WILLIAM T. GRANT FOUNDATION

Improving After-School Program Quality

Robert C. Granger
William T. Grant Foundation

Joseph Durlak
Loyola University Chicago

Nicole Yohalem
The Forum for Youth Investment

Elizabeth Reisner
Policy Studies Associates, Inc.

Acknowledgements

Improving After-School Program Quality

The after-school[1] field has been expanding for 20 years because parents and other taxpayers believe the field will deliver on four goals: improvement in the safety and health of our communities and our youth; improvement of students' academic performance; development of their civic, artistic, and other skills; and provision of care for young people while parents work. The expansion has raised important practical and policy questions, particularly as public funding has increased. Can after-school programs deliver on these multiple goals? What are the program features and staff practices that produce good youth outcomes? What are the best ways to hold programs accountable for their work? What are promising approaches to improving program effectiveness?

In this paper we summarize the findings from two recent reports relevant to these policy and practice questions. One is a new review of evaluation studies by Joe Durlak and Roger Weissberg, showing that after-school programs attempting to enhance youth's personal and social skills can improve outcomes that are important to both school and non-school audiences (Durlak & Weissberg, 2007).[2] Durlak and Weissberg find that programs focusing on specific social or personal skills are most successful when they employ sequential, focused, explicit learning activities and active youth involvement. They also find that these programs tend to improve a range of outcomes at the same time. They refer to such programs as SAFE (Sequenced, Active, Focused, Explicit).

The second report, written by Nicole Yohalem and Alicia Wilson-Ahlstrom, describes instruments that measure the quality of youth program practices at the point of service (Yohalem & Wilson-Ahlstrom, 2007).[3,4] Although various teams of researchers and practitioners created the instruments, Yohalem and Wilson-Ahlstrom show that these instruments share a common core and that practitioners believe the instruments capture the practices that define program quality. This convergence suggests that an important consensus is emerging in the field about effective practices.

Our aim here is to help the field consider the implications of these two reports for policy and practice. The reports support the case that after-school programs are capable of improving important youth outcomes. They also support the need to stay focused on improving program quality. We are not yet sure how programs should be designed and implemented for optimal results, but these reports get us closer to the answers. This paper begins with a reprise of the findings in each document. We close with thoughts about implications and some important issues that the reports leave unresolved.

---

[1] We use the term *after-school* here to refer to a broad range of school- and community-based activities in which children and youth participate between 3:00 and 6:00 p.m. during the school year.
[2] The full report and an executive summary are available at www.casel.org.
[3] By measuring practices at the point of service, most of these instruments emphasize the nature of youth's daily activities in the program and the interactions between staff and youth. This is a supplement to approaches that only measure organizational or program practices, such as whether a program hires staff with particular credentials.
[4] The full report is available at http://www.forumfyi.org/Files//Measuring_Youth_Program_Quality.pdf.

Reviewing the Effects of After-School Programs

In January 2007, Joe Durlak and Roger Weissberg issued their review of the effects on youth of participation in after-school programs. Using the best techniques available in the field of *meta-analysis,* where the results of many studies are reviewed and combined using empirical techniques, these researchers summarized the results from evaluations of 73 after-school programs. Many of these studies were released within the past two years, making the review the most recent and comprehensive analysis available to the field. The authors only included studies of programs that were in part trying to improve participants' personal and social development. Programs exclusively focused on academic performance, such as tutoring programs, were not included in the review.[5] All of the studies reviewed had experimental or quasi-experimental designs that estimated effects by comparing outcomes for youth enrolled in the after-school programs to outcomes for a group of similar youth not enrolled.

Durlak and Weissberg grouped program effects into three areas that were further subdivided into a total of eight categories. *School performance* was subdivided into performance on achievement tests, school grades, and school attendance. *Social behaviors* contained three categories: positive social skills such as helping others or leadership skills, problem behaviors such as criminal activity or delinquency, and drug use. *Attitudes and beliefs* included bonding to school and self-esteem. Across all the studies that measured outcomes in each specific category (never fewer than 20 studies), the researchers found positive average effects in every category except school attendance. As further analysis revealed, the programs that had SAFE features were driving these positive findings.

Having found overall positive effects for a range of outcomes, Durlak and Weissberg then classified the studies into two clusters for comparison. In one cluster were studies of programs characterized by SAFE features: programs that used a *sequenced* set of activities to achieve their goals, used *active* learning techniques to help the participants acquire the skills, were at least in part *focused* on personal or social development, and had *explicit* objectives for the personal and/or social skills. Durlak and Weissberg refer to this cluster as programs that use an "evidence-based training approach."[6] We use their acronym—SAFE—while noting that the SAFE design features *only* pertain to the personal and social skills components of a program. When that was one component among many, the review did not capture the design of the program's other components. In the other cluster were studies of programs that did not have 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 the SAFE features showed positive effects for every outcome but school attendance, and the cluster of programs without these features*

---

[5] This review was part of a larger study to examine intervention effects on social and personal development. A companion report regarding the effects of in-school interventions is forthcoming. (Weissberg, Durlak, Taylor, Dymnicki & O'Brien, (2007).

[6] Several prior reviews had found that learning was more likely to occur when an intervention had these features. See Durlak and Weissberg (2007) for a list of citations (2007, p. 11).

showed no effect for any outcome. This means that the results from the programs with SAFE features created the overall positive average for all of the after-school evaluations. Table 1 summarizes these findings.

Statistically Significant Positive Effects for After-School Programs (Durlak/Weissberg, 2007)			
	Programs overall	SAFE cluster[7]	Other cluster[7]
School Performance			
Achievement tests	√	√	
School grades	√	√	
School attendance			
Social Behaviors			
Social skills	√	√	
Problem behaviors	√	√	
Reduced drug use	√	√	
Attitudes & Beliefs			
Bonding to school	√	√	
Self-esteem	√	√	

Note: A check indicates positive effects.

Another striking finding in this review is that positive effects tended to cluster together. For example, in the Durlak and Weissberg review, on average the individual evaluations in the SAFE group showed positive effects for 70% of the outcomes they assessed. In contrast, the studies in the other cluster revealed positive effects for only 25% of the outcomes (and no positive effects when the individual evaluations were grouped together and an average was computed for the group).

Before developing the implications of these findings for policy and practice, we turn to the report on measures of program quality and to a discussion of how the reports fit together.

Measuring the Quality of Program Practices

Until recently, the after-school field was not sure it agreed on what differentiated the practices of high- and low-quality programs. And for those who thought they "knew quality when they saw it," they were not sure how to measure it accurately and

---

[7] 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.

consistently. Yohalem and Wilson-Ahlstrom have documented recent improvements on both of these fronts.

In 2003 the Forum for Youth Investment reviewed 13 statements of standards for youth program quality (Forum for Youth Investment, 2003). Most were developed by organizations or accrediting groups serving a specific subsection of the youth field (e.g., camps, school-age child care, youth leadership). While differences in program purpose and content were visible in the standards, the different frameworks were quite similar at their core and emphasized the importance of interactions among program staff and participants.

In March 2007, the Forum extended this work by releasing a review of nine 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 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 how the program operated on a day-to-day basis. They emphasized interactions among staff and youth, while also assessing social norms, physical and psychological safety, skill-building opportunities, and program routine or structure. The Yohalem review labels these *core concepts*.

For each of the core concepts, the instruments included in the review offer specific definitions or indicators of effective practice that are used to guide observations. For example, in judging the extent to which youth are engaged, observers look for whether staff use open-ended questions during activities or how often youth have opportunities to talk about what they are doing and what they are thinking. In the case of relationships, they may look at how much youth are able to problem-solve together without adult intervention when minor conflicts occur, or how often staff engage in one-on-one conversations with individual youth. While each of the instruments delimits general concepts into observable behaviors, some are more detailed than others and some emphasize certain core concepts more than others.

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 practices, they will score the instrument similarly (known as inter-rater 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 yet has data showing that improved scores on what it measures translate into improved youth outcomes. Information on whether better scores lead to better outcomes will come from use and systematic study.

How Do the Findings From the Two Reports Relate to Each Other?

The SAFE features in the Durlak and Weissberg review predicted whether a program had positive effects. The observational measures of program quality in the Yohalem and Wilson-Ahlstrom review also are trying to specify features of program practice that will positively affect youth outcomes. Does this mean the SAFE features are captured in the core concepts of the observation instruments? The answer, in large part, is yes.

In general, there is congruence between what the instruments measure and Durlak and Weissberg's *active, focused,* and *explicit* features. That is, 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.

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 describe *active* learning techniques. All but one addresses the *focused* feature, with items that call for the "practice or progression of skills," or activities "designed to meet program goals/objectives." Six of the instruments underscore the importance of *explicit* approaches to programming, by pointing to the importance of "clear expected learning goals" and content that is "well developed, detailed, reflects standards."

Agreement around Durlak and Weissberg's *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 and not necessarily following a pre-determined 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.

The Implications for Policy and Practice

Readers of these two reports might identify additional implications for policy or practice that we do not consider here. We draw implications in five areas: the case for

supporting after-school programs, the need to work on program improvement, the choice of outcomes to guide programming, the need for accountability systems, and interventions to consider for improving program quality.

The Case for Supporting After-School Programs

After-school programs have broad, bipartisan appeal among voters and politicians in large part because of the need for a safe and supervised environment between school and the end of parents' work day. By all accounts, programs are delivering on that need. As the field has expanded, a significant portion of new federal and state public funding for after-school goes through the education system. It has been less clear that it is possible for after-school programs to affect certain academic outcomes that are important to superintendents, principals, or classroom teachers. Some prior reviews have suggested that after-school programs create such effects, while some have not found such effects, particularly on standardized achievement tests.

Durlak and Weissberg found that when programs focused on promoting personal and social skills, they also achieved effects on measures of academic performance, including standardized tests. This is an important new finding that should be reassuring to all who are concerned about the potential of after-school programs. Importantly, program effectiveness was not tied to a particular age range or other demographic factor. At the same time, it is overreaching to say we now know after-school programs *will* create such effects. Durlak and Weissberg showed that they can. But in their report they note that many of the programs did not.

It is also not yet proven that youth outcomes will improve if programs adopt the SAFE features. At this point we know that SAFE features are found in programs that create such effects, but like all correlations, this does not mean that the SAFE features caused such effects. Further research will need to determine if and when adding SAFE features to programs makes a difference in youth outcomes. If so, the research will also need to determine if all the features are necessary.

The Need to Work on Program Improvement

One strength of a meta-analysis is that it brings together a number of individual evaluations and looks for the predominant findings when the studies are considered as a group. Because the finding that programs with SAFE features are effective is based on a group of many individual studies, it is more trustworthy than the results of any single study. But the Durlak and Weissberg work also shows that approximately one-half of the evaluated programs did not make a difference for young people. From the details of the evaluations they reviewed, plus the findings from recent research using observational

measures,[8] it is fair to characterize the environments and activities of many programs as safe but uninspiring.

When a community has a large unmet demand for after-school services, it is difficult to argue that the next available dollar should go to improving existing services. Fortunately, in many communities there is a reasonable supply of programs. This makes it more defensible to work on program improvement. The Durlak and Weissberg report clearly shows that some programs need to change to become effective, and the Yohalem and Wilson-Ahlstrom review identifies several tools that are meant to help such efforts.

The Choice of Outcomes to Guide Programming

These two reports do not absolve program staff and funders of the need for clear and candid discussion about program goals and activities. It is hard to imagine a program lasting and having a reasonable claim on resources unless it is working toward goals that meet the interests of those who fund and deliver it, as well as the needs and interests of participating youth and their families.

At the same time, after-school programs are asked to achieve diverse goals, and practitioners in particular have felt that these various goals may be incompatible. For example, if a program is focused on one set of youth skills, can it improve outcomes in other areas? These reports say that the choices regarding goals and outcomes are less stark than how they are often presented.

The Durlak and Weissberg review shows that programs with SAFE features achieved effects on a range of youth outcomes. Even though they only examined programs with a focus on personal and social skills, such programs with the SAFE features produced academic effects, too.

The instruments in the Yohalem review are meant to be useful in a wide range of programs. Even so, the instruments share a common core. Furthermore, validation work on some of the instruments found that a program's mission (e.g., arts, community service, leadership development) was unrelated to how well it scored on the instrument.

All these findings support what some in the field have assumed for a long time. That is, it is likely that a program's specific focus is less important than *having* a focus and then pursuing it well.

---

[8] Public/Private Ventures used an observational measure to assess the quality of the San Francisco Beacons program three years ago (Walker & Arbreton, 2004). Beth Miller and her colleagues at the National Institute on Out-of-School Time recently used program quality measures in their study of a diverse sample of programs in Massachusetts (Miller, 2005). Charles Smith and colleagues at High/Scope Educational Research Foundation observed a range of programs in Michigan during the validation study of their Youth Program Quality Assessment (YPQA) instrument (Smith, 2005). All came to this conclusion.

<u>The Need for Accountability Systems That Track Both Service Quality and Outcomes</u>

Perhaps the most common policy approach to improving the effectiveness of social services is to hold programs accountable for attaining certain participant outcomes. This approach has clear appeal because it lends itself to flexibility—quality control by specifying the ends, not the means.

In one sense, findings from the Durlak and Weissberg review support this approach. Because programs can affect a range of important outcomes, program providers should choose a finite set of outcomes to work toward, align services with those outcomes, and use improvement in outcomes as a basis for gauging program viability. We agree that measuring such improvement should be part of any accountability system.

However, there are disadvantages to measuring improvement in youth outcomes alone. We all know that the general condition of a society—and, more importantly, the experiences in a child's family, neighborhood, peer group, and school day—combine to contribute to every child's development. This makes it easy for programs to look good when the societal trend in youth outcomes is positive, and bad when it is not.

It may be useful to supplement the measurement of youth outcomes with an ongoing assessment of program and staff practices. Perhaps most importantly, this approach gives administrators and line staff valuable information about their practice and holds programs accountable for something they can control.

The newly emerging set of observational measures of program quality makes such accountability systems feasible. The measures of program practices and staff/child interactions give staff guidance about specific behaviors they should adopt, a shared language for talking about their interactions with youth, and specific direction or feedback about their performance to guide improvement activities. Such information seems fundamental to improving service quality.

The details of an accountability system that includes youth outcomes and measures of program/staff practices need more attention. Some worry that a system that requires periodic observation of program practices is not feasible, but there are now enough examples of such systems in practice to ease this concern. For example, in several states, including Massachusetts, Michigan, and Maine, observational assessment is being used in all after-school programs receiving 21st Century Community Learning Center funding from their state departments of education (Wilson-Ahlstrom & Yohalem, 2007).

Less is known about how to structure the consequences embedded in such an accountability system. For example, should funding be tied to the documentation that youth outcomes and staff practices have both improved? While that is the goal, making

such results a condition for future funding may lead to inaccurate reporting. Also, with our current knowledge, it would be difficult to set reasonable expectations about how much improvement should be expected over what period of time and on what aspects of practice. The field needs organizations and funders who will experiment with and study different approaches to accountability.

Interventions to Improve Program Quality

In our view, learning how to intervene effectively to improve programs is now the primary issue facing the after-school field. The availability of after-school programs has grown to the point where using resources to improve programs is ethical and feasible, and policymakers are increasingly looking for ways to strengthen existing programs. We know programs need to improve—and many are engaged in quality improvement efforts—but we need good evidence about the effectiveness of such efforts. Several questions require further study. Are there intervention strategies that are more or less effective for different types of organizations or for organizations that have different levels of capacity? How much can programs improve practices through use of particular curricula? What are effective ways to structure staff development? In short, what are the most promising and practical methods for intervening in programs to improve the quality of youth experiences and youth outcomes?

Fortunately, several funders are supporting studies of interventions with these questions in mind. For example, the U.S. Department of Education via the Institute for Education Sciences has funded two studies testing the effects on youth of using after-school adaptations of mathematics and reading curricula that are effective in the regular school day. The Charles S. Mott Foundation is supporting research on the youth impacts of after-school programs that meet certain quality criteria. Finally, the William T. Grant Foundation is supporting assessments of staff development and curricular innovations on program practices, and has announced an annual request for proposals to identify and fund additional studies on this subject. In all of the foundation-funded studies, observational measures of program practices are central to the research designs.

The after-school field has expanded and matured in the last 15 years. During that period it has been pulled in many directions by different societal needs. Recent research shows that after-school programs that employ intentional, focused designs and aligned activities can deliver results that are important to a range of constituencies. The field is also reaching consensus on a set of core practices, and has developed instruments that measure these practices. While this progress leaves important questions unanswered, the questions are more about tactics than strategy. Our advice is to focus on improving program quality, refining our tactics as we go.

References

Durlak, J.A., & Weissberg, R.P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. Retrieved March 29, 2007 from http://www.casel.org.

Miller, B.M. (2005). *Pathways to Success for Youth: What Counts in Afterschool*. Boston, MA: United Way of Massachusetts Bay.

Smith, C. (2005). *Youth Program Quality Assessment validation study: Findings for instrument validation*. Retrieved March 22, 2007, from http://www.highscope.org/EducationalPrograms/Adolescent/YouthPQA/YouthPQASummary.pdf.

The Forum for Youth Investment. (2003, July/August). "Quality Counts." *Forum Focus*, 1(1). Washington, DC: The Forum for Youth Investment, Impact Strategies, Inc. Available online at www.forumforyouthinvestment.org.

Walker, K.E., & Arbreton, A.J.A. (2004). *After-school pursuits: An examination of outcomes in the San Francisco Beacon initiative*. Philadelphia, PA: Public/Private Ventures.

Weissberg, R. P., Durlak, J. A., Taylor, R. D., Dynmicki, A. B., & O'Brien, M. U. (2007). Promoting social and emotional learning enhances school success: Implications of a meta-analysis. Unpublished manuscript.

Wilson-Ahlstrom, A., & Yohalem, N. (2007). *Building quality improvement systems: Lessons from three emerging efforts in the youth-serving sector*. Washington, DC: The Forum for Youth Investment.

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

Endnote

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

Authors' Note

Robert C. Granger is president of the William T. Grant Foundation; Joseph Durlak is professor of clinical psychology at Loyola University Chicago; Nicole Yohalem is a program director at the Forum for Youth Investment; Elizabeth Reisner is a founder and principal of Policy Studies Associates, Inc.

Correspondence concerning this paper should be addressed to Robert Granger at the William T. Grant Foundation, 570 Lexington Avenue, 18th Floor, New York, NY 10022 or via email to bgranger@wtgrantfdn.org.