

## Student Growth Percentiles

Damian W. Betebenner

National Center for the Improvement of Educational Assessment  
Dover, NH

NCME Training Sessions, New Orleans, LA      April 7th, 2011

## About Us

**Damian Betebenner** Senior Associate National Center for the Improvement of Educational Assessment. Developed student growth percentiles and percentile growth trajectories to help states and educational associations employ student growth in decision making [Betebenner, 2008, Betebenner, 2009]. Currently refining and sharing these techniques with approximately two dozen other states,

**Adam Van Iwaarden** Ph.D. candidate School of Education, Research and Evaluation Methodology Program, University of Colorado, Boulder. Co-developer of the SGP package.

# About Today's Training Session

- An overview of growth models and student growth percentiles
- An introduction to R
- The basics of using the SGP package
- More advanced uses of the SGP package

## Things we'll need for today

- The R software environment installed on your computer. Get from:  
<http://cran.r-project.org/>
- The SGP package.
- Enthusiasm and patience.

# From a Paucity to a Plethora of Data

## The Economist: The Data Deluge

- Historical records of student achievement
- Historical records of student demographics, teachers, schools, educational programs, . . .
- Stakeholder interest in an examination of this longitudinal data

Interest in examining student achievement over time (student growth) derives from data availability.

# Lessons Learned

## On Questions

John Tukey

It is better to have an approximate answer to the right question than a precise answer to the wrong question.

## On Statistical Models

George E. P. Box

All models are wrong but some are useful.

## On Understanding

Aristotle

We understand best those things we see emerge from their very beginnings.

## On Accountability

Sherlock Holmes

It is a capital mistake to theorize before you have all the evidence. It biases the judgment.

## On Theories of Action

Confucius

Tell me and I will forget, show me and I will remember, involve me and I will understand.

# The Attractiveness of Growth

## What is growth and why measure it?

- Student learning is a central goal of education.
- Assessments of student achievement provide evidence of the current status of student knowledge and understanding.
- Learning is demonstrated by growth in student achievement from one point in time to another point in time—not by status at either point time alone.

# Guiding Questions

## Growth models address specific questions

- Different growth analysis techniques are good at answering different questions.
- It is critical to understand these different questions.
- Different questions lead to different conversations which lead to different uses and outcomes.

## Changing conversations about education

- Starting with the right questions simplifies development and motivating the proper use of the growth model results.
- The questions set the table for those conversations.
- Yen (2007) for an excellent list of questions derived from a survey of parent, teachers and administrators



# What are the relevant questions for parents?

Yen (2007), from a state survey of parents, teachers and administrators, compiled a list of frequently voiced questions/concerns by stakeholder group.

## Parent Questions

- Did my child make a year's worth of progress in a year?
- Is my child growing appropriately toward meeting state standards?
- Is my child growing as much in Math as Reading?
- Did my child grow as much this year as last year?

# What are the relevant questions for teachers?

Yen (2007), from a state survey of parents, teachers and administrators, compiled a list of frequently voiced questions/concerns by stakeholder group.

## Teacher Questions

- Did my students make a year's worth of progress in a year?
- Did my students grow appropriately toward meeting state standards?
- How close are my students to becoming Proficient?
- Are there students with unusually low growth who need special attention?

# What are the relevant questions for administrators?

Yen (2007), from a state survey of parents, teachers and administrators, compiled a list of frequently voiced questions/concerns by stakeholder group.

## Administrator Questions

- Did the students in our district/school make a year's worth of progress in all content areas?
- Are our students growing appropriately toward meeting state standards?
- Does this school/program show as much growth as that one?
- Can I measure student growth even for students who do not change proficiency categories?
- Can I pool together results from different grades to draw summary conclusions?

# Growth & Accountability

## Growth versus Status

- Enthusiasm for growth in accountability stems from the belief that growth and teacher/school quality are more closely related than status.
- Enthusiasm for growth also stems from its potential diagnostic uses.

# Growth & Accountability

## Fundamental Premise

“Good” schools/teachers bring about growth in student achievement in excess of that found at “bad” schools/teachers.

- The inherent believability of this premise is at the heart of current enthusiasm to incorporate growth models into accountability systems.
- Note that current NCLB accountability invokes the same premise (with achievement substituted for growth).
- “Good schools/teachers” are commonly referred to as highly effective schools/teachers.
- It is critical to understand the distinction between **growth** and **effectiveness**.

# Value-added Analyses/Inferences

## Who/What is responsible for student growth?

- Often, analyses of student growth attempt to determine the amount of student progress/growth that can be attributed to the school or teacher.
- Called *Value-added* analyses, these techniques attempt to estimate the teacher/school contribution to student achievement.
- Value-added models quantify **effectiveness**—purporting to measure the impact on achievement that this school or teacher would have, on average, upon similar students assigned to them.
- These models and their effectiveness claims are controversial for a number of reasons (EPI (2010), NRC/NAE (2010), Braun (2005), Rubin and Stuart (2004)) .

# Value-added Analyses/Inferences

## Who/What is responsible for student growth?

- Value-added models are better described as **effectiveness** models than as **growth** models.
- Value-added is an **inference** – a causal conclusion drawn from the data.
- Value-added isn't a type of statistical model or type of assessment (if I showed you the statistical model formulation or an assessment, from the model formulation or assessment you wouldn't be able to tell me whether it's value-added).
- All growth models can be used for value-added purposes but not vice-versa.

## Descriptive Questions

- Note that the questions from Yen's (2007) survey are primarily descriptive.
- The questions are only peripherally associated with causality.
- High stakes accountability has transformed questions about student growth into questions about responsibility/cause: Teacher and School Effectiveness.
- Beginning with description doesn't prevent subsequent discussions about responsibility/cause.
- Beginning with responsibility/cause often prevents discussions about description.



## Descriptive Accountability

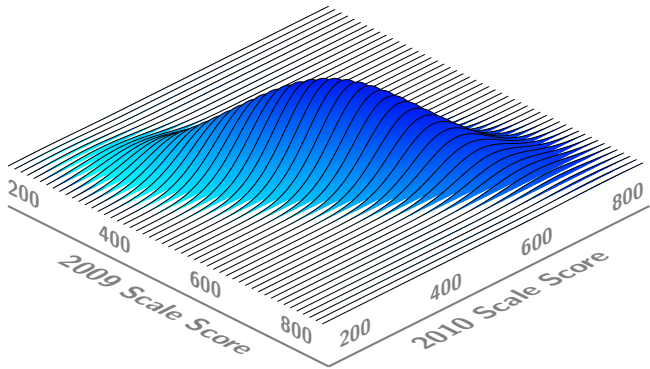
*Accountability system results can have value without making causal inferences about school quality, solely from the results of student achievement measures and demographic characteristics. Treating the results as descriptive information and for identification of schools that require more intensive investigation of organizational and instructional process characteristics are potentially of considerable value. Rather than using the results of the accountability system as the sole determiner of sanctions for schools, they could be used to flag schools that need more intensive investigation to reach sound conclusions about needed improvements or judgments about quality [Linn, 2008, p. 21].*

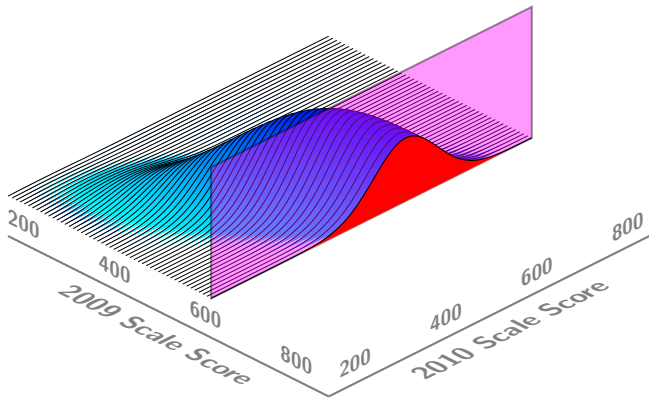
# Causal Explanation versus Causal Attribution

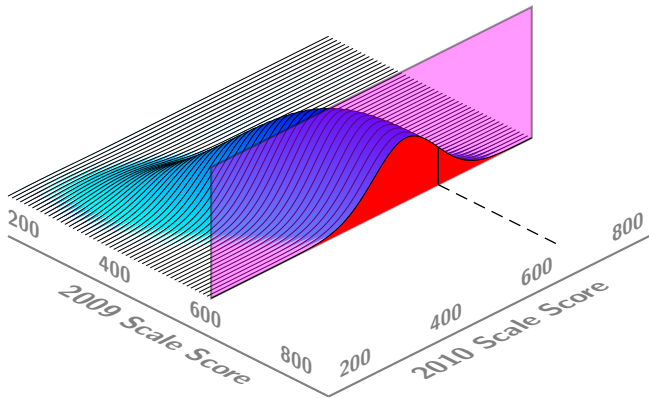
- Hilton (1990) distinguished between causal *explanation* and causal *attribution*.
- Causal explanation is conversational/inter-psychic whereas causal attribution is not.
- The quality of a causal attribution is determined by its truth content whereas the quality of a causal explanation is based upon both *relevance* and truth content.
- Is accountability an enterprise directed toward causal attribution (i.e., a branch of research/program evaluation) or an enterprise enabling causal explanation?

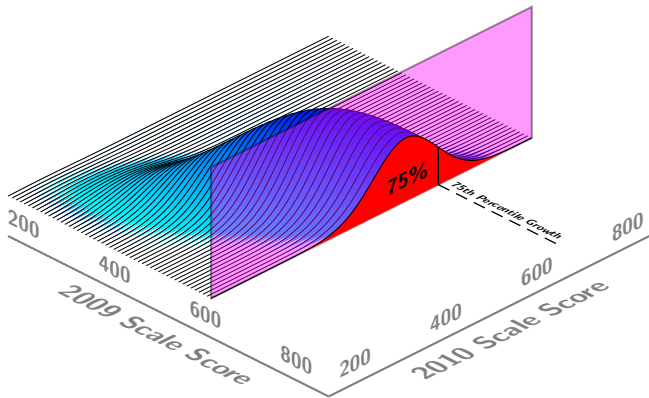
# Describing Student Growth

- Measuring student growth, even with a vertical scale, is not a simple task.
- Some believe a vertical scale simplifies the task of measuring student growth.
- Even with an interval (or ratio) scale, growth is not easy to interpret. Consider, for example, height.
  - A child might grow 4 inches between ages 3 and 4.
  - 4 inches is a well understood quantity.
  - The 4 inch increase becomes really meaningful only when understood alongside the growth of other 3 to 4 year olds.
- **Student growth percentiles** were developed to provide a normative context for describing student growth.









# Student Growth Percentiles

Should we be surprised with a child's current achievement given their prior achievement?

- Given a student's prior scale scores and the associated conditional density, their current scale score corresponds to a percentile of that conditional distribution.
- This percentile is the [student's growth percentile](#).
- Growth percentiles are closely related to estimating the probability of observing a student's current achievement taking account of their past achievement:

$$\Pr(\text{Current Achievement} | \text{Past Achievement}).$$

- As such, growth percentiles describe the rarity of a student's current achievement conditional upon their prior achievement.



# Student Growth Percentiles

Should we be surprised with a child's current achievement given their prior achievement?

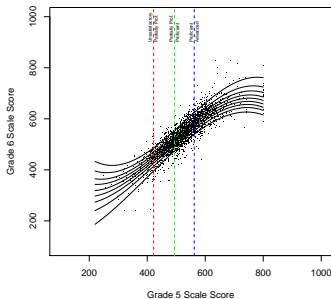
- Student growth percentiles answer this question.
- Consider a low achieving student with 90th percentile growth and a high achieving student with 10th percentile growth.
  - The low achieving student grew at a rate exceeding 90 percent of similar students.
  - The high achieving student grew at a rate exceeding just 10 percent of similar students.
  - The low achiever's growth is more *exemplary* (probabilistically) than the high achiever's.
- Judgments about the **adequacy** of student growth require external criteria.

# Model for Student Growth Percentiles

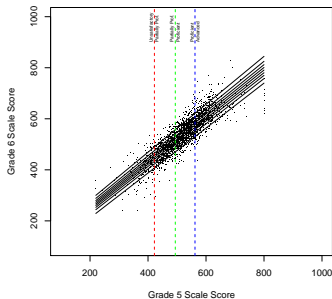
- Student growth percentiles are calculated using quantile regression with B-spline smoothing, a generalized additive model.
- Quantile regression is used to model the complete distribution of the response variable (current achievement)
- B-spline smoothing is used to accommodate non-linearity and heteroscedasticity of the data.
- The model/method is descriptive and all about data fit—it's a data mining procedure.
- The model quantifies *distance* = *rate* · *time* probabilistically.

# Model for Student Growth Percentiles

**Conditional Decile Regression Curves**  
**2006–07 Math: Grade 5 versus 6**



**Conditional Decile Regression Lines**  
**2006–07 Math: Grade 5 versus 6**



# Software for Student Growth Percentiles

- Student Growth Percentiles and Percentile Growth Trajectories are calculated using the open source R Software Environment [R Development Core Team, 2010] and the SGP package [Betebenner, 2011]
- In working with state departments of education, the goal is to build state capacity so states can perform and validate their own analyses using software available and easily used by the state.
- Colorado, Massachusetts, Indiana, and Wisconsin are four states (that I know of) who have state staff capable of generating the student growth percentiles together with aggregations that follow from their calculation.
- State ownership/oversight of the SGP calculations has been a big plus in terms of buy-in from the state, enhancing their ability to explain and work with local stakeholders.

# Description, Inference, and Causality

Richard Berk in *Regression Analysis, A Constructive Critique* (2003) provides an account of the (ab)use of regression

**Three Cheers for Description** Descriptive models are judged by their utility

**Two Cheers for Inference** Inferential models are judged based upon their ability to fulfill statistical criteria associated with generalizing from a sample to a population. What is the chance process?

**One Cheer for Causality** Causal models are judged, in addition to inferential issues, by an external theory which plausibly relates causes/interventions and effects/outcomes.

# Establishing growth-standards based upon growth-norms

- The most common adequacy criterion are judging growth toward an achievement goal (i.e., growth-to-standard).
- Results from student growth percentile analyses can be used to calculate percentile growth trajectories for each student.
- These trajectories indicate what future rates of growth will lead to and are used to make adequacy judgments.
- This growth-to-standard approach was approved as part of Colorado's successful application to the Growth Model Pilot Program.

# Reading

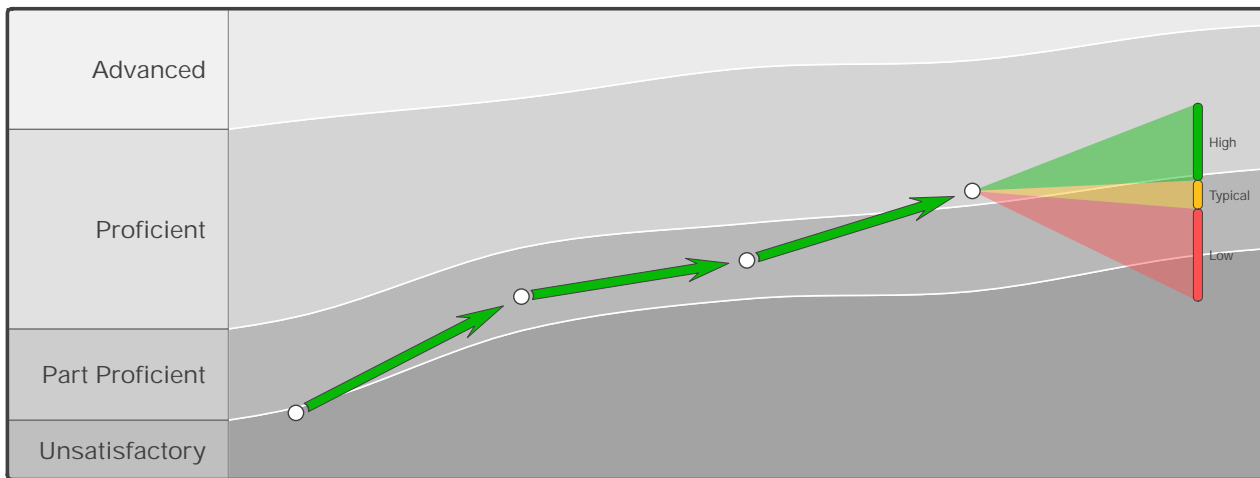
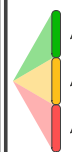
## Achievement



CSAP Reading  
Scale Score

## Growth

| Level   | Percentiles |
|---------|-------------|
| High    | 66th – 99th |
| Typical | 35th – 65th |
| Low     | 1st – 34th  |



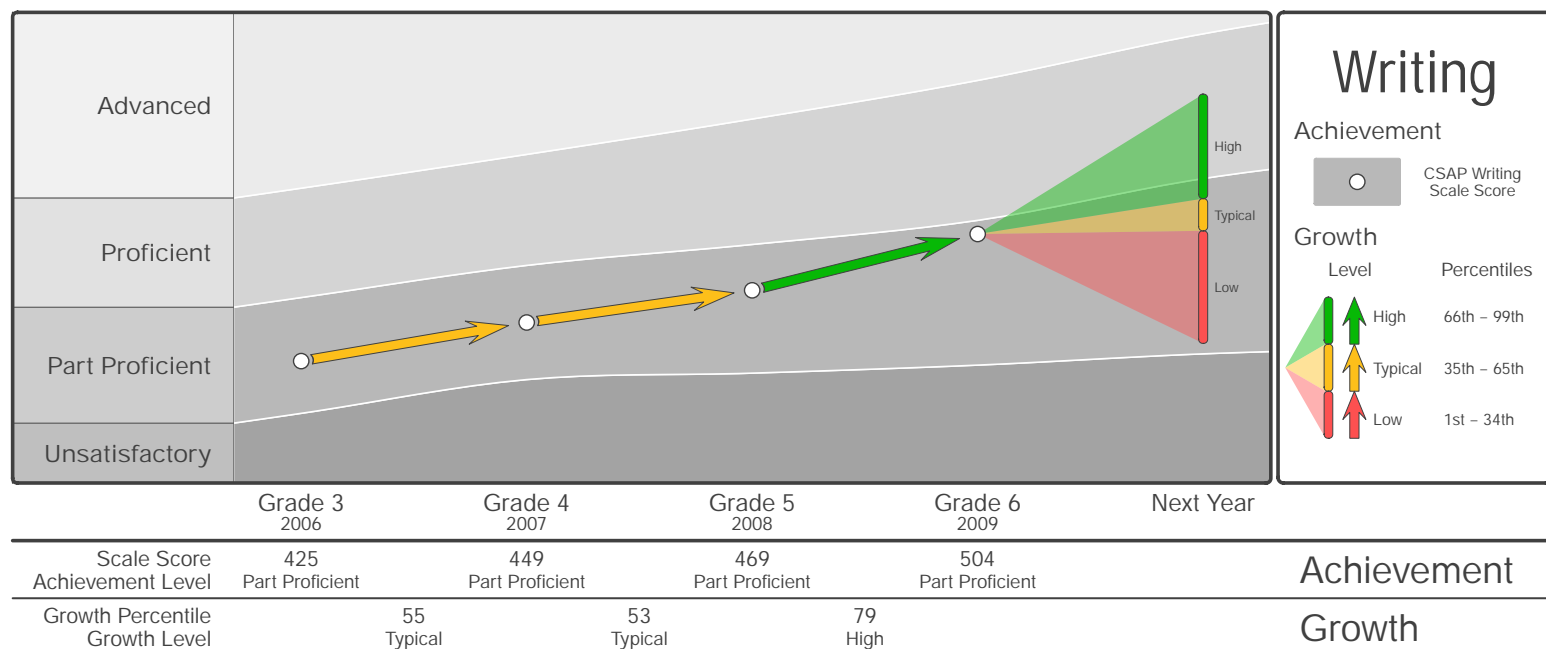
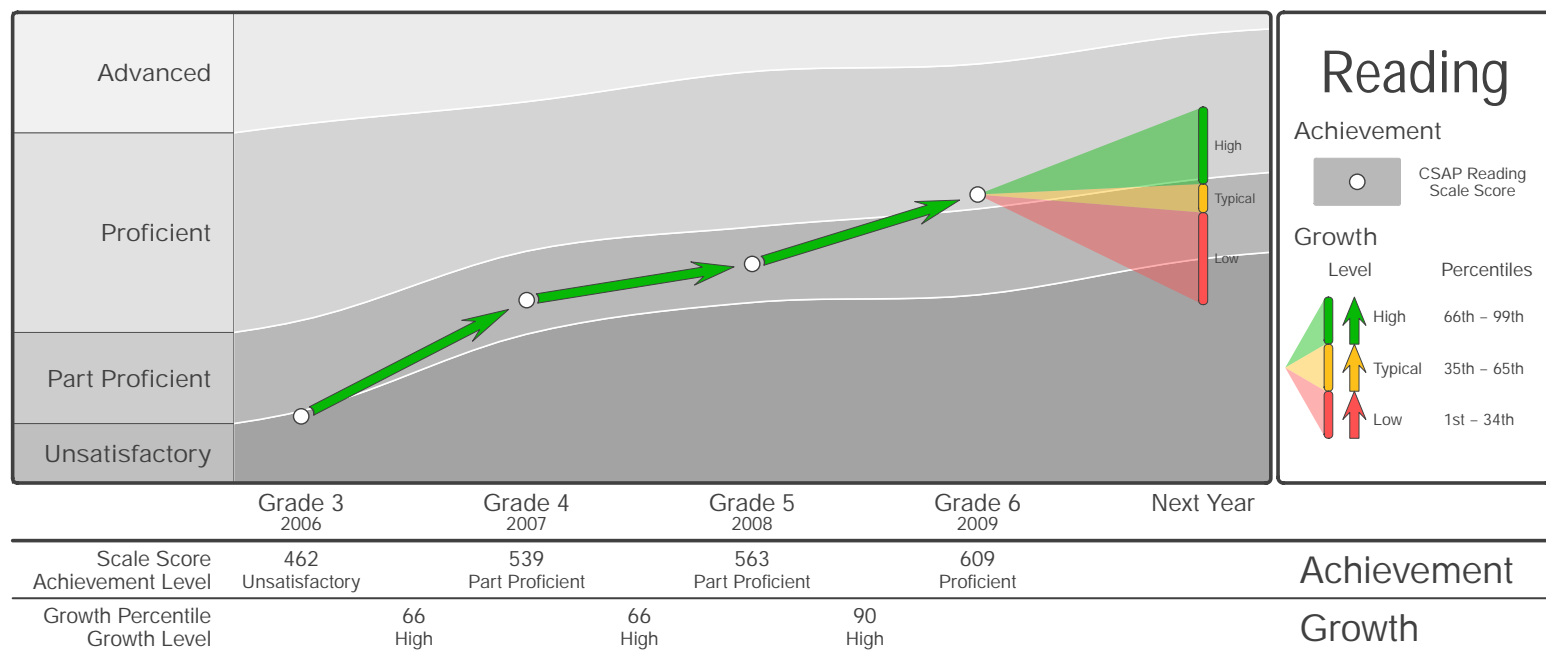
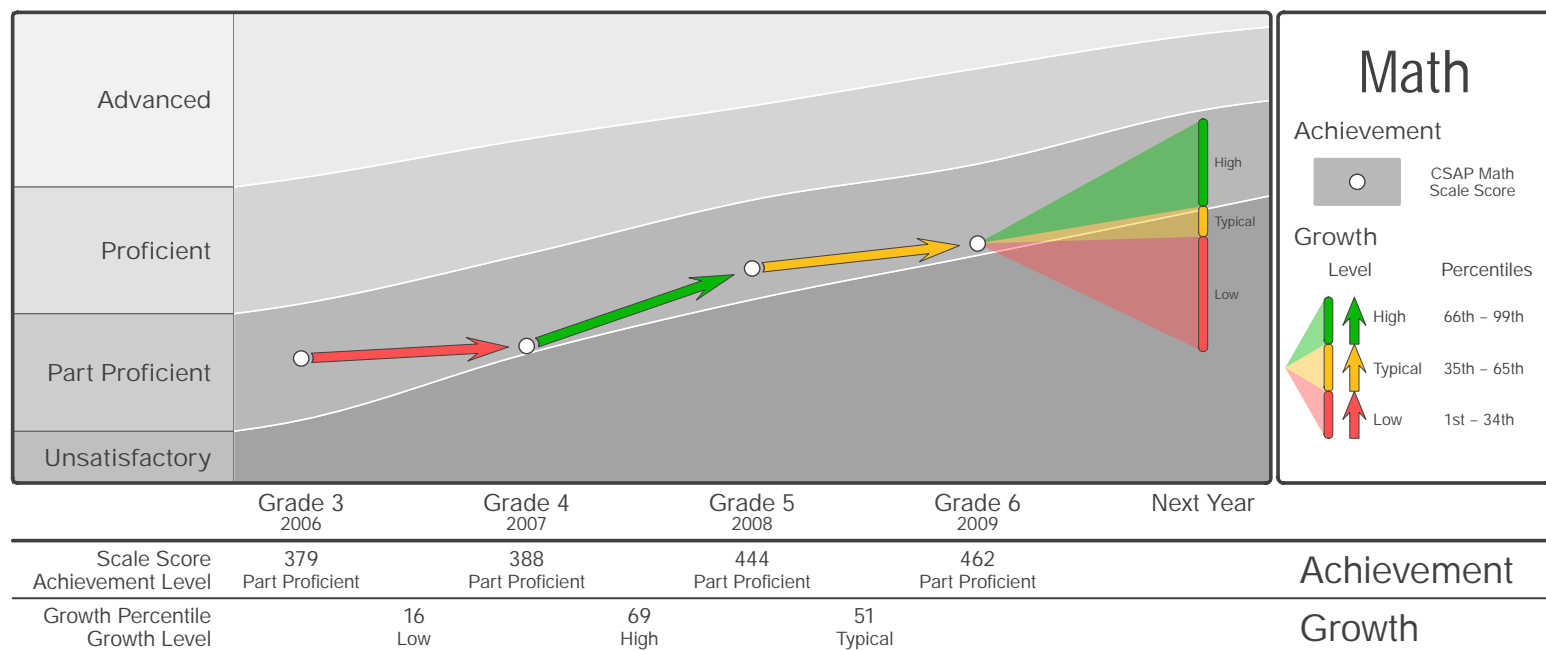
|  | Grade 3<br>2006 | Grade 4<br>2007 | Grade 5<br>2008 | Grade 6<br>2009 | Next Year |
|--|-----------------|-----------------|-----------------|-----------------|-----------|
|--|-----------------|-----------------|-----------------|-----------------|-----------|

|                   |                |                 |                 |            |  |
|-------------------|----------------|-----------------|-----------------|------------|--|
| Scale Score       | 462            | 539             | 563             | 609        |  |
| Achievement Level | Unsatisfactory | Part Proficient | Part Proficient | Proficient |  |

|                   |  |      |      |      |  |
|-------------------|--|------|------|------|--|
| Growth Percentile |  | 66   | 66   | 90   |  |
| Growth Level      |  | High | High | High |  |

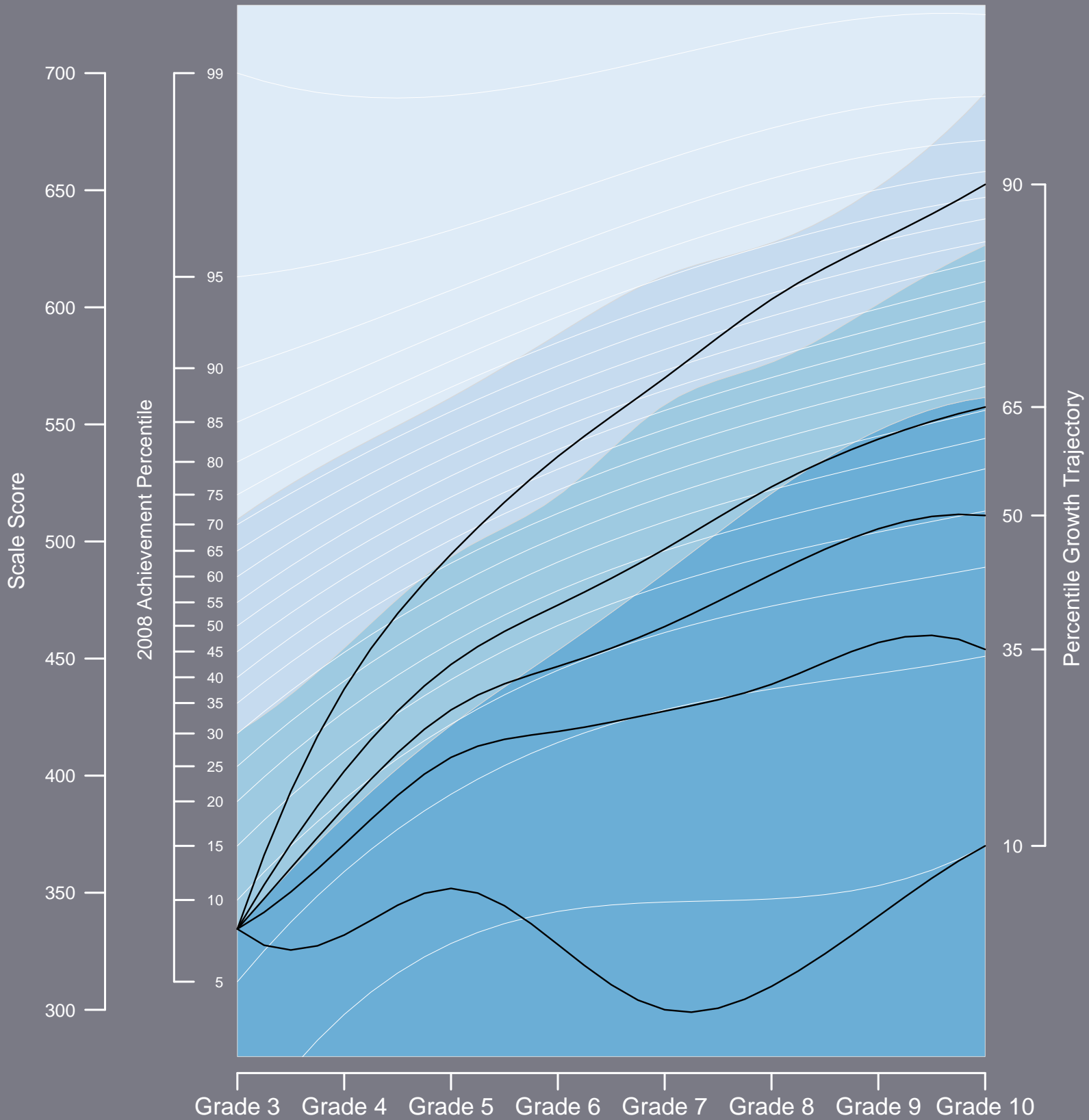
Achievement

Growth





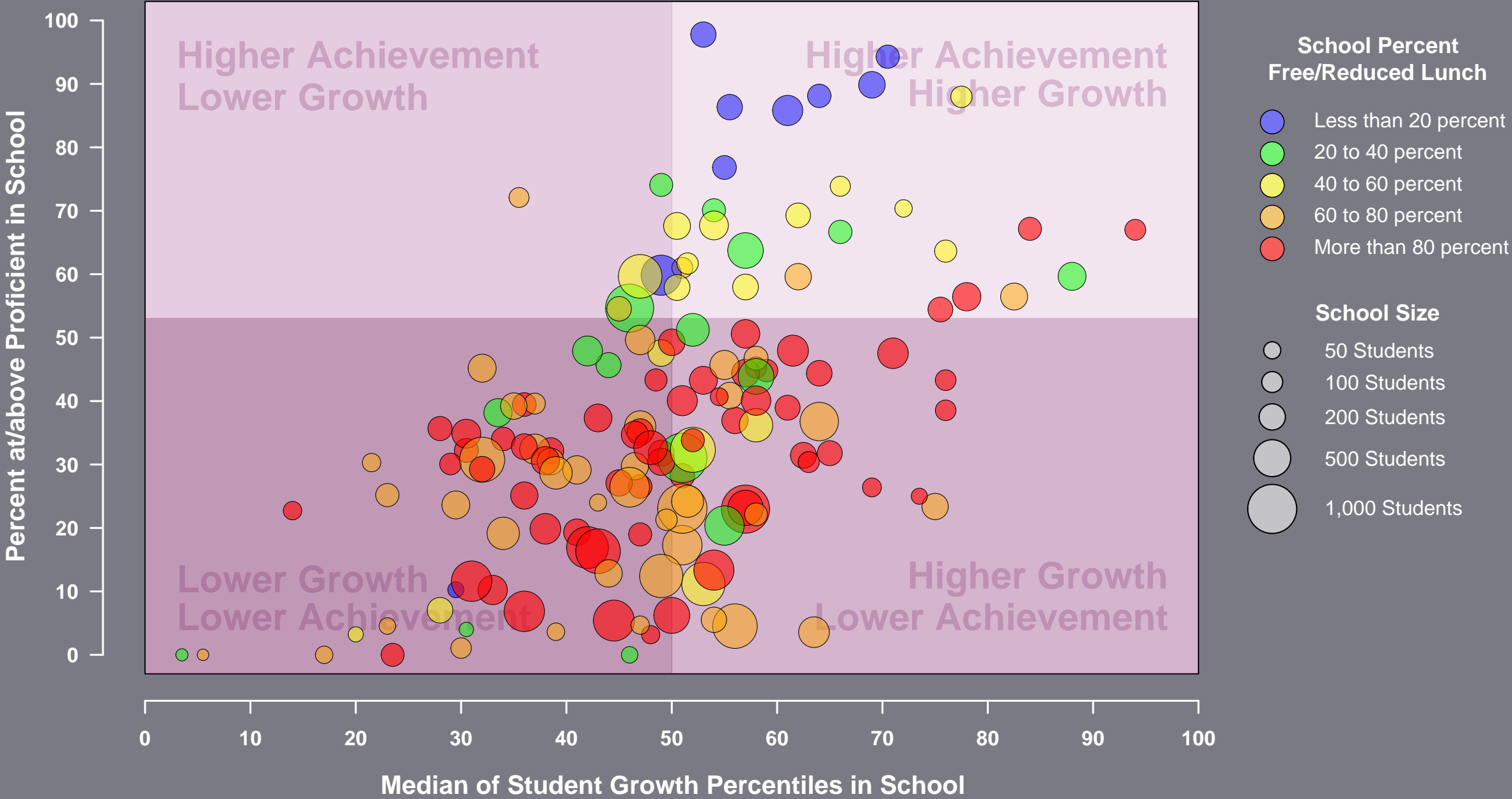
# 2008 CSAP Mathematics Growth and Achievement Chart



## Going from Students to Schools

- It's of interest to examine schools where students demonstrate, on average, extraordinarily high and low student growth.
- To summarize the student growth percentiles associated with a school (or other grouping) calculate the median of the student growth percentiles.
- If students were randomly assigned to schools, expect to see a median of 50.
- Values greatly above or below 50 are of interest in identifying best practices or providing extra support.
- Examining growth with achievement sheds new light on school performance.

*District C: 2008 CSAP Math School Results*  
*Student Growth versus Student Achievement by Percent Free/Reduced Lunch*



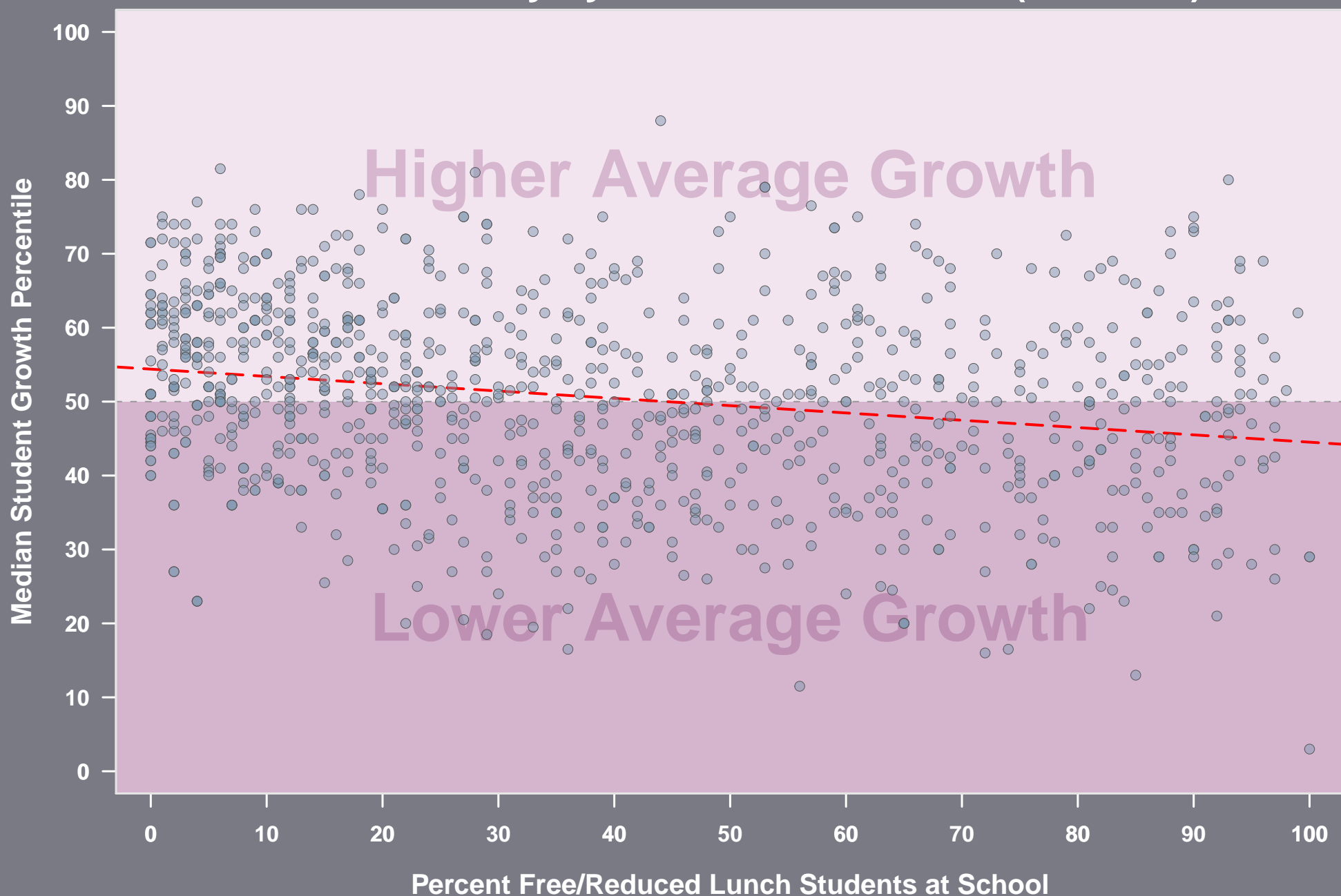
# Growth, Effectiveness, and Value-Added

- “Good schools” are often called highly effective schools.
- What’s the relationship between **growth** and **effectiveness**?
- Effectiveness indicates who/what is responsible for the growth (value-added models).
- RttT has placed teacher responsibility for student progress as a primary issue.
- In teacher performance evaluations, even when large scale assessment results are available, depending upon the state they constitute, on average, about a third of the overall teacher rating.

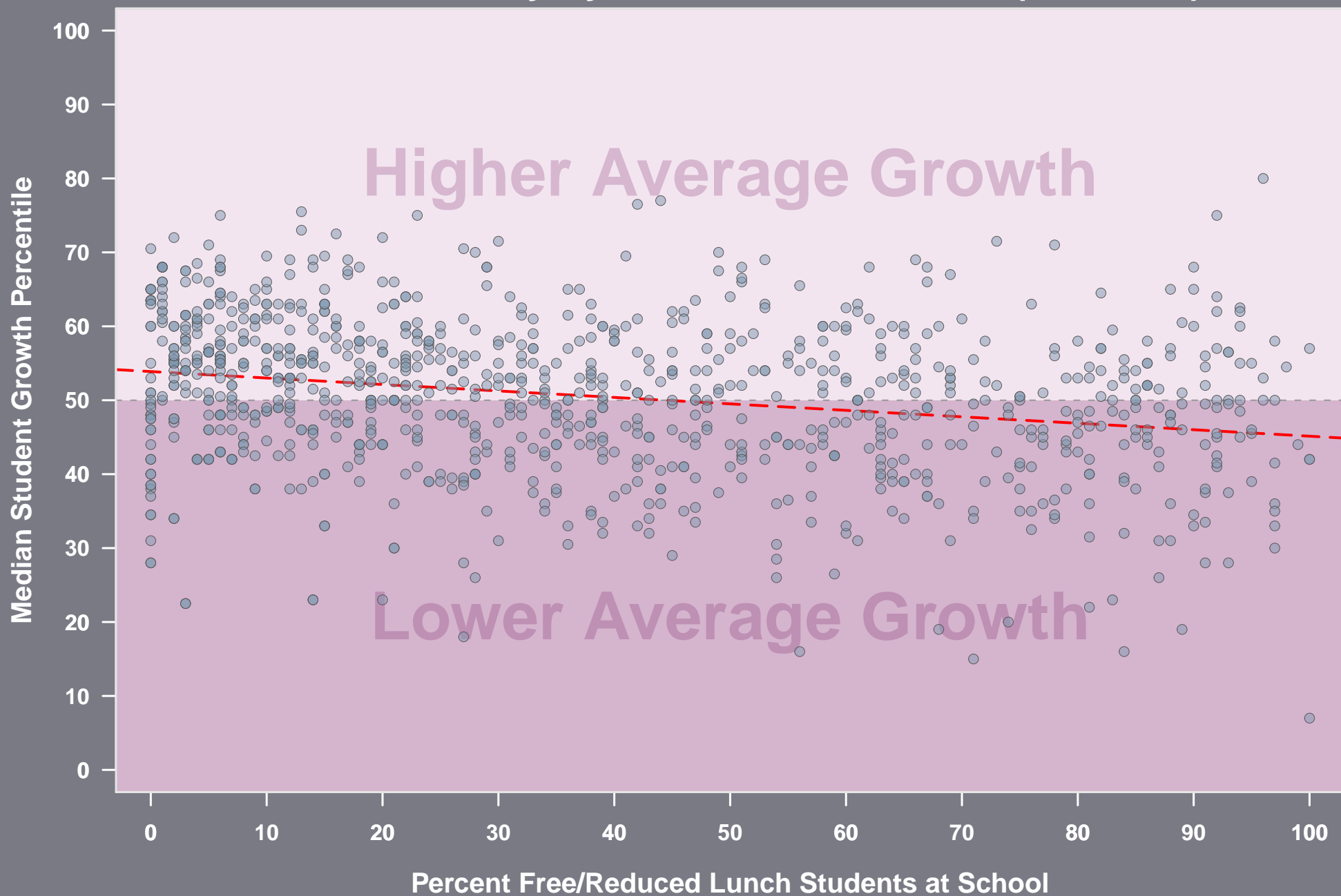
# 2009 CSAP Elementary School Mathematics: School Poverty by School Achievement ( $r = -0.71$ )



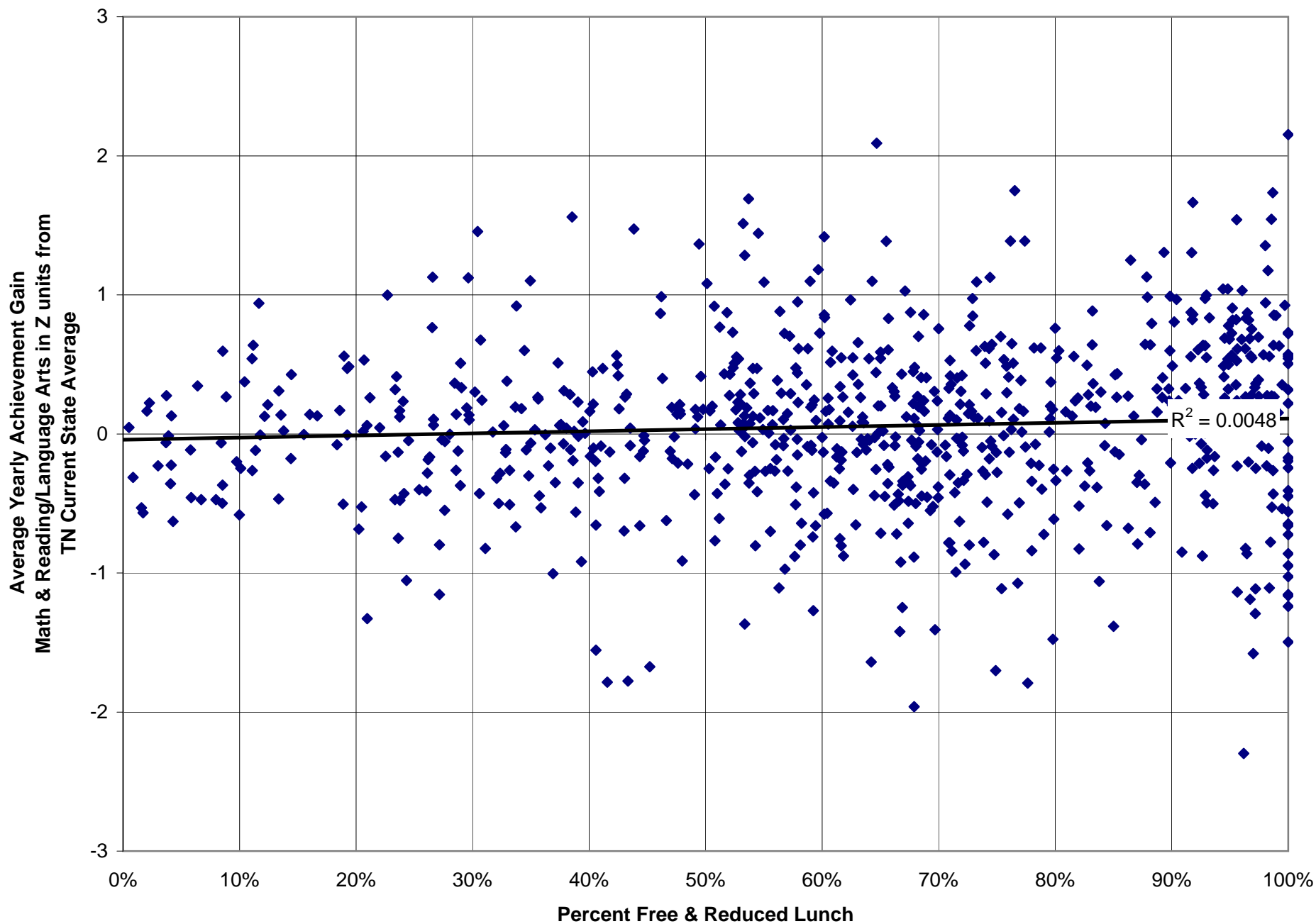
# 2009 CSAP Elementary School Mathematics: School Poverty by School Median SGP ( $r = -0.22$ )



# 2009 CSAP Elementary School Reading: School Poverty by School Median SGP ( $r = -0.25$ )



**Economic Disadvantage vs School Effectiveness**  
**TN Elementary Schools 2008**





## Descriptive Accountability

“This is the difference between a retrospective question of identifying fault as opposed to a prospective strategy to engineer some corrective measure, almost independent of considering whether there was blame-worthiness. And to move away from the blame-worthiness paradigm toward something that is more regulatory in nature where one might seize upon disparities or circumstances that are for some reason deemed unacceptable and engineer the interventions needed to bring about the necessary change. . . . It’s the no-fault gap closing strategy in which the effort is to build a consensus about a vision of an improved society rather than figure out where’s the person we want to pillory.”

C. Edley (2006)

## Web 2.0: Data Visualization and Social Networking

- The Colorado Department of Education and the Center for Assessment have been working for the last two years on developing next generation data visualization to accompany growth model data.
- The goal: Transform conversations about education through active engagement with data (i.e., evidence).
- Our efforts have received tremendous interest and recognition:
  - Recognized by Adobe for innovative uses of their technology as an Adobe Max Award finalist in October, 2009.
  - 2010 NCME Award for Outstanding Dissemination of Educational Measurement Concepts to the Public.
  - 14 states signing MOUs to co-develop a cloud-based reporting platform in a non-proprietary fashion.
  - Colorado recently devoted \$2.5 million of stimulus funds to the development efforts

# The Decade of Data Visualization

**Chart Wars** [http://infosthetics.com/archives/2010/01/chart\\_wars\\_the\\_political\\_power\\_of\\_data\\_visualization.html](http://infosthetics.com/archives/2010/01/chart_wars_the_political_power_of_data_visualization.html)

**Hans Rosling/Gapminder**  
<http://www.youtube.com/watch?v=jbkSRLYSojo>

**Journalism in the Age of Data**  
<http://datajournalism.stanford.edu/>

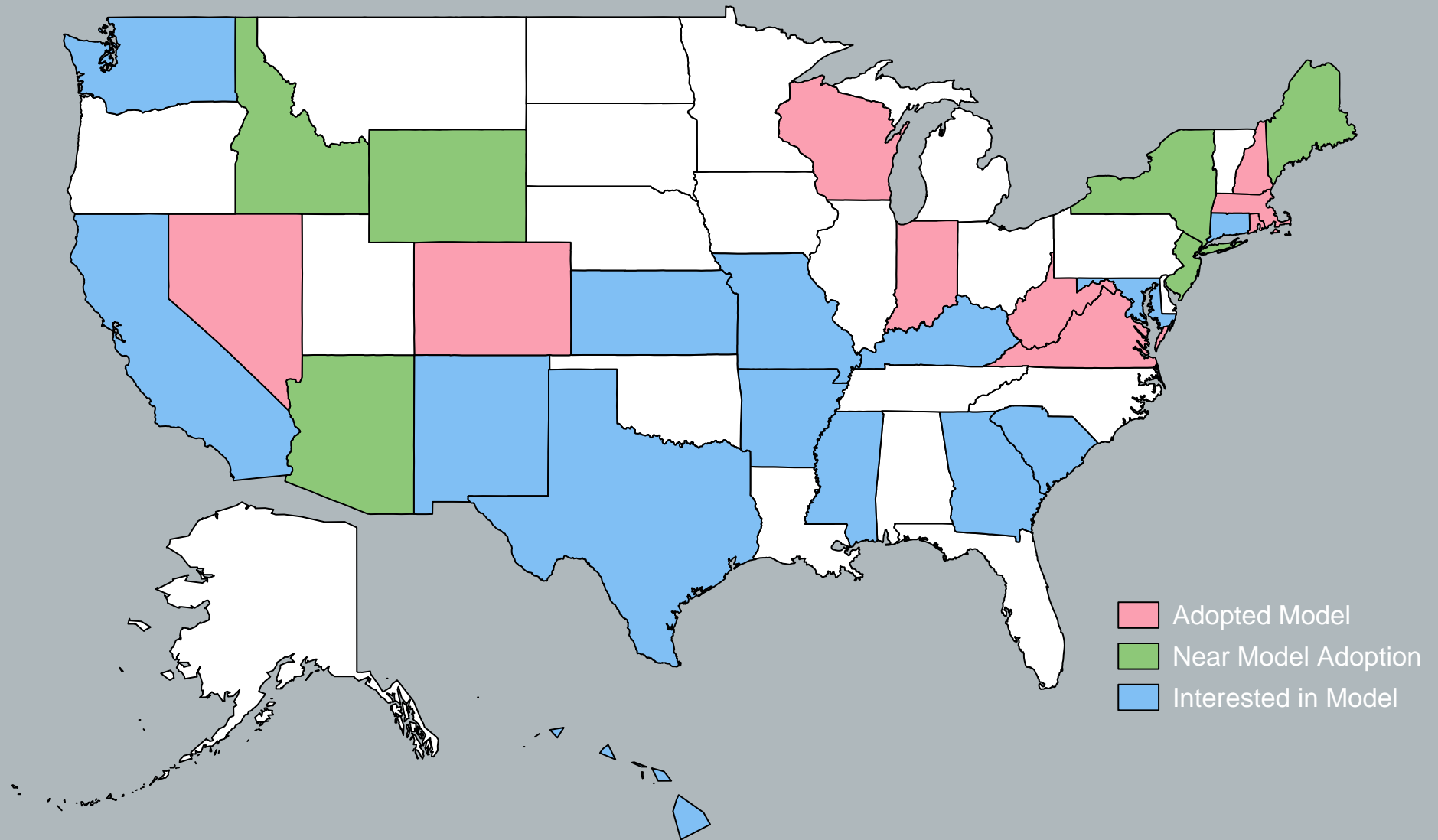
## Web 2.0: Data Visualization and Social Networking

With a collaborative spirit, with a collaborative platform where people can upload data, explore data, compare solutions, discuss the results, build consensus, we can engage passionate people, local communities, media and this will raise—incredibly—the amount of people who can understand what is going on.

And this would have fantastic outcomes: the engagement of people, especially new generations; it would increase knowledge, unlock statistics, improve transparency and accountability of public policies, change culture, increase numeracy, and in the end, improve democracy and welfare.

E. Giovannini, Chief Statistician, OECD. June 2007

# The Colorado Growth Model



# References

1. Berk, R. A. (2003). *Regression Analysis: A Constructive Critique*. Sage, Thousand Oaks, CA.
2. Betebenner, D. W. (2008). Toward a normative understanding of student growth. In Ryan, K. E. and Shepard, L. A., editors, *The Future of Test-Based Educational Accountability*, pages 155–170. Taylor & Francis, New York.
3. Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4):42–51.
4. Betebenner, D. W. (2011). SGP: Student Growth Percentile and Percentile Growth Projection/Trajectory Functions. R package version 0.0-6.7.
5. Braun, H. I. (2005). Using student progress to evaluate teachers: A primer on value-added models. Technical report, Educational Testing Service, Princeton, New Jersey.
6. Economic Policy Institute (2010). Problems with the use of student test scores to evaluate teachers. Technical report, Washington, D.C. Downloaded August 30th, 2010 from [http://epi.3cdm.net/724c89a1eb91c46f8b\\_bwe611j98.pdf](http://epi.3cdm.net/724c89a1eb91c46f8b_bwe611j98.pdf).
7. Edley, C. (2008). Educational "Opportunity" is the highest civil rights priority. So what should researchers and lawyers do about it? Retrieved June 22, 2006 from the World Wide Web: <http://www.softconscience.com/REDA/NRP/258467/#43,618>.
8. Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, 107(1):65–81.
9. Linn, R. L. (2008). Educational accountability systems. In *The Future of Test-Based Educational Accountability*, pages 3–24. Taylor & Francis, New York.
10. National Research Council and National Academy of Education (2010). Getting value of out value-added: Report of a workshop. Technical report, Washington, D.C. Committee on Value-Added Methodology for Instructional Improvement.
11. R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 3-900057-07-0.
12. Rubin, D. B., Stuart, E. A., and Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 29(1):103–116.
13. Yen, W. M. (2007). Vertical scaling and No Child Left Behind. In Dorans, N. J., Pommerich, M., and Holland, P. W., editors, *Linking and Aligning Scores and Scales*, pages 273–283. Springer, New York.