

Translating science to policy under the No Child Left Behind Act of 2001: Lessons from the national evaluation of the 21st-Century Community Learning Centers[☆]

Joseph L. Mahoney^{*}, Edward F. Zigler

Yale University, United States

Available online 12 May 2006

Abstract

Passage of the No Child Left Behind (NCLB) Act of 2001 ushered in a new era of accountability for educational institutions and social programs in the United States. Federal funding became tied to demonstrated results on the basis of scientifically based research. In this paper we recount the science-to-policy translation process for one of the first studies considered under this new mantra, namely, the national evaluation of the 21st-Century Community Learning Centers after-school programs. The recount highlights a range of challenges, problems, and debates in the process, and discusses several lessons learned from this example that may be useful for the science-to-policy translation process in general, and for future studies under the NCLB Act specifically.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Science-to-policy translation; After-school programs; No Child Left Behind (NCLB) Act

1. Introduction

The No Child Left Behind (NCLB) Act of 2001 was passed into law by the 107th Congress on January 8, 2002 ([Public Law 107–110](#)). This Act intended to establish a new era of accountability for Federally-supported programs in the United States. Among other things, the NCLB Act required that funding for education and social programs be tied to evaluation and that evidence of program effectiveness be documented. Effectiveness was to be determined from “scientifically based research.”

The NCLB Act defined scientifically based research as that which:

(i) employs systematic, empirical methods that draw on observation and experiment, (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn, (iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple

[☆] The authors wish to thank the following individuals for their comments on an earlier report that provided a partial basis for this manuscript: Steven Barnett, Lyle Jones, Kathleen McCartney, Elizabeth Reisner, Michelle Seligson, and Deborah Lowe Vandell. We would also like to thank Ron Haskins and Judy Samelson for providing information on funding for the 21st-Century Community Learning Centers and An-Me Chung for discussing the goals of the Charles Stewart Mott Foundation’s enhancement study.

^{*} Corresponding author. Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520, United States.

E-mail address: joseph.mahoney@yale.edu (J.L. Mahoney).

measurements and observations, and (iv) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through comparatively rigorous, objective, and scientific review. (Public Law 107–110, 115 Stat. 1550–1551).

One of the first studies to be considered under the NCLB Act's definition of scientifically based research was the national evaluation of the 21st-Century Community Learning Centers (21CCLCs). The 21CCLCs provide a major source of after-school programming in the United States and serve over 1 million children and families. The evaluation of these centers is conducted by Mathematica Policy Research, Inc.

In this paper we discuss the science-to-policy translation process involving the first-year findings from Mathematica's evaluation carried out under the NCLB Act. In doing so, we consider the 2004 requested funding authorization for the 21CCLCs based on the evaluation, the evaluation's adequacy for informing policy decisions, resulting critiques and debates, and the Senate hearing leading up to the final funding appropriation for the 21CCLCs in 2004. The goal of the paper is to consider lessons that can be learned from this particular instance of translating science into policy that can be applied to future research and evaluation under the NCLB Act.

1.1. Why is the 21CCLC evaluation of interest for Applied Developmental Psychology?

Science should inform policy decisions. The centrality of applied psychological research in the policy process has ebbed and flowed across the history of our field (Zigler, 1998). However, in the current *zeitgeist*, the value and worth of psychological research depends increasingly on an appropriate integration of science with policy (McCall & Groark, 2000). On this score, the principle behind the NCLB Act's emphasis to incorporate science into policy decisions is laudable.

Our interest in Mathematica's evaluation is threefold. First, the evaluation's first-year findings provided the foundation for a proposed 40% reduction (\$400 million) in Federal funding for the 21CCLCs. Had the budget reduction been appropriated, after-school programs supported by 21CCLC funds serving hundreds of thousands of children and families in the United States would have been curtailed or eliminated. Whenever policy is made on the basis of a single study, the scientific underpinnings of the study need to be carefully examined (Zigler, 2003).

Second, the stated findings from the evaluation suggested that the 21CCLC's after-school programs had limited impact on academic or social outcomes. These findings were not congruent with the broader literature that underscores the importance and benefits of after-school programs and activities for young persons (e.g., the Packard Foundation's *Future of Children* (1999) series report *When school is out*, the National Research Council and Institute of Medicine's reports (2000, 2002) on *After-school programs to promote child and adolescent development* and *Community programs to promote youth development*, Public/Private Ventures 2002 report authored by Baldwin Grossman et al. (2002) *Multiple choices after school*, the 2003 Nellie Mae Foundation report *Critical Hours* authored by Miller, the 2003 National Research Council's report, edited by Smolensky and Gootman (2003), *Working families and growing kids: Caring for children and adolescents*, and the 2005 volume on *Organized activities* edited by Mahoney, Larson, and Eccles (2005)). The evaluation's incongruence with this broader literature raises additional questions about a proposed funding reduction made solely on the basis of one report of first-year findings.

Finally, the proposed funding reduction occurred in a historical climate of increased financial investment in, and rapid growth of, after-school programs. Rising rates of maternal employment and recognition of both the dangers and opportunities in how young persons spend their time during the non-school hours have fueled the growth of after-school programs. For instance, in 1991, the National Before and After-school Study reported that 1.7 million children were enrolled in a before- or after-school program (Seppanen et al., 1993). In 1997, data from the National Survey of American Families indicated that 6.7 million children with employed mothers were enrolled in such programs (Capizzano, Tout, & Adams, 2000). State and local expenditures increased markedly¹ and, beginning in 1998, Federal support for after-school programs from 21CCLC grants grew exponentially. Below we provide an overview the 21CCLCs and their funding history.

¹ For example, the Hawaii After-school Initiative (H.R. 69, 2002) provides for the development of quality after-school programs for every school-age child in the state. In 2000, California adopted the After-school Learning and Safe Neighborhoods Partnership Program that distributed \$87.8 million in grants to 947 after-school program sites statewide. More recently, California also passed Proposition 49 that created the After School Education and Safety Program Act of 2002 and earmarks up to \$550 million for universal after-school incentive grants for all public elementary, middle, and junior high schools including charter schools.

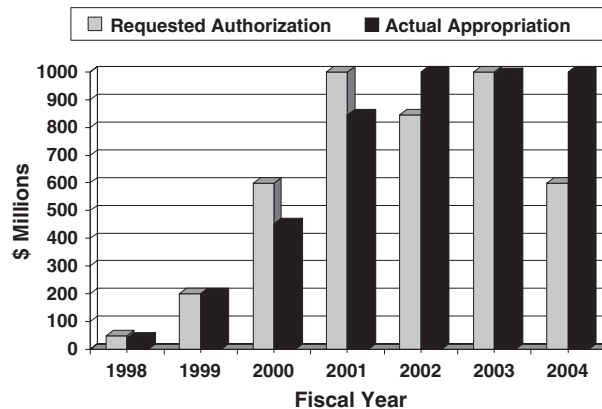


Fig. 1. History of requested funding authorizations and actual appropriations for the 21st-Century Community Learning Centers (1998–2004).

1.2. The 21st-century community learning centers

The definition of a Community Learning Center is an entity that offers academic, artistic, and cultural enrichment opportunities to students and families during non-school hours (before or after school) and periods when school is not in session (including holidays, weekends or summer sessions) (U.S. Department of Education, 2003a). Under the Elementary and Secondary Education Act, each 21CCLC program was required to carry out at least 4 of 13 different authorized activities.²

1.2.1. Overview of 21CCLC funding history

Funding for the 21CCLCs was first authorized under the Elementary and Secondary Education Act as amended by the Improving America's Schools Act of 1994 (U.S. Department of Education, 1994). In 1998, the 21CCLCs received the first substantive funding with grants provided for three years and awarded to local educational agencies (mainly school districts). Funds were restricted to school-based programs (including provisions for before-school and summer programs). The history of requested authorization and appropriated funding from 1998 to 2004 is shown in Fig. 1.

Both funding appropriations and the 21CCLC programs increased markedly. In 1998, \$40 million was appropriated and awarded to 99 grantees in 34 states, supporting programs in approximately 360 schools. In 1999, \$200 million was appropriated and programs served approximately 600 communities and 2100 schools in virtually every state. By 2002, \$1 billion was appropriated and awarded to approximately 1420 communities in 6800 rural and inner-city schools serving over 1 million students.

Funding for the 21CCLCs was reauthorized on January 8, 2002 as Title IV, under Part B of the NCLB Act (Public Law 107–110). The reauthorization resulted in twelve main changes that affect dissemination of funds, program content, and the requirements for program evaluation (e.g., program administration was transferred from the Federal to the State level, eligibility was expanded to public and private entities, requirements for evaluation and performance indicators increased, there was a stronger focus on academic enrichment activities, funds were targeted to low-performing schools in low-income areas) (U.S. Department of Education, 2003a).³

² Authorized activities include: (1) literacy education programs, (2) senior citizens programs, (3) children's day care services, (4) integrated health, social service, recreational, or cultural programs; (5) summer or weekend programs in conjunction with recreation programs, (6) nutrition and health programs, (7) expanded library service hours to serve community needs, (8) telecommunications and technology education programs for individuals of all ages, (9) parent skills education programs, (10) support and training for child day care providers, (11) employment counseling, training, and placement, (12) services for individuals who leave school before graduating from a secondary school, regardless of the age of such individuals, and (13) services for individuals with disabilities. Source: U.S. Code, Title 20, Chapter 70, Subchapter X, Part I—21st Century Community Learning Centers, Section 8245 — Uses of funds (as of 01/02/01).

³ A precursor and apparent impetus for nearly all of the changes in the 2002 reauthorization can be found in the *Examining the 21st Century Community Learning Centers Program* Hearing before the Subcommittee on Early Childhood Youth and Families of the Committee on Education and the Workplace, House of Representatives, on February 10, 2000.

1.2.2. Estimates of supply and demand for 21CCLC funding

In 2002, while approximately \$783 million in requests for new 21CCLC funding were made, roughly \$188 million was available. Thus, the 21CCLC budget was able to meet about 24% of the demand for new funds. The gap in supply and demand is equivalent to more than 848,000 youth who could be served by the 21CCLC programs (Samelson, 2003a).

2. Background and overview of the national evaluation of the 21CCLCs

In November of 1999, Mathematica Policy Research, Inc. was selected by the U.S. Department of Education to conduct a national evaluation of the 21CCLC's after-school programs. As originally proposed, the evaluation sought to identify the "best practices" of the programs. However, following review by the White House Office of Management and Budget (OMB), the evaluation design was reformulated into a two-year random-assignment outcome study that focused on the impact of program participation on academic and psychosocial development. In April of 2000 the OMB approved the evaluation design and provided \$12.7 million in Federal funding to support the two-year evaluation. Baseline data collection for the national evaluation began just a few months later in the fall of 2000. The Charles Stewart Mott Foundation provided an additional \$3.1 million to support an "enhancement study" for the national evaluation that focused on program functioning at six 21CCLC sites. The stated goals of the enhancement study were to assess variation in the context in which the programs operate, program implementation, and family, individual, and community conditioning factors that may affect program impacts (Moore, Dynarski, Mullens, James-Burdany, & Rosenberg, 2000).

The national evaluation involves two age groups – elementary school students and middle school students – and the conceptual framework is similar for these two age groups. In the first year of the study, students were followed from fall to spring during the 2000–2001 school year. Comparison groups of students participating and not participating in the 21CCLC programs were established. Areas of assessment included demographic information, academic functioning, and social behaviors. Information was provided from students, parents, teachers, and school records.

The *elementary school sample* involved 973 students from 18 21CCLC programs at baseline. These sites were selected because the demand for the 21CCLC program was high and exceeded the ability of the programs to serve all families and children. This oversubscription allowed students to be assigned, at random, either to the opportunity to participate in the 21CCLC program or to a control condition that excluded their participation.

The *middle school sample* involved 32 sites⁴ and 4264 students at baseline. A random assignment design was not employed for the middle school sample. Instead, 21CCLC program participants were compared to a group of non-participating students identified through propensity score matching. The matching was based on available data collected at baseline.

2.1. Impact of the 21CCLC national evaluation

The first-year findings from the 21CCLC evaluation were released on February 3, 2003 (U.S. Department of Education, 2003b). The stated findings suggested that – in many respects – 21CCLC participants and non-participants were substantially similar. These findings formed the base of six events that we will describe below: (1) the requested authorization of 21CCLC funding for 2004, (2) reactions and critiques, (3) a debate involving members of the evaluation's Technical Working Group and Mathematica, (4) a Senate Hearing on the 21CCLCs, (5) the Brookings Institution Stakeholders Forum on After-school Programs, and (6) the actual appropriation of 21CCLC funding for 2004.

2.1.1. Requested funding authorization

The most immediate impact of the evaluation was a requested authorization of \$600 million for the 21CCLCs for Fiscal Year 2004, which represented a 40% decrease from the 2003 funding appropriation of \$1 billion (see Fig. 1). The proposed funding reduction was announced on February 3, 2003, the same day that Mathematica's evaluation report of first-year findings was released to the public. The Fiscal Year 2004 Education Budget Summary and Background

⁴ Note that the term "site" refers to one or more centers operating in the same school district (U.S. Department of Education, 2003b, p. xii, footnote 2).

Information (U.S. Department of Education, 2003c) explained the proposed funding reduction as follows: “The decrease in the request acknowledges that the program needs some time to address disappointing initial findings from a rigorous evaluation of the 21st Century Community Learning Centers program. 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.”

Thus, the first-year findings from the national evaluation of the 21CCLCs were the basis for the proposed funding reduction.

2.1.2. Reactions and critiques

Critical commentaries and detailed critiques of both the proposed funding decision and the evaluation on which it was based followed. For example, several notable experts in the areas of child care, after-school time, and program evaluation provided brief critiques of the scientific adequacy of the national evaluation as a basis for policy in Harvard University’s *The Evaluation Exchange* (e.g., Little, 2003; Weiss & Little, 2003). Likewise, advocacy groups (e.g., the *Afterschool Alliance* (Samelson, 2003b) and the *Forum for Youth Investment* (2003)) denounced the proposed funding reduction on scientific grounds and because it contradicted the NCLB Act’s plan that called for increased funding for the 21CCLCs.

Mahoney and Zigler (2003) conducted an independent, comprehensive critique of the 21CCLC evaluation’s first-year findings. Prior to releasing their critical analysis the authors simultaneously invited feedback from several sources including: the evaluation’s principal investigator at Mathematica, two members of the evaluation’s Technical Working Group, and three independent scholars with expertise in evaluation, statistics, child care, and after-school research. The main conclusions from this critical analysis are summarized in Table 1. Below we outline the bases for those points pertaining to the evaluation’s timing and research design. Readers interested in a more comprehensive account are directed to Mahoney and Zigler (2003).

2.1.2.1. A premature evaluation. Timing is very important in any program evaluation. Most of the sites in the 21CCLC evaluation were in the earliest stages of development (i.e., they had been in operation for only one or two years prior to the study). There was evidence – such as low student attendance (only about two days per week) – that many of the sites were not mature. A principal tenet in evaluation research is to hold off conducting outcome evaluations until “the program is proud” (Campbell & Stanley, 1966). This means that the program has completed the start up and innovation phases of program development and is delivering services in the manner intended. As described

Table 1

Summary criticisms of the first-year findings from the national evaluation of the 21st-Century Community Learning Centers

-
- The programs were assessed in the early stages of development and the evaluation is premature.
 - Conclusions about program impacts were based on one follow-up assessment when the evaluation was only half complete.
 - Information on program characteristics and implementation were not connected to the program outcomes evaluated.
 - The rapid initiation of the study contributed to methodological shortcomings.
 - In the middle school evaluation, baseline data were not collected from several key information sources (e.g., teachers, parents, school records).
 - The absence of baseline data from these sources resulted in comparison groups that were not equivalent at the outset of the evaluation. 21CCLC participants were at higher risk.
 - The lack of baseline data from objective sources leaves only subjective student reports – collected at baseline and follow-up – from which to infer program impacts.
 - The extent to which evaluation findings can be generalized is limited.
 - The evaluation was initiated prior to passage of the NCLB Act of 2001. This Act changed the funding, content, and operation of current 21CCLC programs.
 - The elementary school sites in the evaluation are not representative of the population of 21CCLCs serving elementary school students.
 - Attrition was also high in the elementary evaluation, raising further concerns about the representativeness of this sample.
 - Potentially important components of the 21CCLC programs were not evaluated (i.e., the summer and before-school components).
 - Substantial comparison group “cross-over” occurred (e.g., many non-participants actually participated in 21CCLC after-school programs) and this was not dealt with adequately.
 - Several accepted goals of after-school programs were not considered adequately (e.g., positive youth development, achievement test performance, parent satisfaction, facilitation of work, peace of mind).
 - Missing data were treated inadequately.
 - The properties of the research measures were not reported and the statistical procedures used were not sensitive to measuring growth and change over time. As a result of these shortcomings, the study findings are in doubt.
-

in other recent evaluations of after-school programs, program start-up is “noisy” (e.g., Love et al., 2002; Policy Studies Associates, Inc., 2002; Walker & Abreton, 2002). It takes time – often several years – for programs to be fully implemented and they undergo substantial changes in features, operation, and content during the first several years of development (Dynarski et al., 2001a; p. 17). Thus, in contrast to what is known about effective evaluations, the national evaluation of the 21CCLCs was essentially an outcome study and unable to provide an indication of whether the sites were, in general, functioning as intended. Accordingly, Federal support for the national evaluation should have been devoted first to a study of process implementation which would have provided some sense about whether program impacts were ready to be evaluated.

2.1.2.2. A premature decision. Timing is also relevant when drawing conclusions about program impacts. An evaluation must be carried out over a sufficient length of time to observe the expected changes. In this case, however, conclusions were drawn and funding cuts proposed on the basis of one follow-up assessment during the course of a school year. This was done despite the fact that the evaluation is a two-year study and second-year findings were not yet available, and that change in several of the anticipated outcomes were not expected to occur during the first year of the study (e.g., U.S. Department of Education, 2003b; p. 7, Figure I.2). Specifically, the program-related impacts were allowed to take place over only a five-to eight-month period in the evaluation’s first year (Dynarski et al., 2001a; p. 63). Given the knowledge base on longitudinal program evaluation research, it seems unlikely that substantive long-term impacts or sizable short-term effects should be expected. Immediate and long-term program effects can be very different (Maccoby, Kahn, & Everett, 1983). Some of the most robust findings and cost-to-benefit ratios from such programs have been observed years or decades later (Brown, Frates, Rudge, & Tradewell, 2002; Currie & Duncan, 1995; Lattimore, Mihalic, Grotzpetter, & Taggart, 1998; Schweinhart, Barnes, Weikart, Barnett, & Epstein, 1993; Reynolds, 2000).

2.1.2.3. An outcome study. The OMB called for Mathematica to conduct a national evaluation of the 21CCLCs that was largely concerned with program impacts and outcomes. With the exception of the information on program features collected from six sites funded by the C. S. Mott Foundation’s enhancement study, there was little attention to program characteristics or implementation. However, because the enhancement study sites were not linked to the main evaluation sites concerned with outcomes, important connections between program process and outcomes were not made. These connections are important for at least two reasons. First, they provide some understanding about the outcome findings. No one anticipates positive outcomes for poor quality programs. Second, without information on program features that can be linked to the outcomes, the evaluation cannot make evidence-based recommendations on areas where individual programs need to improve.

2.1.2.4. Limitations on generalizability. The national evaluation began in the fall of 2000, prior to passage of the NCLB Act in January 2002. The NCLB Act changed the funding, content, and operation of the 21CCLC programs. Accordingly, some programs assessed in the national evaluation either no longer exist or they operate under substantively different guidelines. As a result, it is unclear whether generalizations to the current 21CCLC programs can or should be made on the basis of the outdated national evaluation.

A second problem with the evaluation’s generalizability pertains primarily to the elementary school sample. The evaluators were able to identify a small number of sites that were experiencing oversubscription and agreed to have students randomly assigned to participate or not to participate in the 21CCLC after-school programs. An obvious strength of the random assignment design is that it allowed for high internal validity. The trade-off, however, was diminished external validity because the program schools that agreed to the random assignment were not representative of the larger population of elementary schools receiving 21CCLC funds (U.S. Department of Education, 2003b; p. 13). The differences were apparent in demographic characteristics such as race/ethnicity and family economic status. While it is exceedingly difficult to conduct an evaluation of social programs that maintains high internal and external validity, it remains a fact that the evaluation findings may not apply to all 21CCLC programs serving elementary school children. Generalizations and policy decisions applied to the national system of 21CCLCs are an extrapolation of the findings and, therefore, may be partly misleading or wrong.

2.1.2.5. Mismatched comparison groups. It was not possible to employ a random assignment design for the middle school sample, so propensity score matching was used to establish comparison groups of after-school program

participants and non-participants. The matching, however, was based on limited information about the students at the baseline assessment. The missing information at baseline included ratings made by teachers and parents, and school records. As a result of the absent baseline data, the identified comparison groups were not equivalent at the start of the evaluation (U.S. Department of Education, 2003a,b,c; p. 124). The after-school participant group was at higher risk in several respects including lower achievement test scores, more behavior problems, and greater socioeconomic disadvantage. Because the baseline data were not collected, these differences could not be adequately controlled for in the analyses.

As noted above, the evaluation was initiated with only a few months time between the approval of the study by the OMB and the start of data collection. The rapid time schedule contributed to the comparison group mismatch (Dynarski et al., 2001a; p. 64, footnote 1; U.S. Department of Education, 2003b; p. 135, footnote 52). From an original pool of over 21,156 middle school students, considerable time and effort would have been required to recruit the sample, identify the 21CCLC program participants and non-participants, train research staff, administer the baseline assessments to multiple informants, and perform the propensity score matching to establish the final comparison groups. Instead, the matching relied primarily on information collected through student surveys. Although convenient and economical, the self-report information provided a limited basis for establishing the comparison groups. Ultimately, an evaluation schedule that allowed for more time to prepare the study and collect the data needed to properly conduct the propensity score matching would have strengthened the design and conclusions from the middle school evaluation.

2.1.2.6. Cross-over of comparison groups. A relatively common problem in social program evaluation is “cross-over” of condition (also known as “contamination”). This usually refers to the inadvertent application of the treatment to the control/comparison group or the inadvertent failure to apply the treatment to persons assigned to receive it. Cross-over did occur in the national evaluation. For example, among students originally relegated to the comparison group (i.e., not involved in the after-school program), 8% of elementary school students and 14% of middle school students participated in a 21CCLC after-school program during the school year (U.S. Department of Education, 2003a,b,c; p. 123). The evaluation does not report the percentage of students that may have participated in a non-21CCLC after-school program. A more detailed examination of cross-over in the national evaluation is available for two of the six sites involved in the enhancement study.⁵ Cross-over was higher in these sites wherein almost half (49%) of the students initially classified as 21CCLC program participants did not actually participate in a 21CCLC after-school program, and most of the students considered as the comparison group (69%) participated in an after-school program (many, 31%, participated in a 21CCLC after-school program).

That cross-over occurred is unremarkable. However, the national evaluation did not address the problem adequately. For instance, in the middle school evaluation, there is no indication that cross-over was considered in the analyses. In the elementary school evaluation, analyses were adjusted only for cross-over involving program participants who did not actually participate in the 21CCLC program (U.S. Department of Education, 2003a,b,c; p. 94, footnote 48), *not* cross-over of non-participants into 21CCLC after-school programs. Cross-over involving participation in non-21CCLC after-school programs was ignored.

2.1.3. A debate: members of the evaluation's Technical Working Group and Mathematica

One of the most salient critiques of the national evaluation came in the form of a statement by several members of its own Technical Working Group (TWiG) (Bissell et al., 2003). The TWiG statement described a variety of specific design and methodological problems with the evaluation, many of which overlap with Mahoney and Zigler's (2003) critical analysis. Furthermore, the TWiG noted that the evaluators at Mathematica were made aware of the limitations well in advance of releasing the final report of first-year findings but did not address or acknowledge many of them.

Mathematica responded to members of the TWiG (Dynarski & Moore, 2003), questioning both the timing of the statement and that only a few members of the TWiG provided substantive feedback to a draft report of the evaluation's first year findings. Mathematica also provided responses to some of the specific methodological criticisms described by TWiG members. Additionally, one TWiG member issued an independent response that argued the statement released by other members of the TWiG did not reflect his own experience (Puma, 2003).

⁵ This information was provided by members of the 21CCLC evaluation's Technical Working Group. The data apply to Year 2 of the national evaluation.

2.1.4. Senate hearing on the 21CCLCs

The funding appropriation for the 21CCLCs was substantively influenced by a special subcommittee hearing of the Senate Labor, Health and Human Services Appropriations Subcommittee on May 13, 2003, Chaired by Senator Arlen Specter (R-PA) (*U.S. Senate Committee on Appropriations*, 2003). This special hearing was devoted to consideration of the budget request by the Department of Education for fiscal year 2004 for the 21CCLCs. Testimony was provided by William Hansen, Deputy Secretary of Education, Russ Whitehurst, Director of the Institute of Education Sciences at the Department of Education, Senator Barbara Boxer (D-CA), Arnold Schwarzenegger, sponsor of California's Proposition 49 to provide universal access to after-school programs, John Stefano, Mayor of New Haven, Havey Sprafka, Chief of Police in Knoxville, Iowa representing the national organization *Fight Crime: Invest in Kids*, and two youth who were participating in 21CCLC after-school programs.⁶

In justifying the requested funding authorization, Deputy Hansen referred to the findings from the national evaluation, describing the study as “rigorous” and “...the strongest to date in the after-school literature...” He added that the Bush Administration would “...fund only what works to help children. We will not pour limited resources into programs that fall short of their mission.” By contrast, Schwarzenegger's testimony emphasized the broader need for and value of after-school programs for youth development, referred to findings from other studies showing social and academic benefits for after-school program participants, and that the national evaluation results were useful insofar as they pointed to areas where the 21CCLCs could be strengthened such as increasing student attendance. Senator Boxer's testimony added that the 40% reduction would affect some 570,000 children and families and that, under the NCLB Act, the 2004 budget actually called for \$1.75 billion in 21CCLC funding (i.e., \$1.15 billion more than the requested authorization).

Senator Specter questioned Deputy Secretary Hansen and Director Whitehurst about some of the limitations in the evaluation. Queries included why it was sound to reduce the program budget before the evaluation was complete and why the 21CCLC evaluation findings were so different from results of several other specific after-school program evaluations. Deputy Hansen was not able to comment on all of the specific studies in question, but following the Hearing he did provide cogent critiques of each study that is now part of the congressional record (*U.S. Senate Committee on Appropriations*, 2003). The criticisms are generally apt and demonstrate that many of the existing after-school program evaluations contain methodological shortcomings. However, the same level of critical analysis was not applied to the Department of Education's own evaluation of the 21CCLCs.

2.1.5. The Brookings Institution Stakeholders Seminar

On July 10, 2003 directors of the Brookings Institution's Welfare Reform and Beyond Initiative (i.e., Ron Haskins and Isabel Sawhill (co-directors) and Andrea Kane (outreach coordinator)) hosted a discussion on the 21CCLC national evaluation of after-school programs. Mark Dynarski, Mathematica's lead evaluator for the 21CCLC study, and Edward Zigler, emeritus professor of psychology at Yale University and known expert in child care research and policy, discussed the 21CCLC evaluation findings.

Dynarski's (2003) remarks were centered on the knowledge that could be gained from the evaluation experience and the soundness of the evaluation. He noted that future efforts to conduct educational research of this sort could falter if studies undergo the same “turbulent experience” as did the 21CCLC evaluation. He pointed to some strengths of the evaluation such as the use of a national sample of 21CCLC after-school program participants, the random assignment component, and that an early-timed outcome evaluation was justified in this case because the program had expanded so rapidly. Finally, he acknowledged that the findings of the 21CCLC evaluation differed from other evaluations but suggested that many of the other studies were limited in terms of sample representativeness and with respect to control for selection influences.

Zigler's (2003) remarks focused on two main points. First, that important policy decisions should not be made on the basis of any single evaluation, no matter how rigorous. Second, that the value of Mathematica's evaluation would be considerable if it were able to elucidate directions for what after-school programs should be like with respect to program theory, goals, and operation. Important functions of the 21CCLCs such as facilitating parental employment, providing parents with peace of mind, and enhancing student engagement and attachment to school need to be

⁶ On June 21, 1983, Senators Specter (R-PA) and Dodd (D-CT) co-chaired the first Senate Hearing of Children's Caucus Policy Forum that focused on the importance of providing supervised care for latchkey children (as they were known then) during the hours following school dismissal (see, e.g., Zigler, 1983). Like the 21CCLC Senate Hearing, children and several experts testified.

considered along with academic and social indicators of child adjustment. Likewise, program practices that do and do not lead to positive child adjustment need to be understood with reference to the stage of program implementation.

2.1.6. Congressional Funding Appropriation

Congress appropriated Fiscal Year 2004 funding for the 21CCLCs at \$1.005 billion — a \$5 million *increase* over the previous Fiscal Year. However, the across-the-board rescission for programs amounted to an actual of appropriation of \$999.07 million for the programs (nearly identical to the previous Fiscal Year). The Senate Appropriations Committee Report of January 22, 2004 (U.S. Congress, 2004) explained the funding appropriation as follows:

The Committee rejects the proposed reduction in the program, which was based on the initial findings from an evaluation funded in the first 3 years of the program. While the Committee appreciates and is committed to the notion that the allocation of Federal resources should be based on high-quality, science-based evaluations, it does not agree that more than 500,000 children and their families should be denied the opportunity to receive educational, developmental and related services based on the first-year findings from the Department's evaluation. The Committee looks forward to subsequent reports, which will include a more complete data set and additional information upon which to make a more informed assessment of the impact of the 21st-Century Community Learning Centers program.

3. Lessons learned

The national evaluation of the 21CCLCs conducted by Mathematica was among the first studies considered under the 2001 NCLB Act's new emphasis on scientifically based research to guide the policy decision-making process. Recounting the science-to-policy process in this instance shows that the process was complicated. Some of the major lessons learned from this process follow.

3.1. Lesson 1: one study is not a sufficient basis from which to formulate policy

Because Mathematica's evaluation was intended to inform policy decisions about the 21CCLCs under the NCLB Act, it is not surprising that the first-year findings were considered in the requested funding authorization. However, it is unfortunate that the requested authorization was based exclusively on one study that was half complete. A great degree of certainty is needed to support policy decisions when the stakes are high. In this case, hundreds of thousands of American families would have been affected by the proposed funding reduction. Because no study is without fault, it is critical that policy decisions consider the body of scientific evidence available to inform their decisions. Relying on any one study without reference to how it fits with the existing literature provides a limited basis for decision making. A much stronger basis is achieved when decisions are based on replicated findings across several independent studies. Accordingly, on its own, the national evaluation did not provide a strong basis for making a policy decision.

This lesson has historical precedents. The most well-known parallel was the impact study of Head Start conducted by the Westinghouse Learning Company (1969). This single study nearly ended a program that is now a cornerstone of early childhood education for low-income children and families in the United States (Henrich, 2004). Although the goals of Head Start were, and continue to be, the facilitation of school readiness by promoting a broad set of social competencies (Zigler & Trickett, 1978), the policy message taken from the Westinghouse evaluation was that Head Start failed to make substantial and lasting changes in the participants' IQ scores (Phillips & Styfco, *in press*). While the debate over Head Start continues, decades of research since the Westinghouse evaluation have shown that Head Start and other early childhood education program can and do have lasting positive impacts on the participants (Phillips & White, 2004).

Based on findings from the broader literature showing positive outcomes for after-school program participants in social, educational, and health areas, it is reasonable to expect that the results from future studies of the 21CCLCs will identify such benefits. This expectation is congruent with the Senate Appropriation Committee Report of January 22, 2004 which concluded it would be premature to diminish the 21CCLC program based on the early findings from one study.

3.2. Lesson 2: a study is not scientific until it passes professional scrutiny through a process of peer review

A basic criterion for what constitutes the scientific record in our field is decided by the peer review process. Indeed, part of the definition of scientifically based research under the NCLB Act requires that the research "has been accepted

by a peer-reviewed journal or approved by a panel of independent experts through comparatively rigorous, objective, and scientific review” (P.L. 107–110, 115 Stat. 1551). The first-year findings from Mathematica’s evaluation were not published in a peer review journal and were subjected to review and approval by only one outside individual through the Institute of Education Science’s editorial review process (Dynarski & Moore, 2003). The variety of criticisms that followed the evaluation’s release suggests that this lone review was not adequate. Science requires a more thorough and rigorous scrutiny of findings, particularly for studies developed to inform policy decisions.⁷ Thus, future evaluations under the NCLB Act need to follow a peer review process that would be acceptable in the scientific community.

3.3. Lesson 3: close collaboration between evaluators and technical working group members for an evaluation is important

The implications of the study and the lack of congruence with the broader literature should have been discussed and reconciled through scholarly debate in advance of a policy decision. In this case, critiques of and broader debates about the report followed, rather than preceded the policy decision. Had Mathematica attended to early feedback from members of its technical working group (TWiG) a certain amount of the controversy may have been eliminated (Bissell et al., 2003). More generally, feedback from an evaluation’s TWiG should be reconciled. Identified weaknesses need to be dealt with, acknowledged, and their implications discussed. Mechanisms for ensuring a close collaboration between evaluators and TWiG members are important to develop.

3.4. Lesson 4: scientists and evaluators have a responsibility to articulate the strengths and weaknesses of their own research

There is no perfect study of after-school programs and there never will be a study that provides “the final answer” on their effectiveness. All scientific investigations contain some flaws, shortcomings, or limitations. As a result, scientists have an obligation to thoughtfully consider and then make explicit how study limitations may affect the interpretation of results. When research interfaces directly with policy, the need for specifying what can and what cannot be concluded from a study is especially critical. Policy makers cannot be expected to have the background or time to perform a critical analysis of stated research findings when considering a policy decision. Therefore, the responsibility to describe findings so that they can be interpreted appropriately rests largely on the investigators and the scientific community.

Some limitations of Mathematica’s evaluation are acknowledged in their report of first-year findings, but limitations are not mentioned in the executive summary, and there is not a separate “limitations” section of the report. To provide an informed basis for policy decisions, reports should have a section devoted to discussing the significance and limitations of evaluation findings, comparability with the existing literature on analogous programs, and the possible use and misuse of the findings for policy. Such a section would have provided an informed basis for drawing conclusions about the 21CCLCs.

3.5. Lesson 5: time and timing are critical in the science-to-policy translation process

The gradual culmination of empirical findings from which science proceeds does not always match well with the time constraints and plurality of knowledge sources under which policy decision are made (Shonkoff, 2000). The asynchrony between the cultures of science and policy has been discussed many times and McCall and Groark (2000) note that policy decisions concerning children and families will be made with or without the use of scientific knowledge. What is apparent from the particular example of the 21CCLCs is that many problems in the science-to-policy translation process stem from less than optimal time and timing. For example, the evaluation focused on programs that were early in their development. The proposed funding reduction was based on a study half complete. Many of the scholarly critiques and debates about the evaluation followed the release of the first-year evaluation report rather than preceding it. The evaluation was initiated rapidly and the schedule contributed to methodological problems

⁷ In this age of the Internet and long review process for journals, more and more respected scholars and institutions are conveying findings and making reports available on the Internet without peer review. While some of these reports have been superb (e.g., Vandell & Wolfe, 2000) the scientific credibility of others is questionable.

(cf., Westinghouse Learning Company, 1969). Finally, the initiation of the evaluation (fall 2000), did not anticipate the January 2002 passage of the NCLB Act that changed the funding process and operation of the very programs evaluated.

In an *ideal* vision, the science-to-policy translation process would consider the following:

- Program evaluations should proceed under a time frame that will yield the strongest scientific evidence as opposed to one that compromises scientific rigor for timely results.
- Impact studies should be undertaken after a process evaluation demonstrates that the programs of interest are mature and operating as intended.
- Policy makers should incorporate science into the decision making process when the research: (1) has reached a mature stage of completion, (2) appropriately articulates limitations with respect to application, and (3) has passed a peer review process undertaken by knowledgeable experts in the scientific community.

Unfortunately, this scenario is often impractical. It is an over-simplification of the complexity of the policy decision process which is based on a myriad of influences, only one of which *may* include scientific research. If scientists want their research to inform policy then it must be relevant and available to the issues at hand. Because of the timetable and manner under which science advances, research cannot always provide timely, certain, or comprehensive answers to immediate problems (Horowitz, 2000). For example, conducting a pilot study, a process evaluation, or following the traditional peer review process can take months or years, not to mention the years often needed to conduct the actual study. Perhaps as a result, some of our best Government-sponsored research institutions employ internal systems for reviewing and disseminating research findings. While many research reports developed through these internal processes have been excellent, it is evident that all have not and that the consequences of basing policy on inadequate science can be far reaching.

4. Concluding comments

Applied developmental psychologists may have experienced a certain amount of déjà vu when reading the above lessons. Some do echo concerns that scholars have raised about the role of program evaluation in policy decisions made decades earlier (e.g., Maccoby et al., 1983). In the light of the NCLB Act's mantra to utilize science in the policy making decision process we believe the lessons take on a new importance. We are hopeful that what has been learned in this particular case will be useful for the science-to-policy translation process in general, and for future studies under the NCLB Act specifically.

References

- Baldwin Grossman, J., Price, M. L., Fellerath, V., Jucovy, L. Z., Kotloff, L. J., Raley, R., & Walker, K. E. (2002). *Multiple choices after school: Findings from the Extended-Service Schools Initiative*. Philadelphia, PA: Public/Private Ventures.
- Bissell, J. S., Cross, C. T., Mapp, K., Reisner, E., Vandell, D. L., Warren, C., & Weissbourd, R. (2003, May 10). *Statement released by members of the Scientific Advisory Board for the 21st Century Community Learning Center Evaluation*. Retrieved August 20, 2004, from http://www.gse.harvard.edu/hfrp/content/projects/afterschool/resources/21stccle_statement.doc
- Brown, W., Frates, S., Rudge, I., & Tradewell, R. (2002). *The costs and benefits of after school programs: The estimate effects of the after school education and safety program act of 2002*. Claremont, CA: Claremont McKenna College, Rose Institute of State and Local Government.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Capizzano, J., Tout, K., & Adams, G. (2000). *Child care patterns of school-age children with employed mothers*. Washington, DC: The Urban Institute.
- Currie, J., & Duncan, T. (1995). Does Head Start make a difference? *American Economic Review*, 85, 341–364.
- Dynarski, M. (2003, July). *It feels fine once you get used to it: Lessons from the 21st Century national evaluation*. Remarks presented at the Brookings Institution's Stakeholder Forum on After-school Programs. Washington, DC. Retrieved September 7, 2004, from <http://www.mathematica-mpr.com/publications/PDFs/feelfine.pdf>
- Dynarski, M., James-Burdumy, S., Mansfield, W., Mayer, D., Moore, M., Mullens, J., & Silva, T. (2001a). *A broader view: The National Evaluation of the 21st Century Community Learning Center Program. Design Report: Volume I*. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved August 20, 2004, from <http://www.mathematica-mpr.com/publications/PDFs/broadviewvol1.pdf>
- Dynarski, M., & Moore, B. (2003). *Response to May 7, 2003, letter from selected members of the technical working group for the 21st Century Community Learning Centers Evaluation*. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved August 20, 2004, from <http://www.mathematica-mpr.com/publications/pdfs/responsetwg.pdf>
- Forum for Youth Investment (2003). *Out-of-school time policy commentary #3: Reflections on system building: Lessons from the after-school movement*. Washington, DC: Author Retrieved August 20, 2004, from <http://www.forumforyouthinvestment.org/comment/ostpc3.pdf>
- Henrich, C. C. (2004). Head Start as a National Laboratory. In E. Zigler, & S. J. Styfco (Eds.), *The Head Start debates* (pp. 517–532). Baltimore, MD: Paul H. Brooks Publishing Co.

- Horowitz, F. D. (2000). Child development and PITS: Simple questions, complex answers, and developmental theory. *Child Development*, 71, 1–10.
- Lattimore, C. B., Mihalic, S. F., Grotpeter, J. K., & Taggart, R. (1998). *Blueprints for violence prevention, Book Four: The Quantum Opportunities Program*. Boulder, CO: University of Colorado at Boulder, Center for the Study and Prevention of Violence.
- Little, P. (2003). The Evaluation Exchange special report on the 21st Century Community Learning Centers program national evaluation. *The Evaluation Exchange*, 9, 11–14 (Retrieved August 20, 2004, from <http://www.gse.harvard.edu/hftp/eval/issue21/special.html>)
- Love, J. M., Eliason Kisker, E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Paulsell, D., et al. (2002, June). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start*. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved September 8, 2004, from <http://www.mathematica-mpr.com/publications/PDFs/ehsfinalsumm.pdf>
- Maccoby, E. E., Kahn, A. J., & Everett, B. A. (1983). The role of psychological research in the formation of policies affecting children. *American Psychologist*, 38, 80–84.
- Mahoney, J. L., Larson, R. W., & Eccles, J. S. (Eds.). (2005). *Organized activities as contexts of development: Extracurricular activities, after-school and community programs*. Mahwah, NJ: Erlbaum.
- Mahoney, J. L., & Zigler, E. F. (2003). *The national evaluation of the 21st-Century Community Learning Centers: A critical analysis of first-year findings*. Retrieved August 20, 2004, from <http://pantheon.yale.edu/~jlm79/Critique-Full.pdf>
- McCall, R. B., & Groark, C. J. (2000). The future of applied child development research and public policy. *Child Development*, 71, 197–204.
- Moore, M., Dynarski, M., Mullens, J., James-Burdany, S., & Rosenberg, L. (2000). *Enhancement of the 21st Century Community Learning Centers Evaluation: A concept paper*. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved August 20, 2004, from <http://www.mathematica-mpr.com/publications/PDFs/enhancing21.pdf>
- National Research Council and Institute of Medicine (2000). *After-school programs to promote child and adolescent development*. Washington DC: National Academy Press.
- National Research Council and Institute of Medicine (2002). *Community programs to promote youth development*. Committee on Community-Level Programs for Youth. In J. Eccles, & J. A. Gootman (Eds.), *Board on Children, Youth, and Families, Division of Behavioral and Social Science and Education*. Washington, DC: National Academy Press.
- Phillips, D. and Styfco, S. (in press). Child development research and public policy: Triumphs and setbacks on the way to maturity. In J. L. Aber, S. J. Bishop-Josef, S. M. Jones, K. T. McLearn, & D. A. Phillips (Eds.). *Child development and social policy: Knowledge for action*. Washington, D.C.: American Psychological Association.
- Phillips, D. A., & White, S. H. (2004). New possibilities for research on Head Start. In E. Zigler, & S. J. Styfco (Eds.), *The Head Start debates* (pp. 263–281). Baltimore, MD: Paul H. Brooks Publishing Co.
- Policy Studies Associates, Inc. (2002, December). *What have we learned from TASC's first three years?: Evaluation of the TASC after-school program*. Retrieved August 20, 2004, from <http://www.nsba.org/site/docs/11700/11700.pdf>
- Public Law 107–110. *No child left behind act of 2001*. Retrieved August 20, 2004, from <http://www.ed.gov/policy/elsec/leg/esea02/107–110.pdf>
- Puma, M. (2003, June 11). After-school study's "unfounded" review [Letter to the editor]. *Education Week*, 22(40), 28.
- Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child–Parent Centers*. Lincoln: University Nebraska Press.
- Samelson, J. (2003, May). *Afterschool in America: The state of afterschool in 2003.: The Afterschool Alliance*.
- Samelson, J. (2003, Februaryb). *Federal afterschool study compromised by negative bias, methodological flaws, and small sample*. Statement retrieved August 20, 2004, from http://www.afterschoolalliance.org/press_archives/mathematica_budget_js.doc
- Schweinhart, L., Barnes, H., Weikart, D., Barnett, W. S., & Epstein, A. S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27 (Monographs of the High/Scope Educational Research Foundation Number 10)*. Ypsilanti, MI: High/Scope Press.
- Seppanen, P.S., Love, J.M., deVries, D.K., Bernsein, L., Seligson, M., Marx, R., & Kisker, E.E. (1993). *National Study of before and after school programs: Final report*. Washington, DC: U.S. Department of Education, Office of Policy and Planning. Portsmouth, NH: RMC Research Corporation.
- Shonkoff, J. P. (2000). Science, policy, and practice: Three cultures in search of a shared mission. *Child Development*, 71, 181–187.
- Smolensky, E., & Gootman, J. A. (Eds.). (2003). *Working families and growing kids: Caring for children and adolescents*. Washington, DC: National Academies Press.
- The Future of Children (1999). *When school is out, Vol. 9* (pp.) Washington, DC: The David and Lucile Packard Foundation.
- U.S. Congress (2004). *Appropriations legislation 108th Congress, 1st session (FY 2004): Labor, HHS, Education Senate Report 108-81*. Washington, D.C.: Author Retrieved August 20, 2004, from <http://www.access.gpo.gov/congress/legislation/04appro.html#labor>
- U.S. Department of Education (1994). *Improving America's Schools Act of 1994*. Washington, D.C.: Author Retrieved August 20, 2004, from <http://www.ed.gov/legislation/ESEA/index.html>
- U.S. Department of Education (2003a). *21st Century Community Learning Centers non-regulatory guidance*. Washington, D.C.: Author Retrieved August 20, 2004, from <http://www.ed.gov/programs/21stcccl/guidance2003.pdf>
- U.S. Department of Education (2003b). *When schools stay open late: The national evaluation of the 21st Century Learning Centers program, first year findings*. Washington, D.C.: Author Retrieved August 20, 2004, from <http://www.ed.gov/pubs/21cent/firstyear/index.html>
- U.S. Department of Education (2003c). *Fiscal Year 2004 education budget summary and background information, February 3, 2003*. Washington, D.C.: Author Retrieved August 20, 2004, from <http://www.ed.gov/about/overview/budget/budget04/index.html>
- U.S. Senate Committee on Appropriations (2003). *Investment in after-school programs. Hearing before a Subcommittee of the Committee on Appropriations, United States Senate, 180th Congress, 1st session. Special hearing, May 13, Washington, D.C.* Retrieved August 20, 2004, from <http://appropriations.senate.gov/hearings/topics.cfm?code=hearings>
- Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Retrieved September 8, 2004, from <http://www.aspe.hhs.gov/hsp/ccquality00/ccqual.htm>
- Walker, K., & Abreton, A. (2002). *Working together to build Beacon Centers in San Francisco: Evaluation findings from 1998–2000. Theory of change process*. Public/Private Ventures. Retrieved August 20, 2004, from http://www.ppv.org/ppv/publications/assets/118_publication.pdf

- Weiss, H., & Little, P. (2003). *Why, when, and how to use evaluation: Experts speak out*. Cambridge, MA: Harvard Family Research Project
Retrieved August 20, 2004, from: <http://www.gse.harvard.edu/hfrp/projects/afterschool/resources/issuebrief5.html>
- Zigler, E. F. (1983, June). Latchkey children: Risks and alternatives. *Testimony on school-age day care for Senate Children's Caucus Policy Forum, Congressional Record 129 (June 21~:88)* Washington, DC: U.S. Government Printing Office.
- Zigler, E. F. (1998). A place of value for applied and policy studies. *Child Development*, 69, 532–542.
- Zigler, E. F. (2003, July). *Remarks presented at the Brookings Institution's Stakeholder Forum on After-school Programs. Washington, DC.*
- Zigler, E. F., & Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist*, 33, 789–798.