Harvard Family Research Project’s (HFRP) Issues and Opportunities in Out-of-School Time Evaluation briefs are short, user-friendly documents that highlight current research and evaluation work in the out-of-school time field. These documents draw on HFRP’s research work in out-of-school time in order to provide practitioners, funders, evaluators, and policymakers with information to help inform their work. This latest brief offers expert commentary on the implications of When Schools Stay Open Late: The National Evaluation of the 21st-Century Community Learning Centers Program, First Year Findings for future evaluation and research.

INTRODUCTION

The 21st Century Community Learning Centers (21st CCLC) program began in 1998 under the Elementary and Secondary and Education Act, with $40 million awarded to 99 grantees in 34 states, supporting programs in about 360 schools. Reauthorized under the No Child Left Behind Act, the program received $1 billion in 2002. On February 3, 2003 the U.S. Department of Education released the first-year findings from the national 21st CCLC program evaluation. Conducted by Mathematica Policy Research, the national evaluation examined the characteristics and outcomes of typical 21st CCLC programs. Simultaneous with the report’s release, the President’s Fiscal Year 2004 Education Budget Summary and Background Information cited the “disappointing initial findings from a rigorous evaluation of the 21st Century Community Learning Centers program” as a rationale to request a decrease in funding for the program by 40 percent. According to the budget summary, “the evaluation indicates that the centers funded in the program’s first three years are not providing substantial academic content and do not appear to have a positive impact on student behavior.”

The decision to recommend a 40 percent reduction in the 21st CCLC budget has stimulated renewed commitment to evaluation as well as efforts to examine this one evaluation in the larger context of other relevant evaluation and research. It also has catalyzed interest in developing a strong future research and evaluation agenda to support program development, improvement, and accountability. This brief provides commentaries from researchers, evaluators, and practitioners about new directions for research and evaluation in order to reframe the “gotcha accountability” game into one of learning for continuous improvement and accountability. Specifically, we asked, Given the recent push for science-based research, coupled with the release of the first-year evaluation findings from the 21st CCLC programs, where do we go from here to use research and evaluation to support the development of high-quality out-of-school time programs?

In a world where the principles of scientific research are increasingly invoked to guide policy, the decision to use the first-year data from an ongoing evaluation to cut funding is producing considerable scientific scrutiny and critique of the study’s methodology and findings—scrutiny that is crucial for scientific as well as policy advancement. We believe there should be an ongoing exchange regarding methodological choices, trade-offs, and issues about how to deal with the inevitable issues of “real-world” implementation in this and other large-scale evaluations. Therefore, this brief includes a commentary on the methodological issues in this particular evaluation and their implications for current policy as well as for future evaluation design, implementation, and use. It is our hope that this will stimulate a larger conversation about a productive research and evaluation agenda, about solutions to methodological problems, and about how to maximize the use of research and evaluation to support policy formation and service improvement.
In 1996, as part of the Congressional reauthorization of the Elementary and Secondary Education Act, I introduced legislation to create “community learning centers.” Our goal was to find ways to more efficiently use school resources, especially in rural and inner-city areas, for all citizens all year. The Clinton administration strategically directed this broad language to create today’s after school program, funded at $1 billion annually in FY 2003.

In the No Child Left Behind Act the law transitioned to a state grant program. Now, the administration seeks to add new standards to all federally funded programs which I call the “three A’s”— academics, access, and accountability. We’d be wise to positively respond to this new focus. After school programming is an important and growing component in the development of today’s youth. Yet we need to target these programs to those most in need, in ways that will enhance a student’s academic progress, assuring limited public dollars meet the test of accountability.

New research and evaluation is desperately needed to improve federal support for this program. Certainly one study (Mathematica’s) does not justify ending the program. But with limited resources and the new focus on academics, we must learn what works—especially for at-risk students. Then, we must restructure our programs to best achieve this goal. So, let’s get on with improving academics, access, and academic support in most programs. Although it is easy to dismiss the effects as small, this conclusion is no doubt premature, especially in light of the fact that this is an ongoing evaluation.

Did the administration have fair and reasonable scientific expectations? Scholars agree that no one should expect the 21st CCLC program evaluation to yield short-term effects on test scores, echoing Zigler’s early warnings concerning Head Start. By what criteria are the findings “disappointing”?

Accountability efforts and scientifically based research can either be used to generate knowledge that informs effective practices or to serve as a political lever to cut programs and expenditures on child and family services. Here we have a sad example of that latter—another case of death by evaluation.

The administration’s proposed cut to the 21st CCLC budget is not surprising. It is a rare elected official who expands rather than downsizes the pet programs of a predecessor from the opposing party. Dozens of social programs will suffer cutbacks in the next budget. What is surprising is that the administration has broken its own rules for bringing science into policy discussions. By announcing the cuts just as it released the report, opportunities for the research and policy community to apply the new rules for scientifically based research—conducting a rigorous, objective, and scientific review, discussing the findings, and debating responses in light of findings from other equally scientific studies—were effectively cut off.

Research should play a more central role in decisions to expand, redefine, or reduce programs. When used correctly, it can be a powerful counterweight to limit the big pendulum swings frequently associated with popular programs, to accelerate the growth of effective programs, and even to curtail the expansion of popular but ineffective programs. The Mathematica report includes promising
findings and valuable lessons that can inform both practice and policy. This and other studies should serve as platforms for much needed conversations about how to augment program quality and encourage longer and more intense participation. By using the study to justify cuts, the administration has curtailed conversation about a range of responsible strategies for improving the program, given these and other findings. Our concern should be the same if the proposed program budget had been doubled.

For a more detailed version of this commentary, visit the Forum for Youth Investment website at www.forumforyouthinvestment.org/resspeech.htm.

MINDY DISALVO
Program Director, Family Technology Resource Center, Decatur, Georgia

Because we want to know if our after school programs are the best that they can be, we eagerly welcomed the opportunity to be a part of the national evaluation of the 21st CCLC program. From the onset of the evaluation we were candid about our strengths and weaknesses, providing honest information about a program in its infancy. As a participant in the evaluation, I have two concerns related to the emphasis and use of the evaluation results.

First, there were very positive findings in the 21st CCLC evaluation—findings that could serve as a road map to improve existing and new programs, not close the doors to them. Based on our concurrent evaluation data, we expanded our curriculum, developed a parent/teacher/student homework completion policy, translated materials into two languages, extended hours of operation, and hired a nurse. We learned how to make our program better by using data. Therefore, it seems that the emphasis should be on how we can learn from the new evaluation report, and not what the report told us about academic impact.

Second, student achievement isn't, and never will be, solely a result of after school programs. Student achievement isn't a result of textbooks either, but we spend a fortune on them and no one is talking about cutting them from a budget! Improved student achievement is a result of a combination of components in a child's life, including how they spend their nonschool hours. Before student achievement becomes a priority for many of our after school programs, a safe place with a caring adult, friends, a healthy snack, and a promise of security comes first.

TIFFANY BERRY
External Evaluator, LA's BEST, Los Angeles, California

LA's BEST uses evaluation data by transforming program outcomes into organizational tools for program improvement. Since the program's inception in 1988, LA's BEST has placed a high priority on evaluation and we encourage feedback from the program's diverse stakeholders. Monitoring of program quality has been accomplished by leveraging internal sources (e.g., random site visits by the board of directors, site activity logs, opportunities to communicate between field and management staff, etc.), as well as data from external sources (e.g., the Center for the Study of Evaluation at the University of California Los Angeles). These data sources have yielded valuable insights, which have been fed back into program operations.

One of the most robust findings of the LA's BEST program relates to the duration and intensity of participation. Our evaluation reports indicate that when compared with nonparticipants, LA's BEST participants have fewer days of absences from their regular school, higher achievement on standardized tests in mathematics, reading, and language arts, and higher language redesignation rates to English proficiency. In addition, we have found that the relationship between participation intensity in one academic year and academic achievement was mediated by regular school attendance. This suggests that participating in LA's BEST resulted in better school attendance, which in turn related to higher academic achievement.

JAMES P. CONNELL
President, Institute for Research and Reform in Education, Toms River, New Jersey

Common sense tells us that public investment in programs serving youth should start with a research-based rationale, or theory of change. This theory should tell us how and why the proposed activities, in this case after school programs, can reasonably be expected to produce the designated academic and social outcomes.

In the absence of alignment between program activities and expected outcomes, the failure of the 21st CCLC program to produce its desired outcomes was virtually preordained. Two remedies present themselves: start with the outcomes you want and change program activities to those with a reasonable shot at achieving the outcomes, or start with the activities you have and adjust your expectations to outcomes they can achieve.

Either course makes some sense. The 21st CCLC evaluators call for remedy number one—enriching after school programs with more research-based activities tied to the desired outcomes. Many commentaries on the evaluation make energetic pleas for remedy number two—holding programs accountable simply for providing positive activities for young people while many of their parents are at work. The following steps could lead to better alignment between after school programs, academic outcomes, and evaluation: (1) develop educational and recreational ac-
As a researcher in the field of after school, as well as a member of the scientific advisory board for the 21st Century Community Learning Centers (21st CCLC) program evaluation, I am concerned about the serious methodological problems that call into question the report’s findings. The authors of the first-year report of the national evaluation of the 21st CCLC program have contended that the evaluation is the “most rigorous examination to date of school-based after-school programs” (U.S. Department of Education, 2003, p. xi) and that the research revealed programs to have “limited academic impact” (U.S. Department of Education, 2003, p. xii). However, neither of these claims is justified because the evaluation has serious methodological problems in both the middle and elementary school samples. These problems make it impossible to draw any meaningful conclusions (positive or negative) about program effects.

**PLAYING BY THE RULES**
The 21st CCLC Program Evaluation Violates Established Research Standards
Deborah Lowe Vandell, Professor of Educational Psychology, University of Wisconsin

As a researcher in the field of after school, as well as a member of the scientific advisory board for the 21st Century Community Learning Centers (21st CCLC) program evaluation, I am concerned about the serious methodological problems that call into question the report’s findings. The authors of the first-year report of the national evaluation of the 21st CCLC program have contended that the evaluation is the “most rigorous examination to date of school-based after-school programs” (U.S. Department of Education, 2003, p. xi) and that the research revealed programs to have “limited academic impact” (U.S. Department of Education, 2003, p. xii). However, neither of these claims is justified because the evaluation has serious methodological problems in both the middle and elementary school samples. These problems make it impossible to draw any meaningful conclusions (positive or negative) about program effects.

### Middle-School Evaluation Issues

The middle-school evaluation used a quasi-experimental design in which students who were enrolled in a 21st CCLC program (the treatment group) were contrasted with other students who were not enrolled in the program (the comparison group). Baseline data were collected for the two groups in fall 2000 and follow-up data were collected in spring 2001. Central to the integrity of any quasi-experimental study is the need to establish that treatment and comparison groups are comparable prior to the treatment (i.e., at baseline). The middle-school evaluation failed to meet this standard.

As documented in the September 12, 2002 Draft Interim Report (U.S. Department of Education, 2002), there were large differences in standardized test scores at baseline between the treatment and comparison groups that favored the comparison group (p<.00***). The mean standardized reading score for the treatment subjects at baseline was the 39.7 percentile, whereas the mean reading score for the comparison subjects was the 50.0 percentile at baseline. The mean percentile score in math for the treatment group at baseline was 34.3 percentile versus 43.6 percentile for the comparison group (p<.00***).

These test score differences indicate that the comparison students were more academically competent than the treatment group on these key variables. At a January 2002 meeting with researchers from Mathematica Policy Research (the research organization that conducted the evaluation) and staff members from the U.S. Department of Education (who commissioned the evaluation), the study’s scientific advisory board advised the research team that these baseline differences represented a significant challenge to the integrity of the evaluation. Unfortunately, the study authors did not address these baseline differences in their impact analyses.

As indicated in Table B.2 of the official first-year report (U.S. Department of Education, 2003, p. 140), baseline test scores were not controlled for in the impact analyses nor were propensity analyses recomputed to select a more appropriate comparison group that was matched to the treatment group on these key variables. Instead, evidence of the baseline test scores was removed from the first-year report (U.S. Department of Education, 2003, Table III.1, p. 55 and Table A.8, p. 127) and the authors did not acknowledge the likelihood that findings were biased because of the baseline differences in standardized test scores. Because the treatment students were much less academically competent at baseline than were the comparison students, the similarities in academic performance of the two groups at follow-up (U.S. Department of Education, 2003, p. 66) may even indicate that the programs had succeeded in narrowing the gap in academic performance between the two groups.

A second overarching issue with the middle-school evaluation pertains to contamination of the treatment and comparison groups. Implementation data, collected by the researchers, indicat-
ed that some schools districts used monies from
the 21st CCLC to fund the programs serving the
treatment subjects and monies from other sourc-
es to fund the same types of programs for the
comparison students. In essence, the programs
differed only in their source of funding. Because
the researchers did not explicitly address this is-

sue in the report, it is not possible to ascertain the
extent that the comparison students were actual-
ly attending similar programs. In other cases, as
the researchers noted, comparison students actu-
ally attended 21st CCLC programs.

Elementary-School Evaluation Issues
The elementary-school evaluation used a random
assignment experimental design to assess program
effects. The first-year elementary evaluation has
four fundamental problems.

1. Baseline data were selectively omitted from
the report, just as they were for the middle-school
evaluation. Baseline and follow-up math stan-
dardized test scores were reported in the Draft
Interim Report (U.S. Department of Education,
2002, pp. 93 and 97), but were omitted from the
official report that was released in February 2003
(U.S. Department of Education, 2003, pp. 93 and
97). In the case of the elementary-school evalua-
tion, the baseline test scores were similar, but the
follow-up scores showed a widening of a perfor-
mance gap that favored the treatment group. The
math scores of the treatment group increased
during the evaluation year, whereas the math
scores in the control group decreased during the
evaluation year. Unfortunately, the study authors
did not test these changes in test scores to deter-
mine if they were statistically significant.

2. Data were collected at only half of the in-
tended sites. Instead of examining impacts in 14
school districts as originally planned, data were
collected in only 7 school districts (a total of 18
different programs). This reduced sample size
substantially limited the investigators’ ability to
detect program effects.

3. A substantial proportion (4 of 18 pro-
gams, or 22 percent) of the programs in the ele-
mentary-school evaluation had only an inciden-
tal focus on academic and developmental
experiences for children. These four programs
were designed to serve adults in the school’s com-

munity (another focus area within the 21st CCLC
program charge) and children attended the cen-
ter only when they accompanied their parent or
grandparent. It is not clear why or how these
adult-focused programs would be expected to di-
rectly impact child outcomes.

4. Sites for the elementary-school evaluation
are not representative of the larger body of 21st
CCLC programs. Consequently, it will not be
possible to use performance data (either positive
or negative findings) as indicators of the success
or failure of 21st CCLC programs generally.

Final Thoughts
The methodological problems found in the first-
year report are so serious that I believe great cau-
tion must be taken in citing or using the study
findings. At the same time, it may be possible to
address some of these issues by a thorough reana-
lysis of the data. I encourage fellow researchers to
do just that when the data set becomes public.
Through further study and analysis of the public
data set, I believe that we will better understand
the circumstances under which the 21st Century
Community Learning Center programs can sup-
port (or fail to support) children’s academic and
social development.

References
U.S. Department of Education, Office of the Under Sec-
When schools stay open late: The national evalua-
tion of the 21st-Century Community Learning
Centers program, first year findings. Washing-
ton, DC: Author.

U.S. Department of Education, Office of the Under Sec-
retary. (2003). When schools stay open late: The
national evaluation of the 21st-Century Commu-
nity Learning Centers program, first year findings.
ed.gov/pubs/21cent/firstyear.

To read more information about the issues raised in this
critique, see the statement prepared by the members of
the scientific advisory board for the 21st CCLC eval-
uation (see Related Resources on page 8). The meth-
odological conversation continues with Mathematica’s
response to these issues (also listed on page 8).
tivities to help students meet a small number of broad academic standards that reflect the schools’ goals for their students, (2) give staff the resources to actively engage young people in these activities in different ways, and (3) assess the quality of implementation of these activities, their intended outcomes, and the connection between the two. We have seen such a theory of change approach help bring both realism and accountability to the work of changing public education. We expect it could do the same for after school programming.

JACQUELYNNE S. ECCLES
McKeachie Collegiate Professor of Psychology, Women’s Studies, and Education, University of Michigan

As Chairperson for the National Research Council (NRC) committee that produced Community Programs to Promote Youth Development,

my comments reflect my concern over the administration’s decision to cut the funds for the 21st CCLC program based on one evaluation report. This seems a very strange decision for an administration that stresses both the need for evidence-based practice and the importance of supporting healthy adolescent development. Our comprehensive report outlined the characteristics of many programs shown scientifically to have positive effects on many different aspects of adolescent development and provided examples of many high-quality programs with rigorously demonstrated effectiveness. We also discussed what is needed for adequate evaluation to improve these programs and make sound policy decisions. We proposed that the real challenge for the field is to increase the availability and sustainability of high-quality programs, especially in the context of unpredictable funding streams. We concluded that increased funding for the 21st CCLC program was one step the federal government could take to help increase the predictability of funding.

Consequently, I was appalled at the decision to cut these funds based on one quite-limited evaluation. Very little attention was paid in this evaluation to the characteristics of the programs being evaluated. Instead great attention was paid to the quasi-experimental and experimental evaluation designs used. As we discussed in the NRC report, these designs are powerful methodological tools, but they are not particularly useful if we do not know the quality of the programs being evaluated. If members of this administration truly value evidence-based practice, then they should pay more attention to the evidence-based reports that are so carefully put together by the NRC rather than use the results of one report to justify funding cuts. The existing evidence suggests to me that the administration should increase the funds for 21st CCLC program, but require better specification of the exact characteristics of the programs eligible for funding.

WHERE DO WE GO FROM HERE?
Heather Weiss, Director, Harvard Family Research Project, Cambridge, Massachusetts

We are now playing in a “new evaluation game” with new players and new rules. The game is different because not only is research and evaluation helping to shape policy, but the reverse also is true. Evaluation has always been played out within a political frame, but the No Child Left Behind Act helped define that frame by setting new rules or standards for research and evaluation with its five principles for scientifically based research (SBR) in education.

Not everyone has agreed that the new evaluation game is being played with the right set of rules. Some would argue that even if the end goal is to comply with this set of rules, there needs to be a learning curve where programs are held incrementally accountable for implementing the new scientifically based standards. For example, the Mathematica impact evaluation, underway well before the new SBR principles were signed into legislation, was prematurely subjected to the new rules and the result has had potentially dire policy consequences. As Pittman, Eccles, and McCartney point out in their commentaries, using one evaluation report to justify a policy decision runs counter to the new rules of SBR—indeed, they argue that subjecting the impact report to the new rules violates some of those rules’ very premises. Further, as Connell points out, in order to accurately assess program impact, there must be alignment between program activities and desired outcomes. In the case of the 21st CCLC programs, old programs were held accountable for new outcomes, thereby almost “preordaining” failure. Moving forward, all players must strive for alignment between desired outcomes and program strategies.

Despite disagreement over the rules, we are now in a position where, like it or not, the new rules are in play and we have to learn how to “get in the game.” So how do we play in this new game? There are at least three approaches.

First, as Gunderson points out, “we must learn what works,” and we must acknowledge the new game rules require after school programs to be accountable, demonstrate results, and improve their quality. However, as noted evaluator Mark Lipsey points out, “individual evaluation studies, however useful they may be to sponsors and stakeholders, yield approximate estimates of intervention effects and the relationships of those effects to the features of the program under assessment.” Further, he points out that perhaps the most useful and informative contribution to program managers and policymakers alike may be the consolidation of our piecemeal knowledge into broader pictures of the program and policy spaces at issue, rather than individual studies of specific programs.
Second, we need to shift from a system of “gotcha accountability” to a system of learning for continuous improvement and accountability. In a recent Washington Post commentary, former U.S. Secretary of Education Richard Riley notes that the original purpose of the evaluation was for continuous improvement, not accountability. He points out that the “Clinton administration did commission this study, but the purpose was to learn how to make after school programs even more effective ... [not to provide] political cover” for proposed budget cuts. Our two practitioner commentators (DiSalvo and Berry) agree that, even in the context of heightened accountability, to be good evaluation players requires a commitment to using data for continuous improvement as well as to show impact. A thorough reading of the Mathematica report reveals many promising implementation findings that need to be brought into the light and used for program improvement.

Third, there is increasing precedent for independent researchers to reanalyze national evaluation data, particularly when there are substantial policy implications. Mathematica’s decision to make public the entire database of results from its New York voucher study to independent researchers for reanalysis led to discovery of methodological problems on the part of Mathematica’s research partner. This, in turn, substantiated Mathematica’s cautions about the study’s findings and refuted their research partner’s conclusion that voucher use among poor black children significantly improved test scores. As Vandell points out in her methodological commentary, some of the issues that she feels plague the recent 21st CCLC study could be addressed through a reanalysis of the data.

Years of evaluation research have taught us lessons that are too expensive to learn again, such as don’t make large-scale investments in evaluation unless you are learning about program implementation along the way and don’t evaluate a program until it is proud. Moving forward, our responsibility as evaluators is to take advantage of the unintended window of opportunity provided to us by both old and new administrations to engage in a dialogue about how to apply these lessons to the new game to ensure that future research and evaluation of after school programs “plays by the rules,” and is used to improve the overall quality of after school programs.

This Issues and Opportunities in Out-of-School Time Evaluation brief was compiled by Heather Weiss, Director, and Priscilla Little, Project Manager of Harvard Family Research Project. All inquiries about this brief can be directed to heather_weiss@harvard.edu or priscilla_little@harvard.edu. We welcome additional methodological commentary and will post them to our website as they become available.

NOTES
3 Ibid.
8 See, for example, commentaries posted on the Afterschool Alliance website at: www.afterschoolalliance.org/voices_budget_cut.cfm.

ACKNOWLEDGEMENTS
Preparation of this brief was made possible through the support of the Charles Stewart Mott Foundation. We would also like to thank Phil Sparks of the Communications Consortium Media Center for his review of this brief.
Additional Commentary

The Afterschool Alliance, a nonprofit organization dedicated to raising awareness for and advocating on behalf of after school programs, has compiled reactions to the administration’s 21st CCLC program budget decisions from voices across the nation, including their own Executive Director, Judy Samelson. To view this set of commentaries go to www.afterschoolalliance.org/voices_budget_cut.cfm


Other Resources


Eccles, J., & Gootman, J. A. (Eds.). (2002). Community programs to promote youth development. Washington, DC: National Academies Press. This report explores the role of youth development programs and how best to design programs that enable youth to develop into healthy, happy, and productive adults. Policy, practice, and research recommendations to address the developmental needs of youth are included. www.nap.edu/catalog/10022.html


