Am J Community Psychol (2010) 45:350–357
DOI 10.1007/s10464-010-9311-3

ORIGINAL PAPER

Inside the Black Box: Assessing and Improving Quality in Youth Programs

Nicole Yohalem · Alicia Wilson-Ahlstrom

Published online: 19 March 2010

Abstract Over the past decade, structured programming for children and youth during the non-school hours has expanded exponentially. A confluence of recent research studies and program evaluations backs the publicly perceived notion that after-school programs *can* positively influence important developmental and learning outcomes. The rapid expansion of the field and the potential of programs to contribute to child and youth development have made defining what high quality programs look like and learning how to improve program quality key challenges facing the field. This paper describes what is known about the relation between youth program quality and youth developmental outcomes, summarizes different quality assessment tools being used in the field, and discusses how such tools are being used to drive systemic quality improvement efforts.

Keywords Youth programs · Quality · Assessment

Estimates suggest that some 6.5 million children are enrolled in after-school programs and that the federal government alone invested $3.6 billion in such programs in 2002 (Afterschool Alliance 2004; Padgette 2003). A growing collection of developmental research and program evaluations suggests that after-school programs *can* improve important developmental and learning outcomes that parents, practitioners, policy makers and researchers agree are worth investing in such as increased pro-social behaviors, reduced problem behaviors and school grades (Vandell et al. 2007; Durlak and Weissberg 2007; Lauer et al. 2006; National Research Council and Institute of Medicine 2002). However, a recent meta-analysis cautions that not *all* programs are designed or implemented in such a way that they will necessarily achieve positive results (Durlak and Weissberg 2007). The positive potential of after-school programs to contribute to youth development, combined with the reality that all programs are not created equal, make defining what high quality programs look like and learning how to improve program quality key challenges facing the field (Granger et al. 2007).

In 2002, the National Research Council released a consensus report developed by a multi-disciplinary committee of experts about community programs designed to support positive youth development (National Research Council and Institute of Medicine 2002). In addition to identifying the personal and social assets that facilitate positive development in youth and affirming the role that youth programs can play in supporting the development of those assets, the committee identified eight specific program features of programs that are important in supporting positive youth development: physical and psychological safety, appropriate structure, supportive relationships, opportunities to belong, positive social norms, support for efficacy and mattering, opportunities for skill building, and integration of family, school and community efforts. Since its publication, this list has been widely adopted and cited in the youth development and out-of-school time fields and has contributed to what we and others have characterized as an emerging consensus about what constitutes program quality (Granger et al. 2007).

Attention to youth program quality is increasing, fueled in part because of growing interest on the part of three important stakeholder groups: researchers, policy makers and practitioners. From a research perspective, more evaluations are including assessments of program quality

N. Yohalem (✉) · A. Wilson-Ahlstrom
Forum for Youth Investment, Washington, DC, USA
e-mail: nicole@forumfyi.org

and many have incorporated setting-level measures (where the object of measurement is the program, not the participants) in their designs. At the policy level, decision-makers are looking for ways to ensure that resources are allocated to programs most likely to have an impact, and building quality assessment and improvement expectations into requests for proposals and program regulations. At the practice level, programs, organizations and systems are looking for tools that capture effective practice and can aid practitioners in assessing, reflecting on and improving their programs.

With this growing interest has come an increase in the number of tools available for assessing and improving quality. The size and diversity of the youth-serving sector is such that it can accommodate a diversity of quality assessment tools. As the field expands, however, it is important that potential users—researchers designing studies, practitioners focused on program improvement, or policymakers interested in accountability—have access to good information about the purpose, structure, content and technical properties of available tools to help guide their decision-making.

In response to this need and in an attempt to summarize progress in the after-school field, the Forum for Youth Investment conducted a detailed review of instruments designed to measure youth program quality (Yohalem et al. 2007). Methods included examining published and unpublished information about each tool, interviewing the developers, and in most cases, interviewing practitioners that have used each of the tools. An earlier scan (Forum for Youth Investment 2003) and a review focused specifically on middle school program standards (Westmoreland and Little 2006) provided a foundation for that inquiry. The 2007 review found that a range of assessment tools exist or are under development that can help advance research and support program improvement. Overall, the review identified significant commonalities in terms of how quality is defined in the different instruments but some important differences in terms of approaches to measurement and technical properties.

Building on the 2007 review, this current paper aims to describe the current state of program quality assessment in the youth-serving sector. In the first section, we describe what is known about the relation between youth program quality and youth outcomes, discussing which specific program features are most promising. Having laid that foundation, the second section provides an overview of tools used to assess after-school quality assessment. We describe common characteristics of several assessment tools, focusing particular attention on how quality is defined within the instruments and how those definitions align with youth development theory. Some differences among tools related to technical properties, structure, purpose, and methodology are also discussed. In the final section of the paper, we use an example to describe how quality assessment tools like those summarized in the previous section are being used to drive systemic quality improvement efforts in the field.

Youth Program Quality and Developmental Outcomes

While consensus is building within the field about what constitutes effective practice, research that links specific program features to youth outcomes is rare. Here we summarize existing evidence about the relationship between program quality and youth developmental outcomes. Research about youth programs often emphasizes the importance of relationships, a supportive climate, and youth engagement—social processes that appear to be linked to positive developmental outcomes. Specifically, Vandell and colleagues have demonstrated that positive emotional climate and peer interactions are associated with school adjustment (Vandell and Pierce 2001). In the Massachusetts After-School Research Study, the level of youth engagement in programs was a predictor of positive outcomes such as behavior, relationships, initiative, and homework completion (Intercultural Center for Research in Education & National Institute on Out-of-School Time 2005).

Levels of engagement and quality of interactions were also significantly related to youth's self-reported feelings of safety, interest, growth and skill building in a validation study of the Youth Program Quality Assessment (Smith and Hohmann 2005) as well as to after-school attendance, reading scores and school-day behavior (Blazevski and Smith 2007). In an evaluation of the Philadelphia Beacons, activity management and positive adult support were important predictors of youth's self-reported enjoyment, engagement and learning (Grossman et al. 2007). Durlak and Weissberg's 2007 meta-analysis found that programs where staff used an active learning approach were more likely to show effects on academic and behavioral outcomes than programs without this feature.

While not as robust as the evidence about the social processes and staff practices described above, some studies suggest that the presence of certain program resources may be important in influencing developmental outcomes. For example, Vandell and colleagues have found that staff ratios, staff education levels and staff training are associated with school adjustment (Vandell and Pierce 2001; Vandell et al. 2004). An evaluation of The After-School Corporation's programs found that having a licensed teacher as site coordinator was associated with improvements in school achievement for program participants (Reisner et al. 2004). However, in a recent study of program quality

in 38 programs in Palm Beach, Florida, supervisor education and experience levels were not strongly related to point-of-service quality (Smith et al. 2008).

Some studies suggest that how resources are arranged within programs can also make a difference. Vandell and colleagues found that variation in activities is associated with school adjustment outcomes (Vandell and Pierce 2001). Durlak and Weissberg (2007) found that programs with explicit, sequenced activities focused on the development of social skills were more likely to show effects on academic and behavioral outcomes than programs without these arrangements.

To summarize, evidence that high quality programs can affect a range of important youth outcomes and that program quality matters is growing. This evidence, combined with the rapid expansion of the field, has led to increased interest among practitioners, policy makers and researchers in finding and developing tools that are designed specifically to assess and improve program quality. Program improvement has become a major focus of the work of state and local intermediary organizations in the youth development field, and public funding sources like the U.S. Department of Education's 21st Century Community Learning Centers program are allocating resources for quality improvement purposes.

Tools for Assessing Youth Program Quality

This section of the paper compares and contrasts different tools that are currently being used in the field to assess and improve program quality. We discuss what purposes the tools have been designed for, how quality is defined within the tools, and different aspects of their structure and methodology. These tools were identified because they share several characteristics: they are (or they include) program-level observational measures of quality; they can be used in a range of school and community-based program settings; they include a focus on social processes within programs; and they are "research-based." Given the relative infancy of observational assessment in the youth development field, the "research-based" label is used fairly broadly to include tools whose development was informed by relevant literature and others where specific technical properties have been studied and established.

Overall, our review of assessment tools suggests that there is a great deal of synergy in terms of what program features are addressed and how quality is defined. This reflects growing consensus in the field about characteristics of effective programs. As one would expect, however, careful examination of the contents of each instrument reveals differences in terms of emphasis, structure and approach to measurement, and technical properties.

Purpose of the Tools

Most of the tools we reviewed were developed with *program improvement* as a primary goal (see Fig. 1). The Youth Program Quality Assessment (YPQA), developed by the High/Scope Educational Research Foundation (2005), is one example of this type of tool. With clear, practitioner-friendly language and a range of associated products and services, this and other tools like it aim to help programs collect and utilize observational data about staff practice with the explicit goal of using that information to reflect on and improve performance. Some of the tools examined evolved within the context of specific monitoring or accreditation efforts, however, such as the Program Observation Tool (POT) developed by the National School-Age Care Alliance (2001). Although this instrument can be used for program improvement efforts, it was designed with the specific goal of collecting data to drive a program accreditation process. Three tools we looked at were designed exclusively for use in research. More tends to be known about the technical or psychometric properties of such instruments, like the Promising Practices Rating System (PPRS) developed by the Wisconsin Center for Education Research and Policy Studies Associates, Inc. (2005), given their purpose and history.

How Quality is Defined in the Tools

The most notable similarity across the nine assessment tools is in terms of their content, or what specific

Assessment Tool	Target Age	Primary Purpose(s)		
	Grades Served	*Improvement*	*Monitoring/ Accreditation*	*Research/ Evaluation*
Assessing Afterschool Program Practices Tool (APT)	Grades K – 8	✓	✓	
Out-of-School Time Observation Tool (OST)	Grades K – 12			✓
Program Observation Tool (POT)	Grades K – 8	✓	✓	
Program Quality Observation Scale (PQO)	Grades 1 – 5			✓
Program Quality Self-Assessment (QSA)	Grades K – 12	✓		
Promising Practices Rating Scale (PPRS)	Grades K – 8			✓
Quality Assurance System (QAS)	Grades K – 12	✓		
School-Age Care Environment Rating Scale (SACERS)	Grades K – 6	✓	✓	✓
Youth Program Quality Assessment (YPQA)	Grades 4 – 12	✓	✓	✓

Fig. 1 Target age and purpose of different program quality assessment tools

components of quality are articulated and measured. All of the tools we reviewed address the following six core concepts, at varying levels of depth: (1) relationships: connections between and among youth and adults in the program; (2) environment: various aspects of the program climate and setting, including physical and emotional safety; (3) engagement: the extent to which children, youth and staff are meaningfully involved in program activities; (4) social/behavioral norms: expectations about and responses to positive and negative behaviors in the program; (5) skill building: opportunities for participants to develop specific skills by participating in intentional learning activities; and (6) routine/structure: various aspects of how the overall program is organized including pacing, transitions and routines.

These six concepts are articulated in slightly different ways in each tool, but the definitions are similar enough to suggest reasonable consensus about the core program features that contribute to quality. For example, the Assessing Afterschool Program Practices Tool (APT), developed by the Miller and Surr (2005), describes engagement as "sharing control or responsibility for activities with youth (2005, p. 11, Appendix E)," and provides the user with observable examples of such staff behavior (e.g., staff encouraging youth to take the lead in selecting and implementing activities). The YPQA articulates the same concept in the item: Youth have opportunities to partner with adults (2005, p. 21). The item further describes staff behaviors very similar to those described in the APT.

In addition to the six concepts described above, four others—youth leadership/participation, staffing (ratios, qualifications), program management, and linkages to family and community—are identified as key components of quality in some, but not all, of the tools we reviewed (see Fig. 2). The six common concepts combined with the additional four that some, but not all, tools measure, align well with current youth development theory (National Research Council and Institute of Medicine 2002; Pittman and Irby 1996; Benson and Saito 2001; McLaughlin 2000). Table 1 illustrates how the ten quality concepts captured in assessment tools parallel the features of positive developmental settings outlined by the National Research Council and Institute of Medicine (2002). The core concepts addressed by the tools are also compatible with Connell and Gambone's youth development framework (2002) which emphasizes the importance of positive relationships, safety, youth involvement and skill building.

When we consider the definitions of quality embedded in each tool, there is a common emphasis on social processes. This is reassuring given growing evidence that relationships, interactions and engagement are associated with positive outcomes. That said, different assessment tools do emphasize unique aspects of after-school settings. Although all of the tools we reviewed include a focus on the social processes that occur between individuals in the program, a subset emphasize program resources and the arrangement of those resources within the program. The School Age Care Environment Rating System (1996) for example, includes 35 items related to program resources and only 9 items about social processes. The Youth Program Quality Assessment (2005), includes 16 items related to resources and 14 related to social processes. The extent to which different tools emphasize different aspects of settings may reflect their purpose; those used for monitoring and regulatory purposes appear more likely to

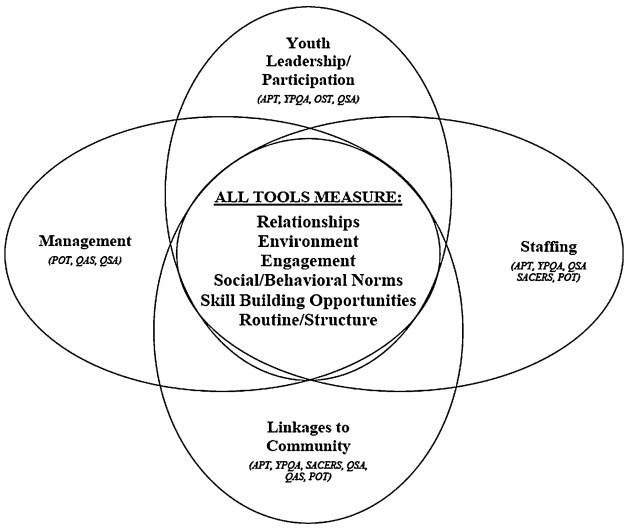

Fig. 2 Key constructs measured by different program quality assessment tools

Table 1 Aligning quality assessment tools with existing youth development literature

Features of positive Developmental settings (National Research Council and Institute of Medicine 2002)	Corresponding concepts from nine program quality assessment tools
Physical and psychological safety	Environment
Appropriate structure	Routine/structure, staffing
Supportive relationships	Relationships
Opportunities to belong	Relationships/environment
Positive social norms	Social/behavioral norms
Support for efficacy and mattering	Engagement, youth leadership/participation
Opportunities for skill building	Skill building opportunities
Integration of family, school, and community efforts	Linkages to community
	Management

address resource issues such as staffing, management, budget and the physical environment.

Tool Structure and Methodology

Across the assessment tools reviewed we found more differences in terms of how quality is *measured* or how the instruments are structured and used, than we did in terms of how quality is *defined*. Four specific areas where differences emerge across the instruments are data collection, types of measures, rating scales and technical properties.

Data Collection

The nine quality instruments reviewed differ in terms of what types of data are collected, who collects it, and the intensity or amount of data collection that is recommended. Five instruments were designed to be used by either external observers or program staff, three were designed for use by external observers only, and one was designed exclusively with self-assessment by program staff in mind. The Program Quality Observation Scale (2006), for example, is designed for use by trained external raters and pairs 90-min observations of the program environment with a time sampling approach that calls for 60 different 30-s observations of interactions between adults and youth or youth and their peers.

Tool developers' recommendations about how much observation time overall is necessary to generate a meaningful assessment of quality range from 2 to 5 h. Although observation is a primary data collection method for all nine instruments we reviewed, several rely upon interview, questionnaire or document review as additional data sources.

Specificity of Measures

The tools we reviewed differ in terms of the level of inference embedded in the measures, as well as the extent to which measures are simply diagnostic or both diagnostic and prescriptive. Low inference measures tend to be very specific, leaving little room for judgment about how to score an item. Some low-inference items, for example, measure the presence or absence of specific behaviors, like the extent to which staff use open-ended questions during activities or how often youth have opportunities to talk about what they are doing or thinking. High-inference measures are less concrete and therefore require more judgment.

The YPQA is an example of a low-inference measure, often about the presence or absence of specific behaviors (e.g., the extent to which staff use open-ended question during activities or how often youth have opportunities to talk about what they are doing and thinking) (2005). On the other hand, the Program Quality Self-Assessment Tool developed by the New York State Afterschool Network (n.d.) represents a higher inference measure. Though it includes open-ended examples of staff behaviors and program characteristics, its 1–4 rating scale is not anchored by specific descriptors that would reduce the amount of judgment required.

Another dimension to consider is the extent to which the measure simply describes or diagnoses what is happening in the program or both describes *and* prescribes what change needs to occur. This is particularly important in the case of tools that have been developed for the purpose of program improvement. Especially given that many youth program staff have relatively little training or experience, the assessment process can offer explicit guidance about effective practice. An example of an item from the YPQA that tells staff how well they are doing in a particular area *and* what they need to do to improve has "the activities do not provide opportunities for all youth to make process choices" as the low point on a scale and "all youth have the opportunity to make at least one open-ended process choice (e.g., youth decide roles, order of activities, tools or materials, or how to present results)" as the high point (High/Scope Educational Research Foundation 2005, p. 22).

Rating Scales

The rating scales included in the instruments we reviewed vary in some notable ways. First, some tools are based on a set of standards that a program either meets or does not meet, while most provide a continuum that the observer is asked to rate the program on. The tools also differ in terms of the level of description embedded in the scales themselves. As discussed in the previous section on specificity, some tools provide general statements and the observer rates whether a practice occurs "most of the time, sometimes, not at all;" other scales explicitly describe what it looks like when a given practice is happening most of the time, sometimes or not at all. In some cases observers are also encouraged to take anecdotal notes to capture evidence or examples that support the score they have applied.

Technical Properties

The instruments reviewed differ significantly in terms of the extent to which developers have studied their technical or psychometric properties. We found seven instruments that have some information showing that if different observers watch the same program practices, they will score the program similarly (i.e., inter-rater reliability). Few, however, have looked at other aspects of reliability

that are of interest when assessing the strength of a program quality measure, such as test–retest reliability. The SACERS, POT, PQO and the Out-of-School Time Observation Tool, developed by Policy Studies Associates, Inc. (2005), had the strongest evidence of inter-rater reliability among the tools we reviewed.

Five of the nine instruments have promising evidence of validity—meaning they have made some effort to demonstrate that the instrument accurately measures what it is supposed to measure. Predictive validity—meaning that scores on the instrument successfully predict related relevant youth outcomes—has only been established for scales on four of the tools we looked at—the APT, the PQO, the PPRS, and the YPQA. Future research needs to carefully attend to this gap in the field.

Improving Youth Program Quality

Tools like those described in this paper are increasingly being used by individual organizations and networks or systems of youth-serving programs around the country as they look for ways to assess and improve their performance. Public funding streams like the federal 21st Century Community Learning Centers (21st CCLC) program, as well as private foundations at the community, state and national levels are allocating resources for capacity building purposes and, in some cases, specifically helping to seed the development of data-driven continuous improvement systems built upon measures like these.

Program improvement is a major focus of local intermediary organizations such as the Providence After School Alliance and Community Network for Youth Development in San Francisco. These and other similar organizations provide tools, training and technical assistance to programs in their networks. Quality improvement is also a primary focus for many state agencies and statewide after school networks across the country.

In the case of 21st CCLCs specifically, state education agencies which disperse funds and monitor programs can use up to three percent of their federally allocated funding to support capacity building and evaluation efforts. Several states are using these "set-asides" to support data-drive quality improvement processes, using a range of tools and approaches. The Michigan Department of Education (MDE) is using a portion of its capacity building set-aside to work with the High/Scope Educational Research Foundation and Michigan State University to build a Quality Improvement System based upon the Youth Program Quality Assessment (YPQA) for all after-school programs in the state that receive 21st CCLC funding. The next section describes the goals and components of that effort as well as preliminary results and challenges.

Michigan After-School Quality System Demonstration

The goals of the Michigan After-School Quality System Demonstration (QSD) are fourfold: increasing knowledge about effective youth development practice among workers, providing a framework for the state to organize professional development opportunities, increasing quality across the state, and increasing youth outcomes related to program quality.

While participation in the QSD is mandatory, it is also intentionally low-stakes from an accountability perspective in that it is based on self-assessment and program funding or other rewards are not contingent upon scores. The focus is on engaging after-school program staff in reflecting on their practice and developing their own professional competencies. With this emphasis in mind, the QSD puts continuous quality assessment and improvement in the hands not just of local sites, but of line staff themselves. Lorraine Thoreson who coordinates 21st CCLC programs for MDE emphasized the value of using a common tool to ground program improvement efforts across the system: "The biggest thing I hear since we began implementation is that everyone gets on the same page, instead of everyone having a different conception of quality" (Wilson-Ahlstrom and Yohalem 2007, p. 43).

The four major stages of the QSD cycle are orientation and training, data collection, data interpretation and planning, and program improvement. Over the period 2005–2008, all 187 21st CCLC program sites in Michigan will be trained to use a customized version of the YPQA developed to assure compliance with Michigan's Out-of-School Time Standards, conduct a self-assessment using the instrument and develop a program improvement plan based on how they score.

Data are collected by teams of staff at each program site over a 2-weeks period. Following data collection, scoring meetings are held in which the team looks across the anecdotal evidence and comes to a consensus on a numerical score for each indicator. This unique self-assessment protocol generates rich dialogue about effective practice and helps facilitate group goal setting. Based on their scores, programs then pick specific items from the YPQA to work on and make plans to change staff behaviors or program structures that relate to the targeted items. At the state level, MDE uses the data—which they only see in aggregate form—to inform system-wide professional development planning.

While QSD results are still very preliminary, several shifts appear promising. First, stakeholders at the system and program level are beginning to use common language to describe programs and effective practice. Second, programs are identifying specific opportunities for improvement based on the data. Areas targeted for improvement

cluster around youth engagement and interactions, and include things like increasing student input into programming, providing more youth choices in daily activities and engaging youth on existing committees and advisory boards. Finally, self-assessment and improvement have become part of the culture as programs face continuing pressure to demonstrate their value.

Despite these promising developments, MDE has encountered several challenges as they work towards building a permanent system-wide continuous improvement system. As one would expect in any complex system, there is uneven capacity among programs to respond to the quality challenge and to implement an assessment and improvement plan. Many sites have raised concerns about the amount of time it takes to conduct observations, score the instrument and facilitate staff discussions. Ensuring programs are able to make the time to successfully engage in this work, particularly when there are no penalties in place for non-compliance, is a challenge. Environmental factors including local resources, politics, and buy-into a specific framework or methodology can also come into play in some communities. The rate of staff turnover in the field is another challenge to the long-term sustainability of this kind of model, and repeat trainings have had to be offered at many sites to ensure a critical mass of staff are on the same page about quality.

Conclusion

As the after-school and youth development fields expand and as more children and youth spend time in formal programs during the non-school hours, consensus is emerging about what program features matter and tools to measure those features are increasingly accessible. Demand for more comprehensive information about how to improve program quality is increasing from practitioners and policy makers, and interest among researchers is also on the rise. Information about the purpose, content, structure and technical properties of different quality assessment tools can help users make informed decisions about what tool may be most appropriate given their goals and interests. As a result, researchers can continue to unpack the relation between specific program features and youth outcomes and practitioners can build data-driven continuous improvement systems designed to ensure the delivery of high quality programming.

References

Afterschool Alliance. (2004). *America after 3 PM: A household survey on afterschool in America*. Washington, DC: Afterschool Alliance.

Benson, P. L., & Saito, R. N. (2001). The scientific foundations of youth development. In P. L. Benson & K. J. Pittman (Eds.), *Trends in youth development: Visions, realities and challenges* (pp. 135–154). Boston: Kluwer.

Blazevski, J., & Smith, C. (2007). *Inter-rater reliability on the youth program quality assessment*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Connell, J., & Gambone, M. A. (2002). *Youth development in community settings: A community action framework*. Philadelphia, PA: Youth Development Strategies Inc.

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.

Forum for Youth Investment. (2003). Quality counts. *Forum Focus, 1*, 1–4.

Granger, R., Durlak, J. A., Yohalem, N., & Reisner, E. (2007). *Improving after-school program quality*. New York, NY: William T. Grant Foundation.

Grossman, J., Campbell, M., & Raley, B. (2007). *Quality time after school: What instructors can do to enhance learning*. Philadelphia, PA: Public/Private Ventures.

Harms, T., Jacobs, E. V., & White, D. R. (1996). *School-age care environment scale*. New York, NY: Teachers College Press.

High/Scope Educational Research Foundation. (2005). *Youth program quality assessment*. Ypsilanti, MI: High/Scope Press.

Intercultural Center for Research in Education & National Institute on Out-of-School Time. (2005). *Pathways to success for youth: What works in afterschool: A report of the Massachusetts Afterschool Research Study*. Boston, MA: United Way of Mass Bay.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275–313.

McLaughlin, M. (2000). *Community counts: How youth organizations matter for youth development*. Washington, DC: Public Education Network.

Miller, B., & Surr, W. (2005). *The assessment of afterschool program practices tool*. Wellesley, MA: National Institute on Out-of-School Time.

National Research Council and Institute of Medicine, Committee on Community-level Programs for Youth. (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.

National School-Age Care Alliance. (2001). *NSACA program observation tool*. Charlestown, MA: National School-Age Care Alliance.

New York State Afterschool Network. (n.d.). *Program quality self-assessment tool.* New York, NY: New York State Afterschool Network.

Padgette, H. C. (2003). *Finding funding: A guide to federal sources for out-of-school time and community school initiatives*. Washington, DC: The Finance Project.

Pittman, K. J., & Irby, M. (1996). *Preventing problems or promoting development: Competing priorities or inseparable goals?*. Takoma Park, MD: International Youth Foundation—US.

Policy Studies Associates, Inc. (2005). *Out-of-school time observation tool* (2nd ed.). Washington D.C.: Policy Studies Associates, Inc.

Reisner, E. R., White, R. N., Russell, C. A., & Birmingham, J. (2004). *Building quality, scale, and effectiveness in after-school programs: Summary report of the TASC evaluation*. Washington, DC: Policy Studies Associates, Inc.

Smith, C., Akiva, T., Blazevski, J. & Pelle, L. (2008). *Final report on the Palm Beach quality improvement system pilot*. Ypsilanti, MI: David P. Weikart Center for Youth Program Quality.

Smith, C., & Hohmann, C. (2005). *Full findings from the youth program quality assessment validation study*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Vandell, D. L., & Pierce, K. M. (2001). Experiences in after-school programs and child well-being. In J. L. Mahoney (Chair), *Protective aspects of after-school activities: Processes and mechanisms*. Paper symposium conducted at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Vandell, D. L., Reisner, E. R., & Pierce, K. M. (2007). *Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afterschool programs*. Washington, DC: Policy Studies Associates.

Vandell, D. L., Shumow, L., & Posner, J. (2004). After-school programs for low-income children: Differences in program quality. In J. Mahoney, J. Eccles, & R. Larson (Eds.), *Organized activities as contexts for development: Extracurricular activities, after-school, and community programs*. Mahwah, NJ: Erlbaum.

Westmoreland, H., & Little, P. (2006). *Exploring quality standards for middle school after school programs: What we know and what we need to know*. Cambridge, MA: Harvard Family Research Project.

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.

Wisconsin Center for Education Research & Policy Studies Associates, Inc. (2005). *Study of promising afterschool programs: Observation manual for site verification visits*. Madison, WI: Wisconsin Center for Education Research & Policy Studies Associates, Inc.

Yohalem, N., Wilson-Ahlstrom, A., Fischer, S., & Shinn, M. (2007). *Measuring youth program quality*. Washington, DC: The Forum for Youth Investment.