

# Balanced Evaluation in Urban Secondary Magnet Programs

Tom Watkins, Saint Paul Public Schools  
Lesa Covington Clarkson, University of Minnesota  
Sheila Arens, McREL  
Geoffrey Borman, University of Wisconsin

# The BioSMART Evaluation

- BioSMART is a Bioscience-focused Federal Magnet grant in two secondary schools in Saint Paul, Minnesota experiencing NCLB restructuring.
- The BioSMART evaluation employs an experimental and a quasi-experimental design implemented by internal and external partners. This and other information from the evaluation helps make sense of the NCLB-related outcomes.

# Saint Paul Public Schools

- Minnesota's second largest district, with 38,500 students and 100+ languages and dialects

**30% African-American**

**29% Asian-American**

**26% Caucasian**

**13% Hispanic**

**2% American Indian**

**40% English Language Learners**

**17% Special Education**

**70% Free/Reduced-Price Lunch**

- Demographics of “target” secondary schools in study: 90% Free/Reduced Price Lunch, and 90-95% students of color.
- See reports and data at ***[datacenter.spps.org](http://datacenter.spps.org)***

# This Presentation

- Part 1: Experimental and Quasi-Experimental Designs in Educational Evaluation
- Part 2: Deeper and Wider than NCLB
- Part 3: Major Lessons Learned



Part 1:

# Experimental and Quasi-Experimental Designs in Educational Evaluation

# Magnet Program Outcomes

- Reduced minority isolation
- Increased student achievement and skills
- Standards alignment
- Innovative practices
- Improved district capacity



# Some Common Challenges for Experimental Designs in Education

- Researchers responding to desire for evidence-based practices
- Providing a promising treatment to one group, while maintaining “business as usual” in the other
- Challenges in perceived parity, propriety, communication
- Generally easier to conduct in situations where schools are over-subscribed

# Technical Progress of BioSMART Experimental Design

- No differential attrition by program condition
- No conflict between random assignment of treatment and missing at random
- Treatment and control sufficiently well-matched at baseline on prior achievement, gender, ethnicity, ELL status, special education status, and free-reduced meal eligibility
- Reasonable external validity: diverse sample, includes nearly all students in school

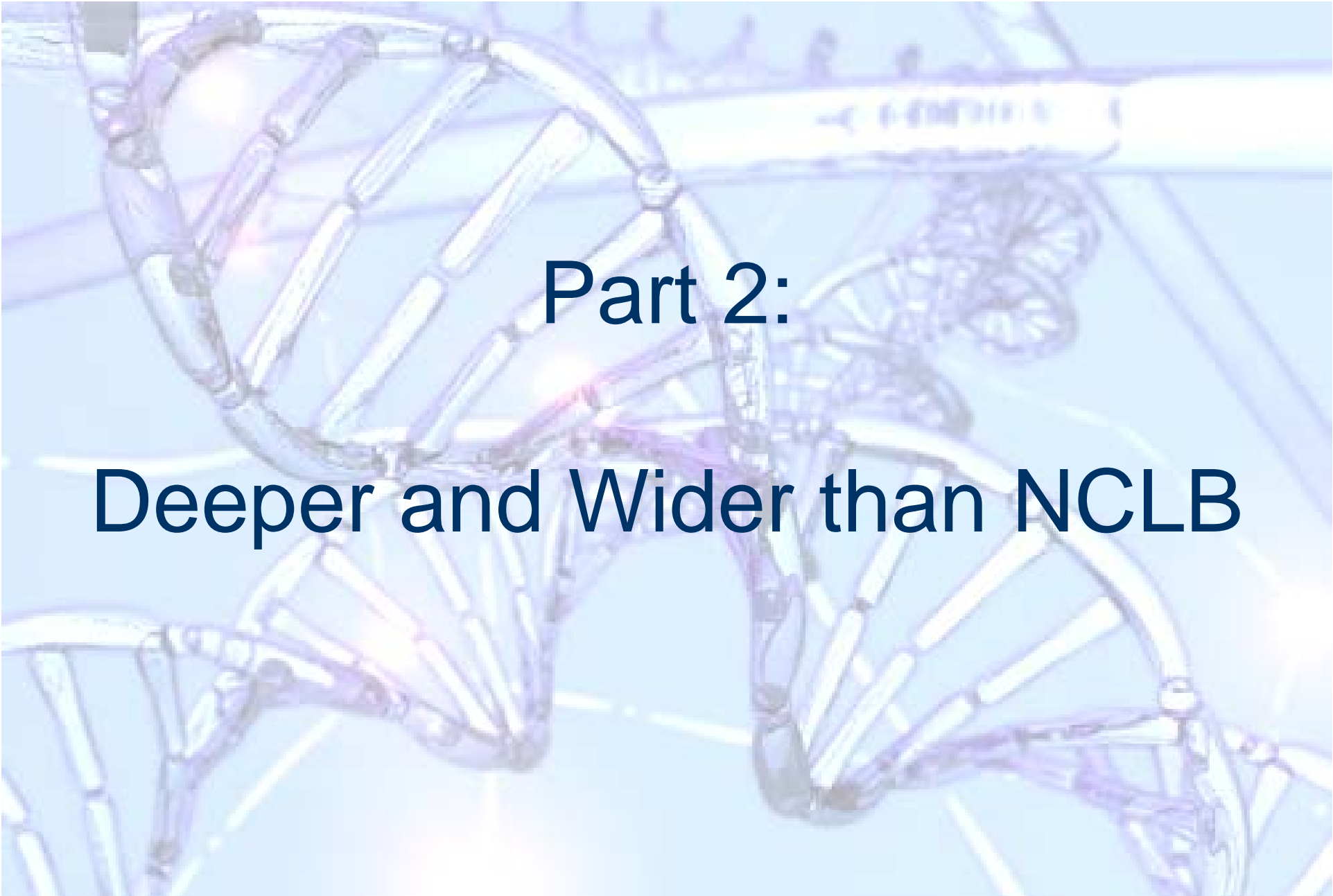


# Technical Progress of BioSMART Quasi-Experimental Design

- The two groups differed initially in regard to prior science achievement, Asian American and Caucasian ethnicity, and free/reduced lunch status.
- Propensity score matching (probit model with achievement and demographics) substantially improved the quality of the comparison
- After matching, no significant differences between the two groups at baseline are observed.
- Treatment status and missing status are independent.

# Key Implementation Steps for Experimental Design

- Engagement of school principal – thorough awareness of, and collaboration on method
- Persistent negotiation with key staff about how to address the most important aspects of experimental design
- Clear and ongoing communication with all staff about what the experimental design means for PLC's, interpreting outcomes, evaluation reporting, etc.
- Determined low risk, sought passive parental consent



# Part 2:

## Deeper and Wider than NCLB

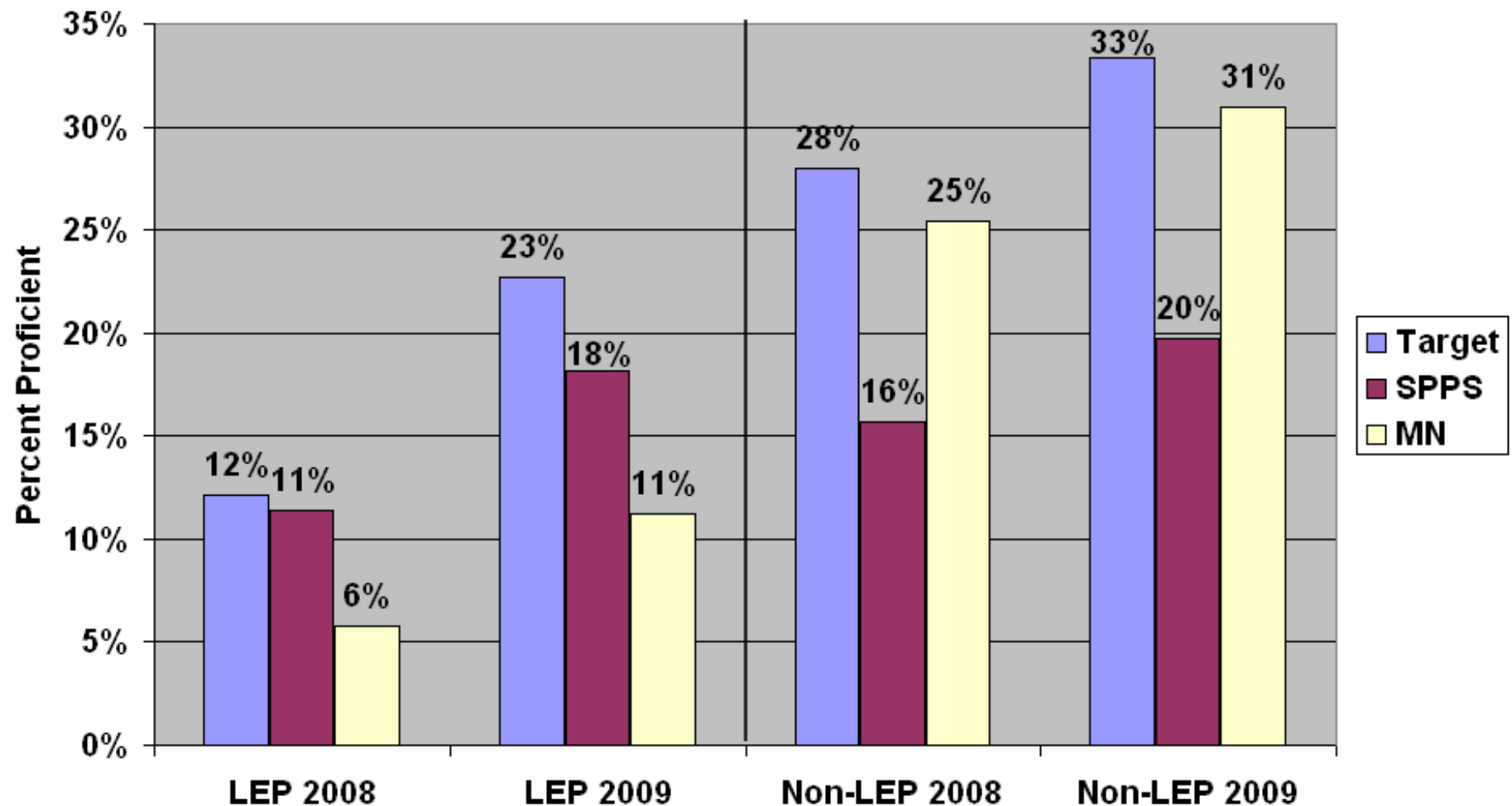
## Linn, R. (2005). Fixing the NCLB Accountability System. CRESST Research Brief #8

- **“set realistic performance targets for AYP”**
- **“consider growth...not just status”**
- **“more meaningful and comparable achievement targets [across states]”**

# **AEA's Public Statement on Educational Accountability (2006)**

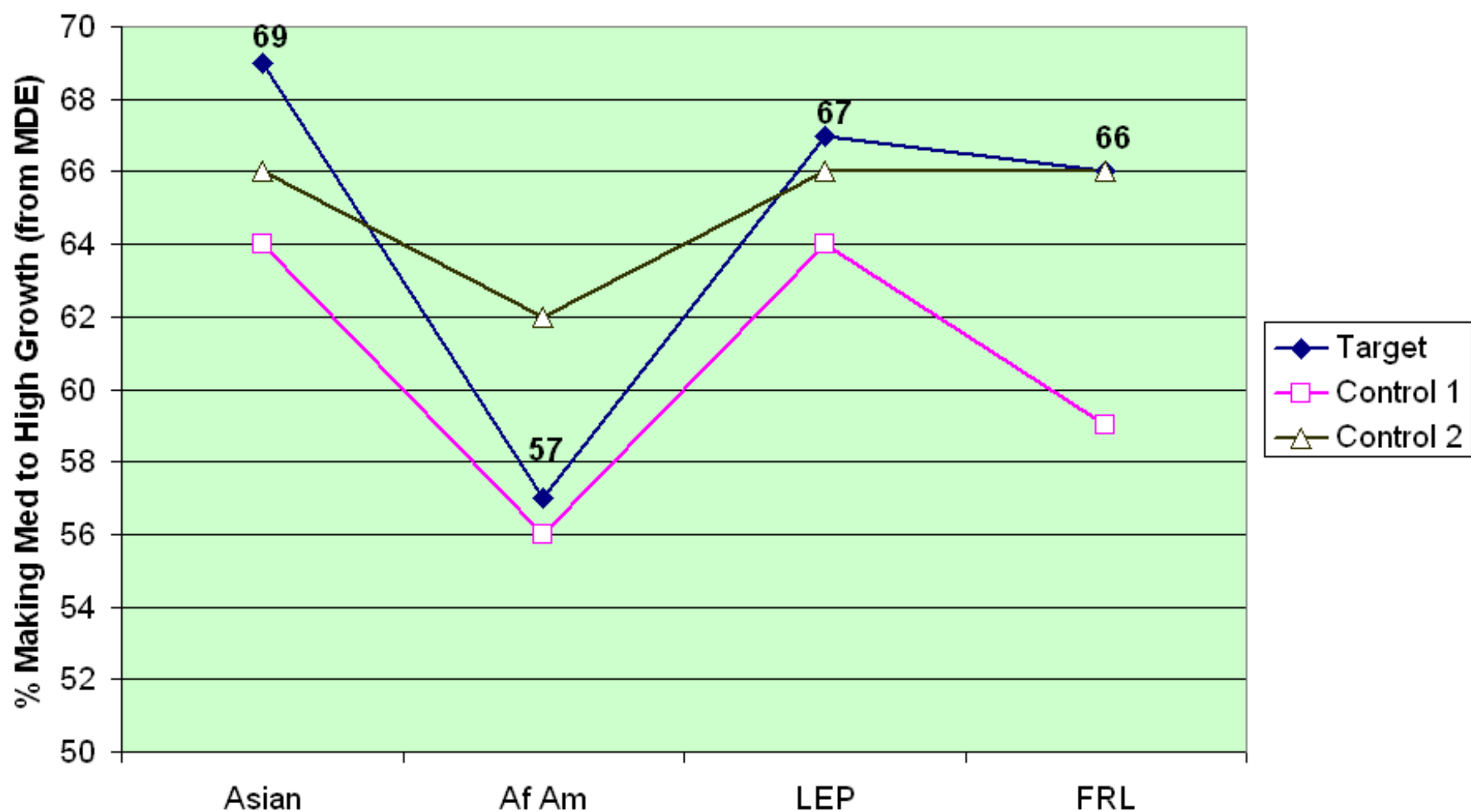
- **Multiple measures**
- **Measurement of individual student progress over time**
- **Context sensitive reporting**
- **Data-based resource allocations**
- **Accessible appeals processes**
- **Public participation and access**

## Minnesota Comprehensive Assessment Science Results for Low Income Students in 2008 and 2009

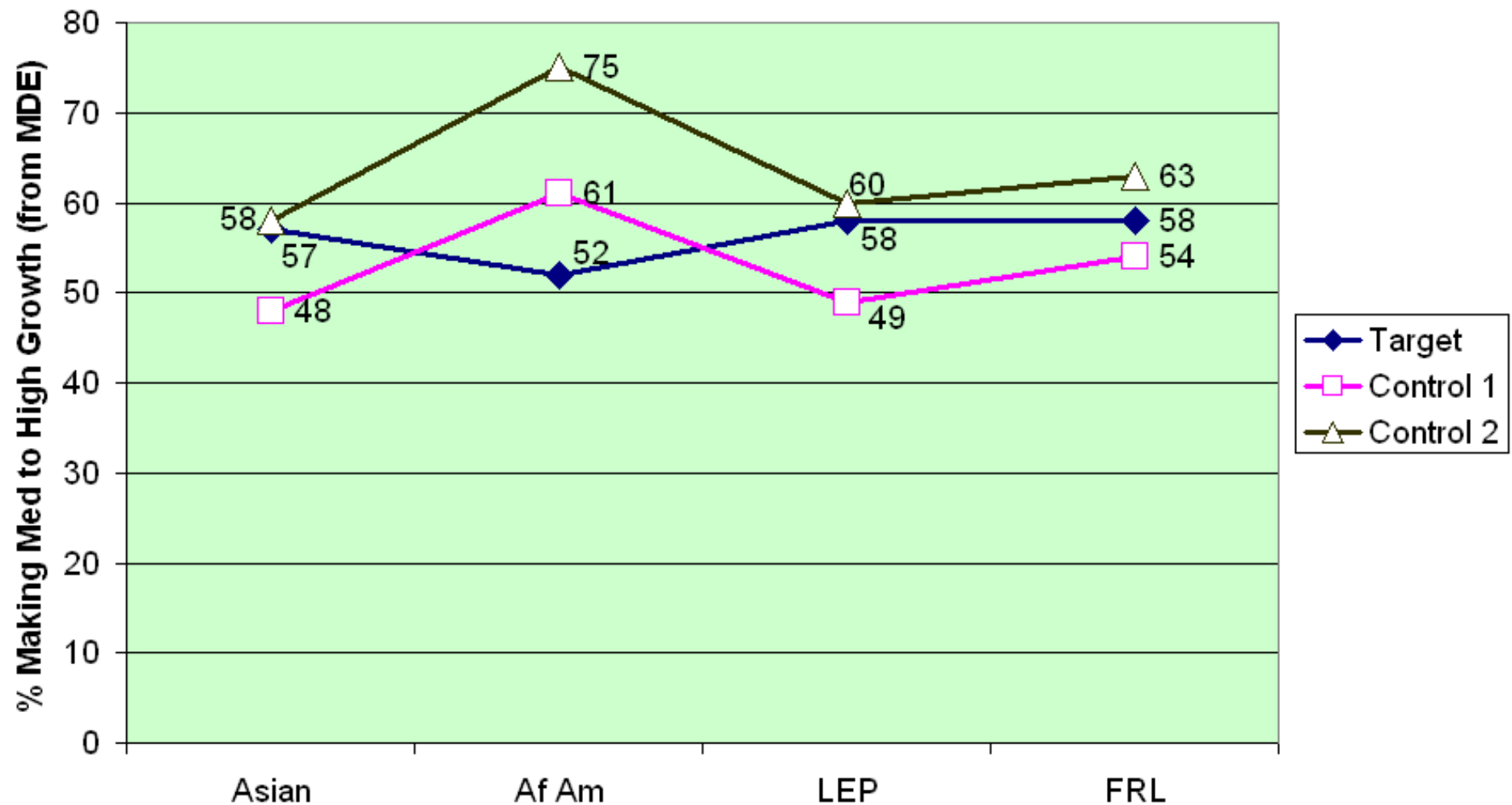




## Percent of Students Making Medium to High Growth on the Minnesota Comprehensive Assessment in MATH 2008-09



## Percent of Students Making Medium to High MCA-II READING Growth 2008-09





# Part 3:

## Major Lessons Learned

# Lesson 1: Internal and external evaluators need to “add value” at the school level as well as the district level

- Begin by taking the time to listen to the needs of leaders and other staff
- Offer feasible, targeted assistance (e.g., data and assessment) to meet needs
- Give opportunities for reflection and input into the evaluation process and tools
- Constantly building evaluation capacity, confidence and engagement – working ourselves out of a job

## Lesson 2: Common and viable assumptions are worth more than a “pretty” logic model

- Help school staff to recognize and articulate assumptions
- Get to know each of the leaders enough to find how, when and where they ask “tough questions” (retreats, 1 to 1, etc.)
- Professional Learning Community process can help with this
- We have pursued this gradually – maybe too slow!

## Lesson 3: Maintain appropriate and effective data practices

- Internal evaluator can quickly and efficiently collect data as needed, through changes in staffing and technologies
- Internal evaluator may have “privileged” access to information that helps to understand the program
- Useful for evaluators to agree about data sharing prior to starting the evaluation
- Both parties need to remain cognizant of their responsibilities to those providing data



## WestEd Magnet Evaluation Toolkit [www.evaluationtoolkit.org](http://www.evaluationtoolkit.org)



# Contact Info

Tom Watkins, Ph.D  
Program Evaluator  
Saint Paul Public Schools  
[Tom.Watkins@spps.org](mailto:Tom.Watkins@spps.org)