

Framing an Evidence-Based Decision about 21st CCLC: How do we see the value?

March 27, 2015

Charles Smith

Karen Pittman

Stephen C. Peck

Gina McGovern

Our investment in the 21st Century Community Learning Centers (CCLC) program is indeed a rare example of a social investment with substantially bi-partisan support and, equally important, a growing footprint of local collaboration with elected officials, school administrators, and community providers. Although the program enjoys wide support, the depth of the investment, \$1.2 billion each year, should lead us to regularly ask questions about social value. Is it worth the money? In this policy commentary, we do some reasoning about how 21st CCLC produces value and discuss the limitations in one particular way of seeing that value – the intent-to-treat impact evaluation design. Because some actors see the intent-to-treat evaluation design as a sufficient source of evidence for high stakes policy decisions about 21st CCLC programs (e.g., Dynarski, 2015; Kremer et al., 2015), we want to take time to understand how the decision to use this powerful evaluation method affects what gets seen and what gets valued.

Fitting the Evaluation Method to the Right Afterschool Theory

What evidence from the Intent-to-treat evaluation design means. Briefly, intent-to-treat evaluation designs entail a sample of study participants in at least two groups, who are equivalent at baseline (preferably through randomization), and who are assigned or matched into comparison and treatment conditions. Typically, one outcome is examined at a time by subtracting the comparison group score from the treatment group score for that outcome. This design has been particularly attractive to policy-oriented evaluators because it doesn't require attention to idiosyncratic qualities across or within programs that may produce an effect for any specific individual. The intent-to-treat design keeps it simple: Pick the

outcome, make sure that your sample is sufficient and outcome measures reliable, figure out how to equate the groups at baseline, and you have a powerful and sensitive impact test.

When applied to afterschool programs, the intent-to-treat design rests on the following theory: Afterschool programs claim to advance a wide range of individual outcomes including academic skills, social and emotional skills, attendance, etc. Therefore, the intent-to-treat evaluation design in search of an average effect across all treatment group subjects is the method of choice. This last part about the “average effect” is important – it assumes that we can take the average score for a single outcome across the treatment group and subtract the comparison group average from it – and should raise some questions. Do all students who attend 21st CCLC programs gain the same skills to the same extent? Does the average effect represent very many actual students?

Getting the afterschool theory of change right: Diverse pathways to diverse outcomes. We would prefer to start with a different theory: Afterschool programs are not really designed to provide the same intervention for all students. Across and within programs, staff may design very different pathways to achieving different school success outcomes, depending on the kids and the local context. Three examples:

- Student A in Program A- Attends intermittently when parent’s work schedule gets busy. The student is at home afterschool with family when it works (a good thing) or in other more specialized activities (i.e., Girls on the Run which 21st CCLC brought to the school) and with the option of a high-quality program when they need it. This program is reducing risk behavior and parent’s stress level and increasing skills related to nutrition and exercise.
- Student B in Program B – Attends regularly due to difficulties in 4th grade math, and parents appreciate the free program with an academic focus. The student is doing homework every night and getting some individualized tutoring in Math. Unfortunately, this program is not very high quality in the specific sense that insufficient attention is given to building the “club” atmosphere and the kids and staff are not always very nice to each other. Conflicts in afterschool for Student B carry over into the school day.
- Student C in Program C – Attends regularly at [Bright Futures](#) because it’s known as a great program and has good marketing in the school. The program has a grit initiative so everyone is setting “grit goals” and doing a semester long photo journalism project to demonstrate those goals – and learning a lot about how to manage their own motivation and the physics of light (cameras).

This student was already getting As and Bs in school, and scoring just above the cut point for acceptable performance on state achievement tests.

Three different patterns of family and student need, three different interventions, and several different types of skills being developed (e.g., health, academic, SEL). All of these skills are likely to have positive effects on schools success. And herein lies the rub – where theory and method don't quite fit. If only some of the students are changing on any given outcome at any given time, then the average amount of change will only be driven by the few students who are changing: Intent-to-treat designs seeking average effects from afterschool participation are likely to yield small effect sizes for any single outcome.¹

To be clear, we agree that school success is a primary stated purpose of the 21st Century program and that school achievement data is often the best performance data we've got - so 21st Century programs should be judged in these terms. Fortunately, there are several [21st CCLC statewide evaluation studies](#), many using state of the art methods to create comparison groups, that have been conducted by Naftzger and colleagues at the American Institutes for Research. Our review of these studies suggests that there is a case to be made for the positive impact of 21st Century programs on a wide range of school outcomes, but our interpretation of the small effect sizes produced through these intent-to-treat evaluations needs to be revisited. Given our theory that afterschool programs do different things for different students, a small intent-to-treat impact estimate on any single outcome should not be automatically interpreted as substantively unimportant.

Program qualities matter for skill change. Let's go a step further. Once we've got the theory right and can name the student skills which are the target of the afterschool intervention, the real challenge for evaluation is to figure out whether or not the program's qualities – the program design, staff practices, and content which make up the afterschool intervention – are actually causing growth in those targeted skills. This is another major problem with the intent-to-treat approach – it doesn't tell us much about how program qualities affect (mediate/moderate) skill growth. We do evaluation, but we don't really get our theories tested, and we don't learn much in terms of how to get better, how to improve performance. Individual evaluation studies and meta-analyses (i.e., summaries of many evaluation studies) sometimes include some “social address” kinds of moderators (e.g., SES, ethnicity). In our view, however, these are unlikely to meaningfully differentiate participants who are more or less likely to advance in specific skill

¹ This is also true for other reasons that we omit here. For example, it is far from certain that any skill learned in the afterschool setting will automatically generalize or transfer to a school day classroom where a completely different adult relationship structures the context that elicits the skill. There are many reasons why students might not demonstrate skills learned in afterschool during the same school year, in particular, a difficult relationship with their school day teacher.

domains, i.e., that subset of students who's large skill change could be buried in an average effect. It is the attributes of the student's background (e.g., stress), the student's program experiences (e.g., interest), and the program's qualities (e.g., warmth, content) that are the key moderator/mediator variables – and these are usually not measured.

Remember: In the well-known Durlak et al. meta-analysis (2007), a wide range of afterschool impacts were completely moderated by the quality indicator. The studies without high quality had nil effect sizes while the studies with high quality had much larger effect sizes – and the average effect, without taking quality into account, was somewhere in the middle. Perhaps the most important finding from the original Mathematica evaluation of the 21st Century program (Dynarski et al. 2004) was the negative effect on behavior – which many suggested at the time was the result of low-quality programs and, paralleling early childhood research, where low quality classrooms have been associated with aggressive behavior. In our opinion, this is where this work really starts to get interesting, and there is an [emerging base of afterschool evaluations](#) that suggest that the quality of afterschool settings may be driving school success outcomes.

Innovation and the Afterschool Social Movement

Dynarski (2015) and Kremer et al. (2015) are the most recent set of policy researchers to marshal intent-to-treat evidence to support an evidence-based decision: The 21st CCLC program should be folded into the federal Child Care Block Grant. On the surface, this argument may have some merits – we may want the federally funded child care programs to better integrate services with 21st Century programs. However, at a deeper level, the suggestion invites a rose-colored view of evidence-based decision making that is hard to accept. This line of thinking seems to neglect completely the other important domains of social value that the federal investment, and the original Mathematica study (Dynarski et al., 2004), has sparked. In addition to providing a lot of afterschool program slots and a lot of jobs, it has joined a larger stream of public and private investment to produce social innovations and enterprises that are laser-focused on improving the quality and effectiveness of 21st CCLC and the broader sector of afterschool services.

Quality improvement systems are an important afterschool innovation. One critical stream of innovation has been *continuous quality improvement systems* or QIS. Leading edge QIS have grown in the wider afterschool sector based on [innovation in state 21st CCLC programs](#). QIS typically consist of four parts: (1) A standard and measures for high quality service, (2) data products that describe performance against

the standard, (3) a continuous improvement cycle, and (4) the training, technical assistance, and coaching necessary to help staff implement the cycle at their site (see [YPOI study](#), [Wallace City Systems guide](#)). The QIS have integrated the culture and the technology of continuous quality improvement in ways that [extend beyond some of the limitations of accountability models](#) built into the No Child Left Behind and federal Child Care subsidy policies. For example, these QIS take seriously the proposition that it is difficult to get anyone to do anything by threatening them with a performance indicator that they do not believe they can move and/or don't consider to be fair. Moreover, these systems produce performance information about the full cascade of practices that ultimately produce changes in student skills, taking seriously the role of processes quality: management practices and the qualities of instruction that *ARE* the afterschool intervention. Google Total Quality Management or ISO 9000 for parallels in the private for-profit sector.

The QIS, in turn, produce innovation at every level – system, organization, classroom - because they put performance information in the hands of the person accountable for that aspect of service production and ask them to figure out how to do better, often with supports (i.e., a “lower stakes” incentive). Examples of the QIS approach to evaluation are available in several state 21st CCLC systems including [Oklahoma](#), [Arkansas](#), [Michigan](#), and [Washington](#). These approaches to statewide evaluation (some of which also include intent-to-treat components) are all about improving the qualities of the service by naming the parts, aligning measures, and empowering local actors to make decisions about where they put their time and passion. The presence of a QIS should mean that the quality and effectiveness of the 21st CCLC service in these states has been increasing in recent years.

21st CCLC has contributed to the growth of an afterschool social movement. Now broaden the view. The QIS are a leading edge innovation in education policy, but their even greater importance is perhaps less intentional: An afterschool social movement. For some, talk of social movements may be anathema to the evidence-based decision-making model (see Kremer et al 2015, p. 618), but we don't think so.

Targeted, steady funding for afterschool programs, combined with an assurance of accountability for service quality, has been a potent combination for superintendents, mayors, local funders, and community-based organizations. It was a welcome push to acknowledge the need for high-quality afterschool services, make public commitments, and craft coordinated responses. In many cities across the country, these [public-private afterschool systems](#) are emerging as a new learning sector that specialize in positive youth development and that directly compliment the missions of public schools. These local education policy partnerships have in turn sparked research-practice partnerships where practitioners

define the theory and researchers apply research designs to describe performance. Visit our [website](#) and those of [AIR](#), [Public Profit](#), [St. Paul Public School Foundation](#), [UC Irvine](#), [Every Hour Counts](#), and [Edvance Research](#) to get a sample of the work.

Today, state and local afterschool leaders are already making careful decisions about the nature and purposes of the afterschool service: Invest in more intensive programs that have a more [explicit focus on a specific skill set](#), or [integrate community partners](#) more deeply into the school day, or step up [data sharing between teachers and after-school staff](#), or inform customers about [quality ratings](#). Very few of these options were even a possibility a decade ago, before schools, mayors' offices, local funders, and community-based organizations began their remarkably rapid journey towards shared accountability. The stability of 21st CCLC program funding, and the heightened demands for accountability that have followed, have raised the performance of the entire afterschool sector and drawn new resources to the needs of children and youth. That is social value.

Conclusions

Summary. In this commentary we've described a mismatch between the afterschool theory of change and the intent-to-treat evaluation design, suggesting that when these powerful evaluation designs are applied to broad developmentally focused programs such as 21st CCLC, the effect sizes are likely to be small but substantively important. We've also suggested that afterschool evaluations need to include description and measurement of critical program qualities and the specific skills these programs are focused on growing. An example of a 21st CCLC evaluation which combines both an intent-to-treat impact design with a substantial effort to understand how all of the pieces fit together is the [Texas 21st CCLC Year 2 Interim Evaluation Report](#). We further suggested that 21st Century has intentionally fostered an important social innovation, the afterschool QIS. Because 21st CCLC represents ethic of accountability for service quality in many states, the program has also sparked a broader social movement that has improved the state of afterschool for all American youth.

A clear policy choice. In the end, we're really left with a pretty clear choice. See the social value that has been created by 21st CCLC and keep pushing the cycle of accountability and innovation. Or, ignore the fact that it takes many years to do anything worthwhile and deconstruct the program. We use the term "deconstruct" intentionally – that decade of funding has created a lot of infrastructure that will cost a lot of time and money to replace. Remember that one of the grandfathers of modern evaluation methods, Donald Campbell, called his important essay the Experiment-ING [i.e., lots of experiments] Society

(Campbell, 1971), and we should heed that implied advice: One type of evaluation design does not entail evidence-based decision making, and scores of intent-to-treat evaluations do not constitute sufficient evidence for all or nothing policy decisions.

References

- Campbell, Donald. (1971). The experimenting society. In William Dunn ed., *The Experimenting Society: Essays in honor of Donald T. Campbell*, Policy Studies Review Annual, 11, pp. 35-68.
- Dynarski, M. (2015). The \$1.2 Billion afterschool program that doesn't work. Posted to Brookings Institution webpage on March 19, 2015.
- James-Burdumy, S., Dynarski, M., Moore, M., Deke, J., Mansfield, W., and Pistorino, C. (2005). When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program. U.S. Department of Education, Institute for Education Science, National Center for Education Evaluation and Regional Assistance.
- Kremer, K., Maynard, B., Polanin, J., Vaughn, M., Sarteschi, C. (2015). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: A systematic review and meta-analysis. *Journal of Youth Adolescence*, 44:616-636.