Title: Exploring Treatment Variation in the Scale-Up of Reading Recovery

Authors and Affiliations:

Henry May
University of Pennsylvania

Jerome D’Agostino
Ohio State University
Abstract Body

Limit 4 pages single-spaced.

Background / Context:
Reading Recovery has been the focus of a number of studies over the course of its 25-year history. Research has consistently shown that most of the low-achieving students who participate in RR achieve at a level similar to the average of their classmates in reading and writing by the end of the intervention (Ashdown & Simic, 2000; Center et al, 1995; D’Agostino & Murphy, 2004; Iversen & Tunmer, 1993; Neal & Kelly, 1999; Pinnell et al, 1994; Schwartz, 2005; Pinnell, 1989; Quay et al, 2001; Rodgers et al, 2004 and 2005). In addition, Reading Recovery produced the largest impacts on student reading skills of any intervention reviewed by the What Works Clearinghouse (WWC, 2008), making it one of the most promising reading interventions for scale up.

Purpose / Objective / Research Question / Focus of Study:
While prior research on Reading Recovery shows that the program’s impacts on student achievement are often large, research also suggests that there is substantial variability in impacts, and that much of this can be attributed to variation in program implementation. The evaluation design for the i3-funded scale-up of Reading Recovery includes a rigorous mixed-methods research design which will support strong causal inferences about program impacts, along with rich descriptions of program implementation and analysis of individual and contextual factors related to variation in program impacts when implemented at scale.

Setting:
This i3 scale-up evaluation involves several hundred elementary schools across the nation.

Population / Participants / Subjects:
Over 1,000 low performing first-grade students from more than 400 schools participated in this study.

Intervention / Program / Practice:
Reading Recovery involves intense one-on-one reading instruction provided to the lowest-achieving first graders in a school. These students receive 12- to 20-week cycles of daily, 30-minute, one-on-one RR sessions. Student progress is monitored daily to ensure that instruction is responsive to changes in student achievement and needs.

Significance / Novelty of study:
The size and scope of the sample, combined with a rigorous multi-site impact evaluation design make this study quite unique. The multi-level design allows for calculation of separate impact estimates for each school, combined with cross-level interactions using school-level factors to explain variation in treatment effects.

Statistical, Measurement, or Econometric Model:
Short-term impacts on students’ reading performance will be estimated by comparing mid-year reading achievement of students randomly assigned to participate in reading recovery at the beginning of first
grade to students randomly assigned to the control condition. Using a three-level hierarchical linear model (HLM) (Raudenbush & Bryk, 2002), the performance of the treatment and control students in each block (i.e., matched pair) will be compared, with blocks nested within each participating school. This HLM will include the pretest scores as a covariate, fixed effects for years (after year-one) at the student level*, along with random effects for blocks, a random effect for overall school performance (i.e., random school intercepts), and a random effect for the impact of Reading Recovery (i.e., random treatment effects across schools). The primary impact analyses will utilize the overall reading composite score from the Iowa Tests of Basic Skills (ITBS) as the posttest outcome measure and the average of stanine scores from the six subtests of the Observation Survey of Early Literacy Achievement (OS) as the pretest covariate.

For longer-term effects based on the regression discontinuity design, program impacts will be estimated by comparing performance of students below and above the original cutoff score for Reading Recovery eligibility. Because there will be a small amount of variability in schools’ cutoff values, the generalizability of results beyond students near a single cutoff score is enhanced. Again using multilevel statistical models, the performance of students above and below the cutoff score will be compared, with students nested within each participating school. This multilevel design will include the pretest assignment variable (i.e., OS scores) as covariates and fixed effects for years (after year-one) at the student level, along with a random effect for overall school performance (i.e., a random school intercept) and a random effect for the impact of Reading Recovery (i.e., a random treatment effect across schools).

**Research Design:**

This impact study involves two parallel designs—a randomized experiment and a regression discontinuity quasi-experiment, both of which are multi-site with thousands of students as the units of assignment, nested within hundreds of schools.

The enormous sample size for these multi-site studies allows for additional school-level contextual analysis of factors associated with variability in program effects. The use of random treatment effects for schools in a multilevel modeling framework allows for estimation of cross-level interactions to explain variability in treatment effects. School-level data from both quantitative and qualitative sources are used as predictors of school-level variability in impact estimates, while student demographic variables are used to investigate variability in effects across subgroups of students.

Indicators of variation in program contexts, supports, and components derived from these data are also used to explore potential predictors of variation in program impacts in subsequent statistical analyses of student reading achievement scores.

* Note that each cohort of RCT students is pooled with other cohorts for a single impact estimate. The year dummies are included solely to capture any annual cohort characteristics.
Appendices

Not included in page count.

Appendix A. References


