Abstract Title Page

Not included in page count.

Title: Evaluating Math Recovery: Assessing the Causal Impact of Math Recovery on Student Achievement

Author(s): Thomas Smith, Paul Cobb, Dale Farran, David Cordray, Charles Munter and Alfred Dunn

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Children enter school at a wide range of mathematical abilities (Baroody, 1987; Dowker, 1995; Gray, 1997; Griffin & Case, 1999; Housasart, 2001; Wright, 1991, 1994a; Young-Loverage, 1989). A study conducted by Aunola, Leskinen, and Lerkkanen (2004) found that, in the absence of intervention, the initial gap in mathematics achievement continues to widen. Zill and West (2000) examined data from the U.S. Department of Education's Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), to describe the nature of this pre-K gap in mathematics achievement. On the one hand, they found that 20% of all kindergarten students in the U.S. in 1998 (approximately 780,000 students) were already beyond counting and reading single-digit numerals, and 4% (156,000) were even doing arithmetic. On the other hand, 42% (1.6 million) could not count up to 20 objects, and 6% (234,000) were unable to count even 10 objects.

Initial gaps in student achievement are persistent and in many cases widen as student progress in school (Aunola et al., 2004; Cockcroft, 1982). Duncan, Claessens, and Engel's (2004) analysis of ECLS-K indicated that pre-K mathematical ability is highly predictive of achievement at the end of first grade. Although teacher- and parent-reported social and emotional behaviors had standardized coefficients between .01 and .05, early mathematics abilities had coefficients in the .35 to .50 range for the subsequent first-grade data. Princiotta, Flanagan, and Germino Hausken's (2006) analysis of ECLS-K data revealed that achievement gaps are still prevalent in fifth grade. They found that 67% of students who scored in the top third in their kindergarten year did so again six years later; and that those among the lowest third in 1998 generally scored low in 2004.

Children's differing levels of mathematical ability when they enter school are related to multiple factors. Children who are less ready for school typically come from families of low SES status, are of racial or ethnic minority backgrounds, have parents who do not speak English in the home, or possess disabilities (Alexander & Entwistle, 1998; Barton, 2003; Berends, Lucas, Sullivan, & Briggs, 2005; Cahalan et al., 2006; Chen, 2005; Crosnoe, 2005; Dahlstrom, 2005; Fuson, Smith, & Lo Cicero, 1997; Griffin, Case, & Siegler, 1994; Griffin, 2004; Marchand, Pickreign, & Howard, 2005; Vandivere, 2004; Walker, 2006; Wilms, 1986). However, it is important to note that ethnicity per se does not predict success in mathematics once it is adjusted for other factors (Thomas, 2000). One potential explanation for the persistence of the initial achievement gap focuses on inequities in the educational opportunities across ethnic, racial, and socioeconomic groups. In this regard, Oakes (1990) found that "[s]chools with large concentrations of low-income and minority students [tend to] offer fewer classroom conditions that are likely to promote active engagement in mathematics and science learning—such as opportunities for hands-on activities and time working with the teacher" (p. 101).

The research findings we have reviewed thus far indicate that differences in early mathematical abilities are relatively stable and can lead to differentiated instruction in the later years of elementary school and in middle school. The findings emphasize the pressing need to equip schools with effective methods for closing the pre-K gap (McWayne, Fantuzzo, & McDermott, 2004).

Purpose / objective / research question / focus of study:
Description of what the research focused on and why.

Our goal was to evaluate the potential of Math Recovery (MR), a pullout, one-to-one tutoring program that has been designed to increase mathematics achievement among low-performing first graders, thereby closing the school-entry achievement gap and enabling participants to achieve at the level of their higher-performing peers in the regular mathematics classroom.

Specifically, our research questions were as follows:

1. Does participation in MR raise the mathematics achievement of low performing first-grade students?
2. If so, do participating students maintain the gains made in first grade through the end of second grade?

Setting:
Description of where the research took place.

The two-year evaluation of Math Recovery was conducted in 20 elementary schools (five urban, ten suburban and five rural), representing five districts in two states. Each was a 'fresh site' in that the program was implemented for the first time for the purposes of the study.

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

Students were selected for participation at the start of first grade based on their performance on MR's screening interview and follow-up assessment interview. The screening is designed to select the lowest achieving first graders (25^{th} percentile and below) in terms of math achievement. The number of students eligible for tutoring ranged from 17 to 36 across the 20 schools. The number of study participants before attrition totaled 517 in Year 1 and 510 in Year 2, of which 172 received tutoring in Year 1 and 171 received tutoring in Year 2. Approximately 50% of participants were males, 48% were non-white and 48% received free or reduced lunch.

We recruited 18 teachers to receive training and participate as MR tutors from the participating districts—all of whom had at least two years of classroom teaching experience. Sixteen of the tutors received half-time teaching releases to serve one school each; two of the tutors received full-time teaching releases to serve two schools each. All tutoring positions were underwritten by their respective school districts.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

MR consists of three components: 1) tutor training, 2) student identification and assessment, and 3) one-to-one tutoring. The first component of the MR program, tutor training, involves 60 hours of instruction provided by an MR leader. The goal of this training is to support tutors' in learning new practices for clinical assessment and intervention teaching in which they use the Learning Framework and the Instructional Framework to adjust instruction based on cognitive evaluations of student responses.

The second component of the program, the tutor conducts an extensive video-recorded assessment interview with each child identified as eligible for the program. The tutor analyzes these video-recordings to develop a detailed profile of each child's knowledge of the central aspects of arithmetic using the MR Learning Framework, which provides information about student responses in terms of levels of sophistication

The third component of the program, one-to-one tutoring, is diagnostic in nature and focuses instruction at the current limits of each child's arithmetical reasoning. Each selected child receives 4-5 one-to-one tutoring sessions of 30 minutes each week for approximately 11 weeks. The tutor's selection of tasks for sessions with a particular child is initially informed by the assessment interview and then by ongoing assessments based on the student's responses to prior instructional tasks. The Learning Framework that the tutor uses to analyze student performance is linked to the MR Instructional Framework that describes a range of instructional tasks organized by the level of sophistication of the students' reasoning together with detailed guidance for the tutor.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

The structure of the MR program allowed us to use the fact that two thirds of the participating students will have their treatment delayed by either 11 or 22 weeks to establish an experimentally assigned control group for each cohort of participants consisting of both students whose treatment has not yet begun and a small number of students who are on a "wait list" for treatment. By randomly assigning the students selected for participation in the study each year to one of the three treatment cohorts or the wait list, we can establish the essential characteristics of an experimental design: a comparison of students' change in mathematics achievement during their 12 weeks of participation in MR to the gains they would have made if they not participated in the intervention.

In each year (2007-08 and 2008-09 academic years), three eligible students from each school were randomly assigned to a tutoring cohort with a different start date (i.e., Cohort A—September, Cohort B—December, Cohort C—March) or to the "waiting list" for MR. In both years students on the randomly ordered waiting list were selected to join an MR tutoring cohort if an assigned participant left their school or were deemed "ineligible" due to a special education placement.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Each of the students participating in the study were assessed using alternating forms of the Applied Problems, Quantitative Concepts, and Fluency subtests of the Woodcock Johnson III Achievement tests (WJ III) subtests, as well as the MR proximal instrument designed in consultation with the program developers, at the start of the study and when each cohort entered or exited tutoring in December, March, and May. Wait list students took the Fluency subtest of the WJ III at the same time as each cohort entering treatment, as well as the full battery of other WJ III and MR proximal assessments at the start and end of the school year.

Our research design allowed us to describe and compare the growth trajectories of treatment and control cohorts across the whole school year, punctuated at the end of each 11-week period by the students completing MR tutoring. We used the estimated growth rate of Cohorts 1B and 1C prior to receiving treatment, as well as the estimated growth rates of students on the wait list who did not receive the MR intervention, to estimate a counterfactual to the growth rate of MR participants. At the end of each of these intervals, a given study participant has one of three statuses: not yet received any MR, just completed MR tutoring, or is post MR. We made different comparisons within this scheme to determine mathematics achievement outcomes immediately at the end of an MR session relative to students who had not received the MR intervention (to test the treatment effect) and outcomes 12 or 24 weeks after completing MR to those who have just completed MR (to test whether MR gains are maintained after the end of treatment).

To estimate these growth trajectories, we used a 3-level hierarchical linear growth models (Raudenbush and Bryk, 2002; Singer and Willett, 2002) with repeated observations of WJ III scores or MR proximal scores indexed by time, time since starting MR, and time since completing MR at level 1, student level demographics at level 2 (e.g., gender, minority status), and school characteristics at level 3. To assess whether gains made in MR tutoring are maintained after the tutoring is completed, a time varying covariate ($POSTMRTIME_{ijt}$) that counts the number of days after a student completes MR. The level 1 equation looks like:

$$WJIII_{ijt}=\pi_{0j}+ \pi_{1j}(Time)_{ijt}+ \pi_{2ij}(MRTIME)_{ijt} + \pi_{3ij}(POSTMRTIME)_{ijt} + \varepsilon_{ijt}$$

Thus, the coefficient π_{2ij} on $MRTIME_{ijt}$ can be interpreted as the treatment effect—the additional daily learning associated with participation in MR relative to non-participants and cohorts who have not yet begun the tutoring program. The coefficient π_{3ij} on $POSTMRTIME_{ijt}$ can be interpreted as the additional daily learning for participants after completing MR compared to their rate of learning when participating in MR tutoring. Although the results presented here are only for the first year cohort in this study, the paper presented at SREE will include end of second grade data for Cohort 1 and end of first grade data for cohort 2. We are particularly interested in testing the hypothesis that the gains made from participation in MR are maintained through the end of second grade.

Findings / Results:
Description of main findings with specific details.

The first year results show a small to moderate effect of participation in MR on WJ III scores and moderate to large effects on the MR proximal assessments. Specifically, differences in the end of first grade mean scores on the WJ III subtests between students selected for tutoring and those on the waitlist ranged in effect size from .21 on the quantitative concepts scale to .28 on the applied problems scale (all differences statistically significant at the $p<.05$ level). Effect sizes on the MR 1.1 screening assessment ranged from .34 on the forward number sequence scale to .92 on the arithmetic strategies measure. These results compare favorably to those reviewed recently by Slavin and Lake (2006), including several cooperative learning programs that had median effect sizes of at least +0.30 in studies using randomized experimental or randomized quasi-experimental designs, including Class wide Peer Tutoring (.33), Student Team Learning (.19-.60), and TAI Math (.28-.38). A meta-analysis of 52 studies on the relationship between tutoring and student achievement (Cohen, Kulik, and Kulik, 1982), however, found average effect sizes greater than .40—higher than MR effects on the WJ III measures but lower than effects on some of the more proximal assessments.

Results from the growth models show increases in mathematics achievement for MR participants across all assessments during the tutoring period (with $p<.05$ in each case), although this growth rate tends not to be maintained after completion of MR. For example, on the applied problems subscale of the WJ III, MR participants gained .063 points per day, on average, during tutoring while students on the wait list for MR gained .038 points per day across the same time period. After exiting MR tutoring, however, participants' growth trajectories reverted back to pre-tutoring rates—a rate of .033 per day on the applied problems subscale. The pattern of strong gains pre to post tutoring, with regression towards the growth trajectories of non-participants was consistent across assessments.

By November 2009, we will have completed processing and cleaning of the Year 2 data for this study, allowing us to test whether MR tutors are more effective in their second year of tutoring than their first. We will also test differences in both the WJ III scores and a second grade version of the MR proximal assessment to determine whether the gains made by participants in first grade are maintained through the end of second grade.

Conclusions:

Description of conclusions and recommendations based on findings and overall study.

The findings of this study have theoretical, practical, and policy significance. Practically, the positive causal effect of MR tutoring demonstrates that programs that are diagnostic rather than scripted in nature can overcome fidelity concerns and have an impact on student early mathematics performance. Theoretically, our findings indicate that investing in tutors' knowledge of student reasoning and pedagogical content knowledge can pay off in terms of improvement in student's mathematical learning, particularly if tutors use carefully designed tools such as the MR Learning and Instructional Frameworks that codify and schematize this knowledge. With regard to policy, our finding that the MR program can reduce some of the pre-K mathematics achievement gap provides an initial indication that the cost of the program per student might be justified, although further work is needed to understand why initial gains made by participants appear to diminish after tutoring ends. It is possible that the forms of arithmetic reasoning that MR develops needs to be further supported in the regular classroom to see the full benefit of this form of tutoring. Longitudinal studies that track MR students and their initially higher performing peers until the end of elementary school are needed to address this question adequately.

.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Aubrey, C., Dahl, S., & Godfrey, R. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal, 18*(1), 27-46.

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 94*(4), 699-713.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*, 141-157.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fenema, E., & Empson, S. B. (1997). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29*, 3-20.

Carpenter, T. P. & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education,15*, 179-202.

Cobb, P., Gravemeijer, K., Yackel, E., McClain, K., & Whitenack, J. (1997). Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom. In D. Kirshner, & J. A. Whitson (Eds.), *Situated cognition theory: Social, semiotic, and neurological perspectives* (pp. 151-233). Mahwah, NJ: Lawrence Erlbaum.

Cockcroft, W. (1982). *Mathematics counts*. London: HMSO.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.

Dowker, A. (1995). Children with specific calculation difficulties. *Links 2*(2), 7-11.

Duncan, G. J., Claessens, A., & Engel, M. (2004). *The contribution of hard skills and socio-emotional behavior to school readiness*. Retrieved October 26, 2006, from http://www.northwestern.edu/ipr/people/duncanpapers.html.

Fuson, K. C. (1992). Learning addition and subtraction: Effects of number words and other cultural tools. In J. Bideaud, C. Meljac, & J. P. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 283-306). Hillsdale, NJ: Lawrence Erlbaum.

Fuson, K. C., Smith, S. T., & Lo Cicero, A. M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education, 28*(6), 738-766.

Gray, E. M. (1997). Compressing the counting process: Developing a flexible interpretation of symbols. In I. Thompson (Ed.), *Teaching and learning early numbers* (pp. 63-72). Buckingham: Open University Press.

Griffin, S. & Case, R. (1999). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3*(1) 1-49.

Houssart, J. (2001). Counting difficulties at Key Stage 2. *Support for learning, 16*, 11-16.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Singer, J. D., & Willet, J. B. (2002). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Slavin, R. E., & Lake, C. (2006). *Effective programs in elementary mathematics: A best-evidence synthesis.* Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.

Steffe, L. P., Cobb, P., & von Glasersfeld, E. (1988). *Construction of arithmetical meanings and strategies*. New York: Springer-Verlag.

Steffe, L. P., von Glasersfeld, E., Richards, J. J., & Cobb, P. (1983). *Children's counting types: Philosophy, theory and application*. New York: Praeger Publishers.

Wagner, S. (2005). *PRIME: PRompt Intervention in Mathematics Education: Executive summary of research and programs*. Columbus, OH: Ohio Resource Center for Mathematics, Science, and Reading & Ohio Department of Education.

Wright, R. J. (1991). What number knowledge is possessed by children beginning the kindergarten year of school? *Mathematics Education Research Journal, 3*(1), 1-16.

Wright, R. J. (1994a). A study of the numerical development of 5-year-olds and 6-year-olds. *Educational Studies in Mathematics, 26*(1), 25-44.

Young-Loveridge, J. M. (1989). The development of children's number concepts: The first year of school. *New Zealand Journal of Educational Studies, 24*(1), 47-64.

Appendix B. Tables and Figures

Not included in page count.