Education and Urban Society

http://eus.sagepub.com/

The Effect of the Experience Corps® Program on Student Reading Outcomes

Yung Soo Lee, Nancy Morrow-Howell, Melissa Jonson-Reid and Stacey McCrary

Education and Urban Society published online 13 September 2010
DOI: 10.1177/0013124510381262

The online version of this article can be found at:
http://eus.sagepub.com/content/early/2010/06/16/0013124510381262

Published by:
SAGE
http://www.sagepublications.com

Education and Urban Society OnlineFirst, published on September 13, 2010 as doi:10.1177/0013124510381262

Education and Urban Society
XX(X) 1–22

DOI: 10.1177/0013124510381262
http://eus.sagepub.com
SAGE

The Effect of the Experience Corps® Program on Student Reading Outcomes

Yung Soo Lee[1], Nancy Morrow-Howell[1], Melissa Jonson-Reid[1], and Stacey McCrary[1]

Abstract

A randomized field trial involving 883 students at 23 schools in three urban cities assessed the effectiveness of Experience Corps® (EC), a program that places older adult volunteers in elementary schools to tutor students who are poor readers. Students were assessed at the beginning and end of the academic year with standardized reading measures. Program effects were analyzed using Generalized Estimating Equations (GEE) to adjust for clustering effects. Findings demonstrated that EC students made statistically greater gains over the academic year on passage comprehension and grade-specific reading skills. The gains were stronger for students who received at least 35 tutoring sessions. These findings indicate that older community volunteers can be effectively deployed to improve reading achievement in low income, ethnic minority children who are at risk of reading failure.

Keywords

tutoring, civic service, older volunteers

Since the federal No Child Left Behind law took effect in 2002, gains in reading achievement have been marginal (Dillon, 2007). For instance,

[1]Washington University in St. Louis, MO

Corresponding Author:
Yung Soo Lee, MA, George Warren Brown School of Social Work, Washington University in St Louis, Campus Box 1196, One Brookings Drive, St. Louis, MO 63130
Email: ylee@gwbmail.wustl.edu

preliminary results from an evaluation of Reading First, the nation's most comprehensive reading initiative that promotes evidence-based reading curricula and professional development for teachers, suggest that this intervention does not produce statistically significant impacts on reading comprehension for students in first through third grades (Gamse, Jacob, Horst, Boulay, & Unlu, 2008). The 2007 National Assessment of Educational Progress (NAEP) indicates that more than 30% of fourth-grade students are still reading below the basic level (Lee, Grigg, & Donahue, 2007). Achievement gaps exist between students of different races, ethnicities, and social classes. The NAEP, for example, shows that minority, low-income students have consistently performed poor, although there are fluctuations over time (Lee et al., 2007; Vanneman, Hamilton, Baldwin-Anderson, & Rahman, 2009). It is clear that reading achievement remains a national problem and that we are still searching for effective interventions, especially for those students who are at risk of reading failure.

Supplemental tutoring programs have long been implemented to improve reading abilities of young students, and the effectiveness of one-to-one tutoring has been consistently demonstrated by researchers (Cohen, Kulik, & Kulik, 1982; Elbaum, Vaughn, Hughes, & Moody, 2000; Wasik & Slavin, 1993). For example, a meta-analysis conducted by Elbaum and her colleagues (2000) showed that students who received one-to-one tutoring outperformed the comparison group with an average effect size of .41. Recent studies also identified a positive, longer term effect of one-to-one tutoring on at-risk beginning readers (Vadasy, Sanders, & Abbott, 2008) or on language-minority students (Ehri, Dreyer, Flugman, & Gross, 2007). Although tutoring programs implemented by certified teachers were found to be more effective (Wasik & Slavin, 1993), the cost of those programs may prohibit more at-risk children from accessing this help. Furthermore, due to comprehensive federal initiatives such as the America Reads Challenge Act (1997) and the No Child Left Behind Act (2001), there is a growing interest in tutoring programs using community volunteers as a cost-effective way to improve reading outcomes.

The present study evaluates the effect of the Experience Corps® (EC) program on student reading. The EC program is designed to bring older adult volunteers aged 55 and above into public elementary schools to provide tutoring for children at risk of reading failure. Currently, more than 2,000 EC tutors are serving approximately 20,000 students across 23 cities. This national program is growing as more nonprofit and public organizations seek to involve older adults in important community issues like public education. Thus, this program seeks to build its evidence base to guide program

dissemination and implementation. This study focuses on EC programs in three cities, New York, Boston, and Port Arthur, Texas.

Background

Effectiveness of One-to-One Tutoring Using Community Volunteers

Several studies have examined the effectiveness of various volunteer tutoring programs. Older reviews (Elbaum et al., 2000; Wasik, 1998) suggested that while evaluation methodologies were weak, tutoring interventions showed promising results. More recently, Ritter, Denny, Albin, Barnett, and Blankenship (2006) reviewed 21 studies that used randomized field trials to investigate the effectiveness of volunteer tutoring programs. The results of this meta-analysis revealed that the average effect size of volunteer tutoring on reading outcomes for elementary students was .30. The effects did not differ by types of tutors (e.g., parents, college students, or community volunteers) or grade levels of tutees.

More recent experimental studies on the effectiveness of volunteer tutoring programs showed a positive effect on beginning readers. Burns, Senesac, and Symington (2004), for example, evaluated the effectiveness of the Helping One Student to Succeed (HOSTS) program and found that at-risk students in the treatment group (n = 129) significantly gained over those in the control group (n = 127). The mean effect size was found to be .26 across reading measures. In addition, they documented longer term positive effects through a 2-year follow-up study (Burns, Senesac, & Silberglitt, 2008).

A few studies focused exclusively on the effectiveness of tutoring programs that use older adult volunteers (Meier & Invernizzi, 2001; Rebok et al., 2004) and found positive effects on reading outcomes. For example, Meier and Invernizzi (2001) evaluated the Book Buddies model that used older adult volunteers as tutors for first graders in a high-poverty urban setting. Results from this study demonstrated that the children who received 40 Book Buddies lessons were significantly stronger in reading skills than the comparison group. Rebok and his colleagues (2004) examined the effectiveness of the Baltimore EC program in a pilot field trial where elementary schools were randomized to the EC program. Students at the treatment schools made more improvement than students in the control schools on a state-wide measure of academic achievement assessed at the school level.

Contribution of This Study

In sum, a review of literature suggests that one-to-one tutoring programs using community volunteers produce positive effects on beginning readers at risk for reading failure. Although recent studies employed experimental designs, sample sizes were often small and reading outcomes were assessed through locally developed measures related to specific curricula. Furthermore, prior samples were drawn from single local programs rather than assessing a program model that spans several schools, districts, and cities. This study, in contrast, focuses on a national program that operates in multiple cities and school districts. The sample includes more than 20 schools in three cities in three different states. Also, reading outcomes were measured using standardized and normed instruments that did not correspond directly to any one school district or tutoring curriculum in attempt to assess a model that spans these program variations.

The EC program is unique in that it specifically targets older adults, rather than incorporating older adults into a multiage volunteer pool. As the U.S. population ages, the number of volunteers aged 65 and older is expected to increase 50% over the next few years (Corporation for National and Community Service, 2007), and there is increasing potential to engage this pool of volunteers in public education. With a growing emphasis on community service by older adults (Serve America Act, 2009), it is important to establish the effectiveness of programs that recruit this pool of volunteers.

EC Programs in Boston, New York, and Port Arthur

EC is a national program, sponsored at the city level by various nonprofit or public agencies. In Boston, Generations Incorporated, Inc., a nonprofit organization, operates the program and in Port Arthur, Texas, the program is hosted by the Southeast Texas Regional Planning Commission, a voluntary association of local governments. In New York, the Community Services Society operates the program. In all cases, districts and elementary schools choose to participate in the program. To be part of the EC national program, local sites adhere to program missions and standards, participate in national training and support activities, and receive assistance with fund-raising and expansion efforts. Although the tutoring programs at all EC schools share common elements, they vary from school district to school district, reflecting local preferences.

We chose programs in three cities—Boston, New York, and Port Arthur—to represent the commonalities as well as the diversity of EC programs across

the country. We also considered other elements in selecting these programs: each provided access to a diverse population of students, was large enough to ensure an adequate sample size, and was among the oldest of the EC programs, well established and expected to be stable over the evaluation period. The EC programs in these three cities differ in the specific curriculum used by the tutors, program size, and some other program features. However, they all share essential elements of the national EC program: the presence of a designated staff providing coordination between the volunteers and classroom teacher; training and monitoring provided to the tutors; and the presence of structure in the tutoring sessions (Wasik, 1998). The following description demonstrates that the essential elements are in place in these three program sites and that there are also differences at the local level.

EC staff members and coordination efforts. Designated coordinators are paid staff members who recruit, train, and monitor the tutors and coordinate the program within the schools. They meet regularly with teachers and principals to handle any concerns associated with integrating EC tutors into classroom routines. In New York, EC staff has daily meetings with the team of tutors in each school. In Boston, on-site coordinators are available during tutoring hours, and EC staff visit the site at least three times a week. In Port Arthur, EC staff monitors tutoring activity at least every other week or more if needed.

Training and monitoring. Older volunteers participating in EC are trained and monitored regularly. Across the three programs, initial training ranges from 15 to 32 hr before placement in the schools. The training focuses on employing a specific reading curriculum, building relationships with the children, and dealing with challenging behaviors.

The structured curriculum. The EC national model calls for each program to use a specific curriculum to structure tutoring sessions, but programs can use the curriculum suited to local preferences. The New York EC program uses the Book Buddies program which consists of reading, writing, and phonics and tutoring sessions include rereading familiar books, word study, writing, and reading a new book (Invernizzi, Juel, & Rosemary, 1997; Meier & Invernizzi, 2001). In Boston, Reading Coaches is used, where tutors use books that fit with student's ability and interest (Houston, 1998). During reading activities with students, tutors focus on building student's oral vocabulary and increasing reading comprehension by asking prediction questions, discussing, and writing about the story. The Brigance Inventory of Basic Skills and related study materials were used in Port Arthur to structure tutoring time (Torian, 1998). An initial assessment highlights areas of focus for the tutors, including word recognition, comprehension, and word analysis. Based on this

assessment, worksheets and activities from the Brigance materials are selected with guidance by the EC coordinators and teachers. More details about Book Buddies, Reading Coaches, and the Brigance Inventory and related materials are not presented here as this evaluation was concerned with the effectiveness of the EC program across sites and therefore across these curriculums. An essential element of effective tutoring programs (Wasik, 1998) and the EC model standard is the presence of a curricular structure for the tutoring session, and the reading programs used in New York, Boston, and Port Arthur meet that standard.

Based on these structured curricula, the EC tutors spend two to four sessions a week with the student, averaging about 30 to 40 min per session. EC staff coordinators track program participation, and attendance records, lesson plans, and daily lesson logs are kept for each student. During this evaluation, the average number of tutoring sessions over the academic year was 48 for New York, 35 for Boston, and 58 for Port Arthur. Across all three cities, 75% of the students received at least 35 sessions.

Method

Participants

Human subject approval was obtained from the university institutional review board and the school districts. A total of 23 schools that used the EC program in New York, Boston, and Port Arthur participated in the study. Of these, nine schools in Boston and six schools in New York participated in the study during the 2006-2007 school year, and eight schools in Port Arthur participated during the 2007-2008 school year. All of the elementary schools in Port Arthur and Boston implementing the core program (one-to-one tutoring with first through third graders) at the time of evaluation planning were included. In New York, 6 of the 16 EC schools participated. We included schools where the EC programs were administratively stable and that were geographically dispersed. Furthermore, we used district records to ensure that these EC schools were similar to the other EC schools in terms of school size, percentage of students who were free/reduced lunch, and teacher/student ratios (see Morrow-Howell, Jonson-Reid, McCrary, Lee, & Spitznagel, 2009 for more details).

The EC program relies on the professional judgment of teachers for the identification of students who need assistance with reading, and each EC program uses a classroom-specific referral form. A total of 1,100 students in need of assistance with reading were referred by teachers in these three cities

during the project period. Parental consent to participate in the evaluation was obtained on 81% of the referred students. Random assignment was conducted for these referred students within each school, and resulting sample size for the EC program was 430 and the control group was 453. All students were pretested at the beginning of the school year and 825 were located and reassessed at the end of the academic year.

Two observations were dropped due to incorrect birth dates because the reading assessment was standardized by birth date. Missing data were imputed using the Markov Chain Monte Carlo (MCMC) multiple imputation method. Therefore, the final sample size used in the analysis was 881, with 430 in the EC program group and 451 in the control group. The sample consists of first graders (41%), second graders (36%), and third graders (23%). Fifty-one percent of students were male and 49% were female. In terms of racial composition, African American students accounted for 58% and Hispanic students for 36%, with the remaining 6% representing non-African American, non-Hispanic races, such as White or Asian. A vast majority of students, about 94%, were enrolled in the free lunch program. In sum, most students participating in this study were ethnic minorities and of low socioeconomic status.

One-to-one tutoring was provided by 174 EC volunteers across three cities: 81 in Boston, 52 in New York, and 41 in Port Arthur. Although specific descriptive data on these 174 tutors are not available, we have descriptive data on tutors for all EC programs across the country. Of this sample, 87% were female, 38% were married, and the remainder was widowed or nonmarried. In terms of racial composition, 43% were White and 49% were African American. About 20% had high school diplomas, and more than 75% had at least some college education. The age of the tutors ranged from 50 to 93 years, with the average age being 65 years. Most commonly, the older volunteers lived in the neighborhood of the schools they served.

Data Collection

Data collection services were provided by Mathematica Policy Research, Inc. (MPR). Data on student demographics, reading, and behavioral measures came from three sources: interviews with students; surveys completed by teachers; and school records. MPR staff assessed reading ability at the beginning of the school year by interviewing students. Students were taken from the classroom at times approved by their teachers and completed 30-min face-to-face interviews. Surveys distributed to teachers at the beginning and end of the school year provided another measure of reading ability,

with teachers providing information on 84% of the students. At the end of the academic year, school personnel abstracted school records, capturing student demographics and school behavior.

Measures

Student Demographic and Behavioral Measures

Student demographic variables obtained from school records included gender, race, grade, site, special education status (the presence of an individualized education plan [IEP] in the school record), and limited English proficiency (LEP) status. These variables were dummy coded and included in the model as covariates. To measure student's classroom behavior, a modified version of the Sutter-Eyberg Student Behavior Inventory–Revised (SESBI-R) was included in the teacher survey. The SESBI-R is a brief teacher rating scale designed to measure disruptive behavior problems (Eyberg & Pincus, 1999). The SESBI-R, consisting of 38 items with a 7-point Likert-type scale, was modified for this study to include five additional questions focusing on positive student behavior. In addition, the scale was changed to a 5-point Likert-type scale to ease teacher burden of completing the scale. To ensure the validity and reliability of this revised measure, an exploratory factor analysis was conducted and a Cronbach's alpha calculated. A factor analysis with principal axis factoring indicated that the one-factor model was preferred, and interitem correlation was .97.

Reading Outcomes

To assess reading abilities of students, this study used three standardized reading tests as well as teacher assessment of grade-specific reading skills. Standardized measures were chosen that were independent of any particular tutoring curricula. Standardized reading tests included the Woodcock Johnson Word Attack (WJ-WA), the Woodcock Johnson Passage Comprehension (WJ-PC), and the Peabody Picture Vocabulary Test III (PPVT-III). The WJ-WA and the WJ-PC are subtests of the Woodcock-Johnson III Tests of Achievement (WJ-III ACH), designed to measure intellectual abilities and academic achievement with tests on written language, oral language, and academic knowledge (Gunn, Biglan, Smolkowski, & Ary, 2000; Woodcock, McGrew, & Mather, 2001). The WJ-WA subtest assesses phonemic awareness skills, and the WJ-PC subtest assesses overall skill at understanding text. The validity and reliability for the WJ-III ACH has been

found to meet or exceed the criterion (Woodcock et al., 2001). The PPVT-III, which is age normed for 2.5 years to 90+ years, measures receptive or hearing vocabulary for Standard American English and estimates verbal ability. Test-retest reliability, internal consistency reliability, and criterion-related validity have been established (Dunn & Dunn, 1997). The PPVT-III was chosen for this study in part because it had been used in a previous study of EC (Rebok et al., 2004).

Teachers' perceptions and expectations of student ability have been linked to later performance (Kuklinski & Weinstein, 2003). To assess teacher perceptions of the students' abilities, a measure of grade-specific reading skills was developed for the purpose of this evaluation. This measure was based on a scale developed to assess self-efficacy of young readers, and it was adapted for use by teachers in this evaluation (Chapman & Tunmer, 2003; Pajares, 2002). The measure consisted of 10 grade-and task-specific questions drawn from various curricula standards and reviewed by reading consultants at MPR. As students referred for tutoring are likely to be behind grade level, questions were chosen from one grade level below the student's current grade. For example, first grade teachers evaluated their students on skills like sounding out letters (a kindergarten-level skill) whereas second grade teachers were asked about sounding out a word (a first-grade-level skill). Each skill was assessed on a 4-point scale. The result of a factor analysis with principal axis scoring confirmed the unidimensionality of these measures at both pre- and posttest. Interitem correlation was .90 for the pretest and .92 for the posttest.

Teacher Perception of the EC Program

To examine how teachers perceived the EC program, we included two questions asking about the program in the teacher survey: (a) "To what extent do you agree with the following statement: the EC program is beneficial to the students that participate?" and (b) "How would you rate the level of burden to teachers of the EC program?" Four-point Likert-type scales were used to assess teacher perception of the EC program.

Analytic Procedures

About 7% of participating students dropped out during the tutoring intervention, and there were additional missing data ranging from 0% to 20%, mostly from variables in the teacher survey and school records. These missing data were imputed using the MCMC multiple imputation method (Rubin, 1987;

Schafer, 1997). This method is effective in handling dropouts in randomized field trials (Shaffer & Chinchilli, 2007) and produces unbiased estimates even with a large fraction of missing data (Schafer, 1997; Schafer & Olsen, 1998). Missing data were imputed separately for the treatment and control group, and five imputed data sets were created. Therefore, parameter estimates reported throughout this study are combined from the five imputed data sets.

To estimate the impact of the EC program on student's reading abilities, we used the following model for each reading outcome:

$$Y_{ij} = \beta_0 + \gamma_0 Y_{\text{-}1ij} + \beta_1 T_{ij} + \Sigma_k \gamma_{1k} X_{kij} + \varepsilon_{ij}$$

where, $Y_{\text{-}1ij}$ = the pretest score for student i from classroom j; T_{ij} = one if student i from classroom j is assigned to the EC program and zero otherwise; X_{kij} = the k other covariates for student i from classroom j, including gender, race, grade, site, classroom behavior, IEP and LEP status; β_0 and ε_{ij} represent intercept and error term, respectively. In the regression equation above, β_1 represents the program impact controlling for pretest scores and other covariates. Furthermore, to investigate whether any subgroups of students benefit more from the EC program than others, we added the interaction terms between program participation (T_{ij}) and other covariates (X_{kij}) to the model.

The data used in the current study have a clustered structure (i.e., students are clustered within classroom and schools). In these clustered data, outcomes of individuals within a cluster are likely to be correlated, and a failure to incorporate within-cluster correlations into the analytic model leads to incorrect coefficients and standard errors (Ballinger, 2004; Peters, Richards, Bankhead, Ades, & Sterne, 2003). Based on this notion, the regression model presented above was estimated using the generalized estimating equation (GEE) introduced by Liang and Zeger (1992). Among two possible clustering effects at the classroom and school levels, variance component analysis revealed that between-classroom variance was substantial and significant compared with between-school variance. We therefore considered the classroom-level clustering effect in the GEE model, and an *exchangeable* working correlation matrix was used to specify classroom-level clustering. GEE has been reported to produce asymptotically unbiased estimates given the clustered structure of data, especially with enough clusters. Furthermore, GEE does not require strict distributional assumptions, and estimates from GEE are robust to misspecification of a correlation structure within clusters (Feng, McLerran, & Grizzle, 1996; Peters et al., 2003; Zeger & Liang, 1992). Another popular analytic procedure for clustered data is the hierarchical

linear modeling (HLM). We analyzed program effects using HLM (where classroom-level random intercepts were allowed) to check robustness of our results, and the results were consistent in terms of both coefficients and significance levels (results of the HLM analysis available on request).[1] We reported effect sizes based on Hedge's G (Rosenthal & Rubin, 1986).[2]

Findings

Table 1 presents descriptive statistics for the sample at baseline. Randomization of the students into the EC program and the control group was effective in creating two equal groups in terms of main demographics and other variables, such as school absences and classroom behavior. Also, standardized reading scores of the two groups at baseline were equivalent. Reading scores of the students referred to the EC program were very low at baseline. On the WJ-PC, 92% were below the nation-wide mean, with 50% being one standard deviation below the nationwide mean and 12% being two standard deviations below. Similarly, 62% of the students scored one standard deviation below the nationwide mean on vocabulary (PPVT-III) and 20% were two standard deviations below this mean. This finding indicates that the students being referred to EC are being correctly identified for the program and are in need of reading assistance. One quarter of the students referred to the program spoke English as their second language and 14% were receiving special education services.

Table 2 presents our main findings on the impact of the EC tutoring program. Unadjusted posttest scores for each group are presented in the first column of the table, which indicate that EC students had higher posttest scores than students in the control group on three reading measures, WJ-WA, WJ-PC, and grade-specific reading skills. More importantly, these posttest scores were adjusted for pretest scores as well as other covariates, including gender, ethnicity, grade, program site, classroom behavior, and IEP and LEP status; and program impacts were estimated using GEE to adjust for the clustering effect across classrooms. According to the results presented in the second column of the table, on the WJ-PC and grade-specific reading skills, changes in EC students' scores were significantly greater than the changes in control group students' scores ($b = 1.52$, $p < .05$ for the WJ-PC; $b = 0.10$, $p < .01$ for grade-specific reading skills). Effect sizes associated with these gains were .13 and .16, respectively. For WJ-WA, the difference in adjusted posttest scores between the two groups was marginally significant ($b = 1.59$, $p = .07$), with an associated effect size of .10.

Table 1. Sample Description at Baseline

	Total (*N* = 881)	EC (*N* = 430)	Control (*N* = 451)
Reading outcomes			
WJ Word Attack (standardized score)	91.89 (20.45)	91.89 (20.31)	91.89 (20.61)
WJ Passage Comprehension (standardized score)	84.41 (13.72)	83.99 (13.59)	84.84 (13.85)
PPVT (standardized score)	80.95 (13.98)	80.87 (14.30)	81.03 (13.65)
Grade-specific reading skills	2.37 (0.59)	2.36 (0.57)	2.38 (0.62)
Demographics			
Gender			
Male	451 (51%)	209 (49%)	242 (54%)
Female	430 (49%)	221 (51%)	209 (46%)
Race			
African American	511 (58%)	259 (60%)	252 (56%)
Hispanic origin	321 (36%)	145 (34%)	176 (39%)
Others	49 (6%)	26 (6%)	23 (5%)
Grade			
1st grade	363 (41%)	180 (42%)	183 (40%)
2nd grade	318 (36%)	162 (38%)	156 (35%)
3rd grade	200 (23%)	88 (20%)	112 (25%)
Age	7.09 (1.10)	7.07 (1.07)	7.12 (1.14)
School events			
Free lunch			
Yes	829 (94%)	401 (93%)	428 (95%)
No	52 (6%)	29 (7%)	23 (5%)
IEP			
Yes	131 (15%)	64 (15%)	67 (15%)
No	750 (85%)	366 (85%)	384 (85%)
LEP			
Yes	207 (23%)	97 (23%)	110 (24%)
No	674 (77%)	333 (77%)	341 (76%)
Student behaviors			
Classroom behavior	3.58 (0.76)	3.57 (0.75)	3.59 (0.77)

Note: EC = Experience Corps®, WJ = Woodcock Johnson, PPVT = Peabody Picture Vocabulary Test, IEP = individualized education plan, LEP = limited English proficiency. None of the differences between the EC and control group are statistically significant.

Figure 1 graphically demonstrates the gains made by EC students compared with controls presented in Table 2. The dark bars represent posttest scores for the EC group adjusted by pretest scores, gender, grade, site, race,

Table 2. Unadjusted and Adjusted Posttest Reading Scores and Program Impact

	Unadjusted posttest mean		Adjusted posttest mean[a]			
Outcome variable	EC (N = 430)	Control (N = 451)	EC (N = 430)	Control (N = 451)	Program impact	Effect size
WJ Word Attack	95.67 (15.94)	94.36 (16.55)	95.79 [0.63]	94.20 [0.70]	1.59† [0.88]	0.10
WJ Passage Comprehension	88.40 (11.88)	87.30 (12.18)	88.69 [0.45]	87.17 [0.57]	1.52* [0.73]	0.13
PPVT	82.55 (12.93)	82.74 (12.26)	82.72 [0.39]	82.70 [0.35]	0.03 [0.51]	0.002
Grade-specific reading	2.75 (0.60)	2.66 (0.66)	2.76 [0.03]	2.66 [0.03]	0.10** [0.03]	0.16

Note: EC = Experience Corps®, WJ = Woodcock Johnson, PPVT = Peabody Picture Vocabulary Test. Numbers in the parenthesis are standard deviations and those in the bracket are standard errors.

a. Coefficients and standard errors were estimated with generalized estimating equations (GEE) to adjust for classroom-level clustering. Posttest mean was adjusted on pretest scores, gender, race, site, grade, individualized education plan, limited English proficiency, and classroom behavior.

†$p < .10$. *$p < .05$. **$p < .01$.

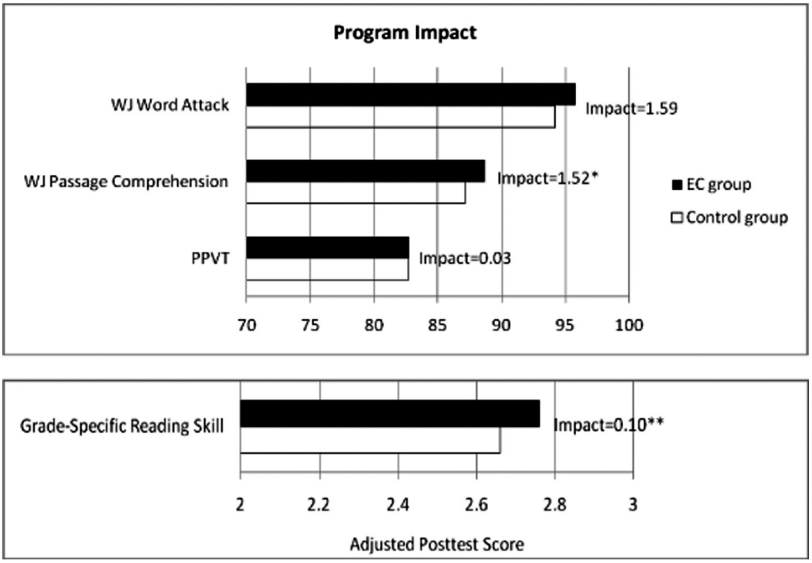

Figure 1. Adjusted posttest scores and the impact of EC tutoring
Note: Posttest scores are adjusted for pretest score, gender, site, race, grade, classroom behavior, and IEP and LEP status. Each dark bar represents adjusted posttest scores for the EC group and each light bar represents those for the control group.

classroom behavior, and IEP/LEP status. The light bars illustrate adjusted posttest scores for the control group and the difference represents the program impact of the EC tutoring program. Figure 1 shows that, over the school year, students in the EC group made more than 60% more progress on the WJ-WA and WJ-PC and 40% more gain on grade-specific reading skills.

To explore program effects based on dosage of the intervention, we used a subset of the sample to explore the students who received 35 or more tutoring sessions, which we considered the full intervention. EC students who had at least 35 sessions made greater gains than control group students on three of the four reading measures (WJ-WA, WJ-PC, and grade-specific reading skills). Effect sizes were .13, .17, and .17, larger than those on the full sample presented in Table 2.

Differential impacts of the EC program. To test whether pretest scores, gender, grade, ethnicity, site (city), classroom behavior, IEP status, or LEP status moderated program effects, we added interaction terms between EC participation and these variables into the model separately. The results show that EC was equally effective across gender, ethnicity, grade, classroom behavior, and English proficiency. However, there were two moderating conditions. EC students in New York made greater gains on the WJ-WA than those in Boston and Port Arthur. Furthermore, IEP students in the EC tutoring program did not benefit from the program as much as non-IEP students with regard to the WJ-PC.

Teacher perception of the EC program. Teachers were asked their overall impression of the program. More than 97% of the teachers agreed that EC was beneficial to their students. The majority (85%) of the teachers rated the EC program as no or low burden to them.

Discussion and Implications

Limitations of the current study need to be considered. First, we were unable to randomly select participating schools in New York, mainly because some schools were not in a position to participate due to administrative/financial uncertainties. (As described above, selection was not necessary in Port Arthur and Boston, where it was feasible to include all schools participating in the core EC program at the time of the evaluation.) Yet the New York EC schools participating in this study were similar to other New York EC schools in terms of school characteristics. A second limitation is that the teachers, in general, liked the EC program and thus may have been biased in the posttest assessment of grade-specific reading skills. Finally, we only tested short-term effects of EC over a single school year. It is important to determine whether gains of the program persist over time.

Despite these limitations, this study has several strengths. The findings from this study are based on a large-scale randomized field trial across multiple schools and school districts, enabling us to examine effects across sites that employed different curricula with different student populations. To produce unbiased estimates for the program impact, this study used GEE based on multiply imputed data sets.

Elementary school students who participated in the EC program made statistically greater improvements over the academic year on WJ-PC and on grade-specific reading skills. In addition, the gains on WJ-WA were marginally significant. Although program effects were detected in the full sample,

program effects were stronger for the subset of EC students who received at least 35 sessions.

As indicated by pretest standardized reading scores, students referred to the EC program were very poor readers and clearly in need of assistance. By virtue of the program's focus on large urban school districts, students participating in this study were mostly ethnic minorities and from families of low socioeconomic status and thus at higher risk of reading problems (Lee et al., 2007). The findings from this study suggest that this program effectively reaches vulnerable students and offers an intervention that can reduce reading disparities.

To place the impact of the EC program in context, we present effect sizes from studies of other reading interventions. Reading Recovery® (RR), a one-to-one intensive tutoring program employing certified teachers specifically trained in the intervention, reported effect sizes around .80 (Institute of Education Sciences, 2007). The Tennessee Star program, which reduced class size to improve academic achievement, reported an effect size of .26 (Mosteller, 1995; Nye, Hedges, & Konstantopoulos, 2000). Reading First, a national initiative that promotes the use of evidence-based reading curricula, provides professional development for teachers, and monitors reading progress through routine testing, did not produce a statistically significant impact on reading comprehension for students in first through third grades (Gamse et al., 2008).

The effect sizes reported in this study are smaller than those reported in other studies of volunteer tutoring programs (e.g., Burns et al., 2004; Ritter et al., 2006). However, it is important to note that this study sample spanned many schools and multiple districts and took place under realistic conditions. Also, we used standardized measures, which tend to produce smaller effect sizes than locally developed measures directly related to a specific curriculum (Cohen, Kulik, & Kulik, 1982; Elbaum et al., 2000; Slavin & Madden, 2008). The larger effect sizes mentioned above were often associated with use of certified teachers or actual changes in the size of classrooms. In this context, the magnitude of the reading improvements associated with the EC program are notable, given that the intervention is delivered by trained volunteers.

We did not find a significant program effect on the PPVT and there is no clear explanation for this finding. We selected this measure because it had been used in a pilot study of the EC program in Baltimore and it picked up a trend toward more improvement in the EC students (Rebok et al., 2004). Thus, we sought to build on this pilot work with a larger sample but we were

not able to replicate the finding. Perhaps, gains in vocabulary depend more on the specifics of the tutoring curriculum.

We did not find evidence that the program was differentially effective with various subgroups of students. This suggests that improvements in reading did not differ on several dimensions that put some children at increased risk—ethnic minority status, LEP, poor classroom behavior. Furthermore, the findings imply that it is not necessary to target on gender, ethnicity, grade, LEP, or classroom behavior to maximize program impact. However, findings do suggest that EC students participating in special education made less improvement. The program may benefit from reviewing its training and methods in working with students who have disabilities. The finding that the New York EC program had a greater effect on word attack is not surprising, given that the specific tutoring curriculum used at this site focused on phonetics (Invernizzi et al., 1997). This finding is a useful reminder that the tutoring curriculum selected at the program-site level can have differential impacts on certain reading skills. A review of all curricula used across the EC programs nationwide and their alignment with both program and school district goals may be useful.

Teachers overwhelmingly rated the EC program as beneficial to students and not burdensome for them. Although these results derived from only two questions about teachers' overall perceptions of the EC program, they are relevant to replication and program expansion. If teachers do not have positive perceptions of a program—despite objective measures of positive program outcomes—program effectiveness and sustainability are threatened.

The contribution of this study is its focus on a national program that exclusively relies on older volunteers who generally live in the local community of the school. The involvement of older community volunteers is significant for several reasons. First, the demographic changes in this society have doubled the size of the older population. The "Baby Boomers" will further increase this pool of volunteers (Freedman, 1999). There is ample evidence that these older adults will seek ways to contribute to their communities and that serving younger generations is a high priority (Peter D. Hart Research Associates, 2002). Thus, supplemental tutoring programs may be well served to target this type of volunteer. In addition, participation by older volunteers may have payoffs at the policy level. Older adults are known to be more active politically and more experienced in community affairs (Binstock, 2006). They are more likely than younger volunteers to become valuable advocates for the social causes that service programs connect them to (Morrow-Howell & Tang, 2007). EC volunteers self-report that participation in the program changes their attitude toward public education and that they are more likely

to advocate for and support funding for public education (Morrow-Howell, McCrary, Hong, & Blinne, 2008). Thus, the civic engagement of older adults in public education may have broad benefits for public schools.

Acknowledgments

The authors would like to acknowledge the staff members at Mathematica Policy Research for data collection services as well as the New York, Boston, and Port Arthur Experience Corps® staff members for their cooperation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed that they received the following support for their research and/or authorship of this article: This project was funded by The Atlantic Philanthropies.

Notes

1. The difference between GEE and HLM is in the treatment of between-cluster variation in the model. Whereas HLM explicitly models the between-cluster variation and incorporates this into calculating standard errors, GEE does not model the between-cluster variation; instead, GEE estimates the within-cluster similarity of the residuals and uses this correlation to reestimate the coefficients and standard errors (Hanley, Negassa, Edwardes, & Forrester, 2003). Thus, one of the disadvantages of GEE over HLM is that it produces no estimates of the variance–covariance structure between and within clusters (Cheong, Fotiu, & Raudenbush, 2001). This disadvantage diminishes in this study as our primary purpose is to produce unbiased coefficients and standard errors for program effects given the clustering effect and not to examine the structure of between and within clusters. As GEE is computationally efficient and does not impose strict distributional assumption, we presented GEE results in this study.
2. The methods of the current study were reviewed by the What Works Clearinghouse, Institute of Education Sciences, Department of Education and found to be consistent with the Clearinghouses' evidence standards (http://ies.ed.gov/ncee/wwc/pdf/quickreviews/expcorps_020910.pdf).

References

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7*(2), 127-150.

Binstock, R. (2006). Older people and political engagement. *Generations, 30*(4), 24-30.

Burns, M. K., Senesac, B. J., & Silberglitt, B. (2008). Longitudinal effect of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction, 47*, 27-37.

Burns, M. K., Senesac, B. J., & Symington, T. (2004). The effectiveness of the HOSTS program in improving the reading achievement of children at-risk for reading failure. *Reading Research and Instruction, 43*(2), 87-104.

Chapman, J. W., & Tunmer, W. E. (2003). Reading difficulties, reading-related self-perceptions, and strategies for overcoming negative self-beliefs. *Reading and Writing Quarterly, 19*, 5-24.

Cheong, Y. F., Fotiu, R. P., & Raudenbush, S. W. (2001). Efficiency and robustness of alternative estimators for two and three-level models: The case of NAEP. *Journal of Educational and Behavioral Statistics, 26*, 411-429.

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.

Corporation for National and Community Service. (2007). *The health benefits of volunteering: A review of recent research*. Washington, DC: Office of Research and Policy Development.

Dillon, S. (2007, September 26). Math score rise, but reading is mixed. *The New York Times*. Retrieved February 26, 2008, from http://www.nytimes.com/2007/09/26/education/26scores.html

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition: Manua*l. Circle Pines, MN: American Guidance Services.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal, 44*, 414-448.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary school students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605-619.

Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory–Revised*. Odessa, FL: Psychological Assessment Resources. Available from http://www.parinc.com

Feng, Z., McLerran, D., & Grizzle, J. (1996). A comparison of statistical methods for clustered data analysis with Gaussian error. *Statistics in Medicine, 15*, 1793-1806.

Freedman, M. (1999). *The kindness of strangers: Adult mentors, urban youth, and the new volunteerism*. New York: Cambridge University Press.

Gamse, B., Jacob, R., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading first impact study: Final report (NCEE 2009-4038)*. Washington, DC: National Center for

Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 34*(2), 90-103.

Hanley, J. A., Negassa, A., Edwardes, M. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology, 157*, 364-375.

Houston, M. (1998). *Reading coaches: Creating a session* (Basic Training series). Franklin, KY: Simpson County Schools.

Institute of Education Sciences. (2007). *What works clearinghouse intervention report: Reading recovery*. Washington, DC: U.S. Department of Education.

Invernizzi, M., Juel, C., & Rosemary, C. A. (1997). A community volunteer tutorial that works. *Reading Teacher, 50*, 304-311.

Kuklinski, M. R., & Weinstein, R. S. (2003). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72*, 1554-1578.

Lee, J., Grigg, W., & Donahue, P. (2007). The Nation's report card: Reading 2007 (NCES 2007-496). Washington, DC: U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics.

Liang, K. Y., & Zeger, S. L. (1992). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 12-22.

Meier, J., & Invernizzi, M. (2001). Book buddies in the Bronx: Testing a model for America Reads. *Journal of Education for Students Placed at Risk, 6*, 319-333.

Morrow-Howell, N., Jonson-Reid, M., McCrary, S., Lee, Y., & Spitznagel, E. (2009). Evaluation of Experience Corps: Student reading outcomes (CSD Research Report 09-01). St. Louis, MO: Washington University, Center for Social Development.

Morrow-Howell, N., McCrary, S., Hong, S.-I., & Blinne, W. (2008). *Experience Corps: The volunteer experience* (CSD Research Brief 08-31). St. Louis, MO: Washington University, Center for Social Development.

Morrow-Howell, N., & Tang, F. (2007). Youth service and elder service in comparative perspective. In A. McBride & M. Sherraden (Eds.), *Civic service worldwide* (pp. 157-180). New York: M. E. Sharpe.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. The Future of Children: Critical Issues for Children and Youths, *5*(2), 113-127.

Nye, B., Hedges, L., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*, 123-151.

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice, 41*, 116-126.

Peter D. Hart Research Associates. (2002). *New faces of retirement*. San Francisco: Civic Ventures.

Peters, T. J., Richards, S. H., Bankhead, C. R., Ades, A. E., & Sterne, J. A. (2003). Comparisons of methods for analyzing cluster randomized trials: An example involving a factorial design**.** *International Journal of Epidemiology, 32*, 840-846.

Rebok, G. W., Carlson, M. C., Glass, T. A., McGill, S., Hill, J., Wasik, B. A., et al. (2004). Short-term impact of Experience Corps participation on children and schools: Results from a pilot randomized trial. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 81*(1), 79-93.

Ritter, G. W., Denny, G. E., Albin, G. R., Barnett, J. H., & Blankenship, V. L. (2006). *The effectiveness of volunteer tutoring programs: A systematic review*. Campbell Collaboration Reviews of Intervention and Policy evaluations (C2-RIPE). Philadelphia, Pennsylvania: Campbell Collaboration.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400-406.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in survey.* New York: John Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* New York: Chapman and Hall.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.

Shaffer, M. L., & Chinchilli, V. M. (2007). Including multiple imputation in a sensitivity analysis for clinical trials with treatment failures. *Contemporary Clinical Trials, 28*, 130-137.

Slavin, R. E., & Madden, N. A. (2008, March 3-4). *Understanding bias due to measures inherent to treatments in systematic reviews in education*. Paper presented at the annual meetings of the Society for Research on Effective Education, Crystal City, VA.

Torian, S. (1998). Building blocks of a strong literacy project: Kansas City seniors for schools. Tutor, summer 1998, 1-4.

Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading, 12*(1), 51-89.

Vanneman, A., Hamilton, L., Baldwin-Anderson, J., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the national assessment of educational progress* (NCES 2009-455). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Wasik, B. A. (1998). Volunteer tutoring programs in reading: A review. *Reading Research Quarterly, 33*, 266-292.

Wasik, B. A., & Slavin, R. A. (1993). Preventing early reading failure with one-to-one tutoring: A review of five great programs. *Reading Research Quarterly, 28*, 179-200.

Woodcock, R. W., McGrew, K. S., & Mather, M. (2001). *Test review. Woodcock Johnson III Test*. Itasca, IL: Riverside. Retrieved October 1, 2008, from http://www.riverpub.com/products/wjIIIAchievement/scoring.html

Zeger, S. L., & Liang, K. Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine, 11*, 1825-1839.

Bios

Yung Soo Lee, MA, is a doctoral candidate at the George Warren Brown School of Social Work, Washington University in St. Louis, Missouri. His research interests include the effect of social policy on children's economic and educational outcomes and civic engagement among youth.

Nancy Morrow-Howell, PhD, MSW, is the Ralph and Muriel Pumphrey professor of social work at the George Warren Brown School of Social Work at Washington University in St. Louis, Missouri. Her research interests include productive and civic engagement in later life.

Melissa Jonson-Reid, PhD, MSW, is professor of social work and director of the Brown Center for Violence and Injury Prevention at the George Warren Brown School of Social Work at Washington University in St. Louis, Missouri. Her primary research interests include services and outcomes for abused and neglected children and school-based services for at-risk children.

Stacey McCrary, MSW, is a project manager of the Center for Social Development and Center for Mental Health Services Research at Washington University in St. Louis, Missouri. She is currently working on several projects including the Experience Corps® project.