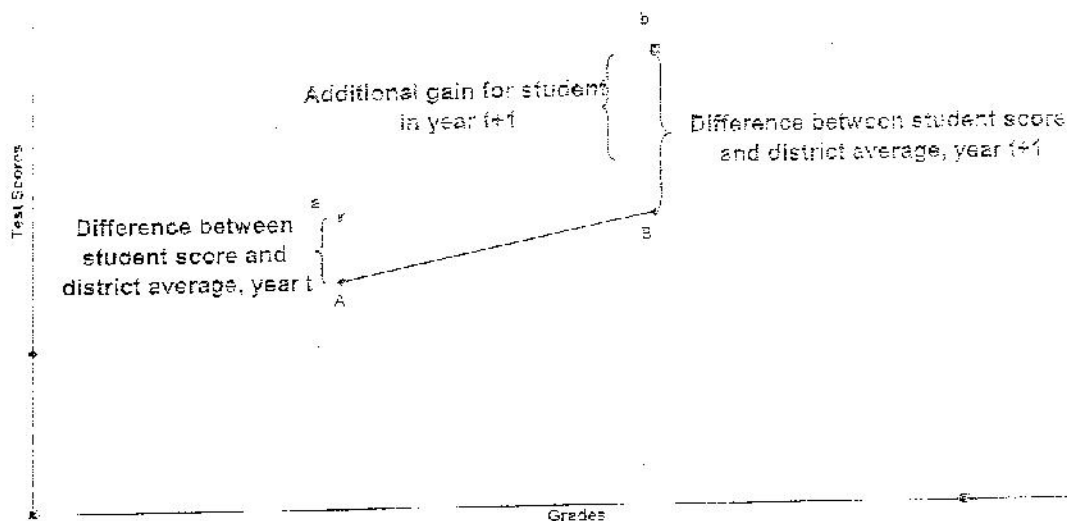
as long as the tests are vertically sequenced (i.e., measure cumulative knowledge of a single subject).

Special features of the TVAAS:

- A teacher's effect is assumed to be the same for all students. There are no interactions of teacher effect with student characteristics.
- Teacher effects persist undiminished over time. ("layering")
- Teacher effects are assumed to be independent across years and subjects for the same teacher.
- Student-level errors are independent across students.
- u_t^k (teacher effects) are assumed independent of e_t^k (student-level errors)

Heuristic Explanation of TVAAS

Figure: TVAAS Student and Teacher Effects



Crude estimate of teacher value-added: additional gain in year $t+1$. Modified for

- Additional gains of other students in the same class
- Inter-year correlations < 1
- Multiple other scores in different years, subjects
- Adjusting difference between student's scores and district average for effect of other teachers
- Protecting against putting too much weight on data for a single teacher (shrinkage)

Three Issues: Bias, Imprecision, Issues in Making Interdistrict Comparisons

- Estimated value-added is free of bias if teachers have an equal chance of being assigned each student in the district.
- If each teacher is assigned to the universe of students, and these students' achievement is measured without error, then teacher effects will be estimated exactly (no imprecision).
- Imprecision "averages out." Bias does not.
- Interdistrict comparisons: bias in disguised form.

Data

Large Tennessee District (Ballou, Sanders, and
Wright, 2004)

Tables: Characteristics of Teachers' Classes

Descriptive Statistics: Level and Gain
Scores, 1996-2001

CHARACTERISTICS OF TEACHERS' CLASSES

Grade	# Teachers	# Students per Teacher		Percent Eligible for FRPL		Percent Non-White	
		Mean	Std dev	Mean	Std dev	Mean	Std dev
4	1205	18	7	56	25	53	21
5	833	23	13	59	24	54	20
6	693	27	14	58	24	54	19
7	331	45	18	56	22	54	17
8	316	45	19	49	23	51	19

Table 1: Descriptive Statistics, Level and Gain Scores, 1996-2001

		Number of students	Percentage of students with gain scores	Mean Gain	Std. Dev.
Reading	Grade				
	3	22400	0.0	-----	-
	4	21907	85.9	14.1	23.1
	5	20047	86.5	9.4	22.6
	6	19476	87.1	13.0	21.6
	7	19533	81.9	9.9	21.5
	8	17711	86.0	12.7	20.9
	Language Arts			-----	-----
3	22400	0.0	-	-	
4	21907	85.7	15.0	24.0	
5	20047	86.4	10.3	22.8	
6	19476	87.0	9.9	21.3	
7	19533	82.0	10.5	22.8	
8	17711	85.9	12.1	21.9	
Math	3	22400	0.0	-----	-
	4	21907	85.8	23.9	25.4
	5	20047	86.2	18.2	23.8
	6	19468	86.8	18.9	22.8
	7	19474	81.9	17.1	24.5
	8	17471	85.8	16.5	23.5

Bias

Focus on student SES and demographics (sex, race).

- Are SES and demographics related to achievement gains?

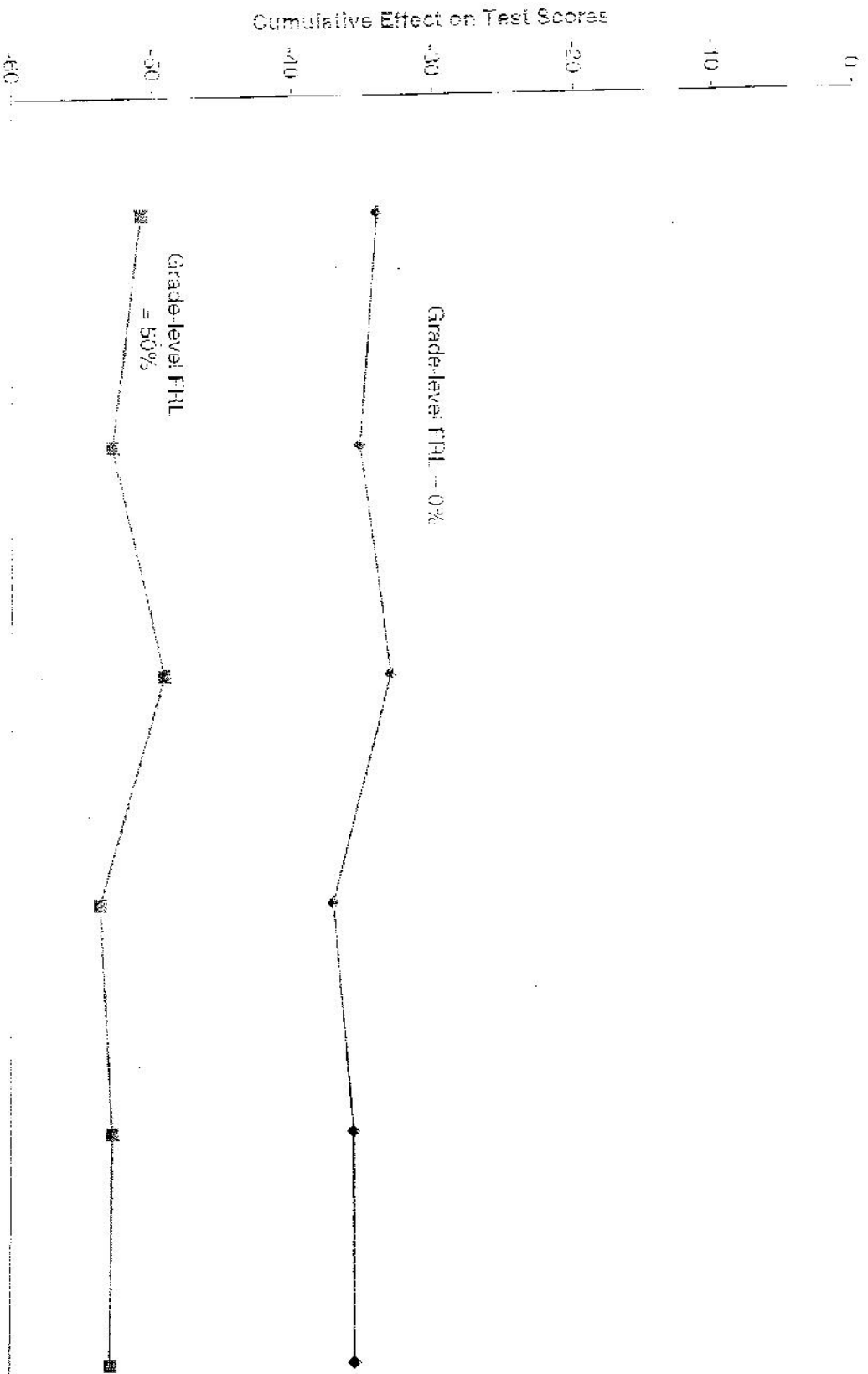
**Figures: Reading: Black, Male, FRL-Eligible
& Math: Black, Male, FRL-Eligible**

- How strong is this relationship compared to typical teacher effects?
- Do longitudinal data substitute for controls for SES, demographics?
- What happens to teacher value-added estimates when we control for SES, demographics?

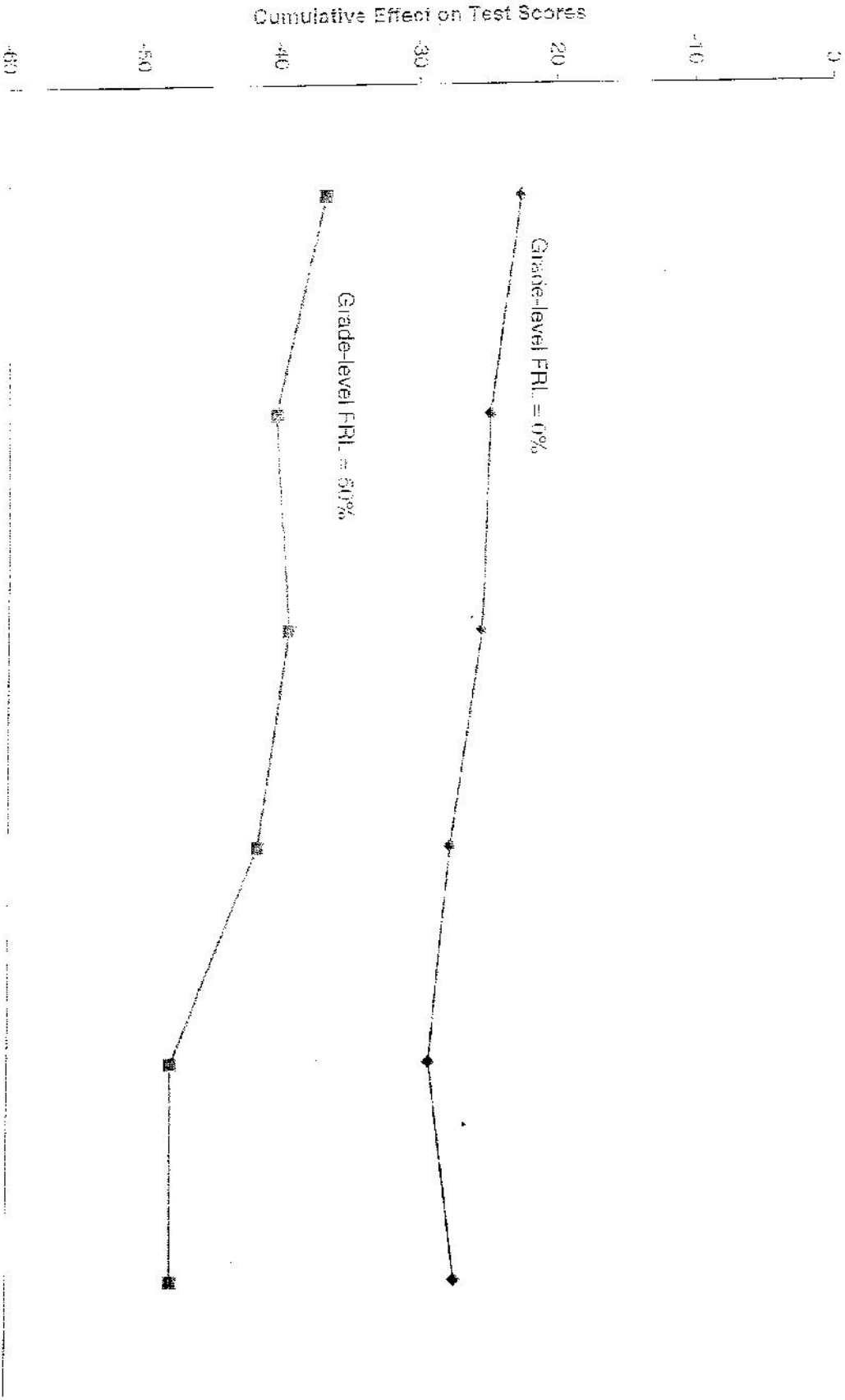
Table: Correlations between original and modified TVAAS

- Do these findings generalize?
- Can we control for SES, demographics without introducing new sources of bias?

Reading: Black, Male, FRL-Eligible
(No Teacher Effects)



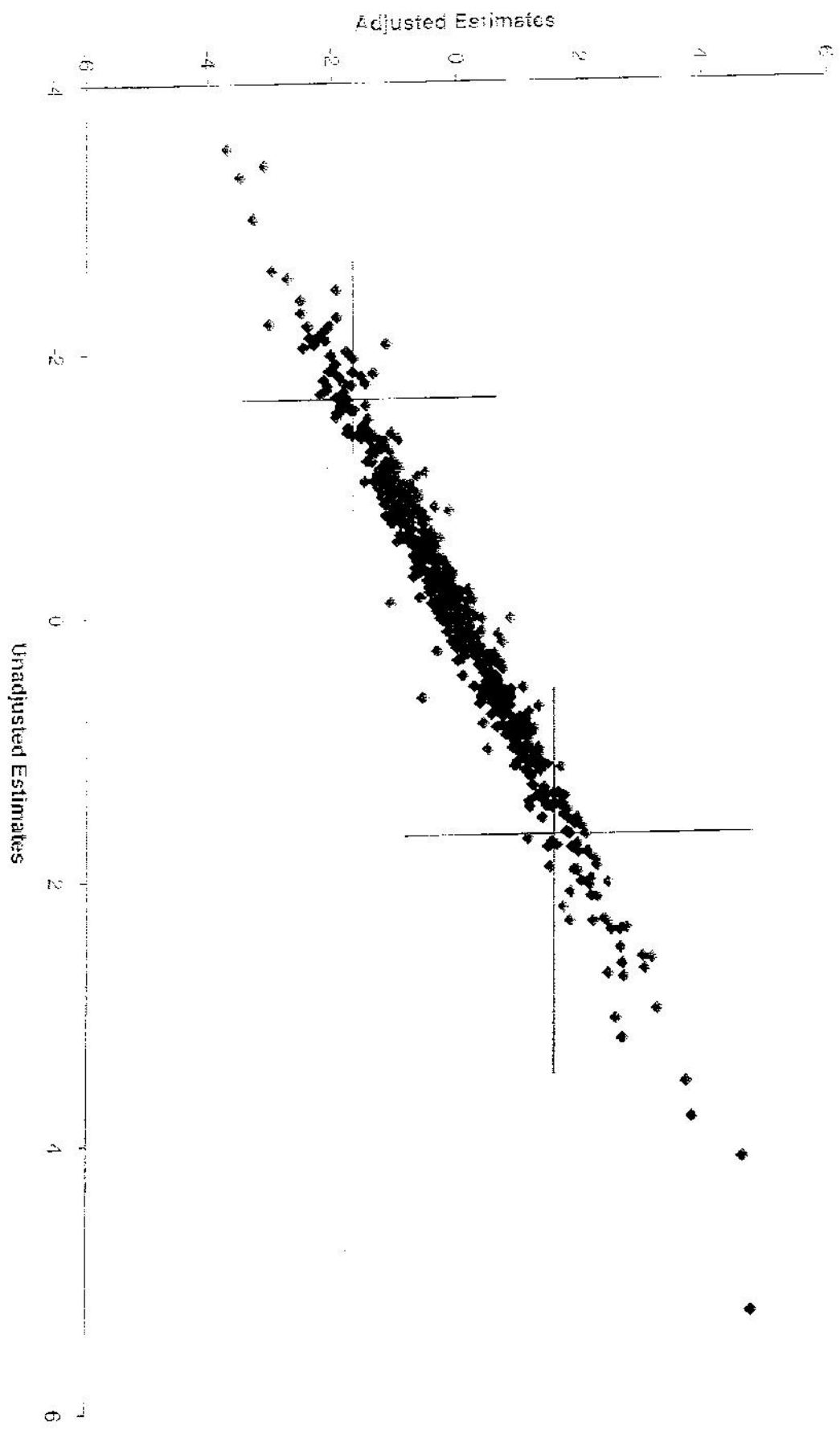
Math: Black, Male, FRL-Eligible
(No Teacher Effects)



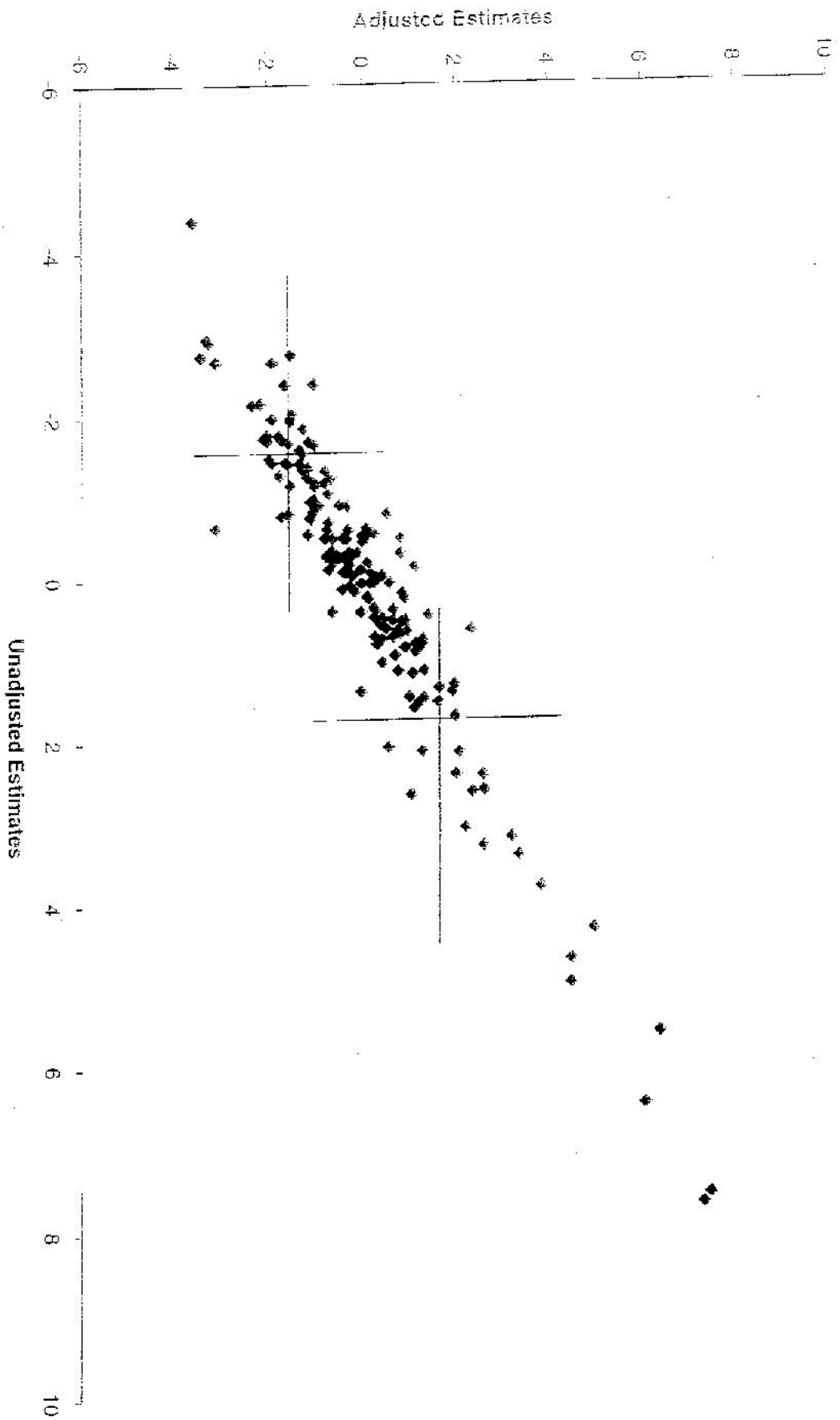
Correlations between Original and Modified TVAAS
Teacher Effects

Grade	Reading	Math
4	0.93	0.98
5	0.84	0.98
6	0.88	0.98
7	0.91	0.95
8	0.94	0.95

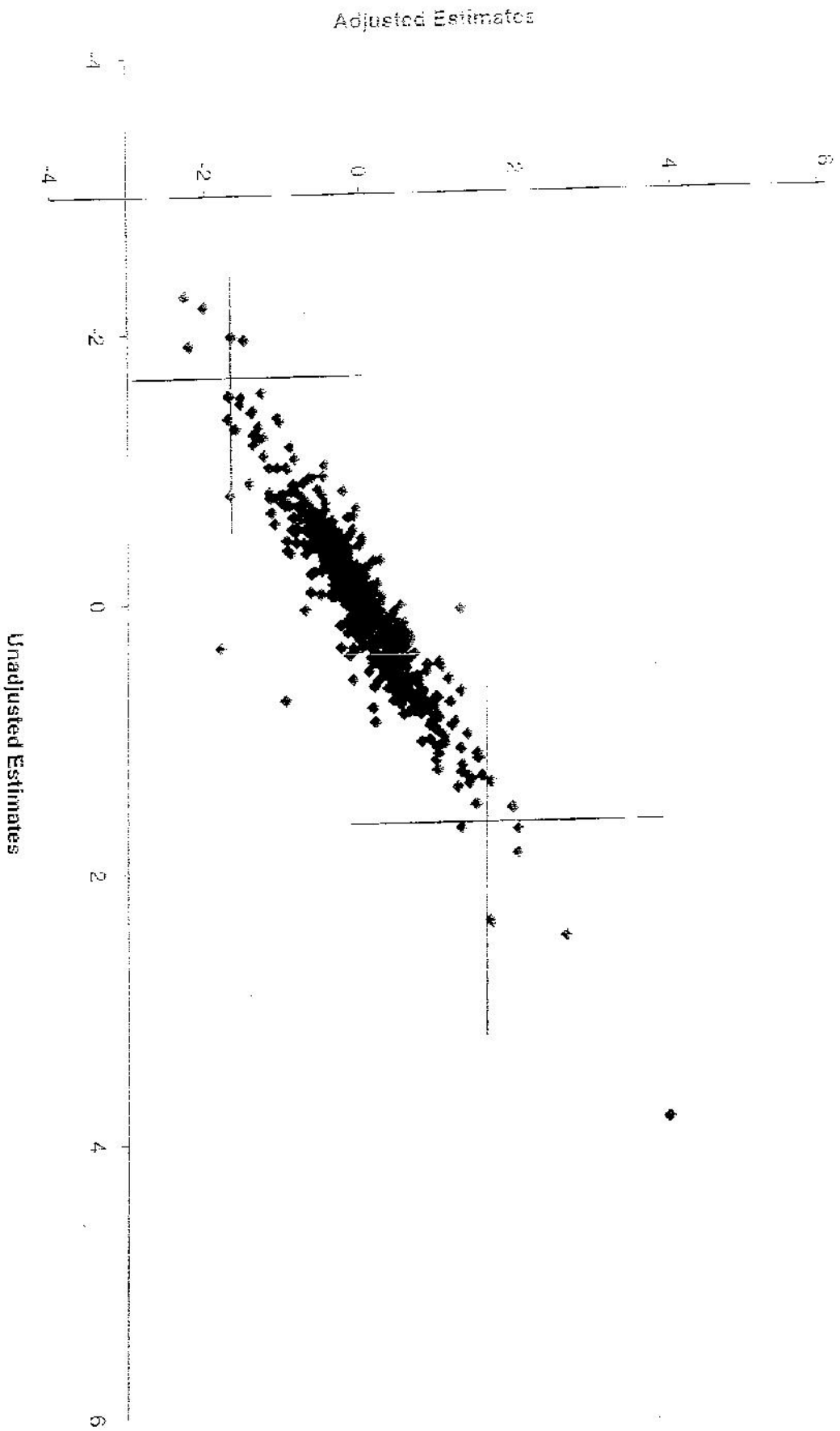
Standardized Teacher Effects, 4th Grade Math,
With and Without Covariate Adjustments



Standardized Teacher Effects, 8th Grade Math,
With and Without Covariate Adjustments

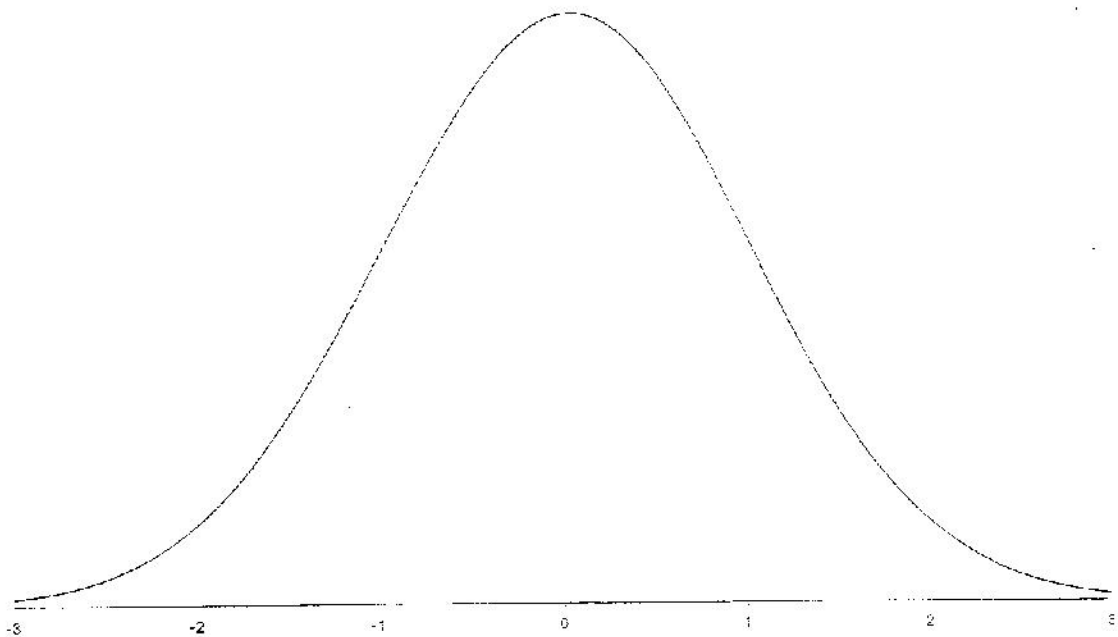


Standardized Teacher Effects, 4th Grade Reading,
With and Without Covariate Adjustments



Imprecision

Teacher effects are random. TVAAS estimates the distribution of each teacher's effect:



"Teacher effect" is the mean of this distribution.

From the distribution, we can determine a confidence interval for the true effect. The width of the confidence interval is a measure of the estimate's imprecision.

- TVAAS confidence intervals are wide: most teachers are not distinguishably different from average at conventional levels of statistical significance.

Figure: Estimated Teacher Effects, with 90% Confidence Intervals: 5th Grade Math

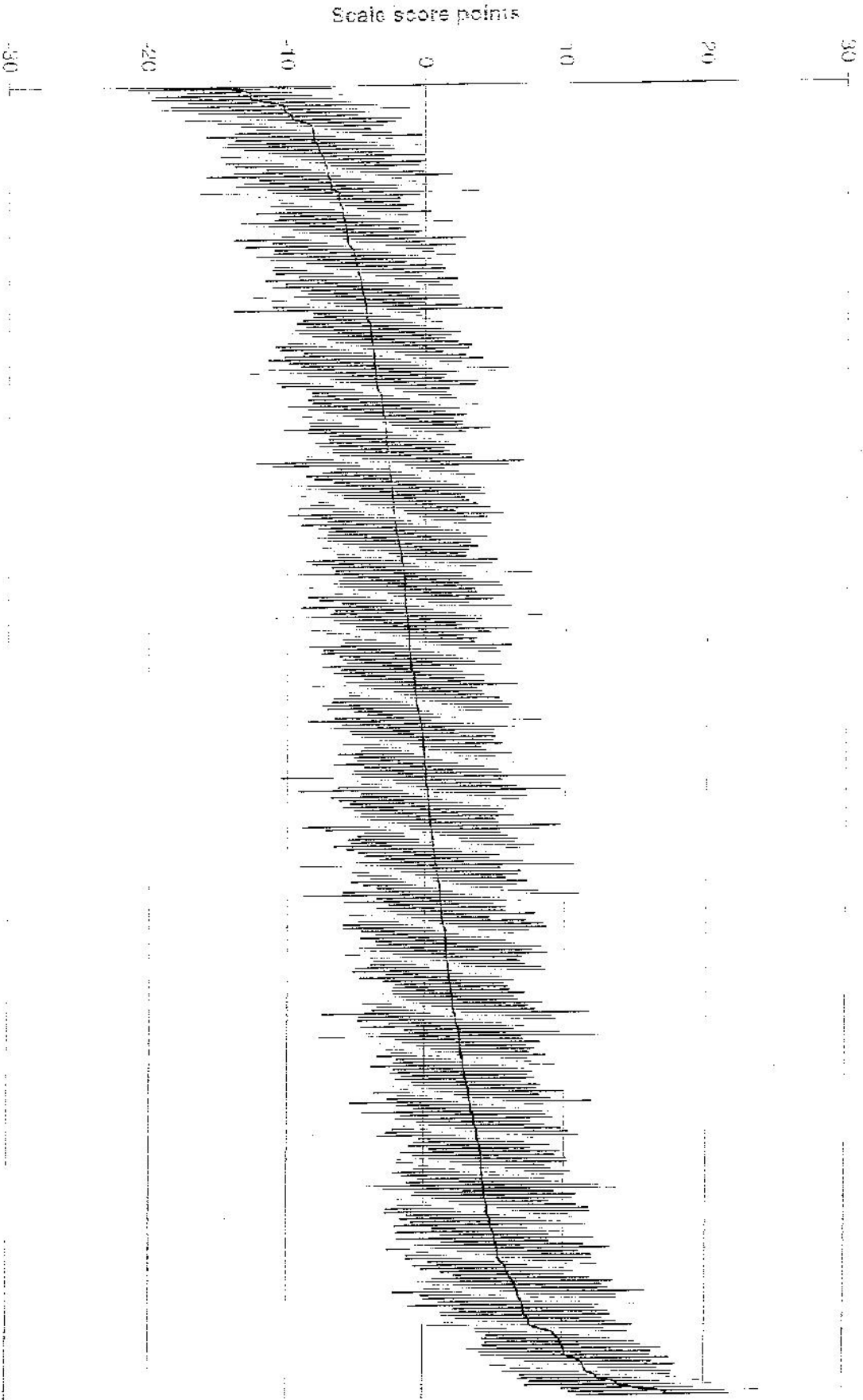
- Width of the confidence intervals is closely related to the amount of data for the teacher.

Table: Percentage of Teachers Significantly Different from Average

- In principle, confidence intervals could be made smaller by exploiting intra-teacher and intra-classroom covariances.
- Confidence intervals are "off" due to model misspecification.
- Confidence intervals aside, a teacher's value-added estimates vary across years, in part due to imprecision.

Figures: Stability of Teacher Effects: Math & Reading

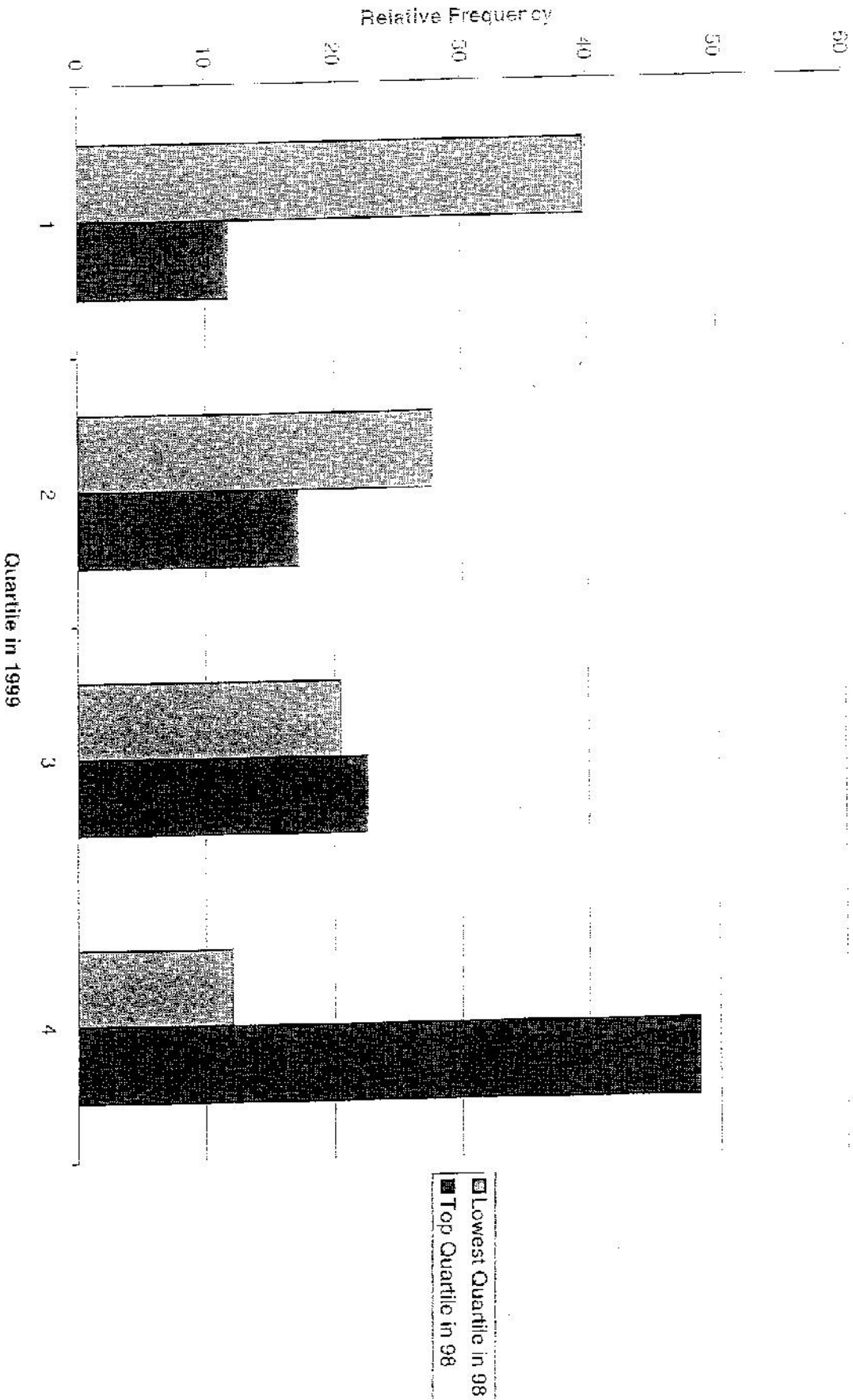
Estimated Teacher Effects, with 90% Confidence Intervals: 5th Grade Math



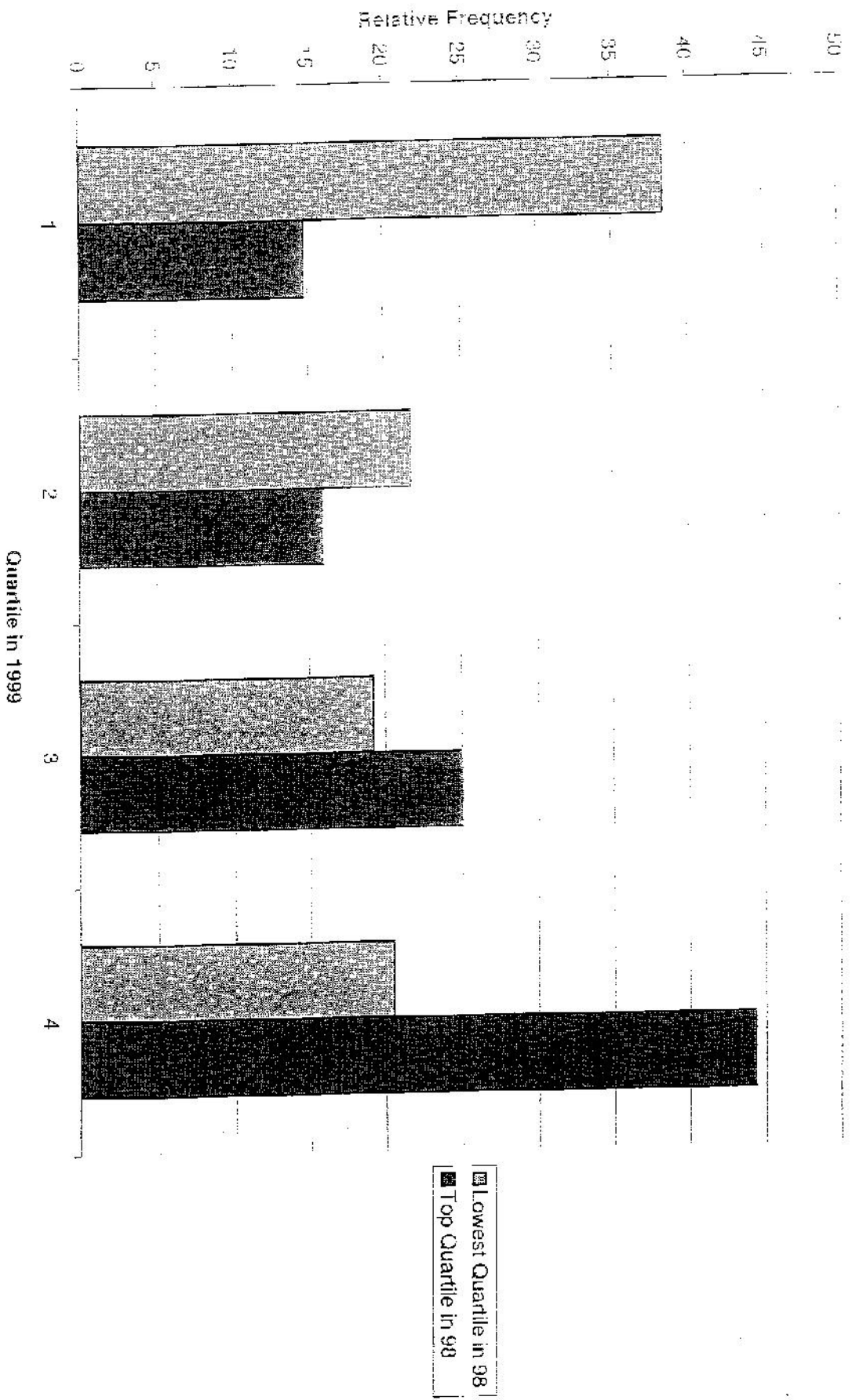
Percentage of Teachers Significantly Different from Average (10% level)

Estimates based on:	Grades	Reading		Mathematics	
		% significant	N	% significant	N
Single year	4-6	2.5	1385	17.0	1313
	7-8	7.6	276	30.4	332
Three-year average	4-6	3.7	782	22.0	732
	7-8	10.8	185	37.8	185
Three-year average (with 3 years' data)	4-6	6.5	199	30.1	203
	7-8	11.1	27	58.0	50

Stability of Teacher Effects: Math



Stability of Teacher Effects: Reading



Interdistrict Comparisons

TVAAS value-added estimates are relative to district mean (i.e., they are centered on zero).

District means capture two things:

- (1) District policies, resources, student characteristics that affect learning, independent of teachers.
- (2) Average quality of teachers in the district.

If district mean gains are added to teacher effects, new estimates will include (1).

If district mean gains are not added to teacher effects, new estimates will exclude (2).

Example: Proposed Pennsylvania plan to penalize teachers whose students do not make AYP, using a value-added measure of AYP.

**Table: Mean District Growth, Math and Reading:
1999-2001, Selected Tennessee Systems**

Mean District Growth, Math and Reading, 1999-2001, Selected Tennessee Systems

System	4M	4R	5M	5R	6M	6R	7M	7R	8M	8R
Alamo	26.1	5.8	20.7	14.9	25.2	5.3				
Alcoa	24.2	15.1	16.9	16.2	24.4	14.3	18.9	12.7	16.4	13.5
Anderson	30.8	10.9	20.0	13.3	15.2	1.9	15.6	10.6	13.5	11.3
Athens	14.4	5.6	27.0	12.1	23.5	6.9	11.4	6.0	15.3	13.8
Bedford	27.0	10.9	14.9	11.3	18.4	4.6	8.9	8.7	17.3	10.8
Belis	34.8	17.4	21.5							
Benton	24.3	12.3	17.7	17.5	20.0	6.4	13.5	12.0	16.5	11.0
Biedsoe	29.3	11.9	21.2	16.3	11.2	5.3	15.8	10.4	17.4	7.9
Blount	27.7	11.7	21.3	13.4	17.4	5.6	8.7	10.0	17.2	10.4
Bradford	36.0	11.8					4.6	12.1	10.2	11.2
Bradley	30.2	11.5	20.3	11.7	16.5	6.3	6.4	8.0	17.4	11.9
Bristol	28.8	11.5	27.8	15.7	20.7	10.1	14.7	5.3	13.1	10.8
Campbell	24.6	5.6	22.7	14	17.0	5.4	7.0	10.7	19.2	10.2
Cannon	24.1	9.7	29	16.8	16.1	6.2	15.5	12.4	15.7	10.0
Carter	21.8	6.6	15.5	13.9	15.9	6.1	13.0	11.6	16.0	12.4
Median standard errors, u.g.	3.5	2.2	3.4	1.9	3.0	1.8	2.7	1.7	2.9	1.5