1-10-2019

Impact of Experience Corps® Participation on Children's Academic Achievement and School Behavior

George W. Rebok

Jeanine M. Parisi

Jeremy S. Barron

Michelle C. Carlson

Ike Diibor

See next page for additional authors

Impact of Experience Corps® Participation on Children's Academic Achievement and School Behavior

Authors

George W. Rebok, Jeanine M. Parisi, Jeremy S. Barron, Michelle C. Carlson, Ike Diibor, Kevin D. Frick, Linda P. Fried, Tara L. Gruenewald, Jin Huang, Sylvia McGill, Christine M. Ramsey, William A. Romani, Teresa E. Seeman, Erwin Tan, Elizabeth K. Tanner, Li Xing, and Qian-Li Xue

HHS Public Access

Author manuscript

Prev Sci. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

Prev Sci. 2019 May ; 20(4): 478–487. doi:10.1007/s11121-018-0972-8.

Impact of Experience Corps® Participation on Children's Academic Achievement and School Behavior

George W. Rebok, PhD, MA[1,9], **Jeanine M. Parisi, PhD**[1], **Jeremy S. Barron, MD, MPH**[2], **Michelle C. Carlson, PhD**[1,9], **Ike Diibor, PhD**[3], **Kevin D. Frick, PhD**[4], **Linda P. Fried, MD, MPH**[5], **Tara L. Gruenewald, PhD**[6], **Jin Huang, MS**[9], **Sylvia McGill, BA**[7], **Christine M. Ramsey, PhD**[1], **William A. Romani, PhD**[8], **Teresa E. Seeman, PhD**[10], **Erwin Tan, MD**[11], **Elizabeth K. Tanner, PhD, RN**[9,12], **Li Xing, MS**[9], and **Qian-Li Xue, PhD**[2,9]

[1]Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

[2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

[3]Office of Assessment and Accountability, Baltimore City Public School System, Baltimore, MD

[4]The John Hopkins Carey Business School, Baltimore, MD

[5]Mailman School of Public Health, Columbia University, New York, NY

[6]Department of Psychology, Chapman University, Orange, CA

[7]Greater Homewood Community Corporation, Baltimore, MD

[8]AARP Experience Corps, Baltimore, MD

[9]Center on Aging and Health, Johns Hopkins University, Baltimore, MD

[10]The University of California at Los Angeles, Los Angeles, CA

[11]Policy, Research and International Affairs, AARP, Washington, DC

[12]Johns Hopkins University School of Nursing, Baltimore, MD

Abstract

This article reports on the impact of the Experience Corps® (EC) Baltimore program, an intergenerational, school-based program aimed at improving academic achievement and reducing disruptive school behavior in urban, elementary school students in Kindergarten through third

Address Correspondence to: George W. Rebok, PhD, MA, Department of Mental Health, Bloomberg School of Public Health, The Johns Hopkins University, 624 North Broadway, Room 891, Baltimore, MD 21205-1901, Phone: 410-955-8550, Fax: 410-955-9088, grebok1@jhu.edu.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest:

The authors declare that they have no conflicts of interest.

Ethical approval:

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent:

For this type of study formal consent is not required.

grade (K-3). Teams of adult volunteers aged 60 and older were placed in public schools, serving 15 hours or more per week, to perform meaningful and important roles to improve the educational outcomes of children and the health and well-being of volunteers. Findings indicate no significant impact of the EC program on standardized reading or mathematical achievement test scores among children in grades 1–3 exposed to the program. K-1st grade students in EC schools had fewer principal office referrals compared to K-1st grade students in matched control schools during their second year in the EC program; second graders in EC schools had fewer suspensions and expulsions than second graders in non-EC schools during their first year in the EC program. In general, both boys and girls appeared to benefit from the EC program in school behavior. The results suggest that a volunteer engagement program for older adults can be modestly effective for improving selective aspects of classroom behavior among elementary school students in under-resourced, urban schools, but there were no significant improvements in academic achievement. More work is needed to identify individual- and school-level factors that may help account for these results.

Keywords

academic achievement; school behavior; childhood education; early intervention; older adult volunteers

A growing number of studies have examined the potential benefits of volunteer tutoring and mentoring programs on children's educational outcomes. Some of these studies have focused exclusively on the effects of older adult volunteers on children's reading in a program called Experience Corps® (Gattis et al., 2010; Lee, Morrow-Howell, Johnson-Reid, & McCrary, 2012; Rebok et al., 2004). Experience Corps® (EC) is an intergenerational program that is designed to create generative and productive roles for older adults while simultaneously meeting unmet educational needs of children in grades K-3 (Fried et al., 2004, 2013). EC is a national program, now run by the AARP Foundation, and operates in 29 cities across the U.S. (AARP Foundation Experience Corps, 2016). Trained EC volunteers perform meaningful roles that are determined by the schools' principals and teachers as being critical unmet needs, including those related to reading, mathematics, and behavioral self-management. Within the context of the EC Baltimore program, the goal of the current study was to examine the impact of the social capital offered by an aging society on child outcomes, as indicated by performance on standardized achievement tests and school record information on behavior (e.g., principal office referrals, suspensions, and expulsions).

In an initial paper, Rebok and colleagues (2004) examined the effectiveness of the EC Baltimore program on children's reading in a pilot field trial involving over 1,100 students and six elementary schools randomized to the EC program or a non-EC wait-list control condition. They found that third-grade students in EC schools made larger gains in reading on a statewide performance measure than students in control schools. There also was a non-significant trend of improvement in alphabet recognition and vocabulary ability among Kindergarten children in the EC program. However, this early study did not account for the fact that students nested within schools or grades are not independent, and this dependence

may have generated a positive intraclass correlation, which typically inflates test values (e.g., t-values or F-values) and Type I error rate (Murray, 1998; Snijders & Bosker, 2012). More recently, Lee et al. (2012) reported results of a randomized field trial of EC involving 883 students at 23 schools in Boston, New York City, and Port Arthur, Texas, using multilevel analyses that incorporated within-cluster correlations into their analytic models. Students in the EC program made greater gains than nonEC students over the academic year on the Woodcock-Johnson passage comprehension subscale and on grade-specific reading skills like sounding out letters (Kindergarten) and words (1st grade) as assessed by their teachers, with effect sizes of 0.13 and 0.16, respectively, based on Hedges G (Rosenthal & Rubin, 1986). The gains were strongest for students who had received at least 35 one-on-one tutoring sessions over the course of one school year. The authors found no statistically significant effects on measures of vocabulary ability. Gattis et al. (2010) presented further evidence for the effects of EC in their study of the New York City EC program involving one-on-one tutoring, using analyses that adjusted for the correlation between students in a given classroom and in a given school. They reported significant improvements for the EC program in first and second graders' reading abilities using the Phonological Awareness Literacy Screening and the Early Childhood Literacy Assessment System measures compared to students not in the program.

Reading ability, a major focus of the EC program, is critical for success in school (Kern & Friedman, 2008) as it opens other learning opportunities and generalizes to academic success for children and youth. Early mathematical ability, another focal point of the EC program, also has been identified as a primary predictor of academic success later in school (Duncan & Murnane, 2011). However, to date, there have been few analyses of the effects of the EC program on mathematical abilities. In the present study we examined the impact of EC on standardized measures of reading and mathematical achievement.

The third major focus of the EC program is on behavioral management and classroom climate. Reducing aggressive/disruptive behavior and promoting appropriate pro-social behavior is a key area of concern given the rising tide of school violence in recent years (Grossman, Neckerman, Koepsell et al., 1997; Guerra, Huesmann, & Spindler, 2003). Early aggressive behavior and learning problems have been found to be correlated during childhood and have problematic consequences that reach far into the life course (Bradshaw, Schaeffer, Petras, & Ialongo, 2010; Bradshaw, Zmuda, Kellam, & Ialongo, 2009; Huesmann, Dubow, & Boxer, 2009). Within the context of Experience Corps, Rebok et al. (2004) reported that office referrals for classroom misbehavior decreased by about half in EC schools, but remained the same in control schools. However, analyses of the effects of the EC program on behavioral problems in elementary, school-aged children have been limited. Moreover, the perceived school climate of schools that adopted the EC program was significantly better in the first year that the program was implemented as compared to control schools (Parisi et al., 2015).

Early reviews suggested that volunteer tutoring and mentoring programs like Experience Corps showed considerable promise, but that they suffered from weak evaluation methodologies (Elbaum, Vaughn, Hughes, & Moody, 2000; Wasik, 1998). Although a few studies using randomized controlled designs support the effectiveness of school-based

tutoring and mentoring programs (e.g., Burns, Senesac, & Symington, 2004, Burns, Senesac, & Silberglitt, 2008), others have suggested, at best, small effects on academic, behavioral, and social-emotional outcomes (Herrera, Grossman, Kauh, & McMaken, 2011; Wheeler, Keller, & DuBois, 2010). For instance, in their meta-analysis, Ritter and colleagues (2006) reported that the average effect size of volunteer tutoring on reading outcomes for elementary school students was 0.30, a moderate effect. Whereas, findings from a meta-analysis conducted by Wood and Mayo-Wilson (2012) suggested effect sizes for academic achievement, behavioral outcomes, and school attendance were near zero. They did, however, find a small, positive effect for self-esteem. Other studies also have suggested small, positive effects on non-academic outcomes (e.g., students' connectedness to peers, self-esteem, social skills), but not for academic achievement (Karcher, 2008). Such mixed findings may be explained, in part, by the grade level examined, and type (group-based vs. one-on-one) and duration of the program. More recently, Markovitz and colleagues (2014) conducted a randomized controlled trial to evaluate the impact of the Minnesota Reading Corps (MRC) volunteer tutoring program on over 1,300 K-3 students at 23 participating schools. The MRC K-3 program provides one-on-one tutoring where members provide supplemental individualized literacy interventions. They found the largest impact at the end of the Fall semester (8 weeks) on literacy outcomes (letter sound fluency) among the youngest students (i.e., Kindergarten and first grade students) and the smallest effects among the oldest students (i.e., second and third grade students). Follow-up analyses showed that over a longer period of time in the Spring semester (16 weeks), the MRC program produced larger improvements in second- and third-grade students' oral reading fluency. Finally, Reading Partners uses community volunteers to provide one-on-one tutoring to struggling readers in under-resourced elementary schools (Jacob, Armstrong, & Willard, 2015). A recent evaluation of a randomized trial with public elementary school students in grades 2–5 comparing students participating in Reading Partners to students not participating in the program reported significant program-related improvements on all three measures of student reading proficiency, with effect sizes based on Cohen's d (Cohen, 1988) of 0.10 on reading comprehension, 0.09 on reading fluency, and 0.11 on sight reading, which is roughly equivalent to one and a half to two months of additional progress in the reading relative to the control group (Jacob et al., 2015).

In the present study, we report the results of a large, quasi-experimental study of the EC program in Baltimore City, in which volunteers aged 60 and older were randomized to EC or a low-activity control condition, and EC schools were matched to control schools on propensity scores (see Fried et al., 2013). The major aims of the study were to: 1) improve academic achievement of students in grades K-3, and 2) reduce suspensions, expulsions, and principal office referrals. We hypothesized that the direct impact that volunteers have on students resulting from face-to-face mentoring, tutoring, role-modeling, behavior management, and skill coaching will lead to improvements in academic achievement and school behavior.

Methods

Sample

During the course of the Baltimore EC study, a total of 25 Baltimore City public elementary schools were enrolled in four waves across the 2006–2007 to 2010–2011 academic years. For any given year, the total number of enrolled schools varied slightly, with an average of 19 schools participating per year (Fried et al., 2013; Parisi et al., 2015). Patterns of school participation in the Baltimore EC study across academic years are shown in the CONSORT flow diagram in Figure 1. All EC schools were successful in retaining a critical mass of volunteers (10 to 15 older adults) for the duration of program participation.

Selection of Control Schools: Matching Criteria

Control schools were identified via a propensity score matching approach to achieve balance between intervention and control arms with respect to the distribution of observed covariates and to maximize inference validity. The controls were selected from all eligible public elementary and elementary/middle schools in Baltimore City (n = 95) that had never participated in the EC program before or during the study period [2006–2011] and were not charter schools or academies (i.e., private or parochial schools). Title 1 status, elementary school-level attendance rates, student mobility, third-grade enrollment size, percent of students receiving free or reduced price lunch, and percent of African American students were selected as variables in the matching model because they might be related to both schools' assignment (EC vs. non-EC) and students' academic performance at the school level. When individually matching each EC school with a control school, the same base-year data from the control school were used to control for confounding time trends in the matching variables. Linear propensity scores were calculated from a logistic regression model of school assignment (EC vs. non-EC), including the above selected variables. To find the best matching control for each school in the intervention arm, we employed a matching method described by Rosenbaum and Rubin (1985): (a) for each EC school, we found a subset of potential control schools with linear propensity scores that differed from the score of the EC school by less than a specified caliper (0.25 standard deviations of the linear propensity score); (b) from this subset, we selected three schools with minimum difference from the EC school in the Stanford-10 normal curve equivalent (NCE) reading score at grade 2; (c) from these three schools we selected the final matching controls. For one EC school, six schools with a minimum difference from that school in the Stanford 10 NCE were selected because none of the first three was a good match for it. In this case, the fifth school was selected as the matching control. Covariates were compared between intervention schools and propensity-matched control schools to confirm the balance between these two groups at baseline. There were no significant differences between EC schools and control schools on the observed covariates ($p > .05$).

Measures

De-identified data on individual students were obtained through a memorandum of understanding from the Office of Achievement and Accountability (OAA) of the Baltimore City Public School System (BCPSS) for intervention and control schools, as well as city-wide. Academic achievement data were not available for Kindergarten students, but data on

principal office referrals and suspensions and expulsions did include Kindergartners. No outcome data were directly collected from students or teachers in the classroom.

Academic Achievement

Stanford Achievement Test Series, Tenth Edition **(Stanford-10, 1st and 2nd grade) (Harcourt Educational Measurement, 2002)—**The Stanford-10 is one of the most frequently used standardized achievement batteries in the United States, and was administered in spring semester of each school year to students in grades 1 and 2 in the BCPSS. The Stanford-10 is an untimed, norm-referenced, multiple-choice test of academic achievement and has been the tool used for meeting No Child Left Behind and national and state standards in academics. Subtests cover both reading (word study skills, reading vocabulary, reading comprehension) and mathematics (mathematics problem solving, mathematics procedures). The Reading section of the Stanford10 has an alpha reliability of . 87 and the Mathematics section has an alpha reliability of .80-.87. We report on individual-level subtest scores for students in grades 1 and 2.

Maryland School Assessment **(MSA, 2006; 3rd grade)—**The MSA was designed by Maryland educators for purposes of providing information that can be used to improve instruction in schools. It consists of criterion-referenced performance tests in reading, mathematics, writing, language usage, science, and social studies for students in grades 3, 5, and 8 and is administered each Spring. We report on individual-level MSA composite scores for students in grade 3.

Behavioral Performance (Kindergarten-3rd grade)

Principal Office Referrals—Data on principal office referrals for disciplinary reasons were obtained from OAA, but these data are not maintained as part of BCPSS's centralized database. The data were obtained annually from individual principals following a set of standardized record-keeping procedures. For a given experimental year and grade, the number of office referrals was counted for each student who was enrolled in K-grade 3 in either EC or control schools.

Suspensions and Expulsions—These data include suspensions and expulsions at the individual student level. Data on the number and type of suspensions and expulsions were collected for each student from BCPSS's centralized database.

Data Analysis

To address the effect of EC on students' academic performance, given data on academic performance are grade-specific (e.g., Stanford-10 Achievement Tests for 1st and 2nd grade; MSA for 3rd grade), we relied on intention-to-treat analysis to assess intervention effects of the EC program on study outcomes. First, we conducted analysis of academic outcomes by grade and by length of exposure (i.e., Year 1 or Year 2) across the first two years of the EC program in each school. For example, academic records for a school joining the EC study in 2007 would be compared to academic records from the same year in the matched control school. Multilevel models (Goldstein, 2012) (also termed linear random effects models (Laird & Ware, 1982)) were used to account for school-level clustering of student outcomes

via a random intercept for schools, which characterizes the between-school heterogeneity in the mean test score of Stanford-10 (or MSA) during the 1st year or 2nd year of the program. The models included fixed effects for intervention status, sex, and school-level average of the outcome at baseline. An interaction term between intervention status and sex was also added to examine sex-specific effects. The parameters of the model were estimated using the restricted maximum likelihood estimator; and the effect size of EC was calculated as the between-group difference (i.e., EC-Control) in the mean test score divided by the square root of the within-school variance estimated from the multilevel model. The Holm-Bonferroni method (Holm, 1979) was used to adjust for family-wise error in the analyses of subcomponent scores of the Stanford 10 and MSA test. All models were fit using SAS (version 9.4).

To address the effect of EC on students' behavioral outcomes, including principal office referrals, suspensions, and expulsions, separate models were created for Year 1 and Year 2 of the program and for each grade level examined (K-3rd grade). We chose the negative binomial regression model within the family of generalized linear models to account for the non-negative integer-valued behavioral outcomes (Agresti, 2002). We favored negative binomial over the conventional Poisson regression for such data in order to accommodate excess zeros in the outcome data, leading to over-dispersion evidenced by a significant likelihood ratio test (Agresti, 2002). The effect size was expressed as the ratio of the average incidence in the EC group to that of the controls, i.e., incidence rate ratio (IRR). To allow grade- and year-specific effect sizes, the model included all two-way and three-way interactions among dummy variables for intervention assignment, years of exposure, and grade. Multivariate WALD tests were used as a global test to assess the overall significance of the interaction terms. The analyses were also adjusted for sex. To account for effect modification by sex, in models where EC participation had a significant main effect, we tested a subsequent model with an interaction term between sex and EC participation. The school-level clustering was accounted for using Huber-White sandwich estimator of standard errors.

Results

Descriptive Characteristics of Students and Schools

The number of students enrolled in the schools by grade and sex across the first two years of the EC study is shown in Table 1. From grades K-3, there were a total of 6,495 students enrolled in EC schools and 6,461 students enrolled in control schools in the first year of the program. In the second year of the program, there were 5,026 students enrolled in EC schools and 5,068 students enrolled in control schools.

Table 2 presents characteristics of students and schools in the EC Baltimore study, by intervention group. As shown in the table, the groups were well matched at baseline with regard to student sex, race, third-grade school enrollment, and percent students receiving free/reduced lunch.

Academic Achievement

Stanford Achievement Test (1st and 2nd grade)—We compared first- and second-grade Stanford-10 reading and mathematics performance scores on the Stanford Achievement Test. As shown in Tables 3a and 3b, there were no significant group differences for Stanford-10 reading or mathematics scores between EC and control schools after either one year or two years of the EC study (*p's* = 0.33 to 0.99; effect sizes = -0.09 to 0.09 for first grade; and *p's* = 0.09 to 0.99; effect sizes = -0.12 to 0.17 for second grade).

Maryland School Assessment (3rd grade)—We compared third-grade MSA reading and mathematics performance scores, by level (i.e., basic, proficient, or advanced) for EC and control schools. As shown in Table 3c, with one exception (EC < control for Mathematics Scale Score for year 2, $p = 0.027$; effect size = -0.027), there were no significant group differences for either reading or mathematics scores between EC and control schools after either one year or two years of the EC study (*p's* = 0.08 to 0.93; effect sizes = -0.27 to 0.06).

Behavioral Performance (Kindergarten-3rd grade)—The majority of students had zero principal office referrals for disciplinary reasons. This positive skewing was typical for all years and grades. As shown in Table 4, girls had fewer office referrals than boys in all grades and years of the program (*p's* < 0.001 to 0.03). There was a significant intervention effect for office referrals for Kindergarten and 1st grade in the second year of the EC program. For a Kindergarten student in an EC school participating for the second year, their rate of office referral would be expected to decrease by a factor of 0.13 compared to a Kindergarten student in a comparable non-EC school during that same 2-year period ($p = 0.01$). Similarly, for a first-grade student in an EC school participating for the second year, their rate of referral would be expected to decrease by a factor of 0.21 compared to a first-grade student in a comparable non-EC school during that same 2-year period ($p = 0.02$). All two-way (i.e., intervention×year, intervention×grade, grade×year) and three-way (intervention×year×grade) interactions were statistically significant ($p<0.01$). There were no significant intervention-by-sex interactions for these models, indicating that sex did not moderate the effect of the EC intervention (all *p's* > 0.05, data not shown).

As shown in Table 5, girls had fewer suspensions and expulsions than boys in all grades and years of the program (*p's* < 0.001 to 0.02). Second graders in EC schools had fewer suspensions and expulsions during the first year of the EC study than second graders in control schools ($p = 0.03$). For a second-grade student, participating in the first year of EC is associated with a decrease in rate of suspension by a factor of 0.56. The two-way and three-way interactions as a whole were marginally significant ($p=0.051$). There was no evidence of effect modification by sex on rate of suspension (all *p's* >0.05, data not shown)

Discussion

Early literacy is highly important, as children who do not learn to read by third grade are at greater risk for failure at school, dropping out of school, and have limited occupational opportunities. The current study explored whether older community volunteers can be effectively deployed to improve reading achievement in low-income, mostly ethnic minority

children who are at risk of reading failure. Our findings are largely consistent with other reports demonstrating ineffectiveness of school-based mentoring programs for academic outcomes (Herrera, Grossman, Kauh, & McMaken, 2011; Wheeler, Keller, & DuBois, 2010). As such, we found no significant impact of the EC program on standardized reading achievement test scores among children in grades 1 to 3 exposed to the program. We also found no significant effects on mathematical achievement, which is not altogether surprising because Experience Corps is designed primarily as a literacy improvement program.

There are several possible reasons why we did not find the hypothesized effects of the EC program on student reading achievement in the present study. First, not every child in EC schools received volunteer mentoring and tutoring in reading since volunteers are assigned individually to their roles by the Principal, as determined to be needed and appropriate (Fried et al., 2013). For instance, some volunteers assisted in the computer lab, library, and art and music classes; although such activities are incorporated into the school day they did not directly target the academic or behavioral outcomes measured in the present study. Also, the study design did not ensure that all children who did receive mentoring and tutoring in reading got the same number of sessions since the goal was to assess the efficacy of the EC intervention under normative school conditions. It was also up to the discretion of the teacher how volunteers were deployed in their classrooms, with some reading to the entire class or small group and others working with individual children. In this respect, some children may have not been directly exposed to the volunteers, thus limiting the formation of bonds that can enhance outcomes by increasing readiness for learning and motivations/expectations regarding school and learning (Bayer, Grossman, & DuBois, 2015; Fried et al., 2013; Herrera et al., 2000; Rhodes & DuBois, 2008). Third, the standardized assessments of reading achievement (i.e., Stanford-10, Maryland School Assessment) were conducted in the Spring, 2–3 months before the end of the school year, so students in the EC schools did not have the benefit of maximal exposure to the EC program. Fourth, although randomization of schools to the EC program or the control condition was the gold standard to which we aspired, political and community realities in Baltimore City made it impossible to randomize schools, thus introducing possible sample bias.

Although we offer plausible explanations for our lack of significant findings, we acknowledge that it is also possible that there are no stable, reliable effects on academic outcomes resulting from this program. Another possibility is that the program would be effective if we increase its duration, standardize the curriculum, ensure all children receive the same level of exposure, offer additional opportunities for mentor-student interactions, and consider how contextual (e.g., school climate) and individual (e.g., race, motivation, emotional closeness to mentors) factors mediate or moderate responsiveness to intervention (Bayer, Grossman, & DuBois, 2015; McQuillin, Lyons, Clayton, & Anderson, 2018). In a previous study, Parisi et al (2015) reported that participation in the EC program improved overall school climate and several independent dimensions of school climate (e.g., learning environment, school safety, overall satisfaction, parental involvement and communication). This result is important given previous findings showing that even slight shifts toward a more positive school environment may contribute to a host of other academic and behavioral benefits for students (Gottfredson et al. 2005; Hoy & Hannum, 1997). Future analyses will

explore whether school climate moderates the impact of the EC program on the academic and behavioral outcomes reported here.

The findings partially support the usefulness of the EC program for improving behavioral management at school. Kindergarten and first-grade students in EC schools had fewer principal office referrals for disciplinary reasons compared to Kindergarten and first-grade students in non-EC matched control schools during the second year of the EC program. Girls had fewer office referrals than boys in all grades and years in the EC program, but there were no significant intervention-by-sex interaction effects on office referrals. Second-grade students in EC schools had fewer suspensions and expulsions than second graders in control schools during the first year of the program. Girls had fewer suspensions and expulsions than boys in all grades and years in the EC program. Lack of significant effects for Kindergarten and grade one for rate of suspensions and expulsions may be due to the relatively lower numbers in the early elementary grades. There was no effect moderation for sex on suspensions and expulsions. In general, both boys and girls seemed to modestly benefit from the EC program in terms of selective aspects of school behavior. These findings are consistent with the findings of Parisi et al. (2015) that used results from a school-wide survey of school climate to show an improvement in the perceived school climate following the first year that Experience Corps was adopted by schools compared to control schools. Although the same perceived impact on school climate did not exist in year two of program implementation (Parisi et al., 2015), the current findings suggest that the EC program continues to contribute to a more positive school climate as is evidenced by a significant decrease in office referral rate for Kindergarten and 1st grade boys in the second year of the program.

Although findings of selective EC program effects on principal office referrals, suspensions, and expulsions are encouraging, we also need to acknowledge limitations in these measures that may affect their interpretation. It is possible that changes in administrative leadership at the schools could have had a large influence on these measures. Some principals, for example, have been known to tell their teachers only to send the most challenging students to the office, while others want students sent to the office for any sort of disruption. There also may be considerable variation in referral practices across schools. Although we carefully measured and monitored various aspects of fidelity with regard to program implementation (e.g., tracked weekly hours and volunteer activities, weekly team meetings, classroom observations), there were other administrative and contextual changes that we were unable to capture, including changes in school leadership and extent of administrative support for the program. Further research with a larger sample of schools, more comprehensive measurement of fidelity (especially at the contextual level), and closer examination of mediators and moderators of EC intervention effects on behavioral outcomes is needed (Parisi et al., 2015). We do not mean to overstate these findings, as we realize that the few significant associations with behavioral outcomes are in the minority of effects examined. We also acknowledge that the number of statistical tests conducted and sample-specific fluctuations in the data may account for the selective nature of the behavioral outcome effects reported here. Additional work is needed to determine the durability, replicability, and breadth of these effects.

In summary our findings indicate that the EC Baltimore program had statistically significant, but only modest and selective effects on students' school behavior outcomes. These results may highlight the potential usefulness of a volunteer engagement program for older adults for reducing behavioral problems among elementary school children in under-served, under-resourced urban schools. At the same time there is a need to better understand why we did not find the hypothesized effects of the EC program on student reading achievement, as well as how to increase effectiveness of school-based mentoring programs. To this end, there is a need for better understanding the natural variation in mentoring relationships and practices and for examining a broad range of outcomes to get a more descriptive account of why these programs are (or are not) effective, as well as to examine a host of potential outcomes (as opposed to large focus on academic achievements). The impact of the EC program on preventing later academic and behavioral problems, delinquency, and school drop-out also remains to be determined and is the subject of ongoing investigation.

Acknowledgments

Funding for this manuscript was provided by the National Institute on Aging under contract P30AG02133, the Harry and Jeanette Weinberg Foundation, Erickson Foundation, Baltimore City: Baltimore City Public School System, and state and federal grants from AmeriCorps. We would like to thank the Baltimore City Public School System for their efforts in helping us acquire and interpret the academic and behavioral outcomes data. The opinions expressed in this article are those of the authors and do not represent the official position of the National Institute on Aging.

Funding:

National Institute on Aging under contract P30-AG02133

Harry and Jeanette Weinberg Foundation

Erickson Foundation

Baltimore City: Baltimore City Public School System

AmeriCorps

References

AARP Experience Corps (2016). *Frequently asked questions*. Retrieved October 18, 2016. (http://www.aarp.org/experience-corps).

Agresti A (2002). Categorical data analysis. 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Bayer A, Grossman JB, & DuBois DL, (2015). Using volunteer mentors to improve the academic outcomes of underserved students: The role of relationships. Journal of Community Psychology, 43, 408–429

Bradshaw CP, Schaeffer CM, Petras H, & Ialongo N (2010). Predicting negative life outcomes from early aggressive-disruptive behavior trajectories: Gender differences in maladaptation across life domains. Journal of Youth and Adolescence, 39, 953–966. [PubMed: 19688587]

Bradshaw CP, Zmuda JH, Kellam SG, & Ialongo NS (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. Journal of Educational Psychology, 101, 926–937. [PubMed: 23766545]

Burns MK, Senesac BJ, & 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 MK, Senesac BJ, & 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, 87–104.

Cohen J (1988). Statistical power analysis for the behavioral sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Duncan GJ, & Murnane RJ (2011) (Eds.). Whither opportunity? Rising inequality, schools, and children's life chances. New York: Russell Sage Foundation.

Elbaum B, Vaughn S, Hughes MT, & Moody SW (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.

Fried LP, Carlson M, Freedman M, Frick KD, Glass TA, Hill J, . . . Zeger S (2004). A social model for health promotion for an aging population: Initial evidence on the Experience Corps® model. Journal of Urban Health, 81, 64–78. [PubMed: 15047786]

Fried LP, Carlson MC, McGill S, Seeman T, Xue Q-L, Frick K,. Rebok GW (2013). Experience Corps: A dual trial to promote the health of older adults and children's academic success. Contemporary Clinical Trials, 36, 1–13. [PubMed: 23680986]

Gattis MN, Morrow-Howell N, McCrary S, Lee M, Jonson-Reid M, McCoy H, . . . Invernizzi M (2010). Examining the effects of New York Experience Corps® Program on young readers. Literacy, Research and Instruction, 49, 299–314.

Goldstein H (2012). Multilevel statistical models, 4th *edition* UK: John Wiley & Sons, Ltd.

Gottfredson GD, Gottfredson DC, Payne AA, & Gottfredson NC (2005). School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. Journal of Research in Crime & Delinquency, 42, 412–444.

Grossman DC, Neckerman HJ, Koepsell TD, Liu P-Y, Asher KN, Beland K, … & Rivara FP (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. JAMA, 277, 1605–1611. [PubMed: 9168290]

Guerra NG, Huesmann LR, & Spindler A (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. Child Development, 74, 1561–1576. [PubMed: 14552414]

Herrera C, Grossman JB, Kauh TJ, & McMaken J (2011). Mentoring in schools: An impact study of Big Brothers Big Sisters school-based mentoring. Child Development, 82, 346–361. [PubMed: 21291446]

Herrera C, Sipe CL, McClanahan WS, with Arbreton AJA, & Pepper SK (2000). Mentoring school-age children: Relationship development in community-based and school-based programs. Philadelphia: Public/Private Ventures.

Holm S (1979). A simple sequential rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Hoy WK, & Hannum JW (1997). Middle school climate: An empirical assessment of organizational health and student achievement. Educational Administrative Quarterly, 33, 290–311.

Huesmann LR, Dubow EF, & Boxer P (2009). Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: Implications for the adolescent-limited and life-course persistent models. Aggressive Behavior, 35, 136–149. [PubMed: 19189380]

Jacob RT, Armstrong C, & Willard JA (2015). Mobilizing volunteer tutors to improve student literacy: Implementation, impacts, and costs of the Reading Partners Program. New York: MDRC.

Karcher MJ (2008). The study of mentoring in the learning environment (SMILE): A randomized evaluation of the effectiveness of school-based mentoring. Prevention Science, 9, 99–113. [PubMed: 18297400]

Kern ML, & Friedman HS (2008). Early educational milestones as predictors of lifelong academic achievement , midlife adjustment, and longevity. Journal of Applied Developmental Psychology, 30, 419–430. [PubMed: 19626128]

Laird NM, & Ware JH (1982). Random-effects models for longitudinal data. Biometrics, 38, 963–974. [PubMed: 7168798]

Lee YS, Morrow-Howell N, Jonson-Reid M, & McCrary S (2012). The effect of the Experience Corps ® program on student reading outcomes. Education and Urban Society, 44, 97–118.

Markowitz CE, Hernandez MW, Hedberg EC, & Silberglitt B (2014). Impact Evaluation of the Minnesota Reading Corps K-3 Program. NORC at the University of Chicago: Chicago, IL.

Maryland School Assessment. Maryland State Department of Education: 2006.

McQuillin SD, Lyons MD, Clayton RJ, & Anderson JR (2018). Assessing the impact of school-based mentoring: Common problems and solutions associated with evaluating nonprescriptive youth development programs. Applied Developmental Science, DOI: 10.1080/10888691.2018.1454837.

Murray DM (1998). Design and analysis of group-randomized trials. New York: Oxford University Press.

Parisi JM, Ramsey CM, Carlson MC, Xue Q-L, Huang J Romani WA, . . .Fried LP, Rebok GW (2015). Impact of Experience Corps® participation on school climate. Prevention Science, 16, 744–753. [PubMed: 25708453]

Rebok GW, Carlson MC, Glass TA, McGill S, Hill J, Wasik BA, . . . Rasmussen M (2004). Short-term impact of Experience Corps participation on children and schools: Results from a pilot randomized trial. Journal of Urban Health, 81, 79–93. [PubMed: 15047787]

Ritter GW, Denny GE, Albin GR, Barnett JH, & Blankenship VL (2006). The effectiveness of volunteer tutoring programs: A systematic review Campbell Collaboration Reviews of Intervention and Policy evaluations (C2-RIPE). Philadelphia, Pennsylvania: Campbell Collaboration.

Rhodes J, & DuBois D (2008). Mentoring relationships and programs for youth. Current Directions in Psychological Science, 17, 254–258.

Rosenbaum PR, & Rubin DB (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician, 39, 33–38.

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

Snijders TAB, & Bosker RJ (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling, second *edition* London: Sage Publishers.

Stanford Achievement Test Series. Harcourt Educational Measurement, 10th ed.; 2002 Wasik BA (1998). Volunteer tutoring programs in reading: A review Reading Research Quarterly, 33, 266–292.

Wheeler ME, Keller TE, & DuBois DL (2010). Review of three recent randomized trials of school-based mentoring: Making sense of mixed findings. Social policy report. Society for Research in Child Development, 24, 1–21. Retrieved from http://eric.ed.gov/?id_ED519242.

Wood S, & Mayo-Wilson E (2012). School-based mentoring for adolescents: A systematic review and meta-analysis. Research on Social Work Practice, 222, 257–269.

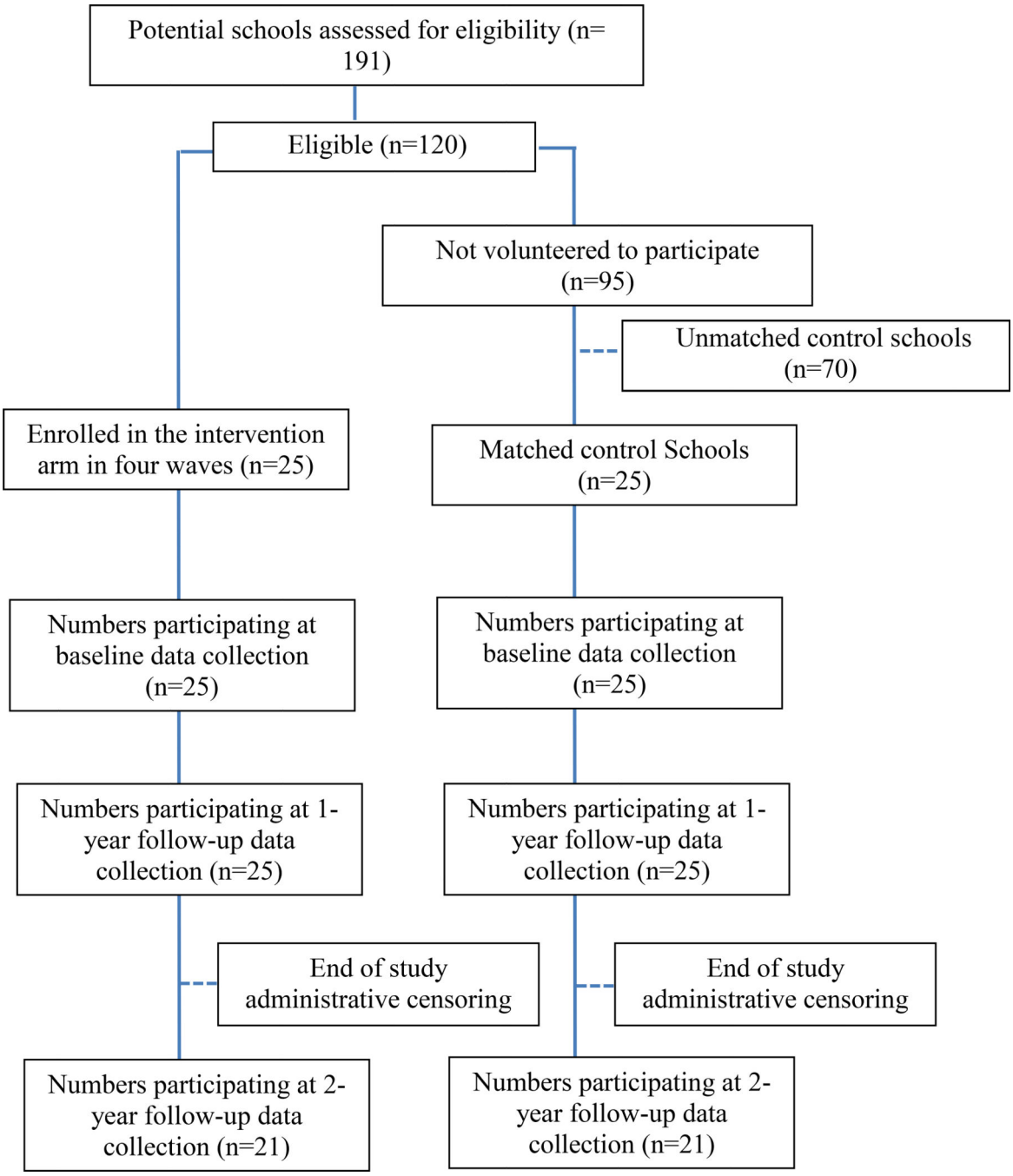

Figure 1:
CONSORT Flow Diagram of School Participation in the Baltimore Experience Corps Study

Table 1

Number of students enrolled in EC and control schools, by grade and sex

Grade	Sex	Year 1		Year 2	
		EC, N(%)	Control, N(%)	EC, N(%)	Control, N(%)
K	Male	824 (50.8)	735 (50.1)	637 (52.3)	585 (49.7)
	Female	799 (49.2)	732 (49.9)	582 (47.4)	592 (50.3)
1	Male	885 (52.0)	974 (53.6)	655 (51.2)	708 (51.7)
	Female	817 (48.0)	844 (46.4)	624 (48.8)	662 (48.3)
2	Male	892 (53.9)	808 (49.6)	642 (50.0)	703 (53.1)
	Female	763 (46.1)	822 (50.4)	643 (50.0)	621 (46.9)
3	Male	745 (49.2)	788 (51.0)	646 (52.0)	596 (49.8)
	Female	770 (50.8)	758 (49.0)	597 (48.0)	601 (50.2)

Table 2

Characteristics of students and schools in the EC Baltimore Study, by intervention group

	EC	Control
Sex[a], %		
Female	48.5	48.9
Male	51.5	51.1
Race, %		
African American	90.6	89.7
European American/other	9.4	10.3
Enrollment (number of third grade students per school)[b], mean (SD)	52.6 (23.0)	54.0 (15.3)
Free/reduced meals, %	82.3 (11.5)	83.5 (9.4)

[a]Results are from the first year after enrollment. All students enrolled in schools were included.

[b]The number of students in third grade was used in the matching algorithm.

Table 3a.

Effects of EC on Stanford-10 scores in first grade

Stanford scores	Year 1				Year 2			
	Effect size	B	95% CI	*P*	Effect size	β	95% CI	*p*
NCE Total Reading	0.08	1.44	−2.40, 5.28	0.4535	0.04	0.73	−3.14, 4.60	0.7044
NCE Word Study	0.03	0.57	−3.69, 4.83	0.7886	−0.01	−0.21	−3.99, 3.57	0.9095
NCE Sentence Reading	0.07	1.13	−2.03, 4.30	0.4751	0.02	0.32	−2.73, 3.37	0.8326
NCE Word Reading	0.09	1.61	−1.67, 4.89	0.3279	0.09	1.45	−1.75, 4.64	0.3656
NCE Reading Comprehension	0.08	1.47	−1.97, 4.91	0.3927	0.05	0.85	−3.33, 5.03	0.6827
NCE Total Math	0.02	0.31	−3.55, 4.17	0.8715	−0.09	−1.73	−7.63, 4.17	0.5565
NCE Math Problem Solving	0.02	0.30	−4.00, 4.60	0.8895	−0.07	−1.28	−6.95, 4.40	0.6515
NCE Math Procedures	−0.01	−0.15	−3.60, 3.31	0.9326	−0.09	−1.89	−7.33, 3.56	0.4874
Total Reading	0.08	3.46	−5.89, 12.81	0.4602	0.04	1.59	−7.73, 10.91	0.7314
Word Study	0.04	1.74	−9.23, 12.72	0.7504	−0.00	−0.24	−10.05, 9.58	0.9614
Sentence Reading	0.07	3.12	−5.97, 12.20	0.4934	0.02	1.04	−7.62, 9.71	0.8088
Word Reading	0.09	4.74	−5.12, 14.61	0.3380	0.09	4.38	−5.26, 14.02	0.3640
Reading Comprehension	0.08	3.76	−5.19, 12.70	0.4024	0.05	2.23	−8.83, 13.28	0.6860
Total Math	0.02	0.75	−6.80, 8.30	0.8425	−0.09	−3.28	−15.00, 8.44	0.5741
Math Problem Solving	0.01	0.50	−8.53, 9.54	0.9117	−0.06	−2.31	−13.95, 9.33	0.6903
Math Procedures	−0.00	−0.07	−7.70, 7.56	0.9862	−0.09	−3.97	−16.56, 8.62	0.5273

Table 3b.

Effects of EC on Stanford-10 scores in second grade

Stanford scores	Year 1				Year 2			
	Effect size	β	95% CI	*p*	Effect size	β	95% CI	*p*
NCE Total Reading	0.09	1.42	−2.14, 4.98	0.4262	−0.07	−1.16	−4.91, 2.59	0.5357
NCE Word Study	0.17	2.93	−0.50, 6.35	0.0918	−0.03	−0.57	−4.64, 3.50	0.7792
NCE Reading Vocabulary	0.09	1.58	−2.38, 5.54	0.4262	−0.01	−0.18	−3.64, 3.29	0.9190
NCE Reading Comprehension	0.03	0.41	−3.02, 3.83	0.8121	−0.12	−1.90	−5.38, 1.59	0.2776
NCE Total Math	0.10	1.86	−3.36, 7.09	0.4765	−0.00	−0.09	−5.17, 4.99	0.9710
NCE Math Problem Solving	0.10	1.88	−3.40, 7.16	0.4766	−0.01	−0.28	−5.21, 4.66	0.9107
NCE Math Procedures	0.07	1.31	−3.60, 6.22	0.5944	−0.00	−0.04	−4.72, 4.64	0.9858
Total Reading	0.09	2.98	−4.66, 10.62	0.4363	−0.07	−2.62	−10.70, 5.46	0.5160
Word Study	0.17	7.15	−1.29, 15.59	0.0948	−0.02	−0.94	−11.13, 9.25	0.8532
Reading Vocabulary	0.09	4.33	−6.22, 14.88	0.4129	−0.01	−0.70	−9.93, 8.53	0.8787
Reading Comprehension	0.03	0.92	−6.84, 8.67	0.8132	−0.12	−4.26	−12.20, 3.68	0.2843
Total Math	0.11	3.87	−6.61, 14.36	0.4607	−0.01	−0.20	−10.50, 10.09	0.9683
Math Problem Solving	0.11	4.12	−6.71, 14.96	0.4473	−0.02	−0.79	−11.01, 9.43	0.8770
Math Procedures	0.07	3.19	−8.80, 15.17	0.5951	0.00	0.14	−11.32, 11.60	0.9810

Table 3c.

Effects of EC on MSA scores in third grade

MSA scores	Year 1				Year 2			
	Effect size	β	95% CI	*p*	Effect size	β	95% CI	*p*
Reading Scale Score	0.01	0.40	−5.75, 6.55	0.8971	−0.19	−6.96	−14.88, 0.97	0.0835
Reading Performance Level	0.06	0.03	−0.06, 0.13	0.4991	−0.12	−0.06	−0.16, 0.04	0.2318
Mathematics Scale Score	0.01	0.34	−7.63, 8.31	0.9315	−0.27	−11.48	−21.58, −1.37	0.0271
Mathematics Performance Level	0.05	0.03	−0.10, 0.16	0.6311	−0.20	−0.12	−0.27, 0.03	0.1050

Table 4

Principal office referrals in EC vs. control schools in the first two years of the trial, by grade and sex

Outcome	Year in Experience Corps IRR (95% CI), p-value	
	Year 1	Year 2
Principal office referrals		
EC		
K	1.84 (0.39, 8.67), 0.44	0.13 (0.03, 0.59), 0.01
1	2.12 (0.36, 12.42), 0.41	0.21 (0.06, 0.79), 0.02
2	0.41 (0.09, 1.90), 0.25	0.72 (0.22, 2.33), 0.59
3	2.90 (0.79, 10.68), 0.11	1.02 (0.33, 3.16), 0.97
Sex		
K	0.07 (0.01, 0.66), 0.02	0.24 (0.07, 0.86), 0.03
1	0.19 (0.08, 0.45), <0.01	0.21 (0.06, 0.74), 0.02
2	0.27 (0.14, 0.50), <0.01	0.32 (0.16, 0.67), <0.01
3	0.19 (0.05, 0.76), 0.02	0.09 (0.03, 0.24), <0.01

Table 5.

Suspensions/expulsions in EC vs. control schools in the first two years of the trial, by grade and sex

Outcome	Year in Experience Corps IRR (CI), p-value	
	Year 1	Year 2
Suspensions/expulsions		
EC		
K	1.80 (0.38, 8.63), 0.46	0.92 (0.15, 5.78), 0.93
1	0.80 (0.41, 1.58), 0.53	0.84 (0.41, 1.72), 0.64
2	0.56 (0.32, 0.95), 0.03	1.27 (0.62, 2.61), 0.51
3	1.33 (0.70, 2.54), 0.38	1.24 (0.63, 2.43), 0.53
Sex		
K	0.07 (0.01, 0.68), 0.02	NA *
1	0.15 (0.09, 0.25), <0.01	0.11 (0.05, 0.22), <0.01
2	0.18 (0.12, 0.28), <0.01	0.17 (0.10, 0.30), <0.01
3	0.19 (0.11, 0.31), <0.01	0.20 (0.11, 0.37), <0.01

* not available because no girl in kindergarten was suspended in Year 2.