Chapter 6: After School Programs
Phillip B. Levine and David J. Zimmerman

1. Introduction

Support for investments in after-school programs is motivated by a few central concerns. Principally, there is a concern that a large number of children end their formal school-day activities and enter a period of time during which they are unsupervised until the time their parents return home from work. As Delaware U.S. Representative Michael Castle stated during a Congressional Hearing on March 11, 2008, "each afternoon, millions of students around the nation leave school with no place to go because they lack affordable, accessible, after-school opportunities."[1] Providing structured after-school programs to these "latch-key" children during this critical time period, it is argued, would enhance children's physical safety, discourage risky behavior, and – depending on the emphasis of the after-school program -- nurture various other desirable outcomes. These benefits might include improved academic outcomes, physical fitness, or artistic creativity.

From the perspective of this volume, after-school programs might then be viewed as a potential investment in poverty-reduction if the programs alter outcomes that either directly or indirectly improve the adult labor market outcomes of the participants.[2] Indeed, after-school programs may be regarded as an essential component of a policy framework promoting equal opportunity. As articulated in the Annual Report of The After-School Corporation – a significant funder of after-school programs in New York:

> *The resources that families with means treat as routine extensions of their kids' education – music lessons, sports, academic help – are out of reach for kids in broad swaths of the city. Kids on the wrong side of the opportunity gap face*

[1] http://edlabor.house.gov/hearings/ecese-2008-03-11.shtml
[2] After-school care may also impact parent's labor supply which may impact family income and hence, children's later outcomes. This link covered in another chapter in this volume.

> *limited possibilities to develop the talents, skills, breadth of learning that would prepare them for college and careers. (TASC, 2007).*

Obviously, the merit of this intervention strategy hinges on several key issues. The marginal benefit of participation will likely depend critically on the subset of the child population that elects to participate in the programs. The economic benefits are likely to be greater if participants come from high-risk families or environments than if they come from stable supportive families or environments. After-school programs, generally voluntary in nature, may have a limited effect if the alternative to a formal supervised after-school program is simply supervised care at home and not unsupervised self-care. Further, potential benefits will hinge on the types of programs provided and the impact they have on children's ability to achieve self-sufficiency later in life. Programs emphasizing recreational activities may foster self-esteem or physical fitness, but may have a less significant effect on academic performance. Lastly, it is critical to understand the cost of after-school programs so that the benefits per dollar spent might compared with other interventions competing for scarce funding. Does an investment in after-school care reap high returns in reducing poverty when the participants attain adulthood?

This chapter reviews the literature on after-school programs with an eye to offering advice on whether these programs are likely candidates for an effective anti-poverty program. The chapter is organized as follows: in the next section we discuss the motivation for interest in after-school programs. We then examine some of the main non-experimental evaluations that have been conducted on "flagship" after-school programs. Next we summarize the evidence on the key experimental evaluations that have been conducted. Finally, we'll discuss the implications of these findings and offer conclusions.

2. Background

Over the past three decades there has been a significant increase in female labor force participation. In 1975 just over half of women with children aged between 6 and 17 were active in the labor force. For women with children under the age of 6 the participation rate was just under 40%. By 2006, almost 80% of women with children aged 6-17 were active in the labor force and 63% of mothers with children six years or younger were working (Census Bureau, 2007). These trends, shown in Figure 1, have implications for the care of children when parents are at work. There is a large literature considering the provision of child care for younger children, but less research has been done on the impact of different child care arrangements for school-aged children (cf. Anderson in this volume). Data from the Survey on Income and Program Participation, administered by the Census Bureau, indicate that roughly 14% (or 5.2 million) of children between the ages of 5 and 14 were spending time in "self-care" on a regular basis during 2005 (Census Bureau, Who's Minding the Kids 2005). Rates of self-care rise with the age of the child and are highest for women who are windowed, separated, or divorced (19.5%) and for those employed full-time (18%). The distribution of regular child-care arrangements in shown in Figure 2.

Concerns over children being unsupervised during the after-school time period of roughly 3 p.m. to 6 p.m. are supported by data on the temporal incidence of crime. Figure 2 shows that violent crimes, sexual assault, and aggravated assault by juveniles show a distinct peak during the after-school hours (cf. Fox and Newman, 1997). Trends for non-students (not shown) do not show this peak. After-school care, when appropriately supervised, would then hold the potential to provide children with a safe environment during the after-school hours and reduce the odds

that they engage in various crimes. Beyond safety, after-school programs are often suggested as a way to help students improve their academic performance, reduce risky behavior such as drug use or sexual activity, or to enhance their social and emotional well-being (cf. Catalano et al., 2004; Durlak and Weissberg, 2007).

This suggests that a prima facie case can be made for the provision of after-school programs of some sort. The federal role in this endeavor comes primarily through the *21st Century Community Learning Centers (21CCLC)* program, which is the only federal funding source, directed solely at after-school programs. The program, which began in 1998, is described as follows by the Department of Education:

> *This program supports the creation of community learning centers that provide academic enrichment opportunities during non-school hours for children, particularly students who attend high-poverty and low-performing schools. The program helps students meet state and local student standards in core academic subjects, such as reading and math; offers students a broad array of enrichment activities that can complement their regular academic programs; and offers literacy and other educational services to the families of participating children. (ED.Gov)*

Funds from the 21CCLC program are granted on a competitive basis by the Department of Education to State Education Agencies, who then grant funds on a competitive basis to eligible organizations. Each state, therefore, funds a variety of programs with these funds.

The 21CCLC program began with $40 million in appropriations in 1998. Funding rose rapidly to $453 million in 2000 and then to approximately $1 billion in 2002. Funding has remained roughly steady in recent years, however pressure for reductions in spending have recently mounted. Figure 3 shows the trends in funding since the programs inception with

predicted appropriations inserted for 2009.[3] Beyond federal funding, after-school programs may receive funding from state and local governments, private foundations, businesses, and fees. Funding from a variety of these sources is common.

3. **Evaluations**

The Harvard Family Research Project provides a national database on a large number of evaluations of after-school programs.[4] Evaluations may be for a particular (singular) program or for a group of programs. Of the roughly 150 evaluations that are reviewed, only 9 programs that were classified as either "after-school", "comprehensive", or "mentoring" were evaluated using an experimental design.[5] The remainder utilized either a quasi or non-experimental framework for their evaluation. This highlights the fact that there is a limited research base from which to draw in forging an assessment of the efficacy of after-school programs.

Non-experimental evaluations

Non-experimental evaluations of particular programs typically contrast the outcomes of participants and "similar" non-participants. These comparisons may utilize a regression

[3] There are a few other federal programs that support after-school programs, though to a lesser extent. Snacks served at after-school programs may qualify for reimbursement from the USDA. The Child Care and Development Fund may be used by states to provide after-school care. Further, TANF funds may be used to support after-school programs if they meet the programs requirements. Government support for childcare, however, is minimal. Indeed, only 4.1% of children under the age of 15 received support from a federal, state, or local government agency, or a welfare office in 2002 (U.S. Census Bureau, 2002).

[4] http://www.hfrp.org/

[5] These programs include the 21st Century Learning Centers – national evaluation, Across Ages Program, Children's Aid Society Carrera-Model Teen Pregnancy Prevention Program, Louisiana State Youth Opportunities Unlimited Summer Program, Quantum Opportunities Program, Woodrock Youth Development Project, Big Brothers Big Sisters of America Program, and the Howard Street Tutoring Program. Some of these programs are significantly more comprehensive than most after-school programs or are only tangentially targeted at academic enrichment.

framework with "program participation" or "self-care" specified as an independent variable or may simply compare the outcomes of participants to a those of a set of non-participants who are selected for their similarity in terms of age, gender, prior-grades, etc. The principal empirical task facing these studies is selection bias: that is, participants and non-participants may differ in a myriad of ways – some of which (e.g., parental concern, child's academic motivation, etc.) may be important predictors of participation and performance, but may be unobservable to the researcher. Selection may bias the effects of program participation up or down depending on its nature. For example, suppose children left in self-care are from homes that, on average, place less emphasis on academic achievement. If these children perform more poorly on a reading test than children in supervised care it may have little to do with "self-care" per se and more to do with other aspects of the child's home environment. Alternatively, parents may only opt for self-care if their children are particularly responsible. In that case, the self-care children may actually outperform children in supervised settings. But, the superior performance may have nothing to do with the particular child care option selected. An experimental protocol, with random assignment into the program, on the other hand, provides the necessary control that participants and non-participants should be similar, on average, except for their participation in the program. Including a variety of control factors in a regression model may not capture these unobservable differences (c.f. Vandell and Corsaniti 1988; Vandell and Posner 1999). Aizer (2001) makes a serious attempt to address these issues by using a family fixed effects model that contrasts siblings who have experienced different child care histories. She finds that adult supervision reduces a range of risky behavior including drug use or school attendance. This estimation strategy, as she notes, would produce biased estimates if "the decision to allocate time to certain children within the family is correlated with the child's propensity to engage in negative

behavior" (Aizer, 2004). Aizer provides some simple tests for this possibility, but they cannot rule out the possibility of selection.

Meta-analyses

A variety of studies attempt to synthesize the large and conflicting literature on after-school programs (cf. Fashola 1998; Eccles 2001; Redd et al. 2002; Scott-Little et al. 2002; Miller 2003; Catalano et al. 2004; Lauer 2006; Durlak 2007; Little et al. 2008). These reviews suffer from a common problem: how can studies of varying credibility be aggregated to form a conclusion? These syntheses often combine evaluations that are scientifically credible with studies that are not methodologically compelling – sometimes then using the results to conduct further evaluations on what program features are likely to be important in constructing an effective program. Clearly, the weights placed on the validity of the various evaluations will play a critical role in any conclusions that are drawn, making a clear synthesis of the literature difficult to accomplish. One survey, for example, limits the population of studies considered to those with "effects demonstrated on behavioral outcomes." Studies with "no-effect" effectively get a weight of zero in the analyses (Catalano et al., 2004). Fashola, in her study of 34 programs concludes "Our review shows that research on after-school programs is at a very rudimentary stage. Few studies of the effects of after-school programs on achievement or other outcomes meet minimal standards of research design" (Fashola, p. 54, 1998). Scott-Little et al., in a comprehensive survey of the literature, notes that most existing evaluation studies "were published outside of peer-reviewed journals" and that "few programs have utilized experimental designs, a problem common in educational research" (Scott-Little et al., pp. 410-412, 2002). Of the 30 studies used in Lauer et al. to investigate the effect of out-of-school time programs on

improving reading, only 4 are published in peer reviewed journals (Lauer et al., 2006). Miller notes that one reason for this shortage of credible information is that the "standards of rigorous scientific research require resources that are not available to most providers" (Miller, p. 85, 2003).

Flagship Non-Experimental Evaluations

Perhaps the best known and largest scale non-experimental evaluations are those done for the Los Angeles' Better Educated Students for Tomorrow (LAB) program and for programs conducted by The After-School Corporation (TASC).

LA's Best

The LAB program, a partnership between the City of Los Angeles, the Los Angeles School District, and the private sector, is a comprehensive school-based after-school program targeted at children aged 5-12 years old. The program began in 1988 and now has over 100 sites located primarily in high risk, low income areas throughout the City of Los Angeles. The program provides children with homework help, recreational activities, snacks, and a variety of enrichment programs through to 6pm each weekday. A brief description of the program and its evaluations can be found in Table 1. The first evaluation used a non-random sample and contrasted a set of outcomes for 80 program participants to those for 66 comparison group members who were selected based on similarity in age, family income, and education of their parents, and their parent's willingness to let them participate in the evaluation (Brooks et al., 1995). The authors note, however, that the "lack of comparability between the control and program children can only in part be compensated by statistical adjustments." They also indicate

a concern about "the representativeness of these groups." The results are difficult to interpret as no statistically significant differences in improvements in math, reading, composition, social studies, or science were found until the sample was adjusted to remove "outliers." As noted in Blau and Currie (2004) "from the pattern of the results, it appears that the effect of deleting these outliers was to raise the mean scores of the LA's Best kids relative to the controls." There were, however, positive effects on attitudinal effects such as feeling "safer during school" and "[liking] school more this year than last year." A second larger study (n=19,322) compared participants to non-participants controlling for ethnicity, gender, language proficiency, eligibility for free/reduced school lunch, and disability status. Participants were differentiated based on their participation (high, medium, low) in the program (Huang et al., 2000). The evaluation showed improvements in standardized tests in math, reading, language arts, and attendance. Differences in grades, however, disappeared by grades 8 and 9. Other work considering the LAB program describes the evidence on academic achievement as "uneven" while reporting reductions in criminal behavior by program participants (Goldschmidt et al., 2007). Importantly, the design utilized in these studies leaves open the possibility that unobserved characteristics that lead the students into the program could be the causal factors behind any differences in outcomes. Further, it is possible that the selection issues are strongest for those students who persist in the program. If, for example, students with more supportive family backgrounds are those that exhibit the most regular attendance, then what might be regarded as a "dosage" effect is really the result of stronger selection effects. Huang et al. recognize this possibility noting "It may be that high-level attenders do so because they and their parents are more highly motivated, and this interest transfers to achievement. But it is equally likely that coming to school and to the LA's Best program regularly is the reason for good performance and persisting impact subsequent to

leaving LA's best." Unfortunately, the research design does not allow us to distinguish between these possibilities. Other research, however, suggests that students with riskier profiles are more likely to drop out of after-school programs (Weisman and Gottfredson, 2001). This suggests that program attrition may taint comparisons using high-level attenders with selection bias.

TASC

The After-School Corporation is a non-profit organization, that began in 1998 and by 2003 spent almost $100 million supporting 50,000 students in 242 after-school programs in New York – with 186 of the projects located in New York City. Programs are typically located in schools serving a high-fraction of "at-risk" students. The programs place emphasis on homework assistance, academic enrichment, reading, fitness/sports, artistic development, and life skills. TASC's objective has been described as follows:

> TASC's mission, in effect, calls for it to demonstrate that high-quality after-school programs can be created, operated, and sustained in partnership with public schools and with other public and private partners. A central proposition of this mission is that after-school programs can attract significant numbers of children on a regular basis and can offer these children important developmental opportunities, all at no out-of-pocket cost to participants or their families. Finally, according to this mission, these programs and the opportunities they offer can increase the likelihood that participants will succeed in school and in life generally (Reisner et al., 2004).

Several evaluations of TASC have been conducted by Policy Studies Associates (PSA) (c.f., Reisner et al., 2002; Reisner et al. 2004). An evaluation conducted in 2003 by PSA summarized impacts for 96 TASC projects over four school years. The analyses employed a regression based approach including a large number of covariates including baseline test-scores, family income, gender, race, and eligibility for special education services. The report argues that

inclusion of baseline test scores and baseline attendance measures should control for any self-selection bias in estimates for grade and attendance respectively. This assumption is true only if the factors generating self-selection are constant over time. If, for example, a child's family actively decided to place more emphasis on education and that emphasis included after-school participation, then selection bias would still be present. The evaluation further distinguished between regular and "active" participants who attended at least 60% of the possible days and attended at least 60 days during the school year. Estimates for reading and math achievement tests were calculated for two separate years for children in grades 3-8. Math test gains were not statistically significant in the first year, but rose by .42 "standardized scale points" in the second year. Gains were higher still for "active" participants – reporting gains of .79 standardized scale points in the second year. Again, reliance on evidence for "active" participants relies on selection not dictating the degree of participation. Interestingly, no significant gains were found for reading tests. School attendance, after two-years, was approximately ½ day more per year for the participants. And, similar to the LAB study, TASC participants report improvements in various attitudinal measures. Participants, for example, are show an increased likelihood in claiming they "like school more" than non-participants.

Experimental Evaluations

21st CCLC Programs

Given the problematic nature of interpreting evaluation evidence gathered in a non-experimental setting, it is important to investigate evidence generated from an experimental design. Certainly the most influential of the experimental studies is a study done of the 21st Century Community Learning Centers Program conducted by Mathematica Policy Research

(James-Burdumy et al., 2005). The 21CCLC study had two components. First, was a study of elementary students based on a random assignment of students from 12 school districts and 26 program centers that had excess demand for their programs in 1999. These schools were not regarded as nationally representative of programs serving elementary school students. The excess demand allowed the use of random assignment in admission to the program to create treatment and control groups for two cohorts.[6] Test scores were administered at the baseline for both treatments and controls. Importantly, this study considered the child-care arrangements of students assigned to the control group. This allowed the researchers to observe the extent to which those students randomly denied access to the after-school program ended up in "self-care." The results were considerably less favorable than those frequently cited for the non-experimental evaluations. While the programs were serving mostly low-income schools, treatments did not differ from controls in frequency of self-care, maternal employment, reading test-scores, math grades, English grades, science or social science grades, TV viewing time, homework completion, or attendance. There were positive benefits measured for in English and Science for low baseline students. Further, behavioral problems were *higher* for the treatment students. And, treatments were more likely to report feeling "not at all safe" after-school. A nationally representative, but non-experimental, evaluation of 4,264 middle school students with 1,782 in 21st Century Learning Center programs – using controls similar to those used in the TASC study – also found no impact on self-care, math, English, or science grades. Social Studies grades were higher in the 2nd year, and school absences were lower for the participants. Again, behavioral problems were *higher* for program participants.

[6] For the 2000 cohort, there were 589 students in the treatment group and 384 in the control group. For the 2001 cohort there were 693 students in the treatment group and 666 in the control group.

Clearly these results offer a very different picture of after-school efficacy than the non-experimental studies. Indeed, rather than observing "latch-key" care for the controls, a full 75% of controls were home with a parent after-school. Only 1% were in self-care three or more days per week. Not surprisingly, a variety of criticisms have been launched against this influential study (cf. Bissell et al. 2003; Dynarski, 2003). Kane (2004) provides an excellent summary and evaluation of the merits of the various criticisms. One possible explanation for the lack of significant impacts is that attendance rates at the 21CCLC programs were quite low with students participating only 1-2 days per week (Kane, 2004). This participation was lower than that at the TASC sites where elementary school participation averaged almost 4 days per week. This suggests that the "treatment" being considered was not very strong. Second, it is apparent that most of the students electing to participate were not latch-key children. The alternative to after-school supervision was most often parental care which might well provide similar impacts on risky behavior or academic enhancement. Third, the sample size may not have been adequate to identify a statistically significant effects on test scores. It is important to distinguish between "no-effect" and an inability to reject the null hypothesis of "no-effect." Kane points out that the typical gain in standardized reading test scores between the fourth and fifth grades is approximately a third to a half a standard deviation. If after-school programs are regarded as adding an additional hour of time-on-task each day (and assuming an attendance rate of 100%), that would imply an addition of approximately $1/6^{th}$ of academic time each day. Assuming after-school academic time impacts learning in the same fashion as time spent during the regular school day, we'd expect an impact on the order of .05 - .08 standard deviations. The sample size

established for the 21st CCLC evaluation, however, was set to capture effects only as small as .20 standard deviations (Kane, 2004).

Other Studies

Other experimentally structured studies have offered somewhat more positive results. A good example of a small scale intervention is provided by the Howard Street Tutoring Program (Morris et al., 1990). Unlike the multi-site evaluations discussed above, this is a careful evaluation of a single program. The Howard Street program began in 1979 and had adult volunteers working after-school one-on-one with low achieving second and third grade readers all attending a poor inner-city school. The mentors met with the students for 1.5 hours after school twice each week and followed a structured lesson with emphasis on contextual reading, word study, writing, and reading to the child. Thirty students each were randomly assigned to the treatment and control groups. The program showed improvements in the children's reading with "a one-half year difference in reading achievement between the tutored and comparison group" being generated by "50 hours per child of well-planned, closely supervised one-to-one tutoring."

While the Howard Street program employed a simple strategy for improving students outcomes other programs have offered significantly more comprehensive interventions. The Quantum Opportunities (QOP) program is a good example of a program that is comprehensive in nature and that was evaluated using an experimental protocol (Hahn et al., 1994). The QOP program was a multi-year program – beginning in the 9th grade -- that included homework help, tutoring, life and family skills counseling (including counseling on alcohol and drug abuse, sex, and family planning), a significant community service requirement, and "caring adult mentors

that stuck with [the participants] no matter what" (Hahn et al., 1994). In addition, students received financial incentives to encourage them to persist in the program. It is important to realize that while QOP was significantly more comprehensive than a generic after-school program it did incorporate regular program activities from 3pm to 6pm.

A pilot study of QOP was launched in 1989 in 5 sites with twenty-five students from disadvantaged families each being randomly assigned to the treatment or control groups. The results, pooled for the five sites, found significant improvements in high school graduation rates (63% for the treatments versus 42% for the controls), reduced drop out rates (23% versus 50%) and higher rates of college attendance (18% versus 5% for four year colleges). Further, participants were less likely to become teen parents and had "less trouble with the police." It is noteworthy that a careful reading of the report shows that the statistically significant pooled results were driven largely, though not exclusively, by results from one of the implementation sites (Philadelphia). Indeed, at the Philadelphia site, over three-quarters of the participants completed high school with 72% of those who graduated from high school attending a post-secondary educational institution. Only 8% of the treatments dropped out of high school compared to 44% of the controls. Other sites showed positive - though often not statistically significant results. The evaluation report attributed the success in Philadelphia to its ability to create a "consistent group identity and design tangible program services to support QOP members throughout their high school years." Based on enthusiasm for the results of the pilot project, a larger scale demonstration – with 580 participants and 489 controls -- of QOP was conducted in seven sites by Mathematica Policy Research (Maxfield et al., 2003) and funded by the U.S Department of Labor and the Ford Foundation. Participants were virtually all African American or Hispanic and entered the program when they were 14 years old. The evaluation

showed heterogeneity across sites in the implementation of the QOP model. Programs sometimes deviated from the intended QOP model in terms of the depth of mentoring or the hours of community service, for example. While some elements of the program were diluted relative to the programs goals, it was still regarded by "school administrators, faculty, and CBO managers [as the] most intensive program they had ever encountered" (Maxfield et al., p. 54, 2003). Unfortunately, the demonstration reported little in the way of program impact. In particular, there we no differences between the treatment and control groups in achievement test scores, grades, high school graduation rates, or behavioral issues in school. These disappointing findings may be caused by deviations from the intended intervention, the depth of the academic disadvantage of the participants, or the larger size of the programs (Milton S. Eisenhower Foundation, 2005). These concerns, of course, raise questions about the scalability of such comprehensive interventions as well as their efficacy in serving highly disadvantaged populations.

The fourth and final report on QOP by Mathematica Policy Research (Schirm et al., 2006) measure impacts when most of the participants are between the ages of 23 and 25 years old. This report echoes earlier reports concluding that QOP did not increase the likelihood that participants had higher grades or achievement scores, were no less likely to engage in risky behaviors, and were not more likely to graduate from high school or engage in post-secondary education or training. In addition, there was no impact on employment or earnings in this latest follow-up. There was some evidence that participants who were 14 or younger when they entered 9th grade may have benefited from participating in the QOP program. This subgroup of younger participants were 7 percentage points more likely to graduate from high school

(significant at the ten percent level). Similar to the earlier evaluations, QOP's impacts were found to vary significantly across sites.

Mentoring, a critical component of the original QOP pilot, has been shown to generate positive results in an evaluation of the Big Brothers Big Sisters Program (Tierney and Grossman, 2000). In this community-based mentoring program, children aged 10-16 were matched with carefully screened voluntary mentors. The majority of these children were from "relatively poor households" -- 44% reported that their family had received welfare. Participants met with the mentors at least 2 -4 times each month for between 2 and 5 hours per meeting. A waiting list of interested youth created the opportunity to randomly assign applicants into treatment and control groups. The data used for the analyses contained 959 youth, with 487 treatments and 472 controls. At the time of the evaluation, program participants had had participated in the program for, on average, one year. The average participant was about 12 years old.

Comparing outcomes for the two groups eighteen months later showed statistically significant reductions in the initiation of drug abuse (-45.8%) and an improvement in grades (+.08). The grade improvement was largely driven by a .17 grade point increase in grades for the female participants. Participants also reported skipping about one-half fewer days of school. No statistically significant impact was found on, stealing, damaging property, or hour spent doing homework or reading. The cost per participant was $1,000 in 1992 (or $1,480 in 2007$).

The promising results found for community-based mentoring programs have led to a variety of school-based mentoring programs. In the school-based programs, mentors are paired with students whom they typically visit in or after school for an hour a week. A study by Public/Private Ventures conducted an experimental evaluation of the Big Brothers Big Sisters school-based mentoring system (Herrera et al., 2007). The study involved 71 schools, 10 Big

Brother Big Sister Agencies, and 1,139 students in grades four through nine. 565 students were randomly assigned to the treatment group and 574 to the control group. A significant portion of the participants were economically disadvantaged with 60% of the participants receiving free or reduced lunch during the first year of the study. About half of the students were identified as experiencing academic difficulties. The mentors were typically younger than in community based mentoring. About half of the mentors were 18 years old or younger with 72% being female. A little under a half were currently enrolled in high school. Further, only 27% of the mentors reported spending "a lot or most" of their time on tutoring or homework help.

At the end of the first year, participant's academic performance had risen relative to the controls. Improvements were seen in Written and Oral Language, Science, and in the Quality of Class Work and Number of Assignments Completed. Overall, academic performance increased by .11 points on a 1-5 scale. Participants were also less likely to skip school or engage in Serious School Misconduct. Improvements were not concentrated on any particular gender/grade/race/ethnicity subgroups in the treatment population. A follow-up at fifteen months suggested that about half the children discontinued mentoring in their second year. For those children, the academic benefits seen at 12 months follow-up, only about half of the treatments were receiving mentoring. The academic gains that had been seen at the first follow-up had now largely disappeared. Indeed, academic performance at the start of the second school year for students who continued mentoring into the second year was not generally statistically different than the controls, though this may have been because the follow-up occurred early in the second school year and most had not met with their mentors over the summer.

4. Discussion and Extensions

Forging a simple assessment of the efficacy of after-school programs is difficult. Rob Hollister, in a survey of several evaluation studies states: "In short, in response to the question … what do we know about what works – our answer has to be: *not much*" (Hollister, 2003). We concur with that assessment. The current literature on after-school programs raises serious concerns about selection bias. This concern makes it difficult to draw lessons from the prevailing non-experimental evaluations of flagship program or from programs evaluated using comparison groups. The concern about selection is likely endemic in programs that are voluntary in nature.

While the experimental evidence helps mitigate concerns about selection, a careful reading of this evidence suggests several other possibilities for why the estimated impacts on academic achievement may be muted. First, programs may pay limited attention to academic goals. If the primary focus of an after-school intervention is physical exercise or recreation, then impacts are not likely to be seen in the academic domain. This would also be the case if the time spent on academics was limited. Further, as noted by Kane (2004), the effects may be too small to detect using the sample sizes selected. Another concern is that control groups gathered from over-enrolled programs may have reasonably good after-school care alternatives compared to the treatment program. If, as seems to be the case, the controls are not simply "home alone" then we might not expect to see differential impacts on academic outcomes. More generally, program effects, may be small if the control population is relatively privileged (as in the 21st CCLC program) or of the treatment group is extremely disadvantaged (as in the QOP demonstration). In the first case, the counterfactual may not differ much from the treatment. In the second case, after-school programs may not be sufficient to overcome other disadvantages. It is also worth noting that, benefits from non-cognitive outcomes may be exist. Several studies we've reviewed

show improvements in various attitudinal measures. Students, for example, may have an improved enthusiasm for school. While this may not translate into any measureable effect on grades it may provide benefits that support their future odds at achieving self-sufficiency. Recent work by James Heckman suggests that these non-cognitive benefits may be substantial.

Of course, these possible explanations for the "no-effect" finding are speculative. It is also possible that there is simply no effect. More research is needed to see investigate whether the concerns raised are substantive. To that end, the evaluation of after-school programs would benefit from something of a "model-evaluation" much like that of the Perry Preschool Project evaluation (Belfield et al., 2006). It would be useful to have an upper-bound on what benefits a "high-end" after school program might provide. It would also be useful to gauge the effect for participants of varying degrees of depravation.

References

Aizer, Anna. 2004. Home Alone: supervision after school and child behavior. Journal of Public Economics 88, pp. 1835-1848.

Belfield, Clive R., Milagros Nores, Steve Barnett, and Lawrence Schweinhart. 2006. "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup." Journal of Human Resources 41(1): 162–190.

Benson,P.L.& Saito,R.N. 2000.The scientific foundations of youth development. In G. Walker & N.Jafee (Eds.),Youth Development: Issues, Challenges and Directions(pp.125-148).Philadelphia:Public/Private Ventures.

Bissell, Joan S., Christopher Cross, Karen Mapp, Elizabeth Reisner, Deborah Lowe Vandell, Constancia Warren, and Richard Weissbourd. 2003. Statement Released by Members of the Scientific Advisory Board for the 21st Century Community Learning Center Evaluation.

Blau, David, and Janet Currie. 2004. Preschool, Day Care, and AfterSchool Care: Who's Minding the Kids. NBER #10670.

Brooks, Pauline, Cynthia Mojica, and Robert Land. 1995. Final Evaluation Report. Longitudinal Study of LA's Best After School Education and Enrichment Program, 1992-94.

Catalano,R.F.,Berglund,M.L.,Ryan,J.,Lonczak,H.S.,& Hawkins,J.D. 1998. Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs.Seattle,WA: Social Development Research Group,University of Washington.

Durlak, Joseph, and Roger Weissberg. 2007. The Impact of After-School Programs That Promote Personal and Social Skills. Chicago: CASEL.

Dwyer, KM, JL Richardson, and KL Daley, et al. 1990. Characteristics of Eighth Grade Students Who Initiate Self Care in Elementary and Junior High School. Pediatrics, 86.

Dynarski, Mark et al., 2003. When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program. Washington, DC: U.S. Department of Education, Office of the Under Secretary.

Eccles, Jacquelynne S. and Janice Templeton (2001). "Community-Based Programs for Youth: Lessons Learned From General Developmental Research and From Experimental and Quasi-experimental Evaluations." Urban Seminar on Children's Health and Safety. John F. Kennedy School of Government, Harvard University.

Fashola, Olatokunbo. 1998. Review of Extended-Day and After-School Programs and Their Effectiveness. Baltimore, MD: CRESPAR.

Goldschmidt, Pete, Denise Juang, and Majorie Chinen. The Long-Term Effects of After-School Programming on Educational Adjustment and Juvenile Crime: A Study of the LA's Best After-School Program.

Grossman, Jean, Karen Walker, and Rebecca Raley. 2001. Challenges and Opportunites in After-School Programs: Lessons for Policymakers and Funders. Public Private Ventures and MDRC.

Hahn, Andrew, Tom Leavitt, and Paul Aaron. 1994. Evaluation of the Quantum Opportunities Program. Did the Program Work? Brandeis University. Center for Human Resources.

Herrera, Jean Baldwin Grossman, Tina J. Kauh, Amy F. Feldman, and Jennifer McMaken with Linda Z. Jucovy. 2007. Making a Difference in Schools. The Big Brothers Big Sisters School-Based Mentoring Impact Study. Public Private Ventures.

Hollister, Rob. 2003. The Growth in After-School Programs and Their Impact. Mimeo.

Huang Denise, Barry Gribbons, Kyung Sung Kim, Charlotte Lee, and Eva Baker. 2000. A Decade of Results. The Impact of LA's Best After School Enrichment Program on Subsequent Student Achievement and Performance. UCLA Center for the Study of Evaluation. Graduate School of Education and Information Studies.

James-Burdumy et al., 2005. When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program. Final Report. Washington, DC: U.S. Department of Education.

James-Burdumy, Mark Dynarski, and John Deke. 2007. When Elementary Schools Stay Open Late: Results From the National Evaluation of the 21st Century Community Learning Centers Program. Educational Evaluation and Policy Analysis. Vol. 29. No. 4. Pp. 296-313.

Kane, Thomas. 2004. The Impact of After-School Programs: Interpreting the Results of Four Recent Evaluations. Working paper: William T. Grant Foundation.

Lauer, P, M. Akiba, S. Wilkerson, H. Apthrop, D. Snow, M. Martin-Glenn. 2006. Out-of-School-Time Programs: A Meta-Analysis of Effectts for At-Risk Students. *Review of Educational Research* Vol. 76, No. 2, pp. 275-313.

Little, P, C Wimer, and H. Weiss. 2008. After School Programs in the 21st Century. Their Potential and What it Takes to Achieve it. Issues and Opportunities in Out-of-School Time Evaluation. Number 10. Harvard Family Research Project.

Maxfield, Myles, Allen Schirm, and Nuria Rodriguez-Planas 2002. The Quantum Opportunities Program Demonstration: Implementation and Short-Term Impacts. Washington, DC: Mathematica Policy Research, U.S. DOL Contract No. K-5547-5-00-80-30.

Maxfield, Myles, Laura Castner, Vida Maralani, and Mary Vencill. 2003. The Quantum Opportunity Program Demonstration: Implementation Findings. Mathematica Policy Research.

Miller, Beth. 2003. Critical Hours. New York: Nellie Mae Foundation.

Milton S. Eisenhower Foundation. 2003. Quantum Opportunities Program 2003 Forum: Lessons Learned.

Morris, Darrell, Beverly Swaw, and Jan Perney. 1990. Helping Low Readers in Grades 2 and 3: An After-School Volunteer Tutoring Program. The Elementary School Journal. Volume 91, Number 2.

Posner, Jill, and Deborah Lowe Vandell. 1994. Low-Income Children's After-School Care: Are There Beneficial Effects of After-School Programs? Child Development, Vol. 65, No. 2. pp. 440-456.

Posner, Jill, and Deborah Lowe Vandell. 1999. After-School Activities and the Development of Low-Income Urban Children: A Longitudinal Study. Developmental Psychology, Vol. 35, No. 3, pp. 868-879.

Redd, Zakia, Stephanie Cochran, and Kristin Moore. 2002. Academic Achievement Programs and Youth Development: A Synthesis. Child Trends. Washington, DC.

Reisner, Elizabeth R., Christina Russell, Megan Welsh, Jennifer Birmingham, and Richard White. 2002. Supporting Quality and Scale in After-School Services to Urban Youth: Evaluation of Program Implementation and Student Engagement in the TASC After- School Program's Third Year. Policy Studies Associates report to The After-School Corporation.

Reisner, Elizabeth R., Richard White, Christina Russell, and Jennifer Birmingham. 2004. Building Quality, Scale, and Effectiveness in After-School Programs. Policy Studies Associates report to The After-School Corporation.

Scott-Litte, Catherine, Mary Sue Hamann, and Stephen G. Jurs. 2002. Evaluations of After-School Programs: A Meta-Evaluation of Methodologies and Narrative Synthesis of Findings. American Journal of Evaluation, 23(4).

Schirm, All, Elizabeth Stuart, and Allison McKie. 2006. The Quantum Opportunities Program Demonstration: Final Impacts. Mathematica Policy Research Inc. U.S. DOL Contract No. K-5547-5-00-80-30. MPR Reference No.: 8279-932.

The After-School Corporation. 2007. Annual Report.

Tierney, Joseph, and Jean Grossman. 2000. Making a Difference. An Impact Study of Big Brothers Big Sisters. Public Private Ventures.

U.S. Census Bureau. 2002. Who's Minding the Kids? Child Care Arrangements: Winter 2002.

Weisman, Stephanie, and Denise Gottfredson. 2001. Attrition from After School Programs : Characteristics of Students Who Drop Out. Prevention Science, Vol. 2, No. 3.

Vandell, Deborah Lowe, and Mary Anne Corasaniti. 1988. The Relation between Third Graders' After-School Care and Social, Academic, and Emotional Functioning. Child Development, Vol. 59, No. 4. pp. 868-875.

Figure 1: Trends in Female Labor Supply

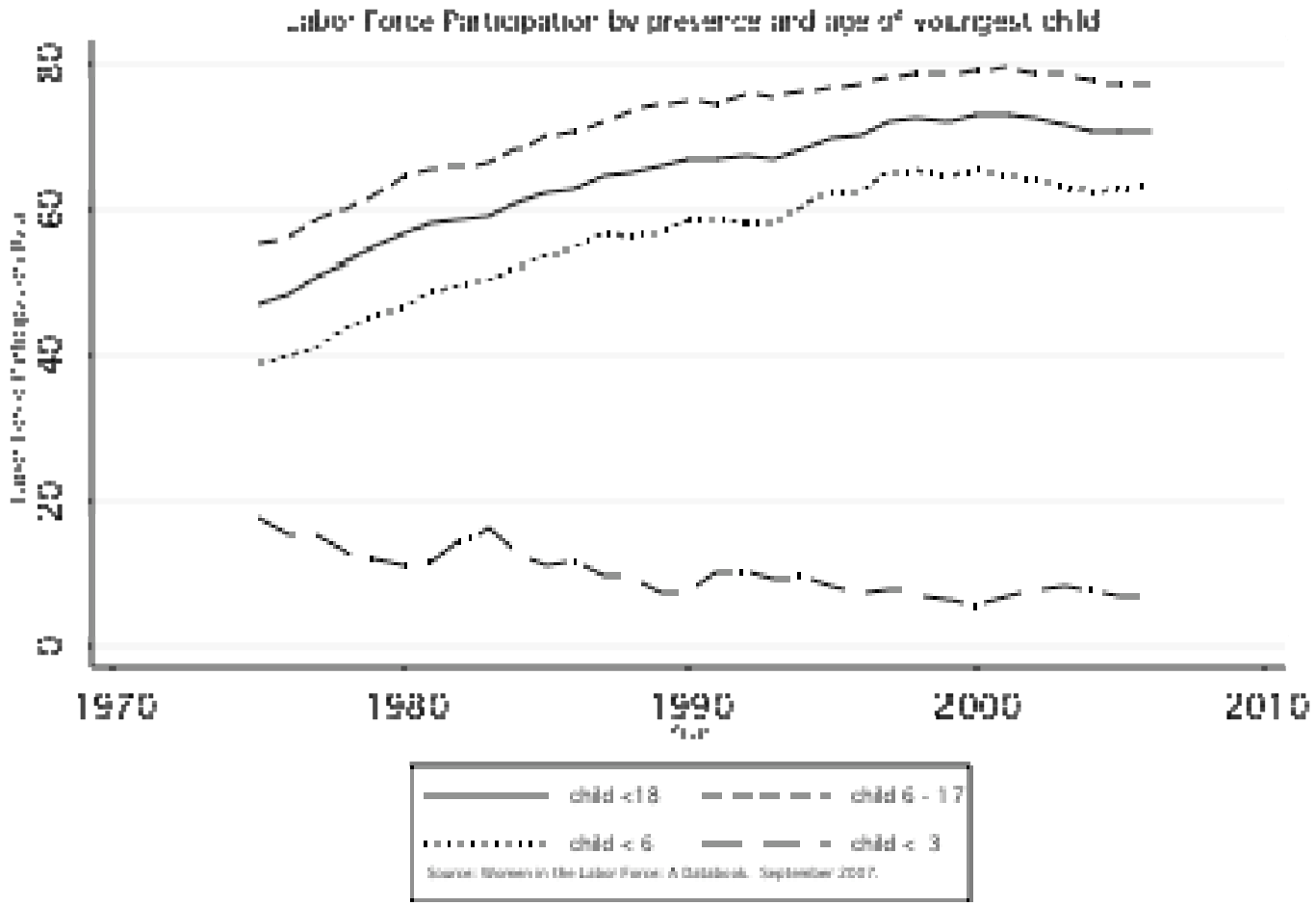
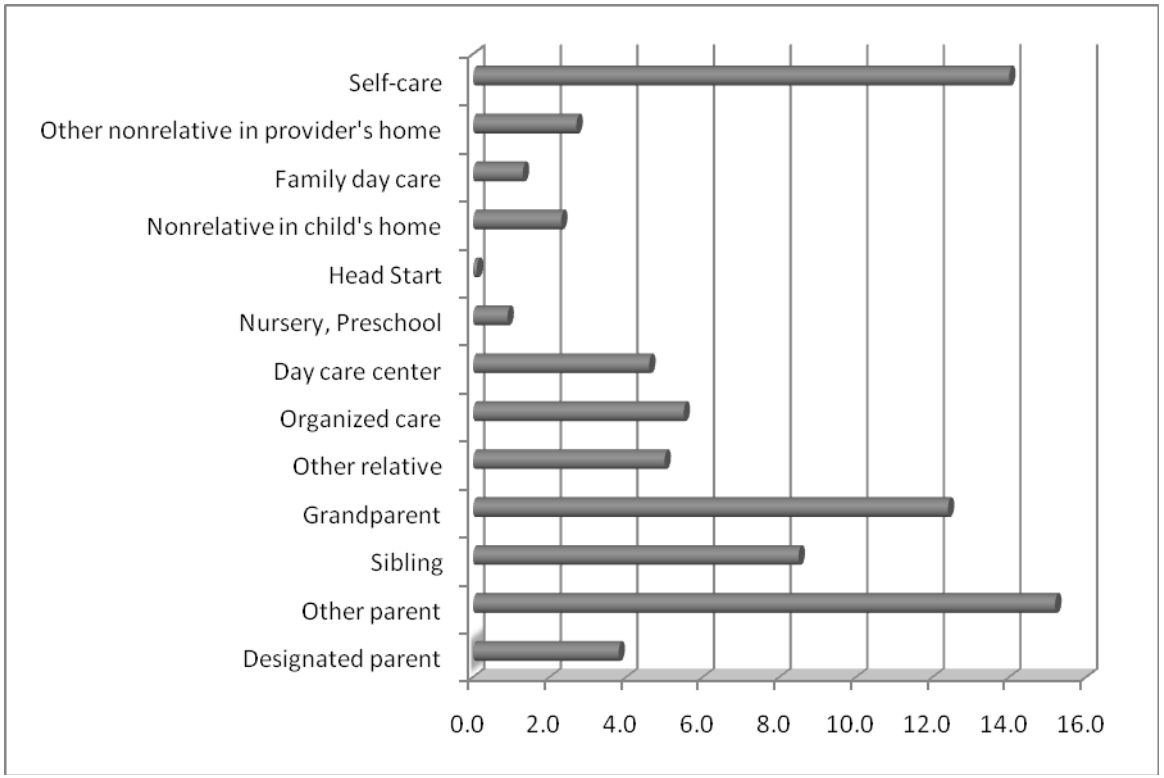
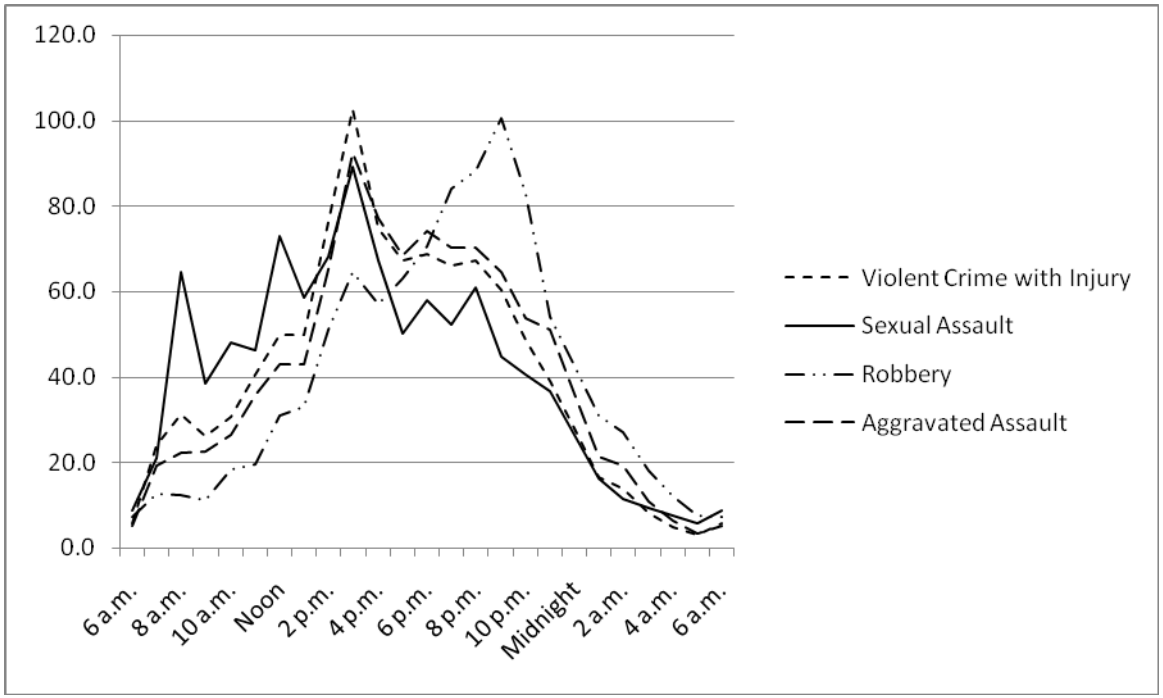


Figure 2: Child Care arrangements of Grade-schoolers 5 – 14 years old living with mother.



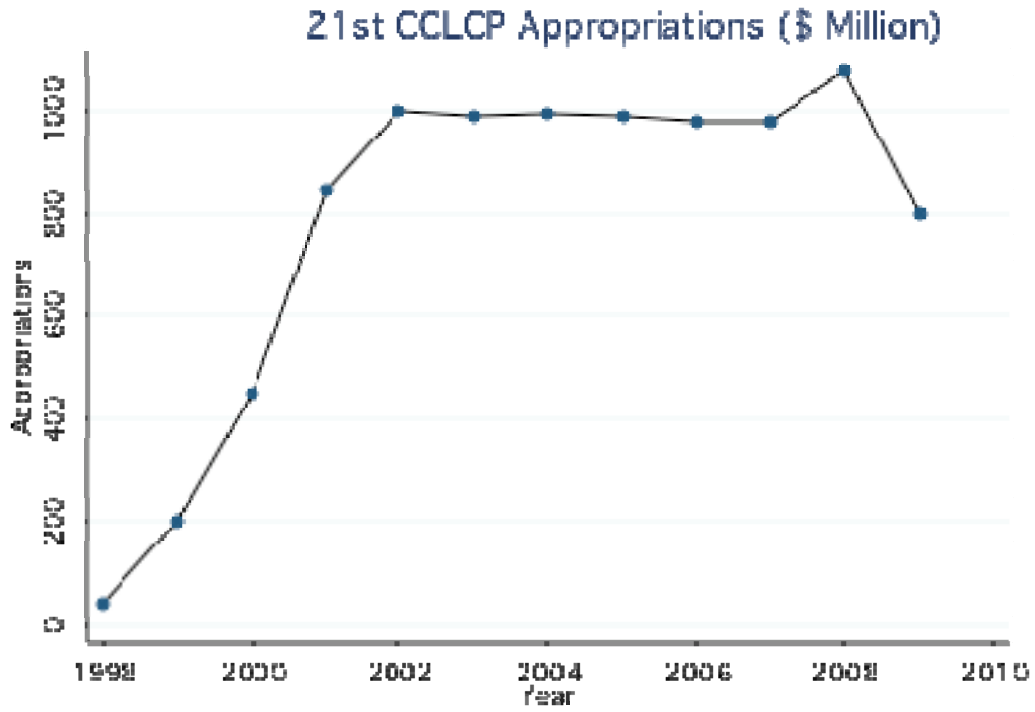
Source: "Who's Minding the Kids" 2005 data. U.S. Bureau of the Census.

Figure 3: Juvenile Crime, Offenders per 1,000 offenders in age group by time of day



Source:
Snyder, Howard N., and Sickmund, Melissa. 2006.
Juvenile Offenders and Victims: 2006 National Report. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Figure 4: Trends in Federal Funding of After-School Programs



Note: 2008/2009 are estimated.
Source: http://www.ed.gov/about/overview/budget/statetables/index.html
and http://www.afterschoolalliance.org/21stcclc.cfm

Table 1: Review of Evidence

Study	Intervention	Evaluation Design	Sample	Outcomes	Effects
			Selected Non-experimental Studies		
Brooks, P.E. et al. (1995)	LA's Best After-School Program	Comparison group design.	Non-random sample. Program participants (n=80 in year 1, n=69 in year 2) were in 5th and 6th grades with 2+ years in program. Comparison group (n=66 in year 1, n=58 in year 2) was similar in terms of age, family income, age and education of parents, but with <3months in program.	Grades in math, reading, composition, social studies, and science.	No significant effect.
				Safety	Improved sense of safety.
				Enjoyment of school and other attitudinal measures.	Improvement in several attitudinal areas.
Huang, D. et al. (2000)	LA's Best After-School Program	Multiple regression. Controls include ethnicity, gender, language proficiency status, eligibility for free/reduced lunch, and disability status.	Non-random sample. Program participants in grades 2-5 (n=4312) and schoolmate non-participants (n=15010). Participants divided into high (75% days present), medium (26%-74% days present) and low (<26% days present) levels of participation.	Comprehensive Test of Basic Skills and Stanford-9 Achievement Test in reading, math, and language arts.	Positive effect (increases with participation)
				School absenteeism	No significant effect by grades 8 and 9.
Reisner, E. et al. (2004)	The After-School Corporation (TASC)	Multiple Regression Controls include baseline test-scores, family income, gender, race, and	Data collected for over four school years from 96 TASC after-school projects and their host schools in New York City with 52,000 after-school participants and 91,000 students who were enrolled in TASC host schools but not	Math	+.06 sigma after 1 year +.42 sigma after 2 years

		eligibility for specialized educational services	participating in TASC projects.		
					Active Participants +.13 sigma after 1 year +.79 sigma after 2 years
				Reading	No significant effect.
			Selected Experimental Studies		
James-Burdumy, S. et al. (2005)	21st Century Community Learning Centers Program	Elementary school: experimental	Mostly low-income schools with large proportions of minority students. Elementary school study based on random assignment of students from 12 school districts and 26 centers that had excess demand for the programs in 1999. First cohort of elementary students First year (2000-2001) had treatment n=589 and control n=384 for 7 school districts. Second cohort (2001-2002) had treatment n=693 and control n=666 for 5 districts. Test scores administered at baseline and follow-up. Average attendance 2.7 days/wk for students continuing into second year. These students were more likely to come from two parent families and had higher baseline reading test scores.	Second year results. (note: participant and intent-to-treat effects are similar) Frequency of self-care. Maternal employment, reading test-score (SAT-9), English grade, math grade, science and social science grades, TV viewing time, homework completion, attendance. Behavioral problems Sense of safety	 No significant effect. Increased for treatment students. Fewer treatments (2.5% vs 7.1%) felt "not at all safe" after school.
	21st Century Community Learning Centers	Middle school: comparison group.	Middle school study is based on a nationally representative sample of program participants (n=1782) and a matched sample of non-	Frequency of self care Math, English, or science	No significant effect

	Program		participants (n=2482). Participants in second year had more educated mothers. Extensive controls.	grade Social Studies grades School absences	No significant effect Treatment higher in second year. More behavioral problems with treatment students.
Morris, Darrell, Shaw, Beverley, and Perney, Jan. (1990) Helping Low Readers in Grades 2 and 3: An After-School Voluntary Tutoring Program.	Howard Street Tutoring Program	Experimental	Bottom 1/3 of second and third grade readers identified by teachers. Students are ranked based on several reading tests administered by outside evaluators and ranked from high to low. Students with similar rank are randomly assigned to program or control groups. Two years evaluated. Treatment: n=30, Control: n=30. Students are from a low-SES school.	Word recognition Basal word recognition and spelling	No significant effect Statistically significant gains.
Hahn, Andrew, Leavitt, Tom, and Aaron, Paul. (1994) Evaluation of the Quantum Opportunities Program (QOP). Did the program work?	Quantum Opportunity Program (QOP)	Experimental	Twenty-five students from each site were randomly selected to participate in the program from 9th -12th grade. Targeted at disadvantaged families. Launched in 5 sites: Philadelphia, Milwaukee, San Antonio, Saginaw, and Oklahoma City.	High-school graduation rate Drop out of school Pursue post-secondary education Attend a four year college Participants less likely to become teen parents (24% vs 38%) (p<.01) Trouble with police	Higher for participants (63% vs 42%) (p<.01) Participants less likely to drop out (23 vs 50%) (p<.01) Participants more likely (42% vs 16%) (p<.01) Participants more likely (18% vs 5%) or two year college (19% vs 9%). (p<.01) Participants had less trouble with police (p=.09).

Maxfield, Myles, Laura Castner, Vida Maralani, and Mary Vencill. 2003. The Quantum Opportunity Program Demonstration: Implementation Findings.	QOP	Experimental	Demonstration evaluation in seven sites conducted by U.S. DOL and the Ford Foundation between 1995-2001. Single cohort of 580 ninth grade in 9th grade program participants and 489 controls. Sites included 6 inner city sites and one rural community. There were100 youth participants in 4 sites, 80 in one site, and 50 in two sites.	Achievement test scores, grades, high school graduation rates, or behavioral issues.	No significant effect.
Schirm, Allen, Stuart Elizabeth, and McKie, Allison. (2006) The Quantum Opportunity Program Demonstration: Final Impacts.	QOP	Experimental	Follow up of U.S. DOL and Ford study. Participants age 23 and 25 years old.	High School graduation or post-secondary education. Earnings or employment.	Participants no more likely to graduate from high school or engage in post-secondary education or training. No impact on earnings or employment.
Tiernay, J, Grossman, J, and Resch, N (1995). Making a Difference: An Impact Study of Big Brothers/Big Sisters. P/PV	Big Brothers Big Sisters of America	Experimental	Sample includes youth aged 10-16 (more than half minority) from 8 sites with excess demand. Agencies were among largest in BBBS organization. Students randomly assigned to treatment (n=571) or to control (a waiting list where they remained for the study duration) (n=567) were surveyed. At one year follow-up survey samples were treatment n=487 and control n=472.	Stealing, damaging property, hours spent on homework, hours spent reading. Initiating drug use, hitting another person, perceived ability to complete homework, skipping school/class, lying to parents. Grades	No significant effect. Significant beneficial impact. Grade effect was .08 (p<.1) (both genders) Grade effect was .17 (p<.05) (girls)
Herrera, Jean Baldwin Grossman, Tina J. Kauh, Amy F.	Big Brothers Big Sisters of America	Experimental design.	Fourth through ninth graders. 565 treatments, 574 controls.	Academic performance	Academic performance was .11 points higher (on 1-5 scale) for treatments at 1 year

Feldman, and Jennifer McMaken with Linda Z. Jucovy. 2007. Making a Difference in Schools. The Big Brothers Big Sisters School-Based Mentoring Impact Study. Public Private Ventures.		follow-up. Impacts largely disappeared at 15 month follow-up.