Service-oriented Interventions' Evaluations as a Case Study in Replication

Session Organizer

Eric Hedberg, NORC at the University of Chicago
Senior Research Scientist
Hedberg-Eric@norc.org

Discussant

A. Brooks Bowden, North Carolina State University
Assistant Professor
Brooks_Bowden@ncsu.edu

Overview

Service-oriented interventions can be defined as intervention programs or approaches that rely, in part or in whole, on volunteerism and other types of community service to deliver an intervention. Reading & Math, Inc. is an organization in Minnesota that utilizes AmeriCorps members (part of the Corporation of National and Community Service) to deliver, in part, reading and math tutoring services for students in need.

In 2014, an evaluation was conducted of the Minnesota Reading Corps to examine the impact on K-3 students using a multisite randomized design. This evaluation randomized students within schools using matched pairs and estimated growth curves based on weekly assessments during the first semester. Based on the estimated growth curves for Program and Control students, a treatment effect was estimated as the difference in end-of-semester scores. This evaluation found significant and large effect sizes.

However a major issue in education research is replication (as is well known, e.g., Easton 2018; Hedges 2018). As Hedges (2018) describes, this problem stems from a number of sources. In this symposium, we focus on *generalizability* by examining several studies that replicate an intervention evaluation across three dimensions: outcome, place, and time. One paper explores how the same intervention evaluated for reading outcomes was then evaluated for math. The next paper revisits the reading outcome in Wisconsin in addition to Minnesota. Finally, using quasi-experimental techniques, the last paper explores whether the program's effectiveness changed from one program to the other. Each of the papers is a form of replication on different dimensions: outcome, place, and time.

Papers

Service-Oriented Educational Interventions: Conceptual, Theoretical, and Empirical Foundations, provides an orientation to the intervention and sets the stage for the three other empirical papers, providing the details on the theoretical foundations that support the subsequent empirical work.

The second paper, *Service-Oriented Interventions in Math: An Extension and Replication*, provides an empirical study of the intervention approach towards a new outcome. This study used a randomized design and found positive effects on some key outcomes.

The third paper, *A Case Study in Geographic Replication: Impact Evaluation of the Minnesota and Wisconsin Reading Corps Programs*, provides the results of two randomized studies of the Reading Corps program, one in Minnesota and another in Wisconsin. This this paper, each studies sample not only

varied by geography, but also in the demographics of the sample. Positive effects were found in each study.

Finally, the fourth paper, *Replication across Time: A Quasi-Experimental Comparison of Two Minnesota Reading Corps Evaluations*, is unique in that it combines the samples from two independent evaluations of the same program to assess whether impacts were different from one study to the next. This cannot be accomplished by simply comparing study results, as the demographics of the samples differed. Thus, using 4-way propensity weighting, thus study compares the means of treatment and control, from each evaluation, using a balanced data.

References

Easton, J. Q. (2018). Improvement Not Evaluation: Using RCTs to Build Usable Knowledge in Education. *Journal of Research on Educational Effectiveness*, *11*(1), 1-21.

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*(1), 1-21.

Service-Oriented Educational Interventions: Conceptual, Theoretical, and Empirical Foundations

Peter Nelson, Serve Minnesota*
Director of Research and Innovation
peter@serveminnesota.org

Background

The practical implications of research are fundamental for applied psychological sciences. Yet the research-to-practice gap is a persistent problem in education and other fields (Cook & Odom, 2013). Advancing theory and empirical inquiry is the purview of researchers, whereas implementing feasible solutions in local contexts is the purview of practitioners. Historically, practitioners have been expected to act as informed consumers of research, overcoming implementation barriers to translate new ideas into applied settings (Carnine, 1997).

Yet a lack of improvement on state and national measures of academic achievement, including a recent worsening of national math scores (National Assessment of Educational Progress, 2018), suggests the traditional approach to translating research to practice is falling short. An increasing evidence-base has not been accompanied by steady improvement. The burgeoning study of implementation science bears testament to the need to better understand how effective practices can be implemented successfully (Norton, Lungeanu, Chambers, & Contractor, 2017). Service-based interventions hold considerable potential for bridging research and practice, but only if they can be demonstrated as effective.

Objective

Service-oriented interventions can be defined as intervention programs or approaches that rely, in part or in whole, on volunteerism and other types of community service to deliver an intervention. To be effective, service-oriented interventions must use evidence-based strategies or procedures designed to improve an educational skill or outcome (e.g., Gersten et al., 2009a, 2009b). It is also common for service-oriented interventions to incorporate data-driven decision-making protocols and specific implementation protocols (e.g., coaching). In this paper, we discuss the conceptual, theoretical, and empirical basis that establishes the potential of service-oriented interventions. Subsequent papers within the panel will provide in-depth examples of service-oriented educational interventions via evaluations that replicate and extend existing research.

Conceptual Foundation

Within the past several decades, prevention and remedial frameworks, often referred to as response to intervention or multi-tiered systems of support (hereafter referred to as 'RTI'), have grown widely popular (Jimerson, Burns, & VanDerHeyden, 2016). These frameworks assume that all students receive high-quality instruction and, for those who are struggling, evidence-based interventions delivered in small-group or one-on-one formats (Jimerson et al., 2016). Unfortunately, educators report little knowledge of such frameworks and limited availability of time and personnel to implement them (Bineham, Shelby, Pazey, & Yates, 2014; Castro-Villareal, Rodriguez, & Moore, 2014). These gaps in knowledge, time, and personnel are amplified by the fact that the evidence-based intervention assumption of RTI frameworks often equates to a dramatic increase in resource demand (as shown in Figure 1). Service-oriented interventions hold potential to augment school capacity for implementing RTI frameworks.

Resources for service-oriented interventions exist. Nationally, nearly six billion hours of service are provided across local communities each year (Corporation for National and Community Service, 2016). By organizing and training a small fraction of this vast human resource, schools could better support the needs of their students. Further, coordination between schools and community partners would be consistent with conclusions from implementation science that inter-organizational alignments are a defensible strategy for implementing evidence-based practices in education (Aarons, Hurlburt, & Horwitz, 2011).

Theoretical Foundation

Child development experts have long used an ecological systems theory (EST) perspective to advocate for inter-organizational alignment between schools and other systems that influence child outcomes (e.g., Gutkin, 2012). However, most EST-informed intervention efforts focus on supporting individual students across the school system and the home system (e.g., Sheridan et al., 2012). The co-location of community-based resources *within* a school system, for the purpose of augmenting the school's capacity to improve academic outcomes, represents a relatively novel extension of EST.

The EST perspective identifies service-oriented interventions as a source for resources from the proximal—or meso-systemic—space surrounding schools (see Figure 2), which can help the school improve its own systemic functions. This parallels in part how EST has been applied to solve student-level issues. With student-level issues, EST compelled educators to recognize that external factors—relative to internal factors—are more malleable and relevant for addressing student difficulties, and thus a more useful focal point for problem-solving (Gutkin, 2012). Similarly, although school systems can change and improve, their internal capacity to readily and effectively meet all student needs within an RTI framework may be limited relative to creating an inter-organizational alignment with resources from a community partner.

Empirical Foundation

The evidence-based interventions that are assumed in an RTI framework and implemented by service-oriented interventions have, by definition, robust research support. They are essentially tutoring, and recent best-evidence reviews identified tutoring as one of the most effective approaches for improving educational outcomes (Inns, Lake, Pelligrini, & Slavin, 2018; Pelligrini, Lake, Inns, & Slavin, 2018). Additional evidence specifically supports tutoring from service-oriented interventions. In literacy, for example, volunteer tutoring interventions were found to have a significant and positive effect on reading and writing outcomes (Ritter et al., 2009), and when volunteer tutoring ensured sufficient dosage outcomes approximated those of paraprofessional tutors or even teachers (Slavin, Lake, Davis, & Madden, 2011). Positive effects for service-oriented interventions are rarer for math, but positive findings for tutoring in math (Pelligrini et al., 2018) suggest service-oriented interventions hold potential across content areas. Finally, due to their reliance on volunteerism and other community-based resources, service-oriented interventions are a cost-effective approach for schools to provide tutoring (Hollands et al., 2016).

Summary

This symposium includes evaluations that add to the evidence-base for service-oriented interventions in several important ways. First, each evaluation was a form of replication research, which is a rare but

important approach to empirical inquiry (Schmidt, 2009). A previous evaluation found positive results for the current service-based intervention in prekindergarten, kindergarten, first, and third grade students, but also identified multiple areas for improvement (Markovitz, Hernandez, Hedberg, & Silberglitt, 2014; 2015). Second, each evaluation was essentially an effectiveness study. No aspect of implementation was directly supported by researchers. Third, one evaluation examined the impact of the service-oriented intervention as extended to a new learning outcome (i.e., math). Across studies, findings highlight both the promise of service-oriented interventions as well as directions for improvement and additional research.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 4-23.

Bineham, S. C., Shelby, L., Pazey, B. L., & Yates, J. R. (2014). Response to intervention: Perspectives of general and special education professionals. *Journal of School Leadership*, *24*, 230-252.

Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children*, *63*(4), 513-521.

Castro-Villarreal, F., Rodriguez, B. J., & Moore, S. (2014). Teachers' perceptions and attitudes about Response to Intervention (RTI) in their schools: A qualitative analysis. *Teaching and Teacher Education*, *40*, 104-112.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, *79*(2), 135-144.

Corporation for National and Community Service (2016). *Volunteering in America—Billions of Hours Served Translates to Huge Impact on our Nation.* Online at https://www.nationalservice.gov/blogs/2011-08-16/volunteering-america%E2%80%94billions-hours-served-translates-huge-impact-our-nation. Accessed 23 Dec 2016.

Gutkin, T. B. (2012). Ecological psychology: Replacing the medical model paradigm for school-based psychological and psychoeducational services. *Journal of Educational and Psychological Consultation*, *22*(1-2), 1-20.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009a). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Gersten, R., Compton, D., Connor, C., Cimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. (2009b). *Assisting students struggling with reading: Response to intervention and multitier intervention in the primary grades. A practice guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides.

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness*, *9*(1), 30-53.

Inns, A., Lake, C., Pellegrini, M., & Slavin, R. E. (2018). *Effective programs for struggling readers: A best-evidence synthesis.* Paper presentation at the 2018 Society for Research on Educational Effectiveness, Washington DC. Retrieved 15 June 2018 from https://www.sree.org/conferences/2018s/program/download/abstract/2361_0.pdf.

Jimerson, S. R., Burns, M.K. VanDerHeyden, A. M. (2016b). From response to intervention to mulit-tiered systems of support: Advances in the science and practice of assessment and intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 1-6). New York, NY: Springer.

Markovitz, C.; Hernandez, M.; Hedberg, E.; Silberglitt, B. (2014). *Impact Evaluation of the Minnesota Reading Corps K-3 Program*. NORC at the University of Chicago: Chicago, IL.

Markovitz, C.; Hernandez, M.; Hedberg, E.; Silberglitt, B. (2015). *Outcome Evaluation of the Minnesota Reading Corps PreK Program*. NORC at the University of Chicago: Chicago, IL.

National Center for Education Statistics. (2018). *2017 mathematics achievement levels, fourth grade. National Assessment of Educational Progress*. Retrieved from https://www.nationsreportcard.gov/math_2017/#nation/achievement?grade=4.

Norton, W. E., Lungeanu, A., Chambers, D. A., & Contractor, N. (2017). Mapping the growing discipline of dissemination and implementation science in health. *Scientometrics*, *112*(3), 1367-1390.

Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis.* Paper presentation at the 2018 Society for Research on Educational Effectiveness, Washington DC. Retrieved 15 June 2018 from https://www.sree.org/conferences/2018s/program/download/abstract/2371_0.pdf.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, *79*(1), 3-38

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90-100.

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Garbacz, S. A., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of conjoint behavioral consultation and the mediating role of the parent-teacher relationship. *School Psychology Review, 41*, 23-46.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1-26.

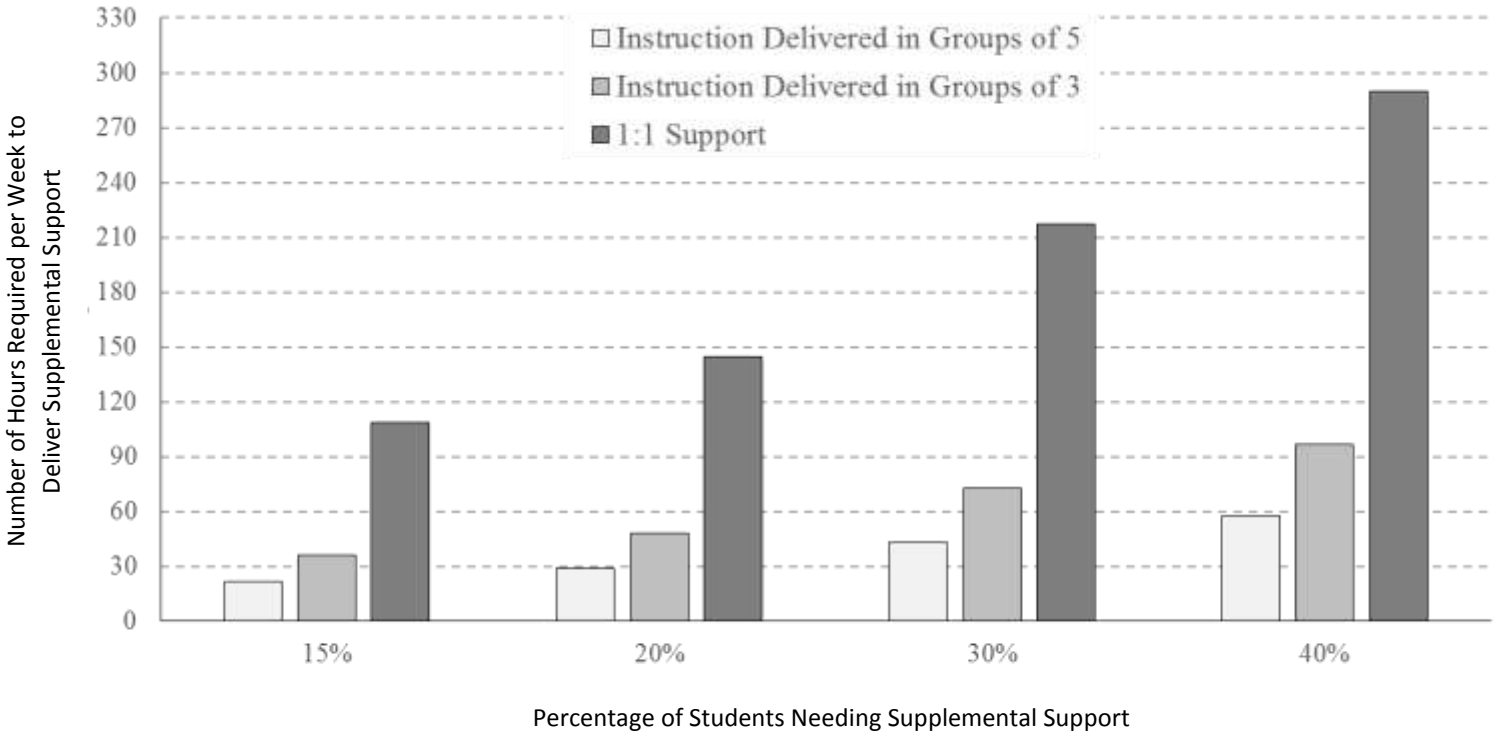

Figure 1. Impact on staff time based on demand for supplemental intervention resources. Assumes weekly support is consistent with evidence-based dosage recommendations.

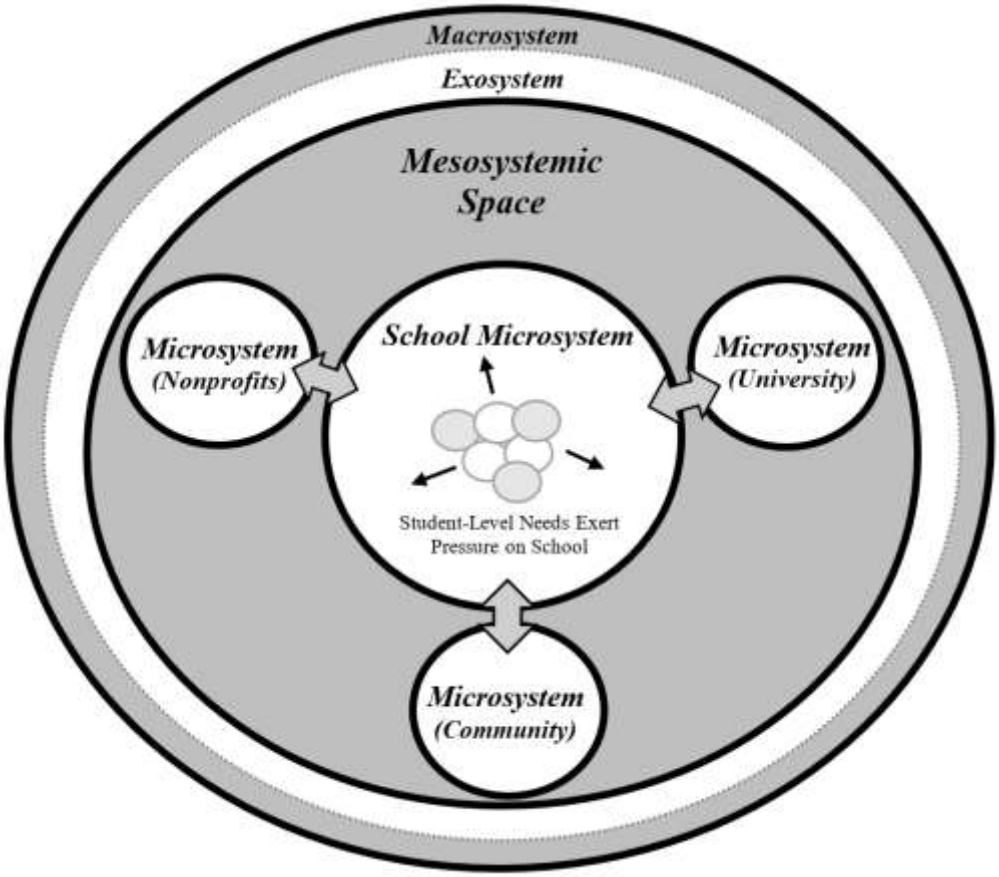

Figure 2. Systems-level orientation of EST.

Service-Oriented Interventions in Math: An Extension and Replication

David Parker, Serve Minnesota*
Vice President of Research and Development
david@serveminnesota.org

Background

A gap exists between observed math performance and proficiency goals within almost every state (National Center for Education Statistics, 2018), suggesting that schools struggle to implement effective instructional practices for at-risk students. Yet empirical guidelines exists to inform practice. For example, there is general consensus that supplemental support focused on whole and rational numbers has great potential to improve student math outcomes (NMAP, 2008; Siegler, Thompson, & Schneider, 2011), and expert panels and meta-analyses have identified instructional practices that support student learning (e.g., Gersten et al., 2009). However, logistical barriers often inhibit schools from aligning practice with research (Noell & Gansle, 2016). A growing evidence-base indicates service-oriented interventions can contribute to addressing this problem, but relatively little research has been conducted in math.

Objective

The purpose of this paper session is to discuss evaluation results from an AmeriCorps program—Math Corps—that coordinated local community members to provide a year of service as full-time math interventionists in their local schools. A recent randomized controlled trial (RCT) of the Math Corps program observed a statistically significant impact on math skills in a single semester (Parker et al., under review), but warranted replication with additional math outcomes aligned with the program's theory of change and implementation across the full academic year. The present study examined the degree to which students randomly assigned to receive Math Corps for a full year differed from students assigned to a control group across three outcomes: computational (math fact) fluency, performance on a computer adaptive test (STAR Math), and the state math proficiency test.

Setting

The evaluation occurred in 20 schools across Minnesota in the 2017-2018 school year (six middle schools, one intermediate school, and 13 elementary schools). Nine schools in the sample were surburban (45%), five were urban (25%), and six were rural (30%). All activities took place during the school day.

Population

Participants included 750 students in grades four through eight. Table 1 shows the distribution of students across grade, gender, and ethnicity. Across schools, 12% of students were identified as English Language Learners and 45% were eligible for free or reduced price lunch.

Intervention

The Math Corps intervention was implemented by 26 AmeriCorps interventionists. Intervention was provided to pairs of students for 90-min each week. All content across grades was directly related to whole and rational number understanding, and the sequence for each grade-level began with precursors to grade-level content and concluded with grade-level content. For each skill target (e.g., double-digit

subtraction), interventionists followed scripted, evidence-based protocols that focused on improving conceptual understanding (Butler, Miller, Crehan, Babbitt, & Pierce, 2003), computational proficiency (Codding et al., 2007), and word problem solving (Montague, Krawe, Enders, & Dietz, 2014).

Research Design

The Math Corps program used fall screening scores on STAR Math (Renaissance Learning, 2015) to identify potential participants, and scores below grade-specific criteria for proficiency denoted eligibility. In the current evaluation, eligible students in participating sites were randomly assigned into treatment (n = 484) or control (n = 266) group conditions. Students assigned to the control group were allowed to receive other school-based services but were not permitted to participate in the program. Analytic samples differed by outcome measure (see Table 2), and included up to approximately 25% attrition (on the fact fluency measure). When missing data exceeded 5% in any cell of the analytical data matrix, chi squared analyses evaluated whether missing data was related to demographic characteristics. The relationship between attrition and gender was non-significant, as was the relationship between attrition and race.

Data Collection and Analysis

All students completed STAR Math and the computational fluency test in the fall. In the spring, all students completed STAR Math, the state math proficiency test, and the fact fluency measure. Each measure had defensible technical characteristics as measured by conventional reliability and validity correlations (Foegen & Deno, 2001; Minnesota Department of Education, 2017; Renaissance Learning, 2015). To evaluate the extent to which students assigned to Math Corps demonstrated improved outcomes, simple regression models following an intent-to-treat (ITT) framework were used to regress post-test (spring) scores from each outcome measure on treatment assignment, pre-test scores, and a variable accounting for the inverse probability of condition assignment (Models A and B). A sensitivity analysis that incorporated trimming for outliers was also used to account for potential pre-test differences (Model C).

Results and Conclusion

Table 3 summarizes results from inferential analyses. A non-significant effect for Math Corps assignment was observed on the state proficiency outcome; however a significant effect was observed for both STAR Math and the fact fluency measure. The unstandardized effect for STAR Math was approximately 19 scaled score units, which given expected weekly growth rates of approximately 1.75 units (*SD* = 0.60) equates to nearly 11 weeks of additional growth. For fact fluency, the unstandardized effect associated with Math Corps ranged from 3.10 to 3.64 problems correct.

The current findings suggest that the Math Corps program produces a positive impact on key components of math skills for students in grades four through eight. However, the apparent lack of benefit on the state proficiency test suggests the program could consider additional practices to improve its impact on a broader set of math outcomes. Alternatively, program implementation factors such as content pacing and dosage could be examined. Despite mixed findings, the current study adds to an increasingly robust literature regarding the effectiveness of tutors serving as math interventionists (Pellegrini, Lake, Inns, & Slavin, 2018) and indicates potential for service-oriented math intervention programs.

References

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research & Practice, 18*(2), 99 –111. doi:10.1111/1540-5826.00066

Codding, R. S., Shiyko, M., Russo, M., Birch, S., Fanning, E., & Jaspen, D. (2007). Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? *Journal of School Psychology*, *45*(6), 603-617.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education*, *35*, 4-16.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Minnesota Department of Education. (2017). *Technical manual for Minnesota's MCA and MTAS assessments. Academic year 2016-17.* Roseville, MN: Author.

Montague, M., Krawec, J., Enders, C., & Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. *Journal of Educational Psychology*, *106*(2), 469-481.

National Center for Education Statistics. (2018). *2017 mathematics achievement levels, fourth grade. National Assessment of Educational Progress*. Retrieved from https://www.nationsreportcard.gov/math_2017/#nation/achievement?grade=4.

National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Math Advisory Panel.* U.S. Department of Education: Washington, DC.

Noell, G. H., & Gansle, K. A. (2016). Assuring the response to intervention process has substance: Assessing and supporting intervention implementation. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 407-420). New York, NY: Springer.

Parker, D.P, Nelson, P.M., Kaiser, P., Foegen, A., Kanive, R. Heisted, D., & Zaslofsky, A. (2018). Evaluation of a math intervention program implemented with community support. *Manuscript accepted, pending minor revisions, in Journal of Research for Educational Effectiveness.*

Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis.* Paper presentation at the 2018 Society for Research on Educational Effectiveness, Washington DC. Retrieved 15 June 2018 from https://www.sree.org/conferences/2018s/program/download/abstract/2371_0.pdf.

Renaissance Learning (2015). *STAR Math: Technical manual.* Wisconsin Rapids, WI: Author.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*, 273-296.

Table 1

Student Demographics

	Math Corps		Control		
Variable	n	%	n	%	X^2
Grade					3.33 (*df* = 4, *p* = .504)
4	78	17.4	44	16.8	
5	101	22.6	54	20.6	
6	130	29.1	79	30.2	
7	67	15.0	51	19.5	
8	71	15.9	34	13.0	
Gender					1.85 (*df* = 2, *p* = .396)
Male	218	48.8	115	43.9	
Female	215	48.1	136	51.9	
Ethnicity					4.29 (*df* = 6, *p* = .637)
White	236	52.8	139	53.1	
African American	110	24.6	61	23.3	
Hispanic	50	11.2	31	11.8	
Asian	38	8.5	18	6.9	
American Indian	4	0.9	3	1.1	
Other/Unknown	9	2.0	10	5.8	
Total	447		262		

Note: Demographic data for 41 students were unavailable.

Table 2

Descriptive outcomes by assessment period, condition, and analytic model.

		A: *MCA: No Adjustment* MC = 402; Control = 236		B: *MCA: Weighting* MC = 402; Control = 236		C: *MCA: Weighting and Trimming* MC = 377; Control = 234	
		M	SD	M	SD	M	SD
Post	Math Corps	41.85	8.71	41.60	8.41	42.15	8.54
	Control	41.43	8.34	41.89	8.68	41.62	8.34
		A: *STAR: No Adjustment* MC = 418; Control = 234		B: *STAR: Weighting* MC = 418; Control = 234		C: *STAR: Weighting and Trimming* MC = 389; Control = 231	
		M	SD	M	SD	M	SD
Pre	Math Corps	645.25	103.14	645.94	103.00	658.30	91.05
	Control	666.53	83.16	666.60	81.32	668.31	80.34
Post	Math Corps	723.33	89.00	723.27	88.65	728.95	91.05
	Control	717.51	81.49	717.52	81.33	718.26	80.45
		A: *Fact Fluency: No Adjustment* MC = 384; Control = 173		B: *Fact Fluency: Weighting* MC = 384; Control = 173		C: *Fact Fluency: Weighting and Trimming* MC = 355; Control = 171	
		M	SD	M	SD	M	SD
Pre	Math Corps	16.53	8.83	16.54	8.91	16.79	8.71
	Control	16.28	8.92	16.11	8.84	16.40	8.43
Post	Math Corps	22.73	9.40	22.78	9.45	23.08	9.32
	Control	19.13	8.45	19.38	8.97	19.18	8.48

Note: Sample sizes for specific outcomes varied as a function of missing data or, in the case of model C, trimming.

Table 3. *Inferential Results across Outcomes and Analytic Models*

	Model A: No Adjustment		*Model B:Weighting*		*Model C:Weighting and Trimming*	
MCA	B	SE	B	SE	B	SE
Intercept	27.18**	2.40	27.54**	2.19	26.97**	2.43
Fall STAR	0.02**	0.00	0.02**	0.00	0.02**	0.00
Math Corps	0.90	0.69	0.77	0.62	0.78	0.62
R^2		*.05*		*.05*		*.05*
MCA; no pre-test control	B	SE	B	SE	B	SE
Intercept	41.48**	0.56	41.60**	0.51	41.62**	0.50
Math Corps	0.36	0.70	0.30	0.64	0.53	0.64
R^2		*0.00*		*0.00*		*0.00*
STAR Math	B	SE	B	SE	B	SE
Intercept	302.81**	17.34	300.89**	15.67	2.33**	0.78
Fall STAR	0.62**	0.03	0.63**	0.02	0.67**	0.03
Math Corps	19.07**	5.11	18.67**	4.59	17.43**	4.46
R^2				*.48*		*.50*
Fact Fluency	B	SE	B	*SE*	B	*SE*
Intercept	8.05**	0.76	8.11**	0.69	7.86**	0.77
Fall Fact Fluency	0.62**	0.03	0.69**	0.03	0.69**	0.03
Math Corps	3.43**	0.64	3.10**	0.58	3.64**	0.64
R^2		*.44*		*.46*		*.45*

Note: Sample sizes for each model equal to those reported in Table 2. *** p < .01*.

A Case Study in Geographic Replication: Impact Evaluation of the Minnesota and Wisconsin Reading Corps Programs

Carrie E. Markovitz, NORC*
Principle Research Scientist
markovitz-carrie@norc.org

Marc W. Hernandez
Principle Research Scientist
hernandez-marc@norc.org

Eric C. Hedberg, NORC
Senior Research Scientist
Hedberg-Eric@norc.org

Background/Context

Reading Corps is the largest AmeriCorps tutoring program in the country, providing trained literacy tutors to implement evidence-based literacy instruction and assessment protocols for students in PreK through grade 3 in school-based settings. Beginning in 2003 with just 250 students in Minnesota, Reading Corps has since expanded to twelve states and Washington D.C. Reading Corps aims to broadly impact literacy outcomes for children. A primary goal for the program is that all children, age 3 to grade 3, who qualify for Reading Corps, will meet reading proficiency standards by third grade.

In the previous 2014 Minnesota Reading Corps study, *all* students (K-3) were followed for a single semester. The results showed positive and large effects for Kindergarten and first grade students. In this new evaluation we reevaluate the Minnesota program and a replication in Wisconsin.

Purpose/Objective/Research Question: Description of the focus of the research, including hypotheses.

The primary research question for the current evaluation is: *For Kindergarten and first grade students in Minnesota and Wisconsin, what is the impact of a single semester of the Minnesota/Wisconsin Reading Corps on program participants compared to similar students who did not receive Reading Corps?*

Setting

Minnesota

There are several hundred Minnesota Reading Corps schools operating in the state. Schools were selected by urbanicity using probability-proportional-to-size sampling (PPS[1]; 16 schools for first grade and 6 school for Kindergarten).

Wisconsin

Of the 12 Wisconsin Reading Corps schools operating in Milwaukee, ten of those schools participated in the study.

[1] PPS sampling allows for schools with larger numbers of eligible students to have a higher likelihood of being selected.

Population/Participants/Subjects

Minnesota

Table 1 presents descriptive statistics for the K-1 students included in the Minnesota evaluation. Demographics include gender, race/ethnicity, English Language Learner (ELL) status, and Fall benchmark scores.

Wisconsin

Table 2 presents descriptive statistics for the K-1 students included in the Wisconsin evaluation. The analytic sample size for the evaluation (N=176).

Intervention/Program/Practice

At the K-3 level, the intervention is focused on the "Big Five Ideas in Literacy" as identified by the National Reading Panel, including phonological awareness, phonics, fluency, vocabulary, and comprehension. AmeriCorps members serve as one-on-one tutors, working with approximately 15-18 K-3 students who scored below the grade-specific literacy benchmark (i.e., Tier 2 students) for 20 minutes each day. Tutors are trained to deliver ten research-based literacy interventions that align with the Big Five literacy targets.

Research Design

In each state, a randomized controlled trial (RCT) design was used to answer the research questions. Students were first sorted into pairs based on fall benchmark and then randomized within pairs to receive program or control.

Data Collection and Analysis

The primary data for the evaluation are comprised of the Fastbridge reading assessments. The evaluation team partnered with the program to utilize the Fall benchmark assessment data that AmeriCorps members collected on all K through 3rd grade students enrolled at the evaluation sites.

The analysis for the impact evaluation employed a mixed model to estimate the treatment effect for each outcome using data from specific grades individually. The model employs student covariates and their school means, including English learner status, gender, race, and most importantly, the Fall benchmark score. Estimates of effect size employ variance components from a model without controls.

Since our Minnesota sample of schools and students are not exhaustive of all eligible schools and students, we employed sampling weights at the school level and attrition adjustments at the student level to improve generalizability.

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

Minnesota (Table 3)

By the end of the first semester, Kindergarten students in the program correctly identified an average of 10.9 more letter sounds correctly in a one minute period than Kindergarten students in the control group (p=0.01, N=60; effect size = 0.85)

First grade students who received Reading Corps tutoring correctly identified in one minute an average of 16.3 more letter sounds embedded within non-real words than first grade control group students identified ($p < 0.001$, N=160; effect size 0.81).

Wisconsin (Table 4)

On average, program students correctly identified 6.5 more letter sounds in a one minute period than the control group after a single semester of the program ($p=0.035$, N=64; effect size = 0.55).

On average, program students correctly identified 8.7 more letter sounds embedded within non-real words than the control group ($p=0.004$, N=112; effect size =0.46).

Conclusions

In sum, the results of the 2018 Minnesota Reading Corps impact evaluation on Kindergarten and first grade students suggest that the program continues to produces large, meaningful effects for its younger students within a single school semester.

The results of the Wisconsin study also showed that Kindergarten and first grade students who received the Wisconsin Reading Corps program achieved significantly higher literacy assessment scores by the Winter benchmark than did control students who did not participate in the program.

An equally consequential result of this study for the broader Reading Corps program is the documentation of their successful replication of the program model in Wisconsin. The implications of this evaluation for research are critical. Much of the literature on the scaling of effective program models is theoretical in nature and mainly offers lessons learned and insights based on experience. Reading Corps has not only been shown through rigorous research to be an effective program in its original location, Minnesota, but now includes evidence of a successful replication of its program model in another state, Wisconsin.

Tables

Table 1. Student participants for the Minnesota Reading Corps Impact Evaluation (Fall 2017)

	Kindergarten (N=60)		1st Grade (N=160)		2nd Grade (N=190)[a]		3rd Grade (N=212)[a]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	52%	—	51%	—	54%	—	48%	—
Race/Ethnicity		—		—		—		—
White	27%	—	66%	—	62%	—	58%	—
Black	23%	—	11%	—	15%	—	18%	—
Asian	12%	—	7%	—	7%	—	6%	—
Hispanic/Latino	37%	—	10%	—	11%	—	12%	—
American Indian or AK Native	2%	—	3%	—	5%	—	4%	—
Multi-racial	—	—	3%	—	1%	—	1%	—
English Language Learner (ELL)	37%	—	31%	—	23%	—	24%	—
Fall Benchmark Score[b]	2.02	2.26	23.04	6.89	0.20	12.74	0.41	18.42

a: 2nd and 3rd grades were pooled for the analysis; Total N=402

b: Kindergarten = Letter sound fluency; 1st grade = Nonsense word fluency; 2nd and 3rd grades = Oral reading fluency

Table 2: Student participants for the Wisconsin Reading Corps Impact Evaluation (Fall 2017)

	Kindergarten (N=64)		1st Grade (N=112)	
	Mean	SD	Mean	SD
Female	45%	—	51%	—
Black	91%	—	79%	—
Asian	2%	—	9%	—
Hispanic/Latino	6%	—	9%	—
Multi-Racial	0%	—	1%	—
English Language Learner (ELL)	6%	—	10%	—
Free and Reduced Price Lunch (FRPL)[a]	82%	—	81%	—
Fall Benchmark Score[b]	2.19	2.48	17.76	10.36

a: School average

b: Kindergarten = Letter sound fluency; 1st Grade = Nonsense word fluency

Table 3: Results Minnesota K-3 Outcome Analysis

Grade and Outcome	Treatment effect [a,b]	p-value[a]	Effect size[c]	Prog. Mean[d]	Control mean[d]	Total N
Kindergarten						
Letter Sounds fluency	10.940	0.011	0.850	26.938	15.995	60
	(4.314)					
First grade						
Nonsense Words fluency	16.303	< 0.001	0.812	61.653	45.350	160
	(2.751)					

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary.

b: Standard errors in parentheses.

c: Effect size reported from unconditional model without control variables.

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: The Minnesota model includes sampling weights for generalizability.

TABLE 4: Results from Wisconsin K-1 Outcome Analysis

Grade and Outcome	Treatment effect [a,b]	p-value[a]	Effect size[c]	Prog. Mean[d]	Control mean[d]	Total N
Kindergarten						
Letter Sounds fluency	6.460	0.035	0.549	27.183	20.723	64
	(3.065)					
First grade						
Nonsense Words fluency	8.735	0.004	0.462	43.814	35.079	112
	(3.032)					

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary

b: Standard errors in parentheses

c: Effect size reported from unconditional model without control variables

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

Replication across Time: A Quasi-Experimental Comparison of Two Minnesota Reading Corps Evaluations

Eric C. Hedberg, NORC*
Senior Research Scientist
Hedberg-Eric@norc.org

Carrie E. Markovitz, NORC
Principle Research Scientist
markovitz-carrie@norc.org

Kelly Holt, NORC
Research Analyst
Holt-Kelly@norc.org

Background/Context

Reading Corps is the largest AmeriCorps tutoring program in the country, providing trained literacy tutors to implement evidence-based literacy instruction and assessment protocols for students in PreK through grade 3 in school-based settings. Over the last decade, there have been several hundred Minnesota Reading Corps schools operating in the state.

In the 2014 and 2018 Minnesota Reading Corps studies, Kindergarten and First Grade students were followed for a single semester. The results in both studies showed positive effects for Kindergarten and First Grade students.

Purpose/Objective/Research Question

The primary research question is whether the program was more or less efficacious in 2018 than it was in 2014 for Kindergartners and First Graders. In this study we estimate difference in treatment effects comparing the impact of the program in 2014 to the impact of the program in 2018.

Setting

Minnesota schools in each study were selected by urbanicity using probability-proportional-to-size sampling (PPS)[2] and agreed to participate in the study (16 schools for first grade and 6 schools for Kindergarten in 2018, 17 schools for first grade and 13 schools for Kindergarten in 2014[3]).

A difference in the population between the 2018 and 2014 studies is that the 2018 study included of only IAL grant recipient schools (i.e., rural and high-poverty urban).

Population/Participants/Subjects

[2] PPS sampling allows for schools with larger numbers of eligible students to have a higher likelihood of being selected.

[3] Four additional schools were selected in 2014 from St. Paul. However, St. Paul did not participate in the study in 2018 and are thus excluded from this analysis.

Tables 1 and 2 presents descriptive statistics for the K-1 students included in the 2014 and 2018 evaluations. Since the sample of students between studies is dissimilar, we employ propensity weights to balance the samples on key covariates.

Intervention/Program/Practice

At the K-1 level, the intervention is focused on the "Big Five Ideas in Literacy" as identified by the National Reading Panel, including phonological awareness, phonics, fluency, vocabulary, and comprehension. AmeriCorps members serve as one-on-one tutors, working with approximately 15-18 K-3 students who scored below the grade-specific literacy benchmark (i.e., Tier 2 students) for 20 minutes each day.

Research Design

We used quasi-experimental methods to assess whether program impact changed as the program changed (2014 to 2018) by examining the interaction between the randomly assigned treatment/control differences between non-randomly assigned cohorts of the program. To maximize the data utilized, we used propensity score weighting to weight the data based on key covariates to produce comparable groups.[4]

We note each of the four treatment groups using two indices: the first is the study index, noted as 0 for 2014 and 1 for 2018, and the second index is assignment to treatment or control, noted as 0 for control and 1 for treatment. Thus, we grouped the data as:

- Y(0,0), the control group in 2014,
- Y(0,1), the treatment group in 2014,
- Y(1,0), the control group in 2018, and
- Y(1,1), the treatment group in 2018.

This study included three tests for each grade, including the treatment effect among the matched 2014 sample using the propensity weights

$$\delta_{2014} = \bar{Y}(0,1) - \bar{Y}(0,0),$$

and the treatment effect among the matched 2018 sample using the propensity weights

$$\delta_{2018} = \bar{Y}(1,1) - \bar{Y}(1,0).$$

However, we were primarily concerned with the following difference in difference effect, which compared the treatment effect in 2018 with that of 2014,

$$\delta_{2018} - \delta_{2014} = [\bar{Y}(1,1) - \bar{Y}(1,0)] - [\bar{Y}(0,1) - \bar{Y}(0,0)].$$

Since we performed three tests from the same estimation procedure, we adjusted the p-values using the Bonferroni method of increasing the rigor of the tests proportional to the number of tests performed. We also clustered the standard errors by school membership.

Data Collection and Analysis

[4] Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. Journal of Econometrics, 155, 138–154.

For propensity weighting purposes, we created several covariates. First, we created an 8-valued group variable that was a combination or race, gender, and ELL status. We also used the Fall benchmark scores in two ways: using both a coarse 3-value quantile grouping variable and then, within those groups, the continuous score. These covariates were also included in the regression model that produced the weighted means.

We fit a multinomial probability model to model membership in each of the four groups using a set of key covariates, created inverse probability weights, and then calculated regression adjusted means to estimate treatment effects. Our weighted sample produced acceptable standardized differences (Table 3). This was executed using Stata's TEFFECTS IPWRA procedure.

Findings/Results

Table 4 presents the treatment effects for the propensity weighted 2014 and 2018 Kindergarten samples, as well as the difference between study samples' treatment effects. In comparing the treatment effects for the 2014 and 2018 study samples, the 2018 study impact differed by 0.28 fewer letter sounds than their 2014 matched counterparts (difference in difference = -0.28; SE=4.308). Given this small difference, there is no statistically significant difference between the sizes of the 2014 and 2018 study samples' impacts (p=1.000). Therefore, both studies demonstrate comparably large impacts on Kindergarten students in the program.

Table 5 displays the treatment effects for the propensity weighted 2014 and 2018 first grade samples, as well as the difference between study samples' treatment effects. There is a statistically significant difference between the 2014 and 2018 study samples' treatment effects, in which the 2018 study impact resulted in an increase of 9.1 nonsense words compared to their 2014 counterparts (difference in difference = 9.1; SE=3.5; p=0.027).

Conclusions: Description of conclusions limitations, and recommendations of authors.

The larger effect size for first grade students in the 2018 study may be due to several reasons. An obvious explanation is that the Reading Corps program's impact on students continues to improve as the program matures and the intervention model is further refined. However, another possible explanation is that the benchmark target scores, which help determine program eligibility and exit criteria, have changed between 2014 and 2018; thus, students may be staying in the program longer and possibly experiencing a larger impact of the program.

Tables

Table 1: Descriptive Statistics of the Kindergarten Students from each Study Group used in the Analysis

Characteristic	Control Mean	Control SD	Program Mean	Program SD
2014 Study Sample				
	N=39		*N=45*	
Male	64%	—	51%	—
non-White	72%	—	47%	—
ELL	10%	—	4%	—
Letter Sounds scores				
Baseline	2.46	2.00	2.78	1.92
Winter benchmark	23.08	16.71	37.36	20.09
2018 Study Sample				
	N=36		*N=31*	
Male	47%	—	55%	—
non-White	81%	—	71%	—
ELL	36%	—	39%	—
Letter Sounds scores				
Baseline	1.69	2.16	2.13	2.29
Winter benchmark	16.47	12.11	27.84	15.62

Table 2: Descriptive Statistics of the First Grade Students from each Study Group used in the Analysis

Characteristic	Control Mean	Control SD	Program Mean	Program SD
2014 Study Sample				
	N=86		*N=128*	
Male	59%	—	55%	—
non-White	47%	—	38%	—
ELL	12%	—	16%	—
Nonsense words scores				
Baseline	23.302	6.53	23.617	6.18
Winter benchmark	51.523	16.29	59.922	16.58
2018 Study Sample				
	N=68		*N=74*	
Male	51%	—	49%	—
non-White	37%	—	45%	—
ELL	21%	—	22%	—
Nonsense words scores				
Baseline	23.147	7.12	23.108	7.14
Winter benchmark	47.176	25.27	64.311	26.22

Table 3: Standardized Differences for 2014 and 2018 Minnesota Reading Corps Evaluations' Kindergarten Treatment Groups (Reference to 2014 Control)

Characteristic	Control		Program	
	Raw Std. Diff.	Weighted Std. Diff.	Raw Std. Diff.	Weighted Std. Diff.
2014 Study Sample				
White, male, non-ELL	Ref.		0.21	-0.02
non-White, female, non-ELL			-0.01	-0.02
non-White, female, ELL			-0.15	0.08
non-White, male, non-ELL			-0.41	0.03
non-White, male, ELL			-0.15	-0.07
Letter Sounds scores				
Baseline Quantile 2			0.01	-0.03
Baseline Quantile 3			0.08	-0.08
Baseline (continuous)			0.16	-0.08
2018 Study Sample				
White, male, non-ELL	-0.39	-0.03	0.04	0.01
non-White, female, non-ELL	-0.10	-0.01	-0.11	-0.04
non-White, female, ELL	0.51	-0.01	0.44	-0.03
non-White, male, non-ELL	-0.28	0.03	-0.57	0.08
non-White, male, ELL	0.30	0.04	0.44	0.05
Letter Sounds scores				
Baseline Quantile 2	-0.20	-0.06	-0.29	0.06
Baseline Quantile 3	-0.30	0.02	-0.08	-0.07
Baseline (continuous)	-0.37	0.06	-0.15	0.02

Table 4: Treatment Effects in Propensity Weighted Kindergarten Samples

Outcome	Treatment effect[a,b]	p-value[a,c]	Control Mean	Program Mean
2014 Study Sample				
Letter Sounds fluency	11.019	0.037	24.791	35.810
	(4.405)			
2018 Study Sample				
Letter Sounds fluency	10.735	0.004	18.080	28.815
	(3.304)			
Difference Between Study Samples' Effects (2018-2014)				
Letter Sounds fluency	-0.284	1.000		
	(4.308)			

a: Reported statistics from predictive model that includes the Fall assessment, gender, race, ELL status

b: Cluster robust standard errors in parentheses.

c: Bonferroni-adjusted p-values

Table 5: Treatment Effects in Propensity Weighted First Grade Samples

Outcome	Treatment effect[a,b]	p-value[a]	Control Mean	Program Mean
2014 Study Sample				
Nonsense Words fluency	8.986	0.002	50.939	59.925
	(2.667)			
2018 Study Sample				
Nonsense Words fluency	18.079	<0.001	46.990	65.069
	(2.006)			
Difference Between Study Samples (2018-2014)				
Nonsense Words fluency	9.093	0.027		
	(3.487)			

a: Reported statistics from predictive model that includes the Fall assessment, gender, race, ELL status

b: Cluster robust standard errors in parentheses

c: Bonferroni-adjusted p-values