

# *Social Policy Report*

*Giving Child and Youth Development Knowledge Away*

Volume XXII, Number 2

2008

## **After-School Programs and Academics: Implications for Policy, Practice, and Research**

Robert C. Granger, Ed.D.  
William T. Grant Foundation

### Abstract

No Child Left Behind (NCLB), with its emphasis on standards-based accountability, has put educators under considerable pressure to improve student academic outcomes. Much of the funding for after-school programs comes from education budgets and is administered by state and local education agencies. Consequently, after-school programs are often expected to incorporate academic achievement as an important goal.

This focus on academic achievement is producing heated debates among after-school practitioners, policymakers, and researchers. Should after-school programs be required to have a positive impact on academic outcomes? Will such an expectation crowd out other important goals and turn after-school programs into an unappealing version of the school day?

This report focuses on the growing program-evaluation literature, observational studies, and commentaries and statements of program standards by practitioners and advocates in the context of this debate. I begin by showing that after-school programs can have positive academic effects, though many do not. To understand the ingredients of an effective program, I examine empirical reviews of program evaluations, observational studies, and practitioner writings. It is clear that to be effective, programs should actively involve participants, be intentional about their goals, and focus on the interactions between youth and staff. If positive academic outcomes are one of those goals, programs may need to include specific activities that are focused on academic achievement, but the approach should build on the opportunities presented by the out-of-school setting. The report concludes by identifying some promising approaches to program improvement and arguing that research on ways to intervene to improve program effectiveness is the highest priority.

## Social Policy Report

### Editor

Lonnie Sherrod, Ph.D.  
sherrod@srcd.org

### Associate Editor

Jeanne Brooks-Gunn, Ph.D.  
brooks-gunn@columbia.edu

### Director of SRCD Office for Policy and Communications

Mary Ann McCabe, Ph.D.  
mmccabe@srcd.org

### Managing Editor

Amy D. Glaspie



### GOVERNING COUNCIL

Arnold Sameroff	Suniya Luthar
Aletha Huston	Ann Masten
Greg J. Duncan	Robert B. McCall
Judith G. Smetana	Ellen E. Pinderhughes
Oscar Barbarin	Elizabeth Susman
Patricia Bauer	Lonnie Sherrod
Marc H. Bornstein	Mary Ann McCabe
Melanie Killen	Alisa Beyer

### POLICY AND COMMUNICATIONS COMMITTEE

Cheryl Boyce	John Ogawa
Dale Farran	Cassandra Simmel
Barbara H. Fiese	Louisa Tarullo
Bonnie Leadbeater	Lonnie Sherrod
Amy Lowenstein	Mary Ann McCabe
Karlen Lyons-Ruth	
Joseph Mahoney	

### PUBLICATIONS COMMITTEE

Anne D. Pick	Joan Grusec
Ann Easterbrooks	Arnold Sameroff
Sandra Graham	Gene Sackett
William Graziano	Judith G. Smetana
Brenda Jones Harden	Lonnie Sherrod
Amy Jo Schwichtenberg	

## From the Editor

The attached article by Robert Granger offers a much needed summary of the research on after-school programs. Federal funding for after-school programs has grown considerably in recent years in part because it was pointed out that youth are likely to get into trouble, including crime, between the hours of 3 and 11 pm—that is, after school. We often forget that after-school programs provide an important child care function. School-age and adolescent children should not be left unmonitored for the hours between schools' closure for the day and when parents get home from work. I am a firm proponent of the Positive Youth Development (PYD) approach to both research and policy. We should strive to promote positive development rather than just prevent negative or risky behaviors, and after-school programs follow this approach. They appropriately seek to use young people's time constructively. However, we should not minimize the importance of their role in helping to keep kids out of trouble; academic and other such gains are "icing on the cake" in my view.

Icing considerably improves the cake so that expecting gains from after-school programs is perfectly appropriate. The issue addressed so effectively in this article is that research has to begin by addressing the question: gains in what area? One of the aspects of this article that is most important is its willingness to consider a variety of different possible positive outcomes of after-school programs. A critical point for research is that the outcomes evaluated in research should map back onto the program's characteristics. We do not fully understand how gains in one area may generalize to others. Art activities and sports could lead to general academic gains. Research on after-school programs provides a venue for examining such questions. However, it is more likely that we will see academic gains if the program has an academic component to its curriculum. Hence a first step in any research study is to spell out the program's theory of change; why do they do what they do and what do they expect to result? Often even asking this question can lead to important program refinement.

Another important point made by this article is the tremendous demand for research, and research studies that allow one to address causality are sorely needed. As I have said, after-school programs did not arise because we knew how to use that time constructively. This accounts in part for the diversity we see across program characteristics. As a result, research must now address this issue by using this diversity to investigate relationships between curriculum or activities and outcomes. Additionally, as Granger points out, studies need also to ask which types of programs are more engaging to young people. Activities are not likely to lead to outcomes if the youth are not actively—and ideally we hope passionately—involved in the activity. This relates to implementation and dissemination which is addressed well in the commentary by Joseph Durlak. The need for staff training is another important point made by this paper, and again we need research that demonstrates what works.

Brooke and I hope that this summary of research by Robert Granger will contribute to setting a research agenda in this field, and thereby contribute to future policy and program development.

## After-School Programs and Academics: Implications for Policy, Practice, and Research

**Robert C. Granger, Ed.D.**  
**William T. Grant Foundation**

The standards movement in K–12 education became prominent in the 1990s, and standards-based accountability is codified in the requirements of the current version of the elementary and secondary education act No Child Left Behind (NCLB).

NCLB and the standards movement have put policymakers and practitioners under considerable pressure to improve student academic achievement. Increased federal and state funding for after-school programs has come through education funding streams, particularly since NCLB's passage in 2002. For example, the 21st Century Community Learning Centers program, the main source of federal support dedicated to after-school, is now funded through NCLB. This has created a collateral push to emphasize academic outcomes in after-school programs.

The emphasis on academic performance has generated heated debates in the after-school field. One argument is that programs can and should have a positive impact on academic outcomes. Another is that a focus on academics will turn a rich field of youth services into a poorly implemented extension of the school day (Halpern, 2004).

The evolving rationale for the recent expansion of after-school programs in California is a good example of how this debate plays out politically. In 2001, Californians debated Proposition 49, a ballot initiative designed to

provide an extraordinary expansion of state support for after-school programs. This ballot initiative proposed adding \$450 million of state money to an existing state program that was funded at \$100 million per year. Advocacy and voter education materials at the time presented a clear and consistent rationale for the program. Expansion was meant to keep children safe in the “prime-time for crime” hours of 3:00 to 6:00 p.m. and to enable parents to work (EdSource, 2002; California Secretary of State, 2002). The legislation received wide support and passed in 2002 as the After School Education and Safety Program Act, with a launch date tied to improvements in the state budget.

In 2006, an improved fiscal situation triggered the Act's implementation, and the state created a planning process to shape the details of its after-school expansion. Given the climate surrounding NCLB, the most contentious debates involved the academic outcomes that would be set as targets for the expanded program (Ames, 2007).

Informing the larger national debate is a growing program evaluation literature on the effects of after-school programs. That literature is augmented by a number of observational studies using either structured observational protocols or less-structured qualitative methods. There are also numerous commentaries and statements of program standards by practitioners and advocates in the field about the features distinguishing effective from less-effective programs. The purpose of this report is to consider this work, focusing specifically on its implications for research, policy, and practice.

I argue the following:

- Well-done syntheses evaluating the effects of after-school programs show that some programs

have a positive impact on academic, social, and emotional outcomes;

- The same research shows that many programs do not improve the academic outcomes of the youth in the program beyond those of youth in a control group who had access to other services in the community;
- There is a consensus among practitioners regarding what effective program practices are, and observational research is beginning to refine that consensus; and,
- There is a great need for research-proven ways to intervene and improve program effectiveness.

This situation creates fertile ground for improving the after-school field and our fundamental knowledge regarding how after-school programs shape youth development.

### **Do After-School Programs Positively Affect Academic Outcomes?**

In the past decade there have been several narrative and empirical reviews of the effects of after-school programs (Bodilly & Beckett, 2005; Durlak & Weissberg, 2007; Fashola, 1998; Hollister, 2003; Lauer et al., 2006; Little & Harris, 2003; Zief & Lauver, 2006) and a “synthesis of the syntheses” is in progress (Cooper, Patall, Tyson, & Valentine, 2008, p. 3). Although the reviews vary in their conclusions regarding academics, the most reliable reviews show that on average programs have positive

impacts on important academic, social, and emotional outcomes.

In coming to this conclusion, I am relying heavily on the results from the three recent empirical reviews that used

the techniques of meta-analysis.

This review approach employs quantitative techniques to statistically combine the results from multiple studies, and such systematic reviews of research are at the top of most hierarchies of evidence regarding the effects of social policies and programs (Flay et al., 2005). The high regard for the approach comes from its transparent, replicable

methods for summarizing results (Cooper & Hedges, 1994; Glass, McGraw, & Smith, 1981; Hedges & Olkin, 1985).

This does not mean that meta-analysts always agree on what research shows. That is the situation with the after-school literature, in which two of the recent meta-analyses found positive effects on a range of outcomes (Lauer et al., 2006; Durlak & Weissberg, 2007) and one found no effects (Zief & Lauver, 2006). One goal of this report is to explore possible reasons for the different conclusions in a way that is informative for policymakers, practitioners, and researchers.

One reason conclusions vary is that the analysts were pursuing related but somewhat different questions. Zief and Lauver (2006) wanted to understand the effects of typical after-school programs that included academic and recreational activities. Lauer et al. (2006) wanted to understand the effects of out-of-school<sup>1</sup> academic

Well-done syntheses evaluating the effects of after-school programs show that some programs have a positive impact on academic, social, and emotional outcomes.

programs for at-risk youth. Durlak and Weissberg (2007) wanted to understand the effects of after-school programs meant to improve personal or social skills. These different goals led the analysts to develop different criteria for the studies included in each review. In effect, each took a different slice of the out-of-school field. Table 1 provides details on the inclusion criteria used by each team.<sup>2</sup>

Their mix of inclusion criteria led Zief and Lauver (2006) to review 5 studies, Lauer et al. (2006) to review 35 studies, and Durlak and Weissberg (2007) to base their main analyses on the effects from 66 studies. The three examinations had some overlap in the studies reviewed. All of the five studies reviewed by Zief and Lauver are in the Durlak and Weissberg review, as are four studies from the Lauer et al. review. There is no overlap in studies between Zief and Lauver and Lauer et al.

### The Average of Effects

Zief and Lauver (2006) found that on average the studies in their review showed no significant effects on academic, social, or emotional outcomes.<sup>3</sup> In contrast, Lauer et al. tested for positive effects in reading and mathematics achievement and found positive results for both. They found a positive effect size for reading of .13 and for mathematics of .17. Durlak and Weissberg (2007) found average positive effects for seven of the eight outcome categories they tested. These positive findings in effect size units were .16 for achievement tests, .11 for school grades, .19 for positive social behaviors, .18 for reduced

problem behaviors, .11 for drug use, .14 for feelings the youth had about school, and .34 for positive youth views of themselves. The only outcome category tested in the Durlak and Weissberg review for which they did not find an average positive effect was school attendance, effect

size .10 and not significant at the .05 level.

While these average effects are not large, they are consistent and positive. They also compare favorably to meta-analyses of other program interventions that may compete with after-school programs for funding, such as

summer school, for which Cooper, Charlton, Valentine, Muhlenbruck, and Borman (2000) found a median effect size of .26, and mentoring programs, for which David DuBois (2002) found an average effect of .14 on academic test performance.

### The Range of Effects

Lauer et al. (2006) and Durlak and Weissberg (2007) found consistently positive effects *on average* across the studies they reviewed. However, the majority of the studies in each review, in addition to the five studies in the Zief and Lauver (2006) review, did not find evidence that the program made a difference when compared to the outcomes for the control group.

The Lauer et al. (2006) review reports on 42 different estimates of effects on reading achievement and 33 estimates for mathematics achievement. Within reading, 11 of the estimates were positive, 3 were negative, and 28 were “null,” suggesting that the program did not

There is a great need for research-proven ways to intervene and improve program effectiveness.

Table 1  
*Comparison of three meta-analyses of the effects of after-school programs*

Research Team	Number of studies in meta-analysis	Inclusion criteria	Main findings
Zief & Lauver (2006)	5	<p>After-school programs operating on a regular basis after school during the school year that combine recreation and/or youth development programming with academic services. Mentoring and tutoring programs excluded.</p> <p>Programs serving youth enrolled in public or private K–12 schools; not specifically youth with special needs.</p> <p>Well-implemented experimental designs.</p>	<p>No effects on average for social, behavioral, or academic outcomes. Of 97 impacts measured across the 5 studies, 84 percent show no significant difference between youth in program and control groups.</p>
Lauer, Akiba, Wilkerson, Apthorp, Snow, & Martin-Glenn (2006)	35	<p>Educational intervention delivered outside the school day. Tutoring and summer school programs were included but not mentoring programs.</p> <p>K–12 students at risk for school failure. The definition of “at-risk” included behavioral indicators (low performance on standardized tests) and status indicators (low SES); not specifically youth with special needs.</p> <p>Experimental and quasi-experimental designs.</p>	<p>Small but statistically significant positive effects on both reading and mathematics; some program features predicted results for some outcomes (e.g., tutoring improved reading achievement). No difference for after-school vs. summer school. About one-third of all estimates were positive, most were null.</p>
Durlak & Weissberg (2007)	73 (most analyses based on 66 evaluations that measured effects immediately following the program)	<p>Programs operating after school during the school year that had as one goal the development of personal or social skills. Excluded programs primarily focused on academics, including tutoring programs, and did not include mentoring programs or summer programs.</p> <p>Programs serving youth enrolled in public or private K–12 schools; not specifically youth with special needs.</p> <p>Experimental and quasi-experimental designs.</p>	<p>Usually small but statistically significant positive effects on seven of eight social, behavioral, and academic outcomes. No significant effect on school attendance. Programs that explicitly focused on specific skills, with a sequenced curriculum and students actively involved, were most successful.</p>

*Note.* From Zief & Lauver, 2006; Lauer et al., 2006; Durlak & Weissberg, 2007.

make a difference. (A null finding indicates that the confidence interval for the estimated average effect included zero). For mathematics, Lauer et al. found 11 positive estimates, 1 negative, and 21 null findings. Thus, less than a third of the estimates were positive and the most frequent finding was that programs made no net impact.

Furthermore, finding effects for one outcome did not assure effects for the other. In the Lauer et al. (2006) review, 24 of the estimated effects were for both reading and mathematics, presumably because the program meant to improve outcomes in both areas (24/42 for reading and 24/33 for mathematics). Of these, 3 analyses showed positive effects on both outcomes and 11 found null effects for both. This implies that 3 programs were uniformly strong and 11 were relatively weak. But for 10 programs, the news was mixed. Four found a positive effect for mathematics and a null effect for reading and six found a positive effect for reading and a null effect for mathematics.

In the after-school field, it is tempting to characterize a program as being of high or of low quality. The Lauer et al. (2006) findings suggest that it is more appropriate to consider quality as something that varies within a program, with many programs (10 out of 24 in this case) being more effective in one area than another.

Durlak and Weissberg (2007) also found considerable variation in effects across studies. They created estimates of effects of all outcomes combined (a grand mean for each study) and also estimated effects for clusters of outcomes that go together conceptually (e.g., various measures of at-risk behavior). Their report has detailed tables for all outcomes combined and for each outcome category.

In contrast to Lauer et al. (2006), Durlak and Weissberg (2007) did not compute confidence intervals for their various effect sizes, so it is harder to separate the estimates into positive, negative, and null findings. As a proxy for those categories, I tallied the number of the estimated effect sizes that were greater than .10, less than -.10, or between .10 and -.10. Table 2 contains the pattern of effects Durlak and Weissberg found for their academic outcomes. Like Lauer et al., the average of effects across studies for academic outcomes is positive, but the modal result is the null finding. This means that the relatively fewer positive effects are large enough to outweigh the more frequent null and negative effects.

...it is tempting to characterize a program as being of high or of low quality... it is more appropriate to consider quality as something that varies within a program.

...the average of effects across studies for academic outcomes is positive, but the modal result is the null finding.

### Why Do Some Programs Create Effects While Others Do Not?

All three of the teams of meta-analysts planned to look

within their analyses to try to identify features of the programs or their circumstances that predicted successful impacts. Having found no effects on average for the five studies in their review, Zief and Lauver (2006) abandoned their search for such factors. However, the Lauer et al. (2006) and Durlak and Weissberg (2007) reviews had enough variation in impacts and large enough samples of studies to explore the predictors of variation. For example, did a program need to operate in a particular way to get positive effects? Did it need a particular content or curriculum?

Lauer et al. (2006) examined moderators<sup>4</sup> that can be shaped by policy and practice (e.g., length of program). Their choices came from prior empirical findings but were not explicitly motivated by a theory regarding why each factor should predict the pattern of results.

Conducting analyses separately for reading and math-

ematics achievement, they examined the degree to which levels of the following features predicted their results: program time frame (e.g., school year vs. summer), grade level of participants, focus (academic vs. academic and social), duration of the program, grouping strategy (e.g., one-on-one or another strategy), and the assessed quality of the study (high, medium, or low). While grade level, focus, program duration, grouping structuring, and study quality were significant moderators, the pattern was not clear across outcomes. For example, one-on-one tutoring and a mixed-group strategy both predicted a positive effect for reading achievement. But for mathematics achievement, one-on-one tutoring was the one strategy that did not predict positive results. In sum, various features seemed to matter at different times, but the lack of a consistent pattern across outcomes suggests that something else is driving the results.

Durlak and Weissberg (2007) took a different approach

Table 2  
*Tally of program effects by academic outcome and effect size*

Outcome	Effect Size		
	> .10	-.10 to +.10	< -.10
School bonding (n=31)	11	13	7
Achievement tests (n=22)	11	9	2
Grades (n=26)	9	15	2
School attendance (n=21)	9	11	1

*Note.* From Durlak and Weissberg, 2007. Estimates for the individual studies can be found in Appendix C of their article.

to their search for the factors that might predict variation in effects. Drawing on developmental theory regarding the importance of active youth involvement and prior empirical reviews showing that focused, skill-based programs are more likely to show effects (Lipsey, 1992), Durlak and Weissberg grouped their studies into two clusters for comparison. In one cluster, they placed programs focusing on specific social and personal skills that employed sequential learning activities to develop those skills and had youth actively involved. They referred to such programs as “evidence-based” given the Lipsey (1992) results and in a subsequent paper (Granger, Durlak, Yohalem, & Reisner, 2007) use the acronym SAFE (Sequenced, Active, Focused, and Explicit)<sup>5</sup>. In the other cluster were the studies of programs that did not have all these features.

Not all the individual evaluations of programs with the SAFE features showed positive effects, while some in the non-SAFE cluster did. However, when grouped together, on average, programs that had SAFE features showed positive effects for every outcome but school attendance, and the cluster of programs without these features showed no positive effects for any outcome. The results from the programs with SAFE features drive the overall positive picture for all programs in the Durlak and Weissberg review. Table 3 summarizes these findings.

The SAFE features offer one explanation for the difference between the findings by Zief and Lauver (2006) and Durlak and Weissberg (2007). Recall that the five studies reviewed by Zief and Lauver are in the Durlak

Table 3  
*Statistically significant positive effects for after-school programs*

Effects	Programs overall	SAFE Cluster <sup>b</sup>	Other Cluster <sup>b</sup>
<i>School performance</i>			
Achievement tests	√ <sup>a</sup>	√	...
School grades	√	√	...
School attendance	...	...	...
<i>Social behavior</i>			
Social skills	√	√	...
Problem behaviors	√	√	...
Reduced drug use	√	√	...
<i>Attitudes and beliefs</i>			
Bonding to school	√	√	...
Self-esteem	√	√	...

*Note.* From Durlak & Weissberg, 2007.

<sup>a</sup>A check indicates positive effects.

<sup>b</sup>The number of evaluations used in each cell of this table was equal to the number of evaluations that measured each outcome. In no case was the number of evaluations for a particular outcome lower than 20. See Table 4 of the full report for the specific findings for these clusters and Table B1 in Appendix B of the full report for details on each of the reviewed programs.

and Weissberg review. None of the five met the SAFE criteria.

Durlak and Weissberg (2007) also present their effect size estimates for each cluster separately and the effect sizes for the SAFE cluster are impressive. For example, the average effect size was .31 for achievement tests and .24 for school grades in the SAFE cluster as compared to .03 for achievement tests and .05 for grades in the cluster of

programs that did not meet the SAFE criteria. Another striking finding in the Durlak and Weissberg review is that positive effects tended to come in bundles. Recall that Lauer et al. found modest consistency in math and reading effects when they had data on both. In the Durlak and Weissberg review, on average, the individual evaluations in the SAFE group showed positive effects for 70 percent of the outcomes they assessed. In contrast, the studies in the other cluster revealed positive effects for only 25 percent of the outcomes (and no positive effects when the evaluations were grouped together and a grand mean was computed for the group).

Like Lauer et al. (2006), Durlak and Weissberg (2007) also examined other features that might predict the variation in effects. They tested the predictive power of the presence of an academic component in the program, active parent involvement, and the grade level of the participants. While the presence of the SAFE factors was a significant predictor for all outcome categories, of these other features, the only positive finding of note was that the presence of an academic component was a strong predictor of positive effects on achievement tests.

Taken together, the findings from these two reviews suggest that the SAFE features are a much better predictor of program effectiveness than other structural features discussed in the literature. One reason this is notable is that the findings from these reviews are stronger than

... the SAFE features are a much better predictor of program effectiveness than other structural features discussed in the literature.

one can produce from analyses in which program features are simply related to measures of youth performance (e.g., students with higher achievement are found in programs with certain characteristics).

The strength of the meta-analytic findings is that they relate features to experimental or quasi-experimental estimates of net changes in performance.

Before discussing the implications of these findings for policy and practice, it is useful to ask if practitioners promulgating program standards and researchers using observational protocols agree that the SAFE features are found in effective programs.

### **The Relationship of the Meta-Analytic Findings to Observational Research and the Views of Practitioners**

In 2003, the Forum for Youth Investment reviewed 13 statements of standards for program quality (Forum for Youth Investment, 2003). Most were developed by practitioner organizations or accrediting groups serving a specific subsection of the youth field (e.g., camps, school-age child care, youth leadership programs). Many of the standards included items addressing features that were found to be important in the Durlak and Weissberg (2007) review. That is, they tended to empha-

size the importance of explicit goals, a clear focus, and activities that actively involve youth in developmentally appropriate ways.

In 2007, the Forum extended this work by releasing a review of nine observational instruments designed to measure youth program quality (Yohalem & Wilson-Ahlstrom, 2007); see endnote for a list of these instruments. To complete the review, Yohalem and her colleagues examined published and unpublished information on the instruments, interviewed the developers and, in most cases, interviewed practitioners who had used each instrument.

Researchers and practitioners worked together to develop most of the instruments in the review. Many of the instruments have their roots in early childhood assessment, while others draw more heavily on the youth development and/or education literatures. All of the instruments rely on observing a program's daily operations. They emphasize interactions among staff and youth, while also assessing social norms, physical and psychological safety, skill-building opportunities, and program routine or structure. The Yohalem and Wilson-Ahlstrom (2007) review labels these *core concepts*.

In general, there is congruence between what the instruments measure and Durlak and Weissberg's (2007) active, focused, and explicit features. Whether they call for activities that are project-based and experiential, or that "involve youth in engaging with...materials or ideas or improving a skill through guided practice," six of the nine instruments emphasize the potential importance of active learning techniques (QAS, 2004-2005; YPQA, 2005, p. 18). All but one addresses the focused feature, with items that call for "practice/a progression

of skills," or activities "designed to achieve program goals/objectives" (OST, 2005, p. 28; QAS, 2004-2005, p. 8). Six of the instruments underscore the importance of clear expected learning goals and content that is "well developed, detailed, and reflects...standards" (APT, 2005; QAS, 2004-2005, p. 24). In sum, the developers of the observational instruments agree that being explicit about program goals, implementing activities focused on those goals, and getting youth actively involved are practices of effective programs.

Agreement around Durlak and Weissberg's (2007) sequenced feature is less clear. In the Durlak and Weissberg review, a program was coded as sequenced if it used a sequential set of activities to achieve its objectives for personal or social skill development. Such an approach was often achieved by using or adapting an established curriculum. While the program might achieve its ends by working with the children's interests, the sequence of activities was largely adult-determined. In contrast, three of the observational instruments include items that emphasize allowing children to choose activities, rather than following a predetermined sequence. These items call for a flexible structure that is "adaptable and responsive to individual wants, needs, talents, moods" or one in which children "move smoothly from one activity to another" at their own pace (PQO, 2006, p. 2; POT, 2001, p. 16).

Although many of these tools are in an early stage of development, the review found that practitioners believe that the measures yield data that can inform program improvement efforts. Because many of the instruments are relatively new, documented information about their technical properties is limited. Most have some data showing that if two different observers watch the same program

## Improving After-School Programs: How do we get there from here?

Joseph A. Durlak

Loyola University Chicago

Robert Granger has offered an excellent synopsis not only of the empirical literature on the impact of after-school programs, but also on the political landscape affecting current practices and the practical challenges that will confront change efforts. For example, on the positive side, recent reviews indicate that after-school programs can significantly improve academic, social, and emotional outcomes, and there is some consensus on several of the components that contribute to effective programs (e.g., explicit goals, a clear focus, and developmentally appropriate activities that truly engage youth). These findings reflect growing scientific support for the role of after-school programs in promoting positive youth development.

On the negative side, however, Granger correctly notes that program impact is strikingly uneven. Some programs are effective while many others are not. Moreover, he notes that efforts to improve current programs must surmount unfavorable working conditions such as a workforce that is predominantly “young, untrained, and prone to frequent turnover.” Granger stresses that the primary issue facing the after-school field is how to improve current programs.

I agree, and in this commentary want to alert readers to the extensive empirical literature on program diffusion that offers helpful guidelines about how to disseminate, implement, evaluate, and sustain innovations in real-world settings. Because more than 5,000 studies have been conducted on program diffusion, I can only hit the highlights here (see Berman & McLaughlin, 1976; Durlak & Dupre, 2008; Fixsen et al., 2005; Greenhalgh, et al., 2005; Rogers, 2003). Yes, we can change settings, but the process of doing so is fraught with potential problems and requires time, patience, necessary resources, and, most important, the willingness and ability to collaborate with front-line providers.

In brief, here are a few of the major themes from research on program diffusion that would apply to the situation in which a consultant or researcher is working to incorporate empirically supported programs or practices into an after-school program. Good intentions are not enough. Although money is an issue, the most important resources in an after-school setting are the people (staff and youth), and their talents and values should be

considered in the change process. If you want to influence real-world practices, you must make a concerted effort to inform and to collaborate with front-line providers and to support and to problem solve with them as new programs or practices are introduced into their setting. The eventual consumers of any after-school program, which means youth and their parents, should have some input. After-school staff must recognize a need for change, reach consensus in terms of their willingness to try something new, and contribute to how any innovations will be conducted. It is critical to find the right balance between fidelity to evidence-based programming and adaptation to fit local needs and values. Staff should have realistic expectations of a program's intended benefits, and be shown how to monitor and interpret change appropriately as it occurs. They need hands-on training and on-going technical assistance that is best provided via personal contact to deal with the inevitable problems that arise whenever something new is tried. They will need more training and assistance as the innovation's complexity and extensiveness increases. The wisdom and experience of front-line staff should be respected, and opportunities for them to assist each other should be maximized.

There are major risks in not applying useful object lessons from the diffusion literature. To the extent that these lessons are ignored, there will be diminishing returns on efforts to improve current programs. For example, there is the likelihood that fewer after-school programs will attempt to initiate any genuine change, fewer of those wishing to change will be able to implement new programming sufficiently to achieve their intended goals, and, in the long run, fewer effective innovations will be sustained after researchers or consultants leave the scene.

### References

- Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., & Ory, M. et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*, 443-451.
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum, 40*, 345-370.
- Durlak, J. A. & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*.
- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved November 1, 2006, from [http://nirn.fmhi.usf.edu/resources/publications/Monograph/pdf/monograph\\_full.pdf](http://nirn.fmhi.usf.edu/resources/publications/Monograph/pdf/monograph_full.pdf).
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). *Diffusion of innovations in health service organizations: A systematic literature review*. Oxford: Blackwell.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th Ed.). New York: Free Press.

practices, they will score the instrument similarly (this is known as interrater reliability). Only a few have data on the extent to which ratings done by the same observer on different days stay the same (test-retest reliability). All of the instruments contain items that practitioners judge as important to assessing program quality (face validity), and several measures have shown a relationship between their scores and youth outcomes (predictive validity). These relationships are encouraging, although no instrument has data showing that improved scores on what it measures translates into improved youth outcomes.

The observational research on practices occurring within after-school programs is a welcome addition to the studies assessing the net effects on youth outcomes reviewed in the meta-analyses.

The observational research on practices occurring within after-school programs is a welcome addition to the studies assessing the net effects on youth outcomes reviewed in the meta-analyses. The standard practice within such program evaluations is to gather data at the participant level to estimate net program effects on youth outcomes. At times, researchers also gather data on program operations to assess the fidelity of implementation of a program model. When multiple studies are combined, that data on program design or operations can be useful. This is the approach Lauer et al. (2006) and Durlak and Weissberg (2007) used to predict variations in program effects. However, many of the features used in such analyses are only proxies for in-program youth experiences. For example, if focused programs tend to achieve positive results, what are the practices and experiences that mediate the effects on youth? The observational studies are starting to refine our understanding of such questions, in part by helping us understand how the

interpersonal interactions occurring in programs may contribute to effects. It is beyond the scope of this report to review this work, but interested readers are referred to the “Program Quality Observation” instrument (Vandell & Pierce, 2006), the “Youth Program Quality Assessment” instrument (Smith, 2005), the previously mentioned review of the nine instruments (Yohalem & Wilson-Ahlstrom, 2007), and the recent ethnographies of youth programs by Reed Larson, Bart Hirsh, and their respective colleagues (Larson, 2006; Larson & Hansen, 2005; Larson, Hansen, & Moneta, 2006; Hirsch, 2005).

### **What Do We Know About Intervening to Improve the Effectiveness of After-School Programs?**

Given the uneven program effectiveness documented in these empirical reviews, I and others have argued that learning how to intervene effectively to improve programs is now the primary issue facing the field (Granger, Durlak, Yohalem, & Reisner, 2007). The availability of after-school programs has grown to the point that using resources to improve programs is now ethical and feasible, and policymakers and practitioners are increasingly looking for ways to strengthen existing programs. Many states, cities, visible networks of programs, and individual organizations are engaged in quality improvement efforts, but we need evidence about the effectiveness of such efforts.

Improving program effectiveness presents some specific challenges. Most line-staff are part-time, hourly wages

are modest, and anecdotal information says that relatively few line-staff qualify for fringe benefits beyond those required by law. This means that improvement efforts need to anticipate a workforce that is predominantly young, untrained, and prone to frequent turnover.

Arguably, the best after-school programs capitalize on the advantages that after-school hours offer compared to the school day (Halpern, 2004). Consistent with the information emerging from practitioners and observational research, these advantages

include a greater opportunity to actively involve youth, project-based activities that can extend many weeks and are not constrained by school-day class schedules, and the use of the surrounding community as a resource and a place to carry out activities. This type of programming is, by design, driven in large part by the students, and as some of the observational instruments reflect, it is inherently hard to codify and sequence this approach.

Some practitioners have approached after-school program improvement by trying to improve the curriculum materials. For example, the U.S. Department of Education via the Institute for Education Sciences has funded two studies testing the effects of using after-school adaptations of mathematics and reading curricula that are effective in the regular school day. Durlak and Weissberg's (2007) findings on sequenced curricula support the wisdom of this approach. We undoubtedly need good work on the effects of after-school curricula, but I am not sanguine that curriculum and curricula-specific training alone will produce the desired positive effects, espe-

cially if the curriculum is sequential and that sequence is fixed. Such an approach runs counter to after-school programming that takes advantage of the strengths of an out-of-school setting, and it fails to acknowledge that many youth have sporadic attendance.

... learning how to intervene effectively to improve programs is now the primary issue facing the field.

Another approach to program improvement is to focus directly on staff/youth interaction through on-site staff development. Many provider networks and organizations are trying this approach. The work often takes the form of ongoing coaching for the line-

staff, perhaps using one of the observational instruments for feedback or as a means of self-assessment. Again, the nature of the after-school workforce poses challenges. Most programs have some permanent, relatively senior staff who can be trained and deployed as coaches. However, the nature and turnover of line-staff means at minimum that such coaching will need to be continuous to maintain any improvements. The William T. Grant Foundation is supporting ongoing experimental tests of this form of staff development, though the results from the first study are a few years away.

### Implications for Policy, Practice, and Research

The after-school field now has strong research reviews showing what many in the field have argued; these programs can have an important impact on academic and other policy-relevant youth outcomes. The research also shows that many programs do not make a greater difference than other services in the community, and



there is good reason to believe that many of the more conventional programs are not going to impact academic, social, or behavioral outcomes.

Research and practitioner views on how to improve programs are less conclusive, though they offer useful direction. Programs should be intentional about what they want to achieve, get youth actively involved, and, if improving academic performance is a program goal, have a component of the programming that is explicitly academic. Furthermore, it is unlikely that programs will improve if they do not commit to an on-site staff development strategy that supports continuous improvement of the line-staff. Learning how to do this well and at scale is a priority for the field. Research needs to provide reliable answers to many basic questions: What type of funder accountability and monitoring supports continuous improvement? How much of the ongoing staff development needs to be delivered on-site, while staff are working with youth? What training do coaches need, and how proscriptive should the coaching model be?

The literature is much less clear on the details of programming. Should programs integrate all learning within larger projects or does specific time need to be devoted to certain skills? Should programs specialize in particular content to attract and motivate youth or can such motivation be built into more generic programs? How can programs capitalize on the heterogeneity of the student participants? How much should programs explicitly include mentoring? Tutoring?

As practitioners and researchers collaborate on questions such as these, policy should be aligned with the best information currently available. If a funding stream has improved academic performance as a goal, I think it is

appropriate for policymakers to expect that after-school programs supported by those funds deliver on that goal. However, policymakers should encourage the sort of student-centered, active, project-based learning that plays to the comparative strengths of the after-school hours. Policymakers should also encourage an ongoing focus on program practices at the point where youth are served, coupled with the on-site staff development that is becoming more common. Such a policy approach complements a focus on individual-level outcomes with a simultaneous emphasis on the program practices that are linked to improvements in those outcomes.

In the next few years, most observers agree that healthcare and social security reform will dominate the domestic policy agenda and discretionary domestic spending. This is likely to increase the pressure on all other domestic services to be as effective and efficient as possible. Now is the right time to work on the unanswered questions regarding how to create and sustain high-quality after-school programs.

## Footnotes

<sup>1</sup>Out-of-school time (OST) refers to all periods of the day and year other than the school day. Thus, OST programs may occur on Saturdays and during the summer in addition to after-school. After-school is a subset of OST, and generally refers to programs operating between 3:00 and 6:00 p.m. on days when school is in session.

<sup>2</sup>In addition to having substantively different inclusion criteria, the three meta-analyses differed in the willingness to mix together experimental and quasi-experimental studies. There is some disagreement in the meta-analysis field about this issue when one wants to assess the effects of social policies and programs. There is a growing body of literature showing that it is not possible to replicate the impact findings from a particular experimental study with quasi-experimental methods (Agodini & Dynarski, 2001; Bloom, Michalopoulos, Hill, & Lei, 2002; Glazerman, Levy, & Myers, 2003). This work raises serious concerns about the reliability of findings in any single quasi-experimental study. This concern led Zief and Lauver (2006) to include only well-implemented randomized trials, and they only found five studies that met this and their other criteria. Other reviewers argue that it is appropriate to include quasi- and fully experimental studies in the review of a group of studies, under the assumption that the lack of reliability of any one study (experimental or quasi-experimental) is averaged out by including many. Such reviewers then typically test to see if the main findings differ for the experimental and quasi-experimental subsets. Lauer et al. (2006) and Durlak and Weissberg (2007) took this approach. Lauer et al. found that program study “quality,” a rating that included factors in addition to the study’s design, predicted effects. Higher quality studies tended to show more positive results. Durlak and Weissberg did not find that research design predicted their findings.

<sup>3</sup>Meta-analysts need to have the results from different studies and different outcome measures in a common form. They do this by converting results into an “effect size.” This is usually the difference between the mean outcome for a program group and a control group, divided by the pooled standard deviation of those scores. A positive value means the program group outperformed the control group. When a study includes several different measures of an outcome, such as assessing student grades through teacher surveys and transcript data, meta-analysts compute an effect size for each measure and then average those effect sizes to get an average effect for the outcome. They then compute the average of those estimates across all the studies that measured the outcome to synthesize the findings from the different studies. The conventions of the field were established in the early 1980s and are evolving as reviewers confront new issues. The three meta-analyses discussed in this article used similar procedures.

<sup>4</sup>In these analyses, a moderator is a feature of the program, its participants, or the circumstances at baseline that predicts the impact findings. For example, “grade level” would be a moderator if a program showed effects for early elementary age students but not for older elementary

students and the difference between impacts for the two groups was statistically significant. Such features do not necessarily cause effects to differ, they predict the variation that is found.

<sup>5</sup>Durlak and Weissberg (2007) coded a program as SAFE if the answers were positive to the following four questions: (1) Does the program use a sequenced set of activities to achieve its objectives relative to skill development? (Sequenced); (2) Does the program use active forms of learning to help youth learn new skills? (Active); (3) Does the program have at least one component focused on personal or social skills? (Focused); and (4) Does the program target specific personal or social skills? (Explicit).

## Endnotes

The following instruments are included in *Measuring Youth Program Quality: A Guide to Program Quality Assessment Tools*.

Assessing Afterschool Program Practices Tool (APT)  
National Institute on Out-of-School Time

Out-of-School Time Observation Tool (OST)  
Policy Studies Associates, Inc.

Program Observation Tool (POT)  
National Afterschool Association

Program Quality Observation (PQO)  
Deborah Lowe Vandell and Kim Pierce

Program Quality Self-Assessment Tool (QSA)  
New York State Afterschool Network

Promising Practices Rating Scale (PPRS)  
Wisconsin Center for Education Research & Policy Studies  
Associates, Inc.

Quality Assurance System (QAS)  
Foundations Inc.

School-Age Care Environment Rating Scale (SACERS)  
Frank Porter Graham Child Development Institute &  
Concordia University, Montreal

Youth Program Quality Assessment (YPQA)  
High/Scope Educational Research Foundation

## References

- Agodini, R., & Dynarski, M. (2001). *Are experiments the only option? A look at dropout prevention programs*. Princeton, NJ: Mathematica Policy Research, Inc.
- Ames, J. (June, 2007). *California's afterschool expansion: The planning for implementation of Proposition 49 and considerations for planners in other states*. New York, NY: William T. Grant Foundation.
- Bloom, H., Michalopoulos, C., Hill, C., & Lei, Y. (June, 2002). *Can non-experimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* New York, NY: Manpower Demonstration Research Corporation.
- Bodilly, S., & Beckett, M. (2005). *Making out of school time matter: Evidence for an action agenda*. Santa Monica: RAND Corporation.
- California Secretary of State. (2002). *Proposition 49 before and after school programs. State grants. Initiative statute*. Retrieved January 14, 2008, from [http://vote2002.sos.ca.gov/2002-vig/propositions2\\_id=222&sid=1.asp](http://vote2002.sos.ca.gov/2002-vig/propositions2_id=222&sid=1.asp)
- Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development, 65(1)*, 1-127.
- Cooper, H., & Hedges, L. V. (Eds.). (1994). *Handbook of research synthesis*. New York: Russell Sage Foundation.
- Cooper, H., Patall, E. A., Tyson, D., & Valentine, J. C. (2008). *Research on out-of-school programs: A synthesis of syntheses*. Manuscript in preparation.
- DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*, 157-197.
- Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills*. Retrieved October 9, 2007, from Collaborative for Academic, Social, and Emotional Learning Web site: <http://www.casel.org/downloads/ASP-Full.pdf>.
- EdSource. (2002, September). *Voter guide: Proposition 49*. Retrieved January 14, 2008, from [http://www.edsource.org/pub\\_abs\\_prop49.cfm](http://www.edsource.org/pub_abs_prop49.cfm)
- Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6(3)*, 151-175.
- Forum for Youth Investment. (2003, July/August). Quality counts. *Forum Focus, 1(1)*. Retrieved October 9, 2007, from [http://www.forumfyi.org/Files//ForumFOCUS\\_Jul-Aug2003.pdf](http://www.forumfyi.org/Files//ForumFOCUS_Jul-Aug2003.pdf)
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.
- Glazerman, S., Levy, D. M., & Myers, D. (2003). Non-experimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science, 589*, 63-93.
- Granger, R., Durlak, J. A., Yohalem, N., & Reisner, E. (April, 2007). *Improving after-school program quality*. New York, NY: William T. Grant Foundation.
- Halpern, R. (2004). *Confronting the big lie: The need to reframe expectations of afterschool programs*. New York, NY: Partnership for After School Education.
- Hedges, L. V., & Oklin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Hirsch, B. (2005). *A place to call home: After-school programs for urban youth*. Washington, DC: American Psychological Association and New York: Teachers College Press.
- Hollister, R. (2003). *The Growth in after-school programs and their impact*. Washington, DC: Brookings Roundtable on Children.
- Larson, R. (2006). Positive youth development, willful adolescents, and mentoring. *Journal of Community Psychology, 34*, 677-689.
- Larson, R., & Hansen, D. (2005). The development of strategic thinking: Learning to impact human systems is a youth activism program. *Human Development, 28*, 327-349.
- Larson, R., Hansen, D., & Moneta, G. (2006). Differing profiles of developmental experiences across types of organized youth activities. *Developmental Psychology, 42(5)*, 849-863.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. A., Snow, D., & Martin-Glenn, M. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76(2)*, 275-313.
- Lipsey, Mark (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-Analysis for Explanation: A Casebook* (pp. 83-127). New York: Russell Sage Foundation.
- Little, P. M. D., & Harris, E. (2003). *A review of out-of-school time program quasi-experimental and experimental evalu-*

ation results. Cambridge, MA: Harvard Family Research Project.

Smith, C. (2005, September 8). *Youth program quality assessment validation study: Findings for instrument validation*. Retrieved October 9, 2007, from High/Scope Educational Research Foundation Web site: <http://www.highscope.org/file/EducationalPrograms/Adolescent/YouthPQA/YouthPQASummary.pdf>

Vandell, D. L., & Pierce, K. M. (2006). Study of after-school care: Program quality observation. Retrieved October 9, 2007, from University of Wisconsin-Madison, Wisconsin Center for Education Research Web site: [http://www.wcer.wisc.edu/childcare/pdf/asc/program\\_quality\\_observation\\_manual.pdf](http://www.wcer.wisc.edu/childcare/pdf/asc/program_quality_observation_manual.pdf)

Yohalem, N., & Wilson-Ahlstrom, A. (2007). *Measuring youth program quality: A guide to assessment tools*. Washington, DC: The Forum for Youth Investment, Impact Strategies, Inc.

Zief, S. G., & Lauver, S. (2006). *The impacts of after-school programs on student outcomes: A systematic review for the Campbell Collaboration*. Retrieved October 9, 2007, from Campbell Collaboration Web site: <http://www.campbellcollaboration.org/frontend2.asp?ID=55>

## About the Authors

**Robert Granger**, Ed.D, has been president of the William T. Grant Foundation since 2003. The Foundation supports research and related activities intended to improve the lives of young people. The Foundation's current focus is on how social settings such as schools, community organizations, and neighborhoods influence young people; how to improve these settings; and how research influences policy and practice. Before joining the Foundation in 2000 as Senior Vice President of Programs, Dr. Granger served as Senior Vice President of the Manpower Demonstration Research Corporation and Executive Vice President at Bank Street College of Education. Dr. Granger also chairs the National Board for Education Sciences. This presidentially appointed advisory panel of the Institute of Education Sciences in the U.S. Department of Education oversees federal activities regarding educational research. In addition, Dr. Granger serves on the editorial board for several professional journals. He received his Ed.D. from the University of Massachusetts.

**Joseph Durlak** is Professor of psychology at Loyola University Chicago. He has written two books on prevention and co-edited a third, and his current interests in addition to meta-analysis are in evaluating the impact of positive youth development programs.

Dr. Granger would like to thank Joe Durlak, Rebecca Maynard, Ed Seidman, Vivian Tseng, Nicole Yohalem, and three anonymous reviewers for comments that helped him improve this manuscript. Thanks also to Krishna Knabe, Sarah Martino, and Julie Wong for editorial and fact checking help.

*Social Policy Report* is a quarterly publication of the Society for Research in Child Development. The *Report* provides a forum for scholarly reviews and discussions of developmental research and its implications for the policies affecting children. Copyright of the articles published in the *Report* is maintained by SRCD. Statements appearing in the *Report* are the views of the author(s) and do not imply endorsement by the Editors or by SRCD.

Electronic access to the *Social Policy Report* is available at the *Report's* website:  
<http://www.srcd.org/spr.html>

## *Purpose*

*Social Policy Report* (ISSN 1075-7031) is published four times a year by the Society for Research in Child Development. Its purpose is twofold: (1) to provide policymakers with objective reviews of research findings on topics of current national interest, and (2) to inform the SRCD membership about current policy issues relating to children and about the state of relevant research.

### **Content**

The *Report* provides a forum for scholarly reviews and discussions of developmental research and its implications for policies affecting children. The Society recognizes that few policy issues are noncontroversial, that authors may well have a "point of view," but the *Report* is not intended to be a vehicle for authors to advocate particular positions on issues. Presentations should be balanced, accurate, and inclusive. The publication nonetheless includes the disclaimer that the views expressed do not necessarily reflect those of the Society or the editors.

### **Procedures for Submission and Manuscript Preparation**

Articles originate from a variety of sources. Some are solicited, but authors interested in submitting a manuscript are urged to propose timely topics to the editors. Manuscripts vary in length ranging from 20 to 30 pages of double-spaced text (approximately 8,000 to 14,000 words) plus references. Authors are asked to submit manuscripts electronically, if possible, but hard copy may be submitted with disk. Manuscripts should adhere to APA style and include text, references, and a brief biographical statement limited to the author's current position and special activities related to the topic. (See page 2, this issue, for the editors' email addresses.)

Three or four reviews are obtained from academic or policy specialists with relevant expertise and different perspectives. Authors then make revisions based on these reviews and the editors' queries, working closely with the editors to arrive at the final form for publication.

The Committee for Policy and Communications, which founded the *Report*, serves as an advisory body to all activities related to its publication.