

## RESEARCH SUMMARY

### Effectiveness of *Measuring Up* as Preparation for the New York State Fourth Grade Test in English Language Arts

#### *A Report of a Randomized Experiment in Mount Vernon City School District*

We were asked by Peoples Publishing Group to find out whether their product, *Measuring Up (MU)*, was more effective in helping a district prepare fourth-grade students for the New York State test of English Language Arts than materials the district already had in place. We conducted a randomized experiment during the 2004-2005 school year in Mt. Vernon, NY.

**Intervention.** *Measuring Up (MU)* is a supplementary, text-based product designed to help teachers in K-12 schools prepare students for their standards-based state exams. *MU* is customized for a range of state standards and assessment practices and goes well beyond simple test prep by systematically addressing the state standards and providing a range of pedagogical tools. After a half-day in-service session led by a PPG staff member, teachers were free to use *MU* as best suited their needs.

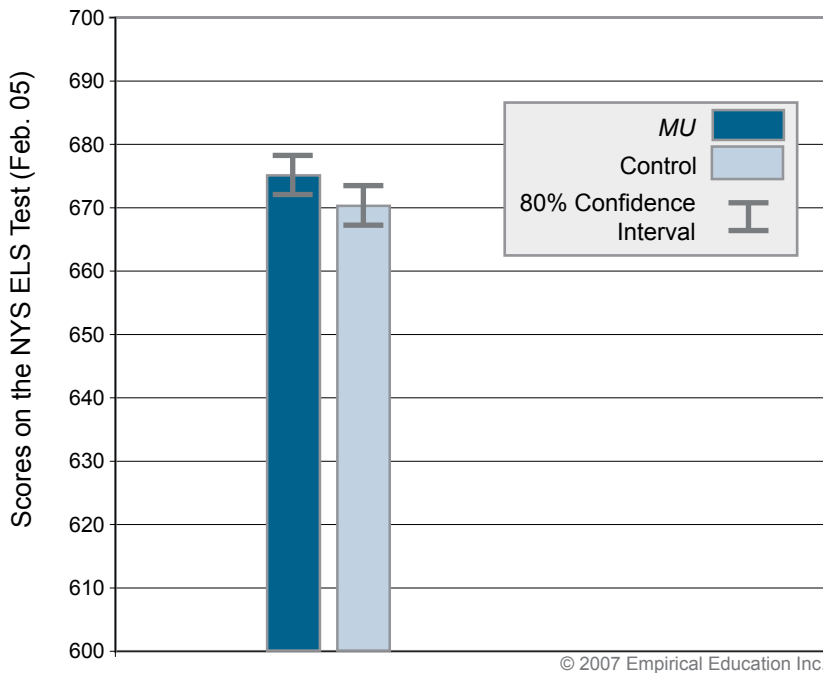
**Setting.** Our research site is a city of approximately 68,000 located in Westchester County just north of New York City, where the median household income is about \$41,000. The district enrolls about 10,000 K-12 students, including 78% African American, 13% Hispanic, and 8% White students. Materials for preparing fourth graders for the ELA test had not been adopted and teachers used a wide variety of materials and methods to prepare their students: teacher-developed materials, enrichment activities, some supplementary products, and segments of their basal text.

**Research design.** Our research design was a randomized experiment (or randomized controlled trial). This type of study is the best way to assure that the new product and not some characteristic of the teachers or students caused the differences observed between groups. Teachers who volunteered to participate were assigned by coin toss to the *MU* group or to the control group. The process of randomly assigning teachers to conditions assured that classes from each school were approximately evenly distributed between conditions and that the distinct populations were represented in each condition. The outcome measure was the state test of English Language Arts. The pretest measure was the TONYSS (Riverside Publishing).

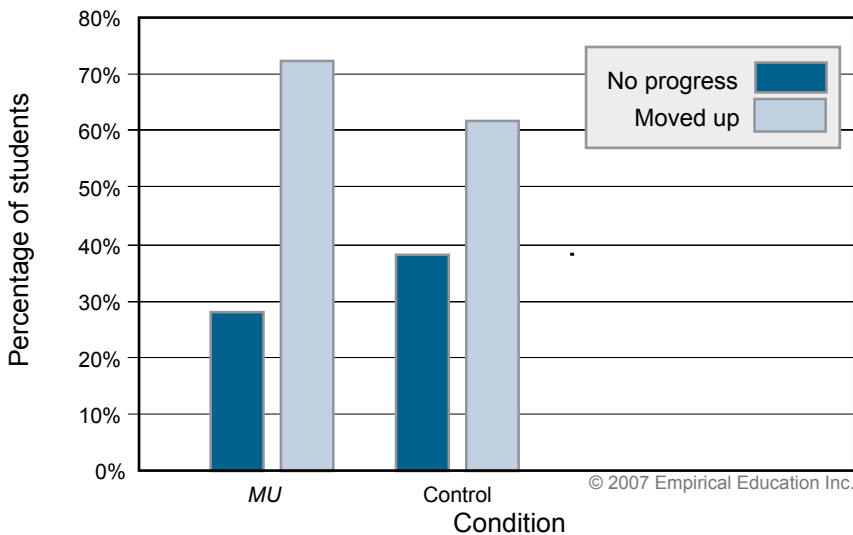
**Participants.** A total of 375 fourth-grade students and 19 teachers participated in the study. Random assignment resulted in two groups containing students who were evenly split on socioeconomic status. Control group students scored slightly higher on the pretest than *MU* students, a difference that was controlled statistically.

**Statistical analysis.** Our method for drawing a quantitative conclusion from the data considered what was known about student demographics and their incoming level of ELA as well as outcome differences between *MU* and control group student scores. We used an analysis of covariance combined with a multi-level analysis to increase the precision of the estimate of *MU*'s impact and to account for the clustering of students in classes. We also tested whether the impact depends on students' incoming skills and other important factors.

**Results.** Teachers in both the control group and the group piloting *MU* reported spending approximately the same amount of classroom time preparing for the state test. Comparison of the means for *MU* and control groups revealed a difference of 5.46 points. Translated into a standardized measure that takes into account the distribution of the scores, we find an effect size of .15. Effect sizes in this range are often found to be educationally meaningful. The bar graph shows this difference between *MU* and control group scores. The bars represent the score that would be predicted for a student performing at the average level in the two groups. At the top of each bar, we have indicated the 80% confidence interval. There is a probability of 80% that the true values for groups lie within their respective confidence intervals.



Viewed in terms of the four proficiency levels established by New York State, we find that a significantly larger portion of students in the classes with *MU* moved to a higher proficiency level compared to those in the control classrooms.



**Conclusion.** For the students and teachers in Mt. Vernon, *MU* was generally more effective than the other products in use for helping to improve student achievement. We can conclude that *Measuring Up* is a valuable option for supplementing English Language Arts instruction where the goal is higher achievement on the New York State test.

**To read the complete report, contact:**

Empirical Education Inc.  
 www.empiricaleducation.com  
 425 Sherman Avenue, Suite 210  
 Palo Alto, CA 94306  
 (650) 328-1734 x211