What Works In Out-of-School Time and How Do We Know?

Brianna Kennedy

Abstract

The burgeoning field of Out-of-School Time provides academic and social services to K-12 students during the time that they are not in school. Hundreds of service providers offer a range of programs that target various student populations. Many tools exist to measure program performance, but no uniform evaluation methods exist to inform funders and community partners about the performance of local programs. The field would benefit from more uniform evaluation efforts, greater articulation between programs, and increased partnerships between service providers and other institutions.

Much attention has been given to the academic achievement gap between students of different ethnicities and socioeconomic status. As schools continue to work to meet the needs of students, often with insufficient resources, other organizations are playing an ever greater role in providing academic and social services to increase student achievement. These organizations and agencies offer such services during the time when students are not in school, also referred to as Out-of-School Time (OST). OST programs include all programs that supplement students' school experiences and that take place during times outside of students' regular school days. OST providers include school districts, non-profit organizations, and community-based groups. While evaluation of Out-of-School Time (OST) programs is gaining ground as a necessary and desirable component of program implementation, no collection of commonly employed techniques and evaluations exists. Most OST programs are not evaluated in a comprehensive manner (Tierney 2002), and evaluation is often thought of as client satisfaction or the number of clients served. However, funders, politicians, and community leaders are putting increased pressure on OST programs to document results in the form of some kind of formal evaluation. By linking specific program components to positive results through rigorous evaluation, programs can focus resources on those activities that are effective.

OST evaluation sits at the nexus of traditional program evaluation[1] and ad hoc program-specific efforts used to justify funding. One challenge of evaluating OST programs is the variety of programs offered to children during out-of-school time.

[1] See *Measuring Program Outcomes: A Practical Approach* published by the United Way for an example of a "how to" approach to traditional program evaluation.

Each program is based on the particular mission of the sponsoring organization, targets a particular group of students, lasts a specific length of time, occupies facilities of varying quality, operates with particular financial resources and personnel, and faces particular obstacles. Therefore, no single set of criteria can be employed to evaluate all programs. However, multiple existing efforts at evaluating such programs can offer a collection of basic strategies that can be employed to push the evaluation of programs beyond a simple account of the number of clients served. While attendance can be one important indicator of program effectiveness, it does not reflect program content or quality. Similarly, client satisfaction surveys or participant reports may indicate that participants are content with the program, but such reports do not provide definitive evidence that the program made actual significant changes in participants' lives (Royse et al. 2006).

In this paper, I first review the existing literature on OST and evaluation. This section is meant to provide a broad overview of how quality has been defined in the OST field and how programs have been evaluated according to these guidelines. The next section of the paper describes obstacles for evaluating OST programs. The bulk of this article provides strategies for those involved in OST evaluation to implement quality programs and then evaluate their efforts. The goal here is to expand the dialogue about quality OST programs across a variety of stakeholders by providing simple tools and a common discourse regarding program quality.

What Do We Know Works?

Since out-of-school time programs vary widely in purpose, funding, facilities, demographics served, and number of students served, defining "what works" leads directly to the question "For whom?" In addition to the variation in programs, the population of students served may vary widely in attendance rates and length of time in the program. Despite the limitations of generalizing across programs and populations, Bouffard, Little, and Weiss (2006) identify four factors that indicate high quality in OST programs: "positive staff-youth relationships; opportunities for skill-building and mastery; opportunities for youth engagement, voice, and decision-making; and positive peer relationships" (p. 2).

These indicators of quality constitute the requirements for programs to "work," or provide programming that meets students' academic, social, and emotional needs in an after-school setting. However, these program attributes are quite vague. Practitioners need a more directive guide to implementing quality programs. Knowing "what works" and having the knowledge, skills, and resources to implement successful program components are two different things. While the literature presents a fairly cohesive picture of activities that generally make kids' lives better, the exact costs and benefits of those activities depends entirely on successful implementation. Since programs often compete for scarce resources, practitioners may be hesitant to collaborate with others. Open dialogue regarding successful implementation strategies would further the OST field, since a gap in the literature exists regarding

implementation. In a rare acknowledgement that quality programs should be scaled up, even if the scaling is done by other organizations competing for similar resources, an after-school program in Los Angeles called LA's BEST published a program replication manual (Freeman and Redding 2006). Since it was designed to assist those who want to create a similar model, one would presume to look there for assistance in implementation. However, while the manual describes many facets of the program, one is still left wondering how these activities look "on the ground" on a daily basis. Nevertheless, descriptions of program components offer a start in identifying program quality.

Multiple practitioner-oriented manuals offer standards by which out-of-school time programs can be evaluated. In the past five years there has been a proliferation of self-assessment checklists with various indicators of success designed for practitioners. In April of 2006, the Harvard Family Research Project compiled a list of forty-three formative evaluation tools that covered fifteen different categories on which OST programs were judged (Harvard Family Research Project 2006). These evaluation tools addressed between four and all forty-three of the following categories:

a) **Staffing**
   - Supervision and behavior management
   - Human relationships

b) **Programming**
   - Positive youth development
   - Family, school, and community involvement
   - Safety, health, and nutrition; program planning and structure

c) **Physical Environment**
   - Safe facilities
   - Engaging and informative wall decorations
   - Clean, accessible, and functional restrooms and drinking fountains
   - Walls free of graffiti

d) **Accountability**
   - Program administration and management
   - Assessment, evaluation, and accountability
   - Organizational capacity
   - Sustainability; equal access
   - Fiscal management (Harvard Family Research Project 2006).

Since so many evaluation tools mention the points above, one can conclude that a general consensus exists that programs should at least address each of these categories. Each larger category listed above contains measurable indicators that are listed on the actual evaluation tools themselves. Certainly there exists no dearth of lists of indicators of successful OST programs.

How Do We Know If It Is Working?

Various tools are used to collect data for program evaluations. The easiest data to collect is attendance statistics. Practitioners can easily count how many students they serve, how often the students attend the program, and how long each day the students stay. Program evaluations can correlate the attributes of student attendance with other identified student outcomes (Chaput, Little, and Weiss 2004). However, the fact that students attend does not necessarily mean that a program is successful. Indicators of quality can be generated, but the implementation of various program components determines whether certain indicators will be met or not. Practitioner manuals provide guidelines for initial program planning and formative evaluation which can help staff members adjust program components to better meet program objectives.

Since programs are ongoing, summative evaluations-the final evaluations that indicate a program's overall success-might take place at the end of a school year, or at some other natural pause in the flow of the program. Summative evaluations are generally used to see if the program was effective during the time period being evaluated. Summative evaluations will test the program against its own identified goals using a particular evaluation design. Program evaluations can consist of standardized measures, school records, client satisfaction surveys, and qualitative data from focus groups, individual interviews, document review, or other methods. The Harvard Family Research Project's OST Evaluation Snapshot, *Measurement Tools for Evaluating OST Programs: An Evaluation Resource (2005)* identifies fifty-eight measures that have been used to gauge changes due to after school programs in the areas of:

- academic achievement
- academic/educational attitudes and values
- future plans
- life events and experiences
- mental health and behavior
- relationships
- self-perception/self-esteem
- drug use and prevention
- program quality/program environment
- multi-component scales

Out-of-school time programs can have an effect on each of these areas of young people's lives, depending on the particular program goals and outcomes. For example, in programs that target high achieving youth, indicators regarding academic achievement, future plans, and program quality may be of foremost importance in goal-setting and evaluation. However, for programs that target at-risk youth, life events, mental health, self-esteem, and drug prevention might be the most pertinent indicators of success.

Obstacles for Program Implementation and Evaluation

While evaluation is clearly necessary, there are difficulties in ascertaining direct causal relationships between program components and outcomes. The Harvard Family Research Project also maintains a database of OST program evaluations that uses various research designs to show which programs are successful and to what degree (Harvard Family Research Project 2005). However, evaluation indicators do not reflect implementation strategies. Plenty of information exists by way of evaluation tool indicators regarding general ideas about what to do, but little is written about how to do it. Descriptions of program components cannot be linked with particular program outcomes. Thus, the variables necessary in program implementation to ensure quality remain a mystery. The greatest challenge for program implementation is addressing the tacit skills and practices that actually account for program success.[2]

Accounting for confounding variables that might be the actual explanations for why a program is successful matters most if the program is going to be replicated in other environments. Such confounding variables might include the age or years of experience of staff members, staff salary, specific ways that staff interact with students, staff attentiveness to students when they are "hanging out," levels of student delinquency in program participants, or any number of other factors that could affect evaluation outcomes but never actually be addressed or measured. Additionally, students and their families could be receiving multiple services from different agencies, which would affect OST program outcomes, but the outcomes would not reflect the direct result of the OST program itself. Evaluation designs must address the challenge of confounding variables in order to demonstrate the program's effects on students' lives. Finally, evaluation results may be further complicated by the fact that diverse children may benefit differently from certain programs (Scott-Little, Hamann, and Jurs 2002).

Conducting Process and Final Evaluations

While plenty of challenges exist for program evaluation, OST personnel should not get discouraged by potential limitations of their evaluations. Instead, the field of program evaluation offers some simple tools to get started in conducting a program evaluation. While many programs will need to hire an external evaluator as a requirement of particular funders, others may wish to use these guidelines to launch new programs or to conduct internal evaluations to determine the effectiveness of certain activities. The following strategies can be used as a sequential guide or as independent tools to increase the complexity and value of efforts that attempt to show a program's impact.[3] See Table 1 for a summary of strategies discussed throughout this article.

[2] See Shevlin, Banyard, Davies, and Griffiths, 2000, and Willig, 1985 for examples of the importance of confounding variables in evaluation design.

[3] Strategies offered in this article omit cost-benefit analysis.

Table 1: Strategies Used to Evaluate Program Effectiveness

Strategy 1	Identify Outcomes
Strategy 2	Identify the Program Activities That Address Each Goal
Strategy 3	Identify Indicators of Success
Strategy 4	Collect Data
Strategy 5	Analyze Results
Strategy 6	Present Results

Strategy 1: Identify Outcomes

By identifying program outcomes, program staff members share the goals toward which they are working. In its simplest form, identifying program outcomes means identifying first how the program will change the knowledge and attitudes of participants (immediate outcomes), then how the program will change the behavior of participants (intermediate outcomes), and finally, how the program will change the quality of participants' lives (long-term outcomes).[4] For example, assume that a program targets at-risk students and wants to help them be more successful at school. An immediate outcome might be that students become excited about attending the program and receiving homework help. An intermediate outcome might be that students turn in their homework more often. A long-term outcome might be that students become more successful in school.

In traditionally structured OST programs-those in which one agency offers a set of services-program outcomes should be aligned with the mission and goals of the sponsoring organization. However, the field of out-of-school time is expanding to include new models of partnerships across universities, community organizations, schools, and families. The Harvard Family Research Project refers to these partnerships as *complementary learning* (Harvard Family Research Project 2005). In a complementary learning model, multiple agencies coordinate efforts to meet client needs. As the university increases its role in K-12 public education, schools must articulate how OST programs can best benefit students. In partnerships in which one partner has significantly more resources or political influence than another, an effort must be made in the collaboration to counterbalance the power dynamic. Those institutions that work most directly with the targeted population should have the most influence in identifying program outcomes. More specifically, if academic outcomes

[4] Gokul Mandayam, 2006, personal communication

are desired, students' teachers should be consulted to see how students' needs might best be met. If social or emotional needs are being addressed by the OST program, school counselors or local social workers could be consulted so that program coordinators and evaluators gain a sense of the issues that the students might be facing and how progress gained might be evaluated.

Examples of successful program outcomes include:

- Improving students' grades in school
- Improving students' attendance in school
- Providing a safe and healthy environment during after-school hours
- Nurturing primary relationships between youth and staff
- Facilitating healthy social development between peers
- Improving students' self-esteem

There are endless possibilities to the combinations of outcomes a program could choose. The above are only examples of a few. Making sure that all staff members are aware of the goals of the program will enhance the possibility that outcomes are being met successfully. Therefore, the staff as a group should identify program outcomes if possible. If program outcomes are already in place when staff arrive, they should be trained regarding the desired outcomes of the program and how the program design reflects those outcomes. Having all staff "on board" regarding program goals will provide a general guide for adjusting program components that seem less than successful and resolving conflicts in ways that further program objectives.

Strategy 2: Identify the Program Activities That Address Each Goal

Each program goal should lead into the development of an activity that addresses that goal. For example, if a program seeks to improve students' grades, some activity that the program implements should address this goal. A common program component that addresses students' grades is tutoring and homework help. While it is difficult to conclude that particular activities can account for related improvements, there must be at least one activity that addresses each program goal. If program staff cannot articulate how program activities reflect the mission and goals of the program, then the staff should revisit their program goals and implementation design. Table 2 lists sample program activities that could address the outcomes listed above.

Table 2: Program Activities That Reflect Particular Program Goals

Program Outcomes	Sample Activities That Address Outcomes
Improve students' grades in school	Tutoring
Improve students' attendance in school	Offer incentives (e.g. field trips) for improved school attendance
Provide a safe and healthy environment during after-school hours	Offer nutritious snacks and ban "junk food"; ban gang attire; make sure that all participants both inside and outside of buildings are supervised
Nurture primary relationships between youth and staff	Mentoring
Facilitate healthy social development between peers	Teach problem-solving skills, anger management, and/or social skills as mandatory classes; have team-building exercises where students are on the same team over time
Improve students' self-esteem	Have rotating student leader positions that determine meaningful program planning

The activities listed here offer ways that a program could attempt to fulfill its goals. In a quality program, each outcome will have at least one activity that addresses it, and the purpose of each activity will stem from the original program goals.

Strategy 3: Identify Indicators of Success

Perhaps the most difficult part of the process of evaluating a program is determining how to measure outcomes. Multiple indicators can and should be used to determine program success. In the example from the first strategy, an indicator of student excitement would be increased attendance rates and homework completed during the program. An indicator of students turning in homework more often might be a check-in with the students' teachers regarding homework completion. An indicator of school success might be grade point average. Table 3 gives examples of indicators that could show whether a program is meeting its goals. The forty-three evaluation tools identified by the Harvard Family Research Project (2006) provide other examples of identified outcomes and indicators.

Table 3: Sample Indicators of Activity Success

Program Goals	Sample Activities That Address Outcomes	Sample Indicator
Improve students' grades in school	Tutoring	Report card indicates improvement
Improve students' attendance in school	Offer incentives (e.g. field trips) for improved school attendance	Report card indicates improvement
Provide a safe and healthy environment during after-school hours	Offer nutritious snacks and ban "junk food"; ban gang attire; make sure that all participants both inside and outside of buildings are supervised	Students report feeling safe; students receive nutritious snacks
Nurture primary relationships between youth and staff	Mentoring	Staff and students report the development of significant relationships
Facilitate healthy social development between peers	Teach problem-solving skills, anger management, and/or social skills as mandatory classes; have team-building exercises where students are on the same team over time	Students show concern for others, resolve conflicts through talking, and do not submit to peer pressure
Improve students' self-esteem	Have rotating student leader positions that determine meaningful program planning	Self-esteem measures increase over time

Strategy 4: Collect Data

The next strategy that can be used in the process of program evaluation is data collection. There are several types of measures that one can use, all of which are explained here. Program evaluations typically consist of quantitative measures, qualitative methods, or both. Quantitative measures are those that reflect progress through numerical gains. Qualitative methods reflect participants' experiences and require that the evaluator interact directly with the participants. Evaluators may choose to use a combination of quantitative and qualitative data to present a more comprehensive picture of the effects of the program. Quantitative measures include

those referred to above as standardized tests that reflect various changes in students. (For examples of measures, see Harvard Family Research Project 2006; Robinson, Shaver, and Wrightsman 1991; or Maddox 1997). Creating your own scales can be useful in formative evaluations that will only be used internally, but conclusive evidence is often held up to the "gold standard" of experimental design that employs measurement instruments tested for validity and reliability. Therefore, in the evaluation that will be published for distribution, it is important to include the quantitative measures used to show program progress and indicate the validity and reliability of those measures.

Qualitative methods include individual interviews, focus groups, observations, and document reviews. Qualitative methods may be dismissed as not adhering to the "gold standard" of research. However, a quality mixed method design can yield more information about whether the program is successful and why than can an evaluation design that only uses one type of method. For example, a quantitative measure like the Self-Esteem Index can measure whether students' self-esteem changed over time. However, the evaluators cannot be sure that the change was the result of the program. They would gather invaluable data by interviewing the students and their families and asking them directly about the change. Together, the results of the Self-Esteem Index and the interviews will present a more complete picture of how the program is doing.

The First Step: Baseline Data

Before enrolling a new cohort of students into the OST program, it is crucial to the evaluation process to capture baseline data. The evaluators will want to find ways to capture the participants' existing development in each area that the program will address. The evaluators can use the same measures that they will use to test them after they have been in the program for a given length of time, or they can use different methods. Advantages and disadvantages to each are discussed in the following section.

Tool 1: Pre- and Post-Tests

One way to gauge student progress, is to use a measure that shows students' scores before they enter the program and then scores from the same measure after they have participated in program activities for a specified duration of time. For example, as in Table 2, if one of the program goals is to improve self-esteem and the program implements a rotating leadership position, each participant may be given the Self-Esteem Index both before and after their leadership rotation. An increase in scores could be at least partially attributed to the program activities. The advantage to using the same measures before and after participation is that the evaluator and participants are familiar with it, and a change in score cannot be attributed to a change in the test. However, the disadvantages to using the same measure is that the evaluator must make attempts to account for students' natural maturation over time, and also the tendency for students to get better at tests over time if the tests remain the same.

Tool 2: Program Rubrics and Scores

Rubrics are an easy way to assess program progress, but lack validity and reliability. The tools compiled by the Harvard Family Research Project (2006) generally consist of locally produced rubrics that give a quick snapshot of how the program is doing according to a particular stakeholder at a particular moment in time. Rubrics use a set of categories that reflect increasing success. For example, here is a possible rubric based on the sample indicators of success in Table 3.

Rubrics can be completed easily and by a variety of stakeholders at the same time to provide an overview of program progress from multiple perspectives. Rubrics are an example of qualitative methods, as they lack reliability and reflect stakeholder opinions which are subject to change. Rubrics frequently ask the respondent to rate the program on each indicator by giving it a score of one through four, rather than using a Likert scale like the one used above. Comprehensive rubric scores can be subjected to descriptive statistics, allowing the evaluator to compute the mean, median, and mode each time the rubric is completed. Improvements in rubric scores can be used internally to generate staff discussions and serve as process evaluations.

Tool 3: Staff/Stakeholder Observations

Observations done at the site during the program can reveal what program activities actually look like during implementation. Rather than providing opinions from staff, observations give the evaluator direct access to program content. Evaluators may note significant occurrences that staff consider commonplace and do not notice. Also, having various stakeholders observe activities can provide multiple perspectives of a program's performance. Observations can be done in a number of ways, but here is a sample protocol that an observer might use during an observation:

1. Describe the physical space, noting anything you find significant.
2. Describe the activities in which each student or group of students is participating.
3. What do you notice about staff and student interactions?
4. What do you notice about student affect?
5. How does the program provide a safe and nurturing environment? What could be improved?

Any variation of questions can be compiled depending on the information that the evaluator hopes to capture. Also, observation protocols are easy enough to generate that different protocols could be created for subsequent observations. Observations are a qualitative method that can offer multi-faceted perspectives on the same activities. The nature of open-ended questions allows for each observer to note things about their experiences with the program that they find significant. Since each observer will have different background experiences and different values, the evaluator will gain multiple insights in compiling the observation notes.

Tool 4: Client Satisfaction Surveys

Measures of client satisfaction provide an additional tool to evaluate program success. Client satisfaction surveys reflect client perspectives on how successful the program is at meeting its goals. In addition to a survey, clients can be interviewed to get their perspectives. While surveys are less time-consuming for stakeholders to complete, interviews will present a more comprehensive picture of stakeholder views. The program rubric above can be adapted to create a stakeholder survey for participants, families, staff members, or other community partners. Sample interview questions for student participants might include:

1. What does this program provide that helps you to improve your grades in school? How successful are those activities? What would make them even better?
2. How often are you absent from school? Does the program help you attend school more often? Explain.
3. How does this program teach you to solve problems between you and other kids? Have you learned things that you use to solve problems outside of the program?
4. Does this program allow kids to come here who get in trouble a lot? How does it affect you when those kids are here?
5. What do you feel like you have accomplished in this program? Has it helped you to feel proud of yourself?

The above questions could also be used as a protocol for focus groups, small groups of participants that meet together and discuss their answers to the questions. Focus groups can generate discussion that may allow participants to identify more ideas than if they are interviewed on their own. However, focus groups can also stifle the voices of students who are less assertive, and must therefore be conducted by a facilitator that encourages input from all participants. Also, the wording of the questions and the length of the interviews should be age-appropriate; the interviewer should be prepared to explain questions in different ways or give examples that might help generate ideas among students. Oftentimes, students do not know what they know and adults must help them to put into words the experiences that they have had.

Strategy 5: Analyze Results

Once the evaluator has determined which tools to use to collect data and completed the data collection, (s)he must determine how to analyze the results of the data. If multiple tools have been used, there will likely be quite a bit of data to analyze. Data can be analyzed using special software that will help generate statistical reports or qualitative themes that arise. One example of software that can be used to analyze quantitative data is the Statistical Package for Social Sciences (SPSS), and one example of software that will trace themes through qualitative data is NUD*IST.

Since most data are fairly straightforward and simple, the evaluator can also do the data analysis by hand. Likert scales can be given values, which can be summed or used to find the mean, median, or mode. Observation and interview data can be read for general themes that the evaluator believes best represent the comprehensive views of the respondents. Regardless of the method chosen, the final presentation of the evaluation results should include a description of how the data was analyzed.

In the description of data analysis, the evaluator will want to be sure to present possible limitations of the evaluation design. The possibility of confounding variables would enter the discussion at this point. For example, the evaluation may reveal that the students who showed the most improvement on the Self-Esteem Index were also all involved in student government during the same time period, or all had a certain teacher at school. Those are two examples of confounding variables that could explain an increase of self-esteem that cannot be directly attributed to the effects of program activities. The evaluation should be designed as comprehensively as possible so that these external factors are revealed by at least one of the methods used.

Strategy 6: Present Results

All stakeholders should be made aware of the results of the evaluation. It is especially important to brief the staff on outcomes of the evaluation. Even if the evaluation was conducted to sum up the program's performance over time, it can still be used to enhance program performance. Different audiences will require different formats for result presentations.

Participants and Families

For participants and families, use layperson's language and consider visual aids. Also keep in mind that not all families will speak English as a first language. As it should be a priority of the program to include all stakeholders in assessment, program results should be made available in the dominant language(s) of the community where the program is located. Consider asking participants and families for feedback regarding the evaluation results and how the program might do even better on its next evaluation.

Program Staff

Sharing the results with program staff is crucial to improved program implementation. Presentations to staff should also be done in layperson's language and provide visual aids where applicable. This presentation should comprise all program elements and give a full picture of how the program is doing in each area. After the presentation, staff should be aware of how well the program activities are serving the program goals and areas in which improvement is needed. After the presentation, the evaluator could facilitate a discussion among staff members that would serve to realign program goals, activities, and outcome measurements. Additionally, staff members should have the opportunity to question any evaluation results that do not seem to align with their experiences. If laypeople are not given the opportunity to provide input regarding program reforms, the evaluation will simply be dismissed as one more thing "coming down" from people "who don't know what we really do here." Staff should also be

encouraged to recognize which program activities seem to show the best results so that they can continue to implement those components. Even if positive results are due to confounding variables, those variables will not be apparent unless staff members continue to implement the activities.

Advisory Boards, Community Partners, and Universities

Boards should be given more information regarding how decisions were made about the evaluation. Ideally, the board was consulted with the original evaluation design before implementation. The post-evaluation report gives the evaluator a chance to share how (s)he completed her job. If the information is to be shared with university or research-oriented partners, information about the choice of measures, validity, reliability, and the trustworthiness of the qualitative data should be included. Community partners can be engaged regarding the design itself as well as the results in order to improve methods used in the future.

Overcoming Obstacles and the Need for Scaling Up

The above strategies provide possible actions for out-of-school time programs to design and implement program evaluation. OST providers need to work together in discussing program goals, activities, and measures of success. By coming together and adopting similar strategies and policies, programs could more easily scale up. The following advantages from such scaling up could result:

- Ease in starting new programs
  With the reassurance that programs draw from a reliable guide for implementation and evaluation, funders will be more likely to support multiple OST programs. A wider distribution of funds would encourage the diffusion of information among service providers and increasing efforts to create replication materials.

- The avoidance of "reinventing the wheel"
  With the availability of materials that have been proven to work in other settings, service providers interested in adding particular activities to their programs can make informed decisions about implementation. Modifications to any material that has been tested in particular contexts will need to be modified in order to suit the particular context of the program. However, experimenting with modifications to existing ideas will be far less time consuming than thinking of new activities, piloting them, and generating preliminary rounds of evaluation.

- Service to larger numbers of kids
  The ease of replicating programs shown to be successful-even as modifications are made according to contexts-will allow more programs to serve more kids. Needs assessments of communities should be conducted prior to the implementation of any new programs to determine what services would be utilized.

- Increased accuracy of program evaluations
  As programs scale up, evaluations of multiple-site programs will have more data to examine possible confounding variables. As programs evolve and learn, evaluation efforts will become stronger and the field of Out-of-School Time will be able to identify more definitively what works and under what circumstances.

- Expansion of partnerships between families, schools, universities, and community organizations
  The multiplication of programs will require the use of facilities, the hiring of new staff, and coordination between new programs and existing community resources. Ideally, existing community institutions-especially universities and other institutions with multiple resources-would sponsor such programs. Partnerships between multiple community institutions would serve to strengthen existing programs and offer strong foundations for the creation of new ones.

Next Steps

Although efforts at generating tools for internal evaluations are proliferating, the field of out-of-school time is still in its infancy in developing comprehensive and proven activities. Multiple obstacles exist that hinder this effort, including a fundamental disagreement regarding the purposes of OST programs. However, the challenge that exists is for programs and community partners with similar goals to form alliances and work toward rigorously evaluating and expanding their efforts.

References

Bouffard, S., P. Little, and H. Weiss. 2006. Building and evaluating out-of-school time connections. *The Evaluation Exchange* 12 (1): 2-6.

Chaput, S., P. Little, and H. Weiss. 2004. Understanding and measuring attendance in out-of-school time programs. *Harvard Family Research Project.* http://www.gse.harvard.edu/hfrp/content/projects/afterschool/resources/issuebrief7.html (accessed Retrieved March 31, 2006).

Freeman, K., and B. Redding. 2006. LA's BEST after school enrichment program. *Replication Manual.* http://www.lasbest.org/resourcecenter/replication.php (accessed December 1, 2006).

Harvard Family Research Project. 2005a. *After school program quality assessment categories of standards.* http://www.gse.harvard.edu/hfrp/content/projects/afterschool/conference/summit-2005-standards.pdf (accessed December 5, 2006).

———. 2005b. *Complementary learning.* http://www.gse.harvard.edu/~hfrp/projects/complementary-learning.html (accessed March 18, 2005).

———. 2005c. *Out-of-school time evaluation snapshot, measurement tools for evaluating out-of-school time programs: An evaluation resource.* http://www.hfrp.org.

———. 2005d. *Out-of-school time program evaluation database.* http://www.gse.harvard.edu/hfrp/projects/afterschool/evaldatabase.html (accessed March 1, 2005).

Harvard Family Research Project. 2006. S*ummer success: Challenges and strategies in creating quality academically focused summer programs.* http://www.gse.harvard.edu/hfrp/content/projects/afterschool/resources/issuebrief9.pdf (accessed December 11, 2006).

Maddox, T., ed. 1997. *Tests: A comprehensive reference for assessments in psychology, education, and business.* Fourth edition. Austin, TX: Pro-Ed.

New York State Afterschool Network. 2006. *Program quality self-assessessment (QSA) tool.* http://www.tascorp.org/policy_resources/advocacy_partners/nysan (accessed November 1, 2006).

Robinson, J., P. Shaver, and L. Wrightsman, eds. 1991. *Measures of personality and social psychological attitudes.* San Diego: Academic Press.

Royse, D., B. A. Thyer, D. K. Padgett, and T. K. Logan. 2006. *Program evaluation: An introduction.* Belmont, CA: Thomson.

Safe and Sound: Baltimore's Campaign for Children and Youth. 2006. *Standards for Baltimore after- school opportunities in youth places.* http://www.bbravo.org/pdf/workbook.pdf (accessed November 1, 2006).

Scott-Little, C., M. S. Hamann, and S. G. Jurs. 2002. Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation* 23 (4): 387-419.

Shevlin, M., P. Banyard, M. Davies, and M. Griffiths. 2000. The validity of student evaluation of teaching in higher education: love me, love my lectures? *Assessment and Evaluation in Higher Education* 25 (4): 397-405.

Tierney, W.G. 2002. Reflective evaluation. In *Increasing access to college*, eds. W. G. Tierney and L. S. Hagedorn. New York: State University of New York Press.

Willig, A. 1985. A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research* 55 (3): 269-317.

Author Information

Brianna Kennedy studies marginalized urban youth and program evaluation at the University of Southern California. She is a research assistant at the Center for Higher Education Policy Analysis in the Rossier School of Education.

Brianna Kennedy
University of Southern California
Waite Phillips Hall-WPH, Room 701A
3470 Trousdale Parkway
Los Angeles, CA 90089-4037
E-mail: blkenned@usc.edu
Telephone: 213-740-1515
Fax: 213-740-3889