# Using Relative Score Changes to Detect Potential Test Fraud

Sarah L. Toton

Dennis D. Maynes

# Background

- Score changes across time are assessed at the examinee and group level
- Anomalous score changes may be indicative of test fraud
  - Educator coaching or tampering
  - Manipulation of test administration rules
  - Examinee pre-knowledge
  - Access to prohibited materials during the test

The Atlanta Journal-Constitution

SUNDAY

OUTSIDE THE NORM MIRACLE OR MASQUERADE?

# State to examine test-score surge

By HEATHER VOGELL
hvogell@ajc.com
and JOHN PERRY
jgperry@ajc.com

A miracle occurred at Atherton Elementary this summer, if its standardized math test scores are to be believed.

Half of the DeKalb County school's fifth-graders failed a yearly state test in the spring. When the 32 students took retests, not only did every one of them pass — 26 scored at the highest level.

No other Georgia fifth grade pulled off such a feat in the past three years. It was, as one researcher put it, as extraordinary as a snowstorm in July. In Atlanta.

Atherton Principal James Berry said the scores were the product of intense tutoring.

But state education officials said last week they will investigate steep gains at Atherton and four other schools as a result of The Atlanta Journal-Constitution's inquiries.

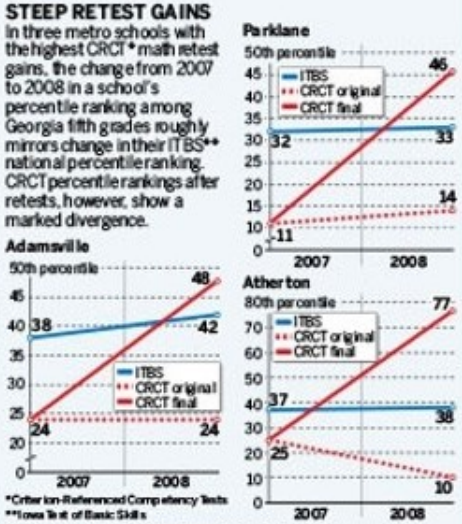
"It's a big red flag," said Kathleen Mathers, executive director of the Governor's Office of Student Achievement. She said officials don't know what caused the Criterion-Referenced Competency Test scores to soar, only that they require explanation.

The state has not routinely mined test data for such anomalies. But officials said it will begin to do so soon, employing widely accepted statistical methods similar to what the AJC used. Atherton's unlikely performance was one of a handful the AJC uncovered by analyzing student scores on the CRCT and retest. The surges

➤ Please see SCORES, A3

**STEEP RETEST GAINS**

In three metro schools with the highest CRCT* math retest gains, the change from 2007 to 2008 in a school's percentile ranking among Georgia fifth grades roughly mirrors change in their ITBS** national percentile ranking. CRCT percentile rankings after retests, however, show a marked divergence.




*Criterion-Referenced Competency Tests
**Iowa Test of Basic Skills
Sources: Georgia Dept. of Education and school district data; research by JOHN PERRY / Staff

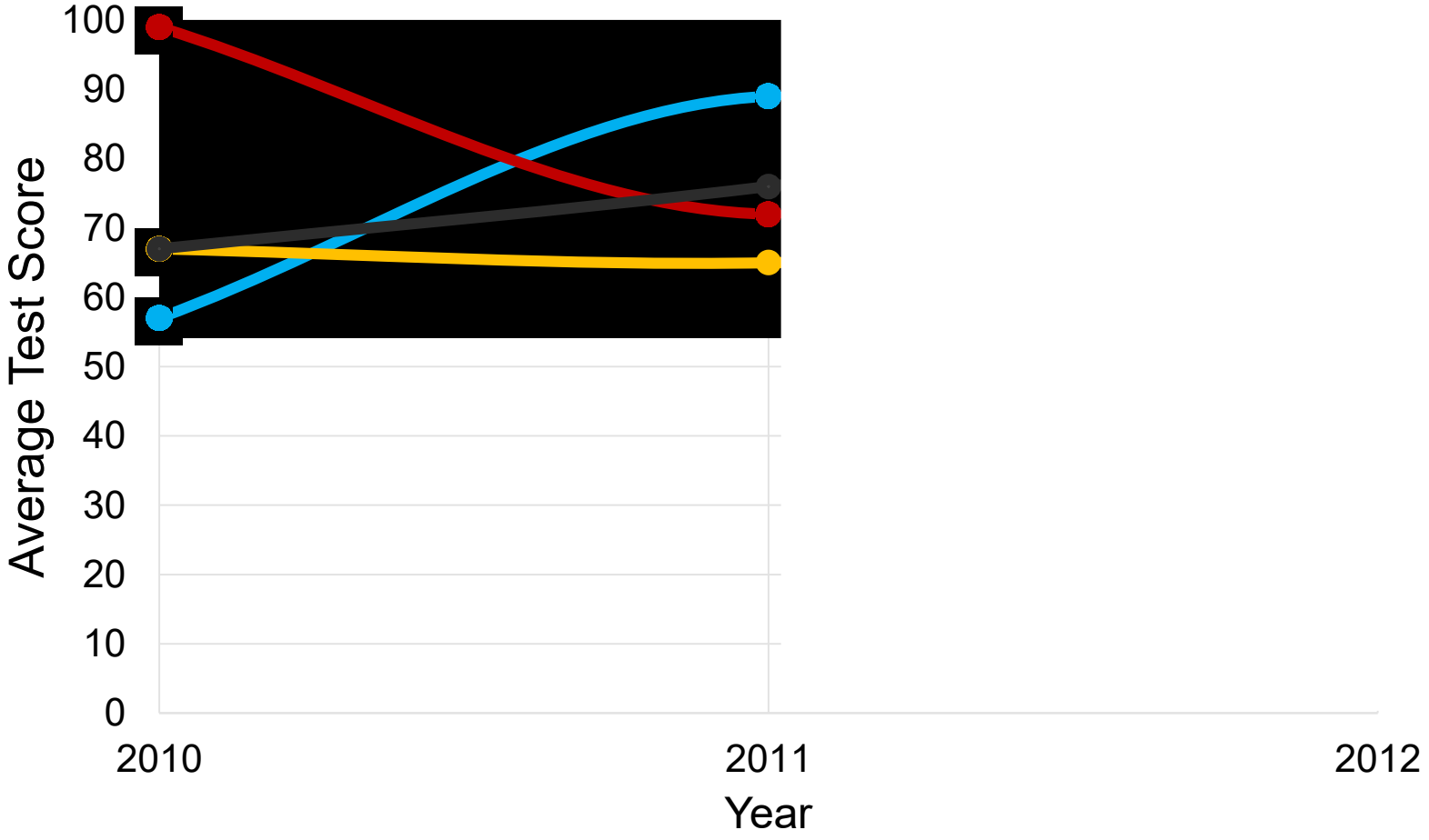
SHANNON PEAVY / Staff

Discovered, in part, because of anomalous score gains

35 educators indicted
11 convicted and imprisoned

"A culture of fear, intimidation and retaliation"

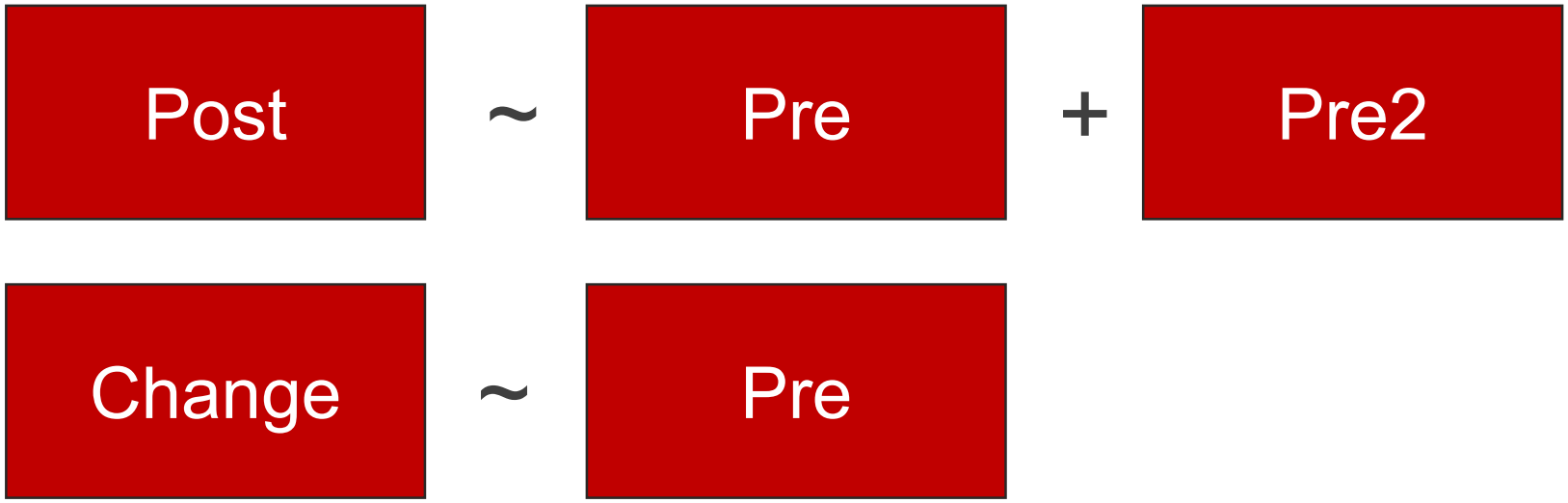
# What Score Changes May Indicate Fraud?



# Considerations in Using Group-Level Gains

- Changes in group demographics over time
  - Ability, ESL, undiagnosed disabilities, accommodations
- Sub-groups may move differently
  - E.g., ESL students in their third year of taking English may show a new rate of improvement
- Small sample sizes
  - Non-normality
  - Outliers

# Current Regression Methodologies




Current treatment: Change/Post is treated as a random variable. Pre is treated as a fixed variable, measured without measurement error and unchanging over time.

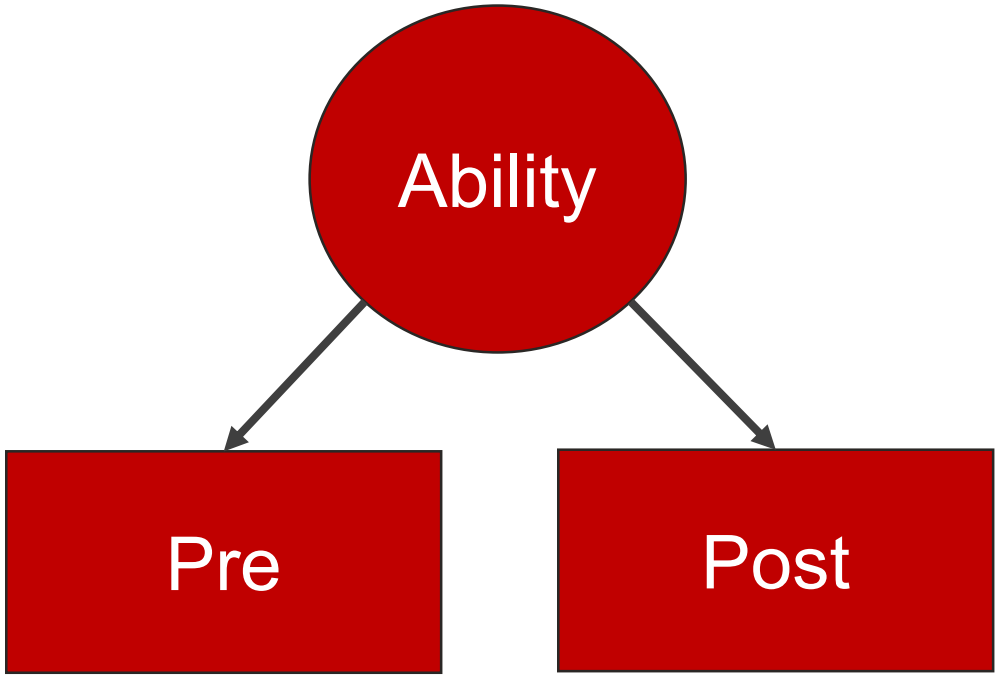
Reality: Both are random variables.

# Statistical Considerations in Using Gains

- Test scores are bounded
  - Scores near the bounds have limited room for change
- Regression to the mean
  - Second observations are often less extreme than initial observations
- Experience shows that consistently high-performing groups are over-flagged

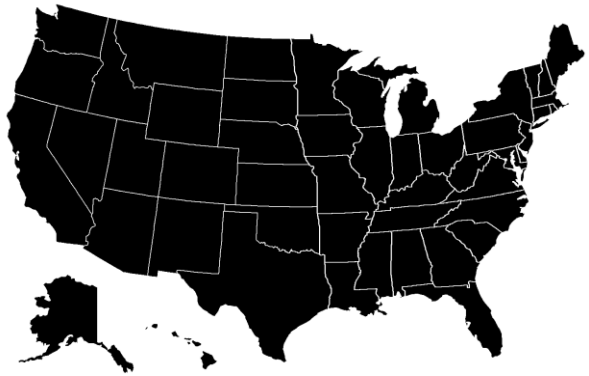
# Proposed Methodology

- True score model



# Special Considerations for State Data

- Entire population measured
- Multiple timepoints measured, attrition or re-tester effects are small



What happens if the data conform to the true score model, but we analyze the data with Post~Pre regression?

# Simulation Method

1. Select the number, n, of bivariate pairs to be generated.
2. Select the distribution for the true scores (e.g., normal distribution with a mean, m, of 500 and standard deviation, $s_t$, of 100).
3. Select a standard error of measurement, $s_e$, or the standard deviation of the pre-test and post-test scores given the true score.
4. For each bivariate pair,
    1. Generate a true score from a standard normal random variable: $T=z*s_t+m$.
    2. Generate the bivariate pair of pre-test (X) and post-test (Y) scores from two standard normal random variables, such that $X=Z_1*s_e+T$ and $Y=Z_2*s_e+T$.

# Simulation Results- Type I Error Rates

| | | Standardized Residuals | | | | | Frequency |
|---|---|---|---|---|---|---|---|
| | | 1.645 | 2 | 2.5 | 3 | 3.25 | |
| Nominal Rate | | 0.0500 | 0.0228 | 0.0062 | 0.0014 | 0.0006 | |
| True Score Range | 200 - 300 | 0.0155 | 0.0047 | 0.0014 | 0.0000 | 0.0000 | 2,132 |
| | 300 - 400 | 0.0243 | 0.0098 | 0.0018 | 0.0005 | 0.0002 | 13,662 |
| | 400 - 500 | 0.0387 | 0.0177 | 0.0038 | 0.0006 | 0.0003 | 34,141 |
| | 500 - 600 | 0.0569 | 0.0263 | 0.0067 | 0.0014 | 0.0005 | 34,104 |
| | 600 - 700 | 0.0815 | 0.0399 | 0.0119 | 0.0024 | 0.0013 | 13,547 |
| | 700 - 800 | 0.1164 | 0.0615 | 0.0211 | 0.0023 | 0.0005 | 2,130 |
| | 800 - 900 | 0.1377 | 0.0942 | 0.0217 | 0.0145 | 0.0072 | 138 |

When data conform to the true score model but are analyzed with Post~Pre regression, the Type I error rates are inflated with high true scores and deflated with low true scores.

Thus, we will have more false-positives for high-performing groups.

We may also miss anomalous gains in low-scoring groups.

# Alternative Inferences

- Score changes can also be caused by unmeasured changes in
  - student learning or learning opportunities
  - test preparation
  - examinee circumstances
  - the quality of teaching
  - the curriculum
  - school or program resources

# Conclusions

- Current score gain methods may result in over-flagging of high ability groups and under-flagging low ability groups

- Using a true score method can address these issues and lead to more accurate flagging