The Growth in After-School Programs and Their Impact

Rob Hollister

February 2003

PREFACE

Interest in after-school programs has increased dramatically in recent years. Yet little is known about the effectiveness of these programs and whether they are a good use of taxpayer dollars. This paper reviews the growth of these programs, the reasons for their growth, and what they hope to accomplish. It also addresses what we know about what works, the costs of the programs, and the implications for policy.

This paper was commissioned by the Brookings Roundtable on Children in an effort to shed more light on these questions. It was written by Rob Hollister, Professor of Economics at Swarthmore College, with background research by Marc Rockmore (made possible by assistance from the Smith Richardson Foundation).

The paper concludes that much of the evidence on these programs is sparse and not very good. However, based on a review of 10 studies that used a relatively rigorous methodology to measure impact on a variety of outcomes such as drug and alcohol use, academic skills, or violence, Hollister concludes that there have been some effective programs.

The Roundtable is making this paper more broadly available in the hopes that it will catalyze further discussion and research on this important topic.

Isabel V. Sawhill
Director, Children's Roundtable
The Brookings Institution

The Growth in After-School Programs and Their Impact

Rob Hollister

I. Recent Rapid Growth in After-school and Youth Development Programs

The National Survey of America's Families by the Urban Institute provides estimates that indicate that in 1997 about 7 percent of youth age 6 to 12 were engaged in some sort of after-school programming. After-school program participation was higher for those children with working mothers. Among 6 to 9 year-olds with a working mother, 21 percent participated in after-school programs. Among 10 to 12 year-olds with working mothers, the rate was 10 percent. Participation also varied by income level: 7 percent of low-income 10-12 year-olds with employed mothers versus 11 percent of high-income children with employed mothers.

In recent years there has been rapid growth in funding of after-school and related youth development programs. Here is my rough accounting of this growth:

- In fiscal year 1998, the federal government provided $40 million dollars to schools to create and run 21st Century Community Learning Centers, which are school-based after-school programs. In fiscal 1999, the funding for this program increased to $200 million, in fiscal 2000 it increased to $450 million, and in fiscal 2001 funding rose to $850 million.
- State and local governments have likewise greatly increased spending on after-school initiatives. During the 1998 school year, New York increased funding for after-school programs from $500,000 to $10 million. Kentucky spends $37 million on extended school services. Maryland passed an After School Opportunity Fund of $10 million. Wisconsin is providing $20 million for after-school programs. The Pennsylvania legislature is considering a bill for $15 million for after-school programs.
- At the local level, in 1998, George Soros's Open Society Institute established The After School Corporation in New York City and it now supports 165 programs in New York City and 20 in other locations in New York State. The Open Society Institute is providing up to $25 million per year on a 3:1 match basis for these efforts. In Boston, the mayor launched his "2:00 to 6:00" initiative that currently funds after-school programs in 57 schools and has recently received a $23 million grant from local foundations, corporations and universities to be used for these purposes. Chicago's Lighthouse provides after-school services to 363 elementary schools. The Wallace-Reader's Digest Funds are making major investments in promoting these programs, as are the Charles Stewart Mott and Annie Casey Foundations.[1]

II. Why Has There Been a Growth in After-School and Youth Development Programs?

In many ways, the growth in after-school programs and the continuing pressure to expand them is, at this point, more a social movement than a policy innovation. As with other social

[1] These data are largely drawn from Grossman et. al. (2001).

movements, the impetus comes from a variety of interest groups: educators, child development experts, community development groups, criminal justice experts, health professionals, and parent associations. They have different concerns and objectives, but coalesce around some structuring of activity in the after-school hours.

Reasons for Growth

After school programs have evolved in response to a set of broader social and economic developments since 1980.[2] In the1980s, proponents of new programs for youth focused on the risky behaviors of youth: increased sexual activity at young ages, drug and alcohol abuse, school dropout, and weak educational performance. The results of these risky behaviors were perceived to be: growth in teen pregnancy and the related growth in single parent families, deeper involvement in drugs and drug selling, growth in gang-activity and crime, and a lack of basic skills such as literacy and numeracy. The perception of increased risky behavior, particularly among youth from low-income families, was given even sharper focus when a Carnegie Foundation report pointed out that a high proportion of crime committed by youth occurred between the hours of 3 p.m. and 6 p.m. in the afternoon. This naturally led to the proposition that filling those hours with after-school programs could reduce the opportunity to, and likelihood that youth would, engage in those risky behaviors.

Some analysts noted that these risky behaviors may have been generated by deeper changes in the economic and social environment. Changes in the economy affected families both at the low-income and middle-income levels. Inequality in earnings grew sharply with declining real wages for middle- and low-income workers. More women entered the labor force in order to try to maintain family income in the face of the declining earnings of their husbands or in response to greater opportunities for themselves.

Poverty became more concentrated in inner city areas. The location of jobs shifted away from the inner city to the suburban fringe (spatial mismatch) and the premium paid for high skills increased (skills mismatch). At the same time, the quality of inner city education was perceived as deteriorating. Due both to "spatial mismatch" and "skills mismatch," inner city youth faced poorer employment prospects and weakened abilities to take up available opportunities.

The earlier emphasis in the 1980s and early 1990s on risky behaviors among inner city youth began to be criticized by experts in youth development. They argued that most youth programming was too focused on negative behaviors on the part of youth and that a more positive stance was needed. Efforts should be made, they argued, to emphasize the programs and activities that were likely to lead to positive outcomes for youth, not just to the diminution of risky behavior.

To summarize: the forces behind increased funding and activity in after-school programming could be characterized in two phrases: "time on task" and "home alone." These two phrases apply *both* to the children of low-income parents and to the children of higher-income parents.

[2] In Appendix A, this complex of factors is discussed in somewhat more detail with references to the appropriate literature.

The phrase "time on task" stands for the increasingly prevalent view that more time spent on an educational or skill-building tasks will result in much improved educational performance. Those who take this view have argued that after-school programs can extend the learning period – the time on task.

The "home alone" phrase reflects worries about "latch-key children." The increase in women's labor force participation across all income levels has led to concerns that more and more children are being left during after school hours in unsupervised situations, particularly during those "3 p.m. to 6 p.m." hours – hours when risky behavior may occur, both among children from low-income families and among children in higher-income families.

The two sets of concerns have been reflected in an increasing tension *within* the after-school movement. There are those who feel the programs should be closely connected to the schools, with programming primarily focused on enhanced educational performance. Others, however, emphasize that after-school programs should provide a "safe place" where children can feel they "belong to something" and can gain in self-confidence in non-academic as well as academic activities. I will return to discussion of this tension below.

From "Prevention" to "Positive Youth Development"

Traditional youth-serving institutions, such as the YMCA, Boys and Girls Clubs, Boy Scouts and Girl Scouts, had virtually never focused on preventing risky behavior by youth.

In the 1960s and 1970s, increased concerns about poverty and the decline of inner cities led to a shift in emphasis toward trying to *prevent* school dropout, substance abuse, youth crime, and teen pregnancy. Researchers tried to isolate factors that predicted later negative behavior.

By contrast, in the late 1980s and early 1990s, greater stress was placed on achieving *positive* outcomes for youth and on programming that would build on the positive assets youth had, or could have, and on "developmentally appropriate" strategies.[3] Attention was refocused on the joint roles of family, schools, and communities in promoting positive youth development.

Central to the ideas of positive youth development programming is the observation that youth in late childhood/early adolescence are dynamically developing their sense of self. They are trying out, to some degree, different persona.[4] Positive youth programming seeks to shape this dynamic in positive directions. Part of the continuing evolution of positive youth development programming concepts has been what Catalano and his colleagues (1999) call "Prevention Science." Prevention science involves attempts to identify "risk factors" that predict the later emergence of problem behaviors and "protective factors" that work to reduce their emergence. "Exposure to increasing numbers of risk factors was found to increase the likelihood of a child's problem behaviors, while exposure to increasing numbers of protective factors was

[3] For a brief introduction to the background of these developments see Catalano et. al. (1999, p. 2-9).
[4] An interesting literary presentation of this process of a youth trying on persona can be found in V. S. Naipaul's A Bend in the River.

found to prevent problem behaviors."[5] Increasingly, most of this work linking early characteristics to later outcomes has been based on longitudinal data and is essentially correlational in nature. Attempts to rigorously establish "causation" are – not surprisingly – lacking. I return to this methodological issue below.

III. Are there "models" for After-School and Youth Development programs?

I have been hard pressed to uncover clear statements setting out a model, or multiple models, of what a youth development program should look like given its purposes or what it is trying to accomplish. In practice, many different program structures have been developed and implemented, but the reasons for the choice of structural elements are not clearly articulated. Some general theories of youth development exist.[6] From these theories, the *outcomes* which might define healthy youth development and the key elements which appear to be necessary to generate those outcomes can be derived.[7]

First, consider this broad description of youth development programs:

> "…most simply, youth development programs are developmentally appropriate programs designed to prepare adolescents for productive adulthood by providing opportunities and supports to help them gain competencies and knowledge needed to meet the increasing challenges they will face as they mature...youth development programs are best characterized by their approach to youth as resources to be developed rather than as problems to be managed and their efforts to help youth become healthy, happy and productive by increasing youth's exposure to external assets, opportunities and supports." [(Roth, Brooks-Gunn et al. 1998) p.427].

In order to make a "model" more specific, sets of *outcomes* which might be measured in order to assess the program's impact are outlined. Here is a short list, representing a consensus reached by practitioners, youth development advocates and youth development researchers:

- Caring and compassion
- Character
- Competence in academic, social and vocational arenas
- Confidence
- Connection[8]

Subsumed in these categories are *outcomes* which youth programmers and evaluators might have focused on previous to the "positive youth development" movement. These might be, for example, academic achievement (test scores), academic attainment (years of formal schooling completed), crime (arrests, incarceration), substance abuse (drugs, alcohol, smoking), and sexual activity (sexually transmitted diseases, pregnancy and child birth). However, the concern to refocus on "positive youth development" so that youth "assets" rather than

[5] Ibid. (p.6).
[6] Eccles (1999). Eccles and Wigfield (2000). Dynarski and James-Burdummy (2001).
[7] Eccles and J. Templeton (2001). Benson (2000).
[8] Roth and Brooks-Gunn (2000). In Appendix B, we provide a more detailed list drawn from Catalano et al. (1999).

"problems" would be stressed, leads to submerging these "problem-oriented outcomes" within the broader constructs.

With measures of outcomes specified, a "model" would then indicate structural features that are designed to affect those outcomes. One discussion of such features is as follows:

- Adequate provisions for physical and psychological safety, developmentally appropriate levels of structure and adult supervision
- Supportive relationships with adults
- Supportive and respectful relationships among peers
- Opportunities to develop a strong sense of belonging
- Opportunities to experience mastery and mattering
- Opportunities to learn the cognitive and non-cognitive skills essential for succeeding in school, work, and other pro-social social and institutional settings
- Strong positive social norms for behavior.[9]

More fully developed "models" seek to specify the broader context in which the youth development program would operate. They try to specify the contextual elements which are likely to affect the stated outcomes independent of the program intervention. They also try to indicate contextual elements that are likely to differentially affect the relevance and efficacy of the various structural elements of the program.[10]

Recently, analysts who have adopted the "theory of change" approach to evaluation press for even greater specificity in the program-outcome model.[11] This approach calls for an even greater articulation of program elements and a specification of the pathways from program components to the outcomes which are the expressed goals of the program.

One major reviewer of evidence on program impacts states: "…the diverse nature of the many community programs for youth makes exact specification of the treatment problematic…Rarely did we find a well-specified model underlying either program design or program evaluations. It is likely that the development of such theoretical models probably requires a prolonged and genuine collaboration between basic researchers, applied researchers, program developers, providers, and program evaluators."[12]

These difficulties in pinning down "models" and links from "theory" to program characteristics are features one might expect of a social movement rather than a policy innovation; interest groups can coalesce where detailed prescription is missing.

[9] Eccles and Templeton (2001).
[10] See Benson (2000, p.137).
[11] Connell et al. (2000).
[12] Eccles and Templeton (2001).

IV. What do we know about what works?

There are now six major, recent reviews of the evidence on the impact of after-school and youth development programming.[13] Several of these cover both in-school and after-school programs. Some of the reviewers seek to derive from observational data implied relationships between program features and later outcomes. I pass over the in-school programs and limit my discussion to assessments of the impacts of out-of-school programs.

These six reviews are based on systematic searches of the evaluation literature from which have been selected out, for summary and discussion, programs and their evaluations which meet criteria established by the reviewers. In general, the reviewers limited their selection to evaluations which provide quantitative impact estimates – as opposed to qualitative or process measures – and have tried to focus on evaluations which have met fairly high standards of rigor in the research design.

In all of the studies reviewed, attempts were made to evaluate the impact of a program intervention on one or more outcomes for youth. Therefore the researchers had to try to develop a "counterfactual" – what the outcome for these youth would have been had they *not* participated in the program. The impact studies reviewed used various types of analytic designs in order to deal with the counterfactual problem. The designs ranged from simple statistical analysis of pre and post measures (through constructed comparison groups) to rigorous random assignment studies.

I diverge here, for a moment, to report on some of the problems that plague attempts to assess the results reported in these reviews. While Catalano et al (1999) describe a number of sensible criteria for including an evaluation in their review, including adequate description of research design methodologies, a final criterion is that the evaluation showed "evidence of significant behavioral outcomes... These included evidence that *positive outcomes* were enhanced or negative outcomes were reduced, or that both occurred" [(emphasis mine) p.22]. Unfortunately, this means that well designed evaluations that found *no statistically significant impact* were not reported. I believe the exclusion of evaluations where there was no statistically significant impact was a mistake, as it is important for us to learn what *doesn't* work as well as what does work.

The issue of "selection bias" looms large in research designs that do not utilize random assignment to program group and control group or among alternative program treatments groups. Without random assignment there is always a chance that there will be a concentration within the program participant group of those with characteristics that affect the outcome (e.g. the program participants may be more motivated than those who are in the comparison group). To the extent that those characteristics are measured it is possible to control for their effects with statistical models. It is the *unmeasured, or unmeasurable, characteristics (like motivation) which create the bias problem*. Note that these biases can go in either direction, leading to overestimation of the program impact (the more motivated are enrolled in the program) or to an underestimation of the program impact (the less motivated are in the program because they want to avoid more

[13] Benson (2000). Catalano et al. (1999). Dynarski and Burdummy (2001). Eccles and Templeton (2001). Fashola (1998). Roth and Brooks-Gunn (2000).

demanding uses of their time). While quasi-experimental methods try to reduce this type of bias, my previous work has convinced me that we do not know when a quasi-experimental estimate is likely to be "close enough to the true impact" to be acceptable.[14] *For these reasons I limit my discussion of impacts to studies of evaluations in which a random assignment design was used*.

Even this highly demanding criterion needs to be carefully applied. For example, Fashola, in the concluding pages of the review where she is discussing selection bias issues, says: "There are solutions to these methodological problems, but they have rarely been applied. The best is to take a list of children applying for a given program and then randomly assign them to the program or to a waiting list control group....The fact of applying and meeting other admission requirements ensures that the waiting list control group is equivalent in all important ways to the treatment group. Of all the programs reviewed in this report, only the Howard Street Tutoring Program … and the Memphis Extended-Day Tutoring Program ... use random assignment of this kind."[15] However, when one turns to the detailed description of the Memphis project one finds: "the treatment group consisted of students who attended the program at least 50% of the time for some of the analyses, and for others, at least 80%. *The students who did not attend, or who had low attendance, were added to the control group*" [(emphasis ours) p.10]. Adding to the control group members of the group initially assigned to the program group but selected out because of non-attendance, or some other reason, seriously undermines the strength of the initial random assignment in avoiding selection bias.

Still more problems arise even within reviews of random assignment studies. The Roth et. al. review includes the Summer Training and Employment Program (STEP) as an example of a well-designed random assignment evaluation. This was indeed a well designed and executed random assignment study, carried out in 8 sites across the country. The program provided summer remedial courses along with part-time summer jobs for the treatment group, with the control group receiving just the summer jobs. Program design was motivated by the fact that the gap in test scores between low-income and higher-income kids *increases* as they move through the school grades. However, a study showed that the increase in the gap at each grade level appeared to have *occurred in the summer time*; during the school year, both groups gained in test scores at about the same rate.[16] The idea was to create a program to stem the summer loss by the low-income kids and thereby prevent the gap from growing.

STEP was characterized by several of the reviewers as a successful summer program; both controls and treatment group members suffered declines in reading and math but by the end of the summer treatment group members had statistically significantly *smaller losses*. The problem is that by the end of the *following school year* (and at graduation after a second summer remedial program exposure) there was *no difference in test scores or graduation rates for the treatment group compared to the control group*. It is difficult to view this program as successful, given that the initial theory being tested was that the summer intervention would have lasting effects, narrowing gaps between low and higher-income students.

---

[14] Hollister and Hill (1995). Connell, et al. (1994). Wilde and Hollister (2002).

[15] Fashola (1998, p. 43).

[16] More recent studies have apparently found similar evidence and were referred to in Alan Krueger's New York Times column (17 Aug 2000, Section C; p.2) in which he advocated voucher's for summer school remediation on the same grounds. See also Entwisle et al (1997).

A final difficult example is the Quantum Opportunities Program (QOP). QOP was a five site random assignment evaluation of a community-based program providing multi-year, year-round academic assistance, life and family skills instruction, college and career planning, and community service and work experiences. In addition, perhaps a unique example in post-1980 programs, participating youth received stipends and bonuses. QOP has been cited widely as a very successful program, resulting in increased test scores and graduation rates, higher post-secondary school attendance, and lower childbearing. Unfortunately, the evaluation report (Hahn 1994) was *very* misleading and requires careful detailed examination to interpret its findings. One site failed completely and was not included in the analysis. All the significant academic outcomes were isolated in the Philadelphia site. There were serious differential attrition problems in the other sites. The lower childbearing did *not occur* at the Philadelphia site and only came from the pooling across sites, several of which suffered from attrition bias. Test scores among the Philadelphia participants were statistically significantly higher than the controls but at the graduation year they were testing only at the eighth grade level. Can higher graduation rates be called a success when they are still achieving only at the eighth grade level? The higher post-secondary school attendance was based on only one follow-up four months after the secondary school graduation date and this short follow-up would also have missed any later secondary graduation among controls or attainment of GEDs. The impact of the stipends – a large cost element – was not evaluated.

The point of this is not just to be "picky," but to emphasize that even generally thorough reviews of multiple studies require very detailed scrutiny of the underlying study reports if one is not to be misled about program impacts.

Summary of Results based on Rigorous Design Evaluations

I will now list the few studies cited in the six major reviews which appear to have met my most rigorous criteria of being a well-designed random assignment evaluation. While the six reviews categorized programs in a variety of ways, I reduce the categories to just three: programs for which mentoring or tutoring appeared to be the major treatment component, programs which focused on remedial skill-building, and programs which stressed broader community and parent involvement in some fashion.

Here are the ten programs with impact evaluations I felt met my criteria for rigor of design. A few of the programs appear twice because they seemed to have major components in more than one category. In Appendix C, I present a table – using primarily the format provided in Roth and Brooks-Gunn – describing the program characteristics, the evaluation authors and design, and the main findings. In this list, I give the name of the program (with an abbreviation used below) so that readers who wish to do so may search out more detail about the program in Appendix C.

<u>Mentoring, Tutoring</u>

Howard Street (Howard)
Big Brothers Big Sisters (BB-BS)
Across the Ages (ATA)
Friendly PEERsuasion (FPEER).

Woodrock (Wdrck)
Quantum Opportunities Program (QOP)- Philadelphia site only

Remedial Schooling
Summer Training and Employment Program (STEP)
Louisiana State Youth Opportunities (LSYO)

Parent-Community[17]
Midwestern Prevention (MWPRe)
Creating Lasting Connection (ClastC)
Woodrock (Wdrck)
QOP

Conclusions from these Ten Studies

The conclusions that follow are very summary and there is no detail about program characteristics. Unfortunately, I could not present salient program features concisely, so the reader must refer to Appendix C for more detail about the individual program characteristics.

For economy of discussion, I broadly group outcomes measured as in-school and out-of-school. In-school outcomes measures would include such measures as test scores, attendance, suspension, and continuation (i.e., not dropping out). Out-of-school outcomes measures would include smoking, drug or alcohol abuse, safe sex practices, pregnancy and childbearing, crime and violence, and employment. I have ignored some of the outcome measures discussed by the six review studies which relate to the youth development constructs listed above. I also ignore what they refer to as "moderators," usually measures of knowledge or attitude which are expected to affect the other behavioral outcomes at some point.

Before turning to the conclusions I feel are well established through these ten studies, I must make several comments about limitations. As I noted above, although the STEP program remediation established a diminishing of summer decline in test scores, since these effects did not translate into long-term differences in school outcomes, I interpret the program as *not* effective. The LSYS was a program similar to STEP in terms of content and it achieved similar results at the end of the summer session. However, because there were no long-term follow up measures taken, it is unclear whether the program had any long-term effects.

As I noted above, interpreting the effects of the QOP program is tricky because the summary of the effects by the evaluators is not reliable. It appears that effects were reliably established *for the Philadelphia site only*, so I limit my conclusions to those based on outcomes at that site.

Mentoring/tutoring appears to have been an effective component having effects both on some in-school outcomes (BB-BS, Howard St., QOP) and some out-of-school outcomes (BB-

[17]Two other programs described by Catalano et al. (1999) might qualify here but there was insufficient detail about the random assignment process for me to determine whether they met my strict criterion. Those programs are fast-track (discussed on p.47) and teen-outreach (discussed on p.58).

BS, Across Ages, FPEER). Both adult and peer (older youth-younger youth) mentors have been found to be effective. Careful selection and matching of mentors may be critical. (BB-BS).

Parent involvement and training have sometimes been effective components for out-of-school outcomes (MWPre, ClastC, Wdrck).

Life skills training curricula may be effective for some out-of-school outcomes (FPEER, Wdrck).

This may seem a meager harvest from the apparently rich field of youth development programming. Others, even the six sets of reviewers I relied on, have been willing to draw somewhat broader conclusions. But that is because they have been willing to include quasi-experimental studies (and, to some degree, strictly correlational and qualitative studies) in their base for drawing conclusions, whereas I have imposed a high hurdle of strict application and implementation of the random assignment evaluation design. I would agree that if one is seeking program structures to test further, then it is reasonable to use less rigorous criteria to draw lessons from these programs. However, my task has been to assess what we know with reasonable certainty about the impacts of youth development programs.

To further shorten my discussion, I quote the conclusions of two of the comprehensive reviews I utilized as basic sources.

Fashola says: "this 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 … After-School programs are increasing rapidly and receiving strong support from the Clinton administration, from Congress, and from state and local policy makers. As is often the case, this enthusiasm and rapid growth is running far ahead of the research base. We need much more research on the effects of all types of after-school programs, especially those intended to enhance student achievement. There is a particular need for development and evaluation of replicable, well-designed programs capable of being used across a wide range of circumstances."[18]

Roth et. al. say: "The review of the evaluation literature highlights the paucity of high quality outcome evaluations of programs fitting the youth development framework. As noted previously, little improvement in the state of program evaluation has occurred since the 1992 Carnegie Report … Nationally, strong interest in expanding adolescents' access to youth development programs exists. However, the current mismatch between the enthusiasm for these programmatic efforts and the empirical evidence calls into question the effectiveness of such efforts."[19]

In short, in response to the question posed at the outset of this section – what do we know about what works – our answer has to be: *not much*.

[18] Fashola (1998, pp. 43-44).
[19] Roth (1998, pp. 441-444).

V. The Costs of Youth Development programs and Estimates of Funding Needs.

As noted at the outset, funding for after-school programming has been growing at a very rapid rate in recent years. Because lobbying efforts to put even more resources into these programs continue to increase, one might want to have some indications of what the costs per individual child would be and how many children would be covered under the most extreme assumptions. That is to say, we might wish to get some indication of where we might be headed if the current trends are maintained.

There have been a number of recent studies which attempt to estimate the universe of needs for after-school or out-of-school time programming. I will report on two such studies in order to give a feel for how substantial an effort would be required to meet the universe of need.

The Universe of Need

One study looked just at Massachusetts and developed an estimate of need for that state.[20] It noted that (at the time of the study) there were an estimated 825,000 school-aged children living in Massachusetts. Of these, about 60,000 were participating in out-of-school-time programs, leaving 765,000 not enrolled in such programs. Of these, it was estimated that 386,000 would enroll in out-of-school-time programs if available and accessible.[21] This means that they are estimating a universe of need in Massachusetts which would cover about 54 percent of all school age children.

The study developing estimates for national needs takes a different approach.[22] The authors first use time budget studies to estimate the amount of time average youth are "without supervision or structure" and how much of that time should be covered with program opportunities. They convert this to hours per year per school-age child who needs program coverage. They then apply this to the total school-age youth population estimate of about 47 million to get an estimate of the universe of need in terms of hours per year to be covered by programs. In one sense, then, this study assumes the universe of need is 100 percent of all school-age children and the estimated number of coverage hours drives their overall cost estimates.

The Costs of Programming

There is a huge diversity of types of after-school programs, but little careful cost accounting information for any of these programs (particularly since donated space, equipment, and volunteer staff are significant in almost all program types). Thus, in order to get cost estimates that are (at best) orders of magnitude, one must rely on considerable imagination and

[20] Wechsler et al. (2001).
[21] The estimate was based on the fact that 73 percent of the total Massachusetts population ages 6 to 17 lived in families where both parents are working. They then assumed that 66 percent of these families would choose to enroll their children in structured out-of-school time programs if they were available and accessible. Subtracting the number of those already enrolled in such programming (an estimated 53,000), they end up with their increment of needed positions.
[22] Newman (2000).

strong assumptions. Be that as it may, these two studies provide two interesting and different approaches to this task.

The Massachusetts study had a working group develop a set of assumptions governing costs and components of a *high quality* out-of-school-time program, including: salaries and wages high enough to retain qualified staff, paid staff development, nutritious meals, supplies, equipment, transportation, insurance, rent, and basic administrative costs. Using these strong assumptions, they get a cost per child for a "school-year only" program of $4,349. Expanding it to cover summer raises the per child cost estimate to $5,989.

The national focus study took a different approach. They gathered cost figures from a group of major programs (Big Brothers/Big Sisters, Teen Outreach Program, The After-School Corporation, Boys and Girls Clubs, Girl Scouts of America) and derived estimates of cost per hour. Then, differentiating by allocation of time to each type of program and the average cost per hour of that program type, they arrive at an average cost per hour of $2.55. Multiplying that number by their time estimate of an average of 1200 hours per year per child, they get an average annual cost $3,060 per child.[23] They then multiply that per child annual cost by the 47 million school age children to get a total annual cost of providing youth development activities for the entire U.S. school age population of $144 billion.[24]

The difference in the cost per year per child between the Massachusetts study and the nationally oriented study clearly arises from the very different methods used to build up the estimates. The Massachusetts authors explicitly state they are estimating the costs of a "high quality" program. The nationally oriented study authors explicitly state they believe that the per-hour costs they have derived are underestimates of true costs. Still, this gives us a feeling about the range of per child expenditures the strong advocates for these programs are suggesting: $3000 to $5000 per child per year. The nationally oriented study also provides us a gross total for national spending of $144 billion or more.

VI. Alternative Strategies

From Theory and Evidence to Design and Testing

Clearly, if we insist on the highest standards of rigor in determining what works and what doesn't in after-school programs, we are left with very little guidance as to whether resources invested in after-school activities are likely to yield social benefits substantial enough to warrant the costs. Further, even ignoring this very tough benefit-cost calculus, there is not much we can

[23] They contrast this with their estimated average yearly cost of public education per child per year of $6,564.

[24] These authors go on to attempt a benefit-cost analysis which is completely wrong and misguided. They use the above described costs, but then generate benefit estimates which are completely made up. They get them by taking the average annual salary of a high school grad, assuming that a grad works for 40 years and gets an annual 3% increase, pays 17% taxes on all these earnings, spends 60% of pre tax income on consumption. They add the taxes and consumption together to get "Total Contributions to Society" which is what they count as the benefits. Comparing this to costs, they conclude "For every dollar invested, society gains $10.51." I will not bother to list all the errors and misleading assumptions that are made but would urge readers to disregard completely this claim of huge "return on investment in youth development."

take from this little bit of current strong evidence to guide us in understanding how to better design after-school programs.

In thinking about the implications of this state of affairs, I believe it useful to draw a parallel between the weak state of knowledge concurrent with extensive program activity in youth development programming and a similar weak state of knowledge concurrent with extensive youth employment and training program activity in the 1970s. In 1985, looking back over the previous 15 years in which there had been extensive total national expenditures on youth employment and training, we found we really could say almost *nothing* about the impacts of these programs.[25] This was because we – the research and policy-making community – had not put ourselves in a position to learn systematically, through rigorous evaluations, the effects of various types of programs on various segments of the youth population. Over the next decade, however, reliable evidence began to accumulate. Rigorous evaluations were provided for Supported Work, Job Start, the Minority Female Single Parent Program and finally the National Evaluation of the Job Training Partnership Act and the National Evaluation of the Job Corps. The results of these evaluations were for the most part *very discouraging about the gains from employment and training interventions for youth* (exceptions being the Job Corps and the Center for Employment Training in San Jose). It could be argued that part of the new Work Investment Act legislation – de-emphasizing structured youth training programs – was largely affected by these evaluation results.

Some of the organizations involved in the evaluations, such as Public/Private Ventures, were so discouraged by these findings that they engaged in deep reconsideration of their approach to youth programming and shifted toward the sort of positive youth development programming that has been discussed here.[26] I fear that if more serious efforts are not made to get rigorous estimates of the impacts of alternative program structures on long-term outcomes, this whole movement may suffer the same fate as did the youth employment and training movement.

In light of these circumstances, there are several stances one might take:

- Abandon the search for hard evidence and justification of these programs
- Capitalize on major evaluation efforts
- Try to draw on theory and observational data for some guidance for improved design of youth development programs
- Try to construct a framework in which knowledge about effectiveness will be accumulated.

I indulge in a few remarks on each of these.

[25]For the findings of a National Academy of Sciences panel on youth employment and training, see Betsey, Hollister and Papageorgiou (1985).

[26]See Connell, Aber and Walker (1994).

Abandon the search for hard evidence and justification.

It could be argued that the changes discussed above in family work patterns, the impact of welfare reform requirements on single parent families, and the decline in informal institutions in the inner city, all imply a need, in some communities, to provide substitutes for some family functions. In these communities, there is a need for some sort of adult-supervised activity for school-age children during the after school hours. The argument here would be largely for a custodial function and the only criterion for program effectiveness would be "do no harm." Indeed, one researcher/evaluator, with deep experience in youth development programming, argued at a recent conference that those involved in the after-school/youth development movement should simply refuse to "play the game" of trying to rigorously establish quantified impacts on "hard outcomes." They should not try, he argued, to justify increased expenditures in terms of a "good quantifiable social return per dollar of investment."[27]

Pressure to produce hard evidence on impacts might be alleviated by the apparent current tendency of after-school program resources to "migrate up the income distribution." As noted several times above, with increases in women's labor force participation in higher as well as lower income families, there are increasing demands from middle and higher-income parents, for programs to fill the 3:00 p.m. to 6:00 p.m. hours until the working parent gets home. Indeed, we have some hard evidence that this "upward migration" is happening with the 21st Century Learning Centers grants. The following data are taken from performance reports to the U.S. Department of Education. The "First Cohort" label applies to the group of grants made in the first round of 21st Century funding and "Third Cohort" to those made in the third round.

RACE AND ETHNICITY OF STUDENTS IN HOST SCHOOLS OF FIRST AND THIRD COHORT 21ST CENTURY LEARNING CENTERS
(percent)

	Overall[a]	First Cohort[b]	Third Cohort[c]
White	43.2	34.5	46.5
Black or African-American	26.4	35.7	22.9
Asian	2.3	2.5	2.2
Native Hawaiian or Other Pacific Islander	0.6	0.5	0.7
American Indian or Alaska Native	3.2	5.2	2.4
Hispanic or Latino	24.3	21.4	25.4
Other	0.1	0.2	0

[a]Number of centers reporting=789
[b]Number of centers reporting=236
[c]Number of centers reporting=553

[27] Speech by Robert Halprin in May 2001, at a John F. Kennedy School of Government conference titled "Urban Seminar on Children's Health and Safety."

Notice that, moving from the first cohort (representing the first round of funding) to the third cohort (the third round of funding) of grantees, the percentage white increased by 12 percentage points and the percentage black or African-American fell by nearly 13 percentage points. Even though the percentage Hispanic or Latino increased somewhat from the second to third cohorts, it was not so sharp a change as in the other groups. I would say this suggests a movement toward families less likely to be low-income.

The following table shows the percentage of students participating in the centers that are eligible for free or reduced price lunch – often used as a proxy marker for poverty.

STUDENTS ELIGIBLE FOR FREE OR REDUCED PRICE LUNCH IN HOST SCHOOLS OF FIRST AND THIRD COHORT 21ST CENTURY LEARNING CENTERS
(percent)

	Overall[a]	First Cohort[b]	Third Cohort[c]
Less than 25 percent eligible	9.7	4.4	12.1
25 to 49 percent eligible	24.0	19.2	26.0
50 to 74 percent eligible	28.9	27.1	29.6
75 to 100 percent eligible	37.4	49.3	32.3

[a]Number of centers reporting=759
[b]Number of centers reporting=229
[c]Number of centers reporting=530

The last row of the table represents the measure of extreme poverty, i.e., 75to 100 percent of the student body eligible for free and reduced price lunch. We can see in this table that from the first to the third cohort there was a sharp decline – a *17 percentage point drop* in the percent of centers who had *75 to 100 percent of participating students eligible for free and reduced price lunch.* At the other extreme – less than 25 percent eligible for free and reduced price lunch – there was an increase of 8 percentage points. In both the first and third cohort, the median is in the 50 to 74 percent eligible category. Thus the centers funded are far from being predominantly centers for the affluent. But the bottom and top row of the table show, there is a distinct drift, from the first to the third cohort, away funding centers in high poverty areas.

I conjecture that as a program moves further away from being identified with poverty the cries for hard data on benefits and costs decline – our toughest evaluation standards are applied primarily to programs for the poor.

Capitalize on major evaluation efforts

This will not be easy to do. A recent publication listed the major evaluations of after-school programs which are underway. Only one of these, the National Evaluation of 21st Century Community Learning Centers, has a random assignment design for at least one component. This evaluation is applying a random assignment design for evaluation of impacts in programs for elementary school children.

There are several privately funded smaller multi-site evaluations. A substantial effort must be made to assure that a major portion of these precious evaluation resources will be devoted to rigorous evaluation designs. For example, there is a strong temptation when programs are largely community-based to resort to a "comparison community" design to try to estimate the impact of the program.[28] I believe that comparison community evaluation designs are a complete waste of resources and that nothing reliable is learned from such designs. Unless the number of communities eligible and applying for the grant is sufficiently large and communities are randomly assigned to the program, little will be learned from such studies, and they can be very expensive in terms of the data collection required in the comparison communities.

Try to draw from theory and observational data some guidance for designs to be subject to more rigorous testing.

There are already several efforts of this type which have been undertaken. Roth & Brooks-Gunn (2000) build upon their review of evaluations, and their extensive knowledge of the literature of youth development theory and evidence, to summarize the implications of the theory and evidence for improved design of youth development programs. The *Future of Children* issue contains several articles that attempt to make the bridge from theory and existing evaluations to suggestions for youth development program design.[29] Benson (2000) makes a very serious effort to outline a program for improving the "Scientific Foundations of Youth Development." Mathematica Policy Research is attempting to draw on the literature to shape the selection of sites for their National Evaluation of 21st Century Learning Centers. Eccles and Templeton include such design suggestions in a preview of The National Research Council Board on Children Youth and Families report. These are all very high quality and encouraging efforts. Thus far, however, the reviews have not developed tight program design models with *specified component variations* that could be adopted by communities and rigorously tested.

Try to construct a framework in which knowledge about effectiveness will be accumulated.[30]

Random assignment evaluations are the first choice vehicle for gathering reliable information about what works for whom. There are a variety of ways to adapt the random assignment process to accommodate both the peculiarities of a specific program and the preferences of funders and program operators, without resulting in selection bias. It is essential that funders and program operators allow the evaluation researchers to implement such adaptations in order to minimize the complications and conflicts that could arise.

When the unit of random assignment is located *above* the individual, e.g. classrooms, schools, communities, the difficulties of carrying out random assignment designs increase. However, I was encouraged to find, in some of the reading for this project, examples of random

[28] Indeed, the National Evaluation of 21st Century Community Learning Centers will use a comparison high school either within or across school districts for its evaluation of impacts of middle school programs. The evaluators found that most of the middle schools had excess capacity in their middle school centers and therefore the usual argument that random assignment is simply allocating scarce places among eligibles could not be applied. This excess capacity is just another manifestation of the problem of enrolling and holding middle school students in after-school programs discussed further below.

[29] e.g. Eccles (1999).

[30] To a large degree, this is exactly what Benson, et al. (2000) attempt to do.

assignment evaluations at the level of units above the individual. Most of these had to do with in-school curricula, often with respect to health-related behavior.[31]

There is an important role for observational data and correlational studies based upon them. That role is to generate *hypotheses* about factors influencing youth development, which are then translated into program structures that are, in turn, rigorously evaluated. As noted above, there are several efforts already underway to do this sort of work.

In the last decade, there has been considerable development of empirical studies related to youth development, particularly those based on longitudinal data. However, commentators and policy developers have often failed to make the distinction between correlation and causation. There is a further consideration that has been largely ignored. It is one thing to establish a link between a "risk factor" and a later outcome. It is quite a different thing to establish a link between a *program-induced change in a "risk factor"* and the later *change in outcom*e that could be expected. For example, a "strong relationship" may be found between lack of contraceptive knowledge among youth and the probability of a teen pregnancy, but, as has been found in many program evaluations, a program-induced improvement in contraceptive knowledge may not induce a change in behavior which lowers teen pregnancy.

In addition, we researchers are often excited to find statistically significant relationships between "risk factors" identified at an early age and later outcomes, but it does not follow that programs designed to alter such "powerful risk factors" will generate significant changes in the program population. For example, from the literature on juvenile delinquency, we could conclude that the relationship between early risk factors and later criminal behavior indicate which types of children are most in need of some sort of intervention. However, it cannot tell us which *individual* children will become serious offenders. The relationship may be statistically significant but the proportion of total variance explained by the risk factors may be low. *Though most adult criminals have been juvenile delinquents, most juvenile delinquents don't go on to become adult criminals.* One of the most "powerful predictors" of juvenile violent behavior is experiencing substantial abuse as a child. Such abuse may raise the probability of violent behavior by a factor of four or more, but most violent juveniles may not have been subject to childhood abuse and most of those subjected to childhood abuse do not exhibit violent juvenile behavior.[32] Thus, for example, in justifying large expenditures on efforts to reduce child abuse (laudable perhaps on other grounds), proponents may project the long term effects of such expenditures in terms of future reductions in violent juvenile behavior on the basis of such correlational studies. The temptation to make such use of correlations should be strongly resisted.[33]

It is all too easy to advocate the development of a rational framework for accumulating knowledge about what works in youth development and for whom. It is quite another matter to figure out how to create the institutional context and resources that would support the realization of such a framework.

---

[31] Peters and McMahon (1996). Cook et al. (1999).

[32] Lipsey and Derzon (1999).

[33] For an example of this type of exaggeration see Newman et al. (2000, p. 24). They claim the QOP showed that every dollar spent on the program produced $3.04 in public benefits. See also Gleason and Dynarski (2002).

VII. Implementation Issues

Finally, I would like to make a few comments on issues regarding the implementation of youth development programming that I have come across in our work for this project.

"Time on Task" versus "Home Alone"

Returning to our earlier reductive phrases, it has become quite clear to me, from talking to researchers and from attending several meetings which included "practitioners," that there is a tremendous struggle between those who believe that after-school programs should be focused on skill development by providing more "time on task" and those who stress the need to provide an atmosphere for growth and adult contact for children who are too often "home alone."

The structure of the 21st Century Community Learning Center grants has exacerbated this conflict. It funds programs through the educational system, *requires* them to be centered in school facilities, but also *mandates* that the wider community be involved in program development and delivery. Those from a youth development background argue for programs centered on positive youth development, with an emphasis on conflict resolution, resistance strategies, and civic involvement. Those from an education background feel that the program should be an extension of the school day and, thus, focus on skill development, helping with homework, and promoting better relationships with teachers. The positive youth development people fear, and with good reason, that they and their approach to youth development will be shut out of the federal funds flow.

Program Location

Closely related to this point is the issue of where after-school youth development programs should operate. It is evident that the school is usually the most significant physical facilities resource in any neighborhood. Equally, the school transportation system can be a critical asset in facilitating broad access to after-school programs and promoting a feeling that the programs (and movement to and from them) represent a "safe place" for the youth. On the other hand, for the most "at risk" parts of the youth population, school may have become a negative setting and they may be less willing to participate in programs located there.

The basic funding stream for schools is probably the most secure facilities support in the neighborhood. However, because control of those funds usually resides with the school board, youth program operators will have to work through that organizational structure, which can be exceedingly difficult. Operating in facilities independent of the schools may impose a much lower organizational burden.

Finally, it is generally agreed that these programs are most needed in predominately low-income urban and rural neighborhoods and it is exactly in such neighborhoods that the schools have been characterized as low performing and organizationally dysfunctional. Attempting to run new after-school programs through such systems may prove unusually challenging.

Participation of Middle School Students

One of the major problems implementers of youth development programs face is creating and sustaining *participation* of low-income youth, particularly as and after they make the transition from elementary to intermediate schools. It is exactly at these transition points that youth development theory suggests there is the greatest need for the additional support these programs are designed to provide. A major contribution that can be made through evaluation studies *not* aimed at measuring the impact on long-term outcomes is to isolate better strategies for boosting and sustaining participation during this transition and continuing into the middle school years.

Targeting

Another issue in program design and implementation is the degree of "targeting" of programs on particular types of youth and of "targeting" program content on particular types of problem behaviors. This is an issue that arises across the spectrum of social policy. As noted above, "prevention science" has stressed the isolation of "risk factors" associated with particular longer-term problem outcomes. This often leads to focusing programs on "high risk" segments of the youth population and shaping program content to reduce the "risk factors" or to promote the "protective factors." In general, limited program resources may be more efficiently utilized if they are focused on the "high risk" groups. Some evaluations have suggested that highly targeted content aimed at specific behaviors may be far more effective than diffuse general youth development content.

On the other hand, researchers have recognized that targeting interventions on particular populations can stigmatize children; in the famous phrase by an English social scientist, "a program for the poor is a poor program." Further, some qualitative evaluations have suggested that targeted programs may create peer groups that reinforce the very negative risk factors that the program seeks to offset; grouping smokers for program purposes may create an environment which validates smoking behavior. We all know the lament that youth in correctional detention are often said to be attending "schools in criminal behavior." I have already noted above the problem that even "powerful risk factors" identified in the correlational literature may do a poor job of identifying individuals whose long-term behavior may be problematic. Indeed, in a recent article, researchers looked at school dropout prevention programs for which they had both pre-program measures of "risk factors" and post-program dropout rates for the same individuals. They conclude: "The findings suggest that dropout prevention programs often serve students who would not have dropped out, and do not serve students who would have dropped out."[34]

[34] Ibid (p.25). More concretely, the researchers say: "The regression risk factor performed best, with a dropout rate among identified students of 42%. Nonetheless, the regression risk factor was wrong more often than it was right. In particular, 58% of students identified as likely to drop out using the regression did not drop out. Considering that the regression risk factor used 40 variables measuring student characteristics and past school performance, dropping out clearly was difficult to predict." (p.37).

What Stage of Development?

The positive youth development literature stresses the importance of "developmentally appropriate youth programming" and this is indeed an important insight which researchers have backed with a much enriched theoretical and empirical literature. One issue that presents a challenge, however, is how to identify the stages of youth development and how to adapt program content to that staging. As I read this literature, it appeared that while stages are generally correlated with age, this is not strictly the case. For example, youth who mature physically earlier may associate themselves with older youths who face similar challenges from their altering physical state. Constructing programs that are flexible enough to respond to these individual differences is a major challenge.

Outcomes Measurement

Another area which deserves attention is how to measure the outcomes which youth development programming should be promoting. Both MacDonald (2000) and Eccles and Templeton (2001) make a good start on this task. The positive youth development movement emphasizes the importance of stressing youth assets rather than youth deficits. They have developed an array of constructs they argue identify positive youth behavior (sketched in the long list in Appendix B). But, in the end, evaluation of most youth development programs has focused on the degree to which the program reduces negative outcomes: lower teen pregnancy rates, lower school dropout rates, and less criminal behavior. There are some outcomes which tend to be evaluated in positive terms – school test scores, advancement to post-secondary education, higher employment rates – but the reduction of negative outcomes seems to weigh more heavily in the policy analysts' assessments of program worth than do the gains in positive outcomes. In my opinion, it will be a challenge to convince policy-makers of the worth of a program intervention if it is based on an increase in a measure of "social competence" or "moral competence," unless it is accompanied by a decrease in the more usual, "harder" negative measures, such as reduced school dropout or reduced arrests.

REFERENCES

Anderson, Elijah (1992). *Streetwise: Race, Class and Change in an Urban Community.* University of Chicago Press.

Benson, Peter and Rebecca N. Saito (2000). "The Scientific Foundations of Youth Development." *Youth Development: Issues, Challenges and Directions*. (Fall). Public / Private Ventures.

Betsey, Charles L., Robinson Hollister, and Mary Papageorgiou, eds. (1985). *Youth Employment and Training Programs: The YEDPA Years.* Washington, D.C.: National Academy Press.

Boo, Katherine (2001). "After Welfare." *The New Yorker*. (April 19).

Carnegie Council Report on Adolescent Development (1992). *A Matter of Time: Risk and Opportunity in the Nonschool Hour.* Carnegie Corporation.

Catalano, Richard, M. Lisa Berglund, Jeanne A. M. Ryan, Heather S. Lonczak, and J. David Hawkins. (1999). "Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs, Social Development Research Group." University of Washington.

Connell, James, Michelle Gambone and Thomas Smith (2000). "Youth Development in Community Settings: Challenges to Our Field and Our Approach." *Youth Development: Issues, Challenges and Directions.* Public/Private Ventures.

Connell, James, J.L. Aber and G. Walker (1995). "How do urban communities affect youth? Using social science research to inform the design and evaluation of comprehensive community initiatives." *New Approaches to Evaluating Comprehensive Community Initiatives: Concepts, Methods and Context.* Connell et al., eds. Queenstown, MD: The Aspen Institute.

Cook, Thomas, Farah-naaz Habib, et al. (1999). "Comer's School Development Program in Prince George's County, Maryland: A Theory-based Evaluation." *American Educational Research Journal* 36(3): 543-597.

Dynarski, Mark and Suzanne James-Burdumy (2001). "A Broader View: The National Evaluation of 21st Century Learning Centers." Design Report Volume 1 (March). Mathematica Policy Research.

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.

Eccles, Jacquelynne S. (1999). "The Development of Children Ages 6 to 14." *The Future of Children.* 9 (2): 30-44.

Eccles, Jacquelynne S. and Allan Wigfield (2000). Schooling's Influences on Motivation and Achievement. *Securing the Future: Investing in Children from Birth to College*. Sheldon Danziger and Jane Waldfogel, eds. New York: Russell Sage Foundation.

Ellwood, David (2000). "The Impact of the Earned Income Tax Credit and Social Policy Reforms On Work, Marriage and Living Arrangements" *National Tax Journal* 53 (December): 1063-1106.

Entwisle, Doris, Karl Alexander and Linda Steffel Olson (1997). *Children, Schools & Inequality.* Bolder, Colorado: Westview Press.

Fashola, Olatokunbo S. (1998). "Review of Extended-Day and After-School Programs and Their Effectiveness." Report 24. Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.

Gleason, Philip and Mark Dynarski (2002). "Do We Know Whom to Serve? Issues in Using Risk Factors to Identify Dropouts." *Journal of Education for Students Placed at Risk.* 7(1): 25-41.

Grossman, Jean, Karen Walker and Rebecca Raley (2001). "Getting School-Based After-School Programming Off the Ground." *Urban Seminar Series on Children's Health and Safety.* John F. Kennedy School of Government, Harvard University.

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

Hollister, Robinson G. and Jennifer Hill (1995). "Problems in the Evaluation of Community-Wide Initiatives." *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts.* James Connell et al., eds. Washington, D.C.: The Aspen Institute.

Holzer, Harry J., Keith Ihlandfeldt, and David Sjoquist (1994). "Work Search and Travel Among White and Black Youth." *Journal of Urban Economics.* 35: 320-345.

Kreuger, Alan B. (1998). "Reassessing the View that American Schools are Broken."*Economic Policy Review* 4(1). Federal Reserve Bank of New York.

Lipsey, Mark W. and James H. Derzon (1999). "Predictors of Violent or Serious Delinquency in Adolescent and Early Childhood: A Synthesis of Longitudinal Research." *Serious & Violent Juvenile Offenders: Risk Factors and Successful Interventions*. Rolf Loeber and David P. Farrington, eds. Thousand Oaks, CA, Sage Publications.

MacDonald, Gary B. and Rafael Valdivieso (2000). "Measuring Deficits and Assets: How We Track Youth Development Now and How We Should Track It." *Youth Development: Issues, Challenges and Directions.* Public/Private Ventures.

Mauer, Marc (1990). *Young Black Americans and the Criminal Justice System.* The Sentencing Project.

Mauer, Marc and Tracy Huling (1995). *Young Black Americans and the Criminal Justice System: Five Years Later*. The Sentencing Project.

Maynard, Rebecca (1996). *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*. The Urban Institute.

Moore, Kristen, Anne Driscoll and Laura Lindberg (1998). *A Statistical Portrait of Adolescent Sex, Contraception and Childbearing*. National Campaign To Prevent Teen Pregnancy.

Morris, Pamela et al. (2001). *How Welfare and Work Policies Affect Children: A Synthesis of Research*. Manpower Demonstration Research Corporation.

Naipaul, V. S. (1979). *A Bend in the River*. New York: Knopf Publishing Group.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*. U.S. Department of Education.

Newman, Robert P., Stephanie Smith and Richard Murphy (2000). "A Matter of Money: The Costs and Financing of Youth Development." *Youth Development: Issues, Challenges and Directions.* Public/Private Ventures.

Newman, Sanford et al. (2000). *America's After-School Choice: The Prime Time for Juvenile Crime, or Youth Enrichment and Achievement*. Washington, D.C.: Fight Crime: Invest in Kids.

Padilla, Felix M. (1992). *The Gang As an American Enterprise*. Rutgers University Press.

Peters, Ray D. and Robert J. McMahon (1996). *Preventing Childhood Disorders, Substance Abuse and Delinquency*. Banff International Behavioral Science Series (3**:** 215-240). Sage Publications.

Roth, Jodie and Jean Brooks-Gunn (2000). "What Do Adolescents Need for Healthy Development? Implications for Youth Policy." *Social Policy Report* XIV(1): 3-19.

Roth, Jodie, Jean Brooks-Gunn, et al. (1998). "Promoting Healthy Adolescents: Synthesis of Youth Development Program Evaluations." *Journal of Research on Adolescence* 8(4): 423-459.

Sullivan, Mercer (1994). *Getting Paid: Youth Crime and Employment in the Inner City.* Cornell University Press.

United States Department of Labor (1991). *What Work Requires of Schools: A SCANS Report for America 2000.* Secretary's Commission on Achieving Necessary Skills, U.S. Department of Labor.

Walker, Gary, & Frances Vilella-Velez (1992). *Anatomy of a Demonstration: The Summer Training and Employment Program (STEP) from Pilot through Replication and Postprogram Impacts.* Public/Private Ventures.

Wechsler, Samantha, et al. (2001). *Meeting the Challenge: Financing Out-of-School Time Programming in Boston and Massachusetts.* Parents United for Child Care.

Wilde, Elizabeth T. and Robinson G. Hollister (2002). "How Close is Close Enough? Testing Nonexperimental Estimates of Impact Against Experimental Estimates of Impact with Education Test Scores as Outcomes." *Institute for Research on Poverty Discussion Papers.* University of Wisconsin.

Appendix A

A complex of factors has led to the growth in youth development programming over the last decade. In each case I cite below, it will be clear why after-school programs might be proposed as a partial remedy for the problem. Therefore, I do not spell out the connection in detail.

- In the 1970s and '80s, there was an increase in the geographic concentration of poverty. Concomitantly, by the early '90s there developed a public perception that there had been collapse in the economic and social structure of inner cities. These changes generated what has been called "spatial mismatch"; job growth occurred largely on the suburban fringe and the unemployed low skill population was trapped in the inner city. The data in support of this hypothesis was strongest for youth.[35]

- There was a feeling that, in poor inner-city neighborhoods, there had been a loss of positive role models because of high unemployment rates (particularly for males) and there had been a decline in civic institutions and civic practices.[36] The growth in drug abuse – particularly the "crack epidemic" in the late '80s – was perceived as a major problem. Accompanying that growth in drug abuse was the growth in the drug trade, particularly involving youth both as users and as purveyors of drugs.[37]

- In the public mind, as reflected in the media, there was a perception of a growing prominence of gangs in the inner city.[38] Neighborhood crime rates increased in the late 1980s. Increased worries about crime led to a huge growth in prisons, stiffer sentences, and to increasing incarcerations – so much so that many studies suggested as much as a third of the African-American male population was either incarcerated or on probation or parole.[39]

- There was a perception of rapid growth in teen pregnancy and unmarried child bearing in the late 1980s and early '90s.[40] In a related development, there was a rapid increase in the portion of all families with children that were single parent families.[41] Concerns about early sexual activity among youth were heightened by the onset of the AIDS epidemic and growth in sexually transmitted diseases.[42]

---

[35]Holzer and Sjoquist (1994).
[36]Anderson (1990).
[37]Sullivan (1989).
[38]Padilla (1992).
[39]In the 1990 report, The Sentencing Project: Young Black Men and the Criminal Justice System, it was reported that almost one in four African American males in the age group 20-29 was under some form of criminal justice supervision. In 1995 follow-up report, Young Black Americans and the Criminal Justice System, Mauer and Huling reported that nearly one in three African American males in the age group 20-29 was under criminal justice supervision (prison, jail, probation or parole).
[40] Moore, Driscoll and Lindberg (1998). Maynard (1996).
[41]Single parent families as a percentage of all families with children: 1970: 12.9 percent; 1980; 21.5 percent; 1992: 29.7 percent; 1998: 31.7 percent. (2000 Green Book. U.S. House Ways and Means Committee).
[42]"… roughly one in four of all young people become infected by an STD by age 21 ... about one in four sexually experienced young people under 20 acquire an STD every year … Furthermore, about one fifth of all AIDS cases in the U.S. are caused by HIV infections contracted while people are teenagers."

- There was a perception, which may deviate from reality, of a decline in the quality of educational institutions.[43] Some of the proposals for increasing educational performance included lengthening the school day and/or the school year.

- There was an increasing concern about skills mismatch, meaning an increase in skills required for jobs combined with an increase in those population elements which were least likely to have the educational experience and attainment to meet the skills requirements.[44]

- In the 1990s, there was an increase in female labor force participation.[45] The increase in the Earned Income Tax Credit appears to have generated a big increase in the labor force participation of single parents.[46] Welfare reform, particularly the 1996 reform, put higher emphasis on work requirements for single parents.[47] All of these developments meant that fewer mothers, whether single parents or in two parent families, were at home providing supervision for children. With welfare reform in particular, it was felt that if the government was to require work, it should also provide institutional programs to supervise the children of working parents during after-school hours.

- In the 1970s, there had been an extraordinary increase in public spending for employment and training programs for youth. However, by the mid 1980s, evaluations of these programs generally concluded that they were not effective in improving the subsequent employment and earnings of young people.[48] Even what seemed a promising educational intervention during the summer months to stem the growth in achievement gaps between low-income and other youth proved a great discouragement and led to a major effort to rethink the structure of youth programming.[49]

- All of these factors may have contributed to an increase in perception of the need for youth programs in the after-school hours. A Carnegie Foundation report found that youth crime occurred primarily in the period from 3 p.m. to 6 p.m. (after school and before returning home) and provided considerable stimulus for concentration on the need for after-school activities.[50]

---

[43] National Commission on Excellence in Education (1984). For an interesting counter-argument about the decline in education, see Krueger (1998).
[44] U.S. Department of Education (1991).
[45] In 1970 the labor force participation rate for females 16 and older was 43.3 percent, in 1980 it was 51.6 percent, in 1990 in was 57.8 percent, and by 2000 it had risen to 60.4 percent.
[46] Ellwood (2000).
[47] Data are now beginning to emerge about the effects of work requirements for women on welfare on their children (Morris et al. 2001). While evidence is limited, the data suggest the effects of increased income and work by single parents may lead to improvement for young children, but to worse outcomes for adolescents. For a striking narrative, see also Boo (2001). These studies lend support to the argument for after-school programs for the children of low-income working parents, particularly single parents.
[48] For a general review, see Betsey, Hollister and Papageorgiou (1985).

[49] For an example of a major discouragement, see Walker and Vilella-Velez (1992).
[50] Carnegie Council on Adolescent Development and Task Force on Youth Development and Community Programs (1992).

- These general problems with regard to the inner city led to an increased concern with "risky behavior" by inner city youth and to a higher interest in programs that promised to reduce the amount of risky behavior. At the same time, the increase in labor force participation spread across income levels so that there were increasing calls for after-school programs for middle-class as well as lower-income youth. The earlier emphasis in the 1980s and early 1990s on inner city youth involved in risky behavior began to be challenged by youth development experts. Efforts should be made, they argued, to emphasize the programs and activities likely to lead to positive outcomes for youth, not just to the diminution of risky behavior.

Appendix B

Below is a list of positive youth development program outcomes presented by Catalano et al. (1999) at the outset of their review. The terms “fosters” or “promotes” applies to what the program’s actions do, whereas the second phrase describes the outcome.

1. Promotes bonding
2. Fosters resilience
3. Promotes social competence
4. Promotes emotional competence
5. Promotes cognitive confidence
6. Promotes behavioral competence
7. Promotes moral competence
8. Fosters self-determination
9. Fosters spirituality
10. Fosters self-efficacy
11. Fosters clear and positive identity
12. Fosters belief in the future
13. Provides recognition for positive behavior
14. Provides opportunities for pro-social involvement
15. Fosters pro-social norms.

Appendix C

Detailed Summaries of Program Evaluations		
About the Program	About the Evaluation	Main Findings
The Howard Street Tutoring Program (Morris, 1990a, b) is a remedial tutoring program created for students in grades two and three who are reading below grade level. When schools become involved in the Howard Street Tutoring Program, a reading specialist or reading teacher becomes the on-site coordinator of the program. This person is trained on how to tutor the students, how to write the lessons and lesson plans to be used by the volunteers, and how to train the tutoring staff. As this is a volunteer program, the staff consists of non-paid adults and college students who must go through the training program before they become tutors. Classroom teachers, using an informal reading inventory, initially assess potential student participants in the program. If the students are performing significantly below grade level, they are placed in the program. Once enrolled, students engage in daily one-hour one-to-one tutoring sessions, which take place every week. (F p.22)	Authors: Morris, D. (1990) and Morris, D., Shaw, B., and Perney, J. (1990). Source: Appalachian State University Reading Clinic and Elementary School Journal. Design: random assignment. (F 22).	The program has been evaluated on a small scale. In two Chicago evaluations, the Howard Street Tutoring Program students outperformed randomly assigned comparison groups in word recognition and word-passage reading. (F 22).
Big Brothers/Big Sisters (BB-BS) of America is the oldest, best-known and most sophisticated mentoring program in the country. Local affiliates create and support one-to-one relationships between adult volunteers and youth living in single-parent households. The program does not specifically target any problem behaviors, but offers a supportive environment and the caring of an adult friend intended to help the youth	Authors: Tierney, Grossman, and Resch (1995). Source: Public/Private Ventures report. Design: Random assignment; eight sites; 18-month follow-up. Sample: 487 program and 472 control youth with follow-up data; youth ranged from age 10 to 16 at baseline (93% between 10 and 14); 23% were girls of color; 15% were White girls, 34% boys of color, and 28% were White boys. (R appendix).	After 18 months of participation, program youth were 46% less likely to start using illegal drugs and 27% less likely to initiate alcohol use during the study than were controls. The results were equally impressive for both boys and girls. The effect was even stronger for the minority participants, who were 70% less likely to initiate drug use. Program youth were 32% less likely to report hitting someone during the previous 12 months.

develop. The local affiliates and matches are governed by carefully established procedures and criteria. National operating standards provide uniformity in recruitment, screening, matching, volunteer training, and match supervision, while allowing for minor variations to accommodate local demands. On average, the youth and adult meet for 3 to 4 hr three times per month for at least 1 year. (R appendix).		Program youth earned moderately higher grades, skipped half as many days of school, skipped fewer classes, and felt more competent about doing their schoolwork than did control youth. The impacts were larger for girls, particularly minority girls. Program participants reported better relationships with peers and parents than did the controls at the end of the study. (R appendix).
The Across Ages program is a comprehensive intergenerational mentoring program for high-risk middle school students. The program was designed to increase resiliency and protective factors for youth in five domains: individual, family, school, peer group, and community neighborhood. The core of the program is the use of older adults as mentors. In addition, the program provides community service activities, classroom-based life-skills training (Positive Youth Development Curriculum) and workshops for parents. (R appendix).	Authors: LoSciuto, Rajala, Townsend, and Taylor (1996). Source: Journal article. Design: random assignment with pre-posttest design; data combined for 3 years. Final sample: 77% of those originally pre-tested, 189 in control, 180 in program without mentoring and 180 in program with mentoring; youth in 6th grade, predominantly African American, and attending three public middle schools characterized by poverty and a high incidence of substance abuse.(R appendix)	The mentoring-added group scored significantly better than the control group on 6 of the 11 scales: attitudes toward school, future and elders, older people, knowledge of elders, reactions to situations involving drug use, and community service. In addition, the mentoring-added group scored significantly higher than the other program group on attitudes toward school, future and elders, and older people. There was a trend favoring the mentor-added group in reported frequency of substance abuse. There was also some positive difference on self reported knowledge of substances, attitudes, and behaviors between mentor group and controls. Mentors may provide an extra benefit in the development of coping and resistance skills. Exceptional involvement with mentors produced the most significant positive outcomes and was correlated with decreased absenteeism. (R appendix).
Girls, Inc.'s Friendly PEERsuasion program is designed to prevent substance use and promote leadership development in young women. The program includes 14 1-hr sessions of hands-on, interactive, and enjoyable activities to teach the short- and long-term effects of substance abuse. The youth are also taught healthy ways to manage stress, recognize media and peer pressure to use drugs and practice skills for making	Author: Girls, Inc. (1993) Source: Girls, Inc. Report Design: Random assignment at four sites to immediate or delayed (control) program participation; pre- and post-measures. Sample: 354 girls ages 11 to 14; 55% African American, 18% White, 14% Latina, and 11% Native American. (R appendix).	At the lead demonstration site, the program was moderately effective in delaying use of harmful substances by the younger participants. This effect was not as strong for delayed participants. At the four sites, there was reduced incidence of drinking among participants who had not previously drank. Early participants were also more likely to leave gatherings where people were drinking, and less likely to have favorable attitudes toward

responsible decisions about drug use to become peer leaders. As peer leaders, youth plan and implement substance-abuse prevention activities for children ages 6 to 10. (R appendix).		drinking. (R appendix).
The Woodrock Youth Development Project (WYDP) addressed eleven positive youth development constructs, including social, emotional, cognitive, and behavioral competencies, bonding, resiliency, self-efficacy, recognition for positive behaviors, prosocial norms, positive identity, and opportunities for prosocial involvement. The program emphasizes life skills and social competence training while also promoting an anti-drug message and providing broad systems support across all three domains. Intervention components include human relations classes, peer mentoring, extracurricular school activities, and structured interactions between students and teachers and children and parents. (C p.76)	Author: LoSciuto, Freeman, Harrington, Altman and Lanphear (1997). Source: Journal of Early Adolescence. Design: The experimental design was a randomized pretest, post-test control-group design. Classrooms within four Philadelphia schools were randomly assigned to program or control conditions. Sample: The sample of 367 students (130 experimental, 237 control) remained after 19% of the original sample of 453 students from ages six through 14 was lost to attrition. Participation in the final measurement sample were 46.9% female, 44.4% Latino, 19.9% Caucasian, 11.4% African American, 11.2% Asian, 9.3% mixed or "other" ethnic identity and 1.9% Native American C p76).	In the published evaluations, the authors reported significant positive differences between the intervention and control groups for the younger sample, and mixed results for the older sample, with one statistically significant (non-behavioral) outcome in the wrong direction. For the younger group, for three of the five dependent variables (outcomes as a function of the intervention) the results showed statistically significant improvements for intervention group compared to the control group. For the older group, there was one positive, statistically significant behavior change for the intervention group compared with the control group. There were no significant differences for drug use in the last year, self-esteem or attitudes about race relations. (C p.76).
Quantum Opportunities Program is a community-based year-round, multi-year, multi-service youth development program for students from families receiving public assistance. Twenty-five students began the program in the 9^{th} grade, and continued with the program until the end of 12^{th} grade. The explicit goals of the program are to foster academic and social competencies. Each year, the students participated in 250 hr of education-related activities (tutoring, computer-assisted instruction, homework assistance, etc.) 250 hr of development activities (community service projects, helping with public events, holding regular jobs). Students received hourly stipends and bonuses for completing each segment of the program. (R appendix)	Authors: Hahn, Leavitt, and Aaron (1994). Source: Unpublished report sent with survey. Design: Random assignment at four sites before recruitment; 5-year longitudinal study, beginning in 9^{th} grade to 1-year post high school. Sample: 100 experimental and 100 control participants at pretest; 88 experimental and 82 control participants at last follow-up. All were from families receiving public assistance. (R appendix)	The rate of differentiation between two groups accelerated after the first two years in the program. By the end of high school, program participants showed significant increases in academic skills and education expectations, but no differences in drop out or childbearing rates, grades, or knowledge about contraceptives and AIDS. There were differences in the extent of success by site. One year after the end of the program, there were significant differences in dropout and childbearing rates: Program participants were significantly more likely to have graduated from high school or received their graduate equivalency diploma (63% vs. 42%), be in a postsecondary school (43% vs. 16%), and have fewer children (24% vs. 38%). Also, they were less likely to have been in trouble

		with the police within the previous 12 months. (R appendix).
The Summer Training and Education Program (STEP), a demonstration project, sought to reduce dropout rates, stop summer learning loss, and prevent teen pregnancy by enhancing the Job Training Partnership Act (JTPA) program. For 6 to 8 weeks during the summer, adolescents worked half-time at jobs (90 hr) and attended academic classes half the day (90 hr). In addition, two mornings a week were devoted to life-skills training. The adolescents were paid for the time spent in the classroom as well as on their jobs. Adolescents were encouraged to participate for two consecutive summers (R appendix).	Authors: Walker and Vileela-Velez (1992); Grossman and Sipe (1992). Source: two Public/Private Ventures reports. Design: Random assignment; five sites; longitudinal study. Sample: 1,263 program and 1,347 control adolescents from two cohorts; 48% male, 86% youth of color, and most 14 and 15 years old; all JTPA eligible and performing below their grade level in reading and math (R appendix).	STEP's summer effects were positive; reading and math test scores after the first summer were about half a grade higher for program youth. In addition, program youth received higher scores on knowledge tests of responsible social and sexual behavior. The program youth had high attendance rates and high return rate (75%) for the second summer. However, the gains did not hold after the summer, during the school year, or long term. STEP youth had the same dropout rate, college entrance, teen pregnancy, and employment rate as controls. However, the treatment youth continued to know more about contraceptives. (R appendix).
Louisiana State Youth Opportunities Unlimited (LSYOU) is a dropout prevention program for 14- to 16-year-olds. For 8 weeks during the summer, participants live on the Louisiana State University campus. They spend half the day in academic instruction in math and reading, and the other half of the day working at sites throughout the campus. Participants earn minimum wage and are required to open a savings account. Other services include counseling, study-skills training, healthcare, recreation, field trips, and speakers. (R appendix).	Authors: Shapiro, Gaston, Hebert, and Guillot (1986). Source: Educational Resources Information Center (ERIC) document. Design: Random assignment to LSYOU or regular JTPA work program; pre- and posttest measures. Sample: 51 control and 94 participant youths; all economically disadvantaged and identified as potential dropouts. (R appendix).	At the end of the summer, program participants scored higher on standardized math tests than they did at the beginning of the summer. In comparison, the control group experienced declines in their math skills during the summer. Although LSYOU participants experienced declines in their reading skills, their declines were significantly less than those experienced by the control group. Program youth expressed a significant increase in their intention to stay in school from the beginning to the end of the program. They also showed greater career maturity at the end of the program than the control youth. (R appendix).
The Midwestern Prevention Program is a social influence based drug prevention program. The program includes mass-media coverage, community organization, parent education about parent-child communication skills, and a 10-session resistance skills training school-based program. Program designed for students in their 1st year of middle school (either	Authors: Johnson et al. (1990). Source: Journal article. Design: Random assignment by class; delayed intervention control pre-, post-test over 3 years. Sample: eight Kansas City communities. Program delivered to 1,607 students in either 6th or 7th grade, with panels through Grades 9 to 10; 1,105 in final analysis; 77% White, 19%	Results revealed significant main effects for tobacco and marijuana use in the last month but no significant effect on alcohol use at the 9th and 10th grade levels. Prevalence rates for all three substances increased over time, but the rate of increase for tobacco and marijuana was significantly less for adolescents in program schools. For Grades 9 and 10, the percentage of students

Grade 6 or 7). The program was introduced sequentially into communities over a 6-year period. (R appendix).	African American, 2% Hispanic, and 1% Asian. (R Appendix).	reporting use over a 30-day period was: 25% cigarette for program versus 31% for control schools; 34% alcohol for program versus 33% or control; and 12% marijuana for program versus 29% for control. These results were measured three years after the administration of the school-based program. In addition, the program was equally effective in both high- and low-risk populations for smoking and marijuana but not alcohol use. (R appendix).
Creating Lasting Connections is a church-based, 5-year substance abuse prevention demonstration program designed to delay the onset and reduce the frequency of alcohol and drug use among high-risk 12-to 14-year-olds by positively impacting resiliency in three domains: church community, family, and individual. The program was administered in suburban, rural, and inner-city settings. Major program components included church community mobilization, parent and youth training, early intervention, and follow-up case management services. (R appendix).	Authors: Johnson et al. (1990). Source: Journal article. Design: Random assignment; pre- and posttest measures. Sample: 49 program parents and 59 program youth, and 48 control parents and 61 control youth; no other information on participants provided. (R appendix).	The evaluation found that the program produced positive moderating effects on alcohol and drug use among youth as a result of conditional relationships, with changes in family-level and youth-level resiliency factors targeted by the program. Statistically significant youth-level moderators included increased communication about alcohol and drug use and school work, youth's reported bonding with parents coupled with decreased conflict, pathology and estrangement in the family, and greater acceptance of conventional values. (R appendix).

Sources: Roth et al. (1998), Catalano et al. (1999), Fashola (1998).