

# Evaluating Federal Social Programs: Finding Out What Works and What Does Not

## Testimony before the Committee on Ways and Means Subcommittee on Human Resources United States House of Representatives

July 17, 2013

**David B. Muhlhausen, Ph.D.**  
**Research Fellow in Empirical Policy Analysis**  
**The Heritage Foundation**

My name is David Muhlhausen. I am a Research Fellow in Empirical Policy Analysis in the Center for Data Analysis at The Heritage Foundation. I thank Chairman Dave Reichert, Ranking Member Lloyd Doggett, and the rest of the committee for the opportunity to testify today on the need to evaluate federal social programs. The views I express in this testimony are my own and should not be construed as representing any official position of The Heritage Foundation.

My testimony is based on my recently published book, *Do Federal Social Programs Work?*<sup>1</sup> This is a simple question. While the question may be straightforward, finding an answer is complicated. As my book demonstrates, the best method for assessing the effectiveness of federal social programs is large-scale, multisite experimental impact evaluations. Unfortunately, these scientifically rigorous assessments are rarely done. By my count, only 20 large-scale, multisite experimental impact evaluations assessing the effectiveness of 21 federal social programs have been published since 1990:

- Early Head Start<sup>2</sup>
- Enhanced Early Head Start with Employment Services<sup>3</sup>
- Head Start<sup>4</sup>
- Even Start Family Literacy Program<sup>5</sup>
- 21st Century Community Learning Centers<sup>6</sup>
- Abstinence Education<sup>7</sup>
- Upward Bound<sup>8</sup>
- Food Stamp (renamed SNAP) Employment and Training Program<sup>9</sup>
- Welfare-to-Work<sup>10</sup>
- Employment Retention and Advancement (ERA) Project<sup>11</sup>
- Building Strong Families<sup>12</sup>
- Supporting Healthy Marriage<sup>13</sup>
- Moving to Opportunity and Section 8 Housing Vouchers<sup>14</sup>
- Job Training Partnership Act (JTPA) programs<sup>15</sup>
- Unemployment Insurance (UI) Self-Employment Demonstrations<sup>16</sup>
- Project GATE (Growing America Through Entrepreneurship)<sup>17</sup>
- Job Corps<sup>18</sup>
- JOBSTART<sup>19</sup>

- Center for Employment Training (CET) Replication<sup>20</sup>
- Quantum Opportunity Program Demonstration<sup>21</sup>

The results of these evaluations are summarized in *Do Federal Social Programs Work?*<sup>22</sup>

The consequence of so few federal social programs being rigorously assessed for effectiveness means that Congress has no credible information on the performance of the overwhelming majority of federal social programs. Faced with this lack of knowledge about the effectiveness of federal social programs, it is past time for Congress to devote serious attention and resources to finding out what works and does not work.

### **Can Effective Programs Be Replicated?**

Policymakers and advocates often assume that a social program that is effective in one setting will automatically produce the same results in other settings. Some proponents of evidence-based policy even make this faulty assumption. For example, advocates of expanding early childhood education programs make scientifically unsupportable generalizations regarding effectiveness based on two small-scale evaluations—the High/Scope Perry Preschool and Carolina Abecedarian Projects<sup>23</sup>—that are nowhere near being the definitive studies on the subject.<sup>24</sup> Policymakers should be very skeptical about the speculated payoffs of implementing these programs on a national scale. The evaluation of the Perry program began in 1962. Despite all the hoopla, the results have never been replicated. In more than 50 years, not a single experimental evaluation of the Perry approach applied in another setting or on a larger scale has produced the same results. The same holds true for the Abecedarian program, which began in 1972. There is no evidence that these programs can produce the same results today.

Many advocates of social programs have adopted the language of the “evidence-based” policy movement. Under the evidence-based policy movement, programs found to be effective using rigorous scientific methods are deemed “effective” or “evidence-based” and held up as “model” programs. The assumption is that the same successful impacts found at a particular setting can be replicated in other settings or on the national scale.

However, many of the programs labeled as “evidence-based”—often by program advocates—have been evaluated in only a single setting, so the results cannot necessarily be generalized to other settings. In addition, these evidence-based programs have often been implemented by highly trained professionals operating under ideal conditions. In the real world, program conditions are often much less than optimal. For example, based upon the results of the Abecedarian Project, Congress created Early Head Start—a national program that serves low-income families with pregnant women, infants, and toddlers up to age three.<sup>25</sup> However, the results of a multisite experimental evaluation of the national program found few initial modest impacts that quickly faded away.<sup>26</sup>

Another excellent example of the federal government replicating an effective local program is the Center for Employment Training (CET) Replication.<sup>27</sup> Of 13 youth job-training programs evaluated, the JOBSTART Demonstration found only one program to have a positive impact on earnings—the Center for Employment Training (CET) in San Jose, California.<sup>28</sup> Based on the results for the CET, the U.S. Department of Labor replicated and evaluated the impact of CET in 12 other sites.<sup>29</sup> The CET model had little to no effect on short-term and long-term employment and earnings outcomes at these other locations. The multisite experimental evaluation of CET, according to its authors, “shows, that even in sites that best implemented the

model, CET had no overall employment and earnings effects for youth in the program, even though it increased participants' hours of training and receipt of credentials."<sup>30</sup>

Just because an innovative program appears to have worked in one location, does not mean that the program can be effectively implemented on a larger scale.

### **What Congress Should Do**

Congress needs to take the lead in making sure that the social programs it funds are evaluated. First, when authorizing a new social program or reauthorizing an existing program, Congress should specifically mandate multisite experimental evaluation of the program. Congressional mandates are necessary because federal agencies often resist performing experimental evaluations. For example, many jurisdictions receiving funding through the Job Training Partnership Act (JTPA) and Job Opportunities and Basic Skills (JOBS) programs refused to cooperate with large-scale experimental evaluations of these programs.<sup>31</sup>

Experimental evaluations are the only way to determine to a high degree of certainty the effectiveness of social programs. Thus, Congress should mandate that all recipients of federal funding, if selected for participation, must cooperate with evaluations in order to receive future funding.

Second, the experimental evaluations should be large-scale, nationally representative, multisite studies. When Congress creates social programs, the funded activities are intended to be spread out across the nation. For this reason, Congress should require nationally representative, multisite experimental evaluations of these programs. For multisite evaluations, the selection of the sites to be evaluated should be representative of the population of interest for the program. When program sites and sample participants are randomly selected, the resulting evaluation findings will have high external validity.

While individual programs funded by federal grants may undergo experimental evaluations, these small-scale, single-site evaluations do not inform policymakers of the general effectiveness of national programs. The success of a single program that serves a particular jurisdiction or population does not necessarily mean that the same program will achieve similar success in other jurisdictions or among different populations. Thus, small-scale evaluations are poor substitutes for large-scale evaluations. In addition, a multisite experimental evaluation that examines the performance of a particular program in numerous and diverse settings can potentially produce results that are more persuasive to policymakers than results from a single locality.

The Building Strong Families (BSF) demonstration project sponsored by the U.S. Department of Health and Human Services is an excellent example of a program that had varying impacts by location. BSF provided counseling services to unmarried couples who were expecting or had recently had a baby in eight sites. The marriage program's intent was to steer low-income unmarried couples with or expecting a child toward marriage.

The eight-site demonstration project underwent an experimental evaluation that reported findings for 15- and 36-month follow-up periods. The 36-month follow-up study concluded: "After three years BSF had no effect on the quality of couple's relationships and did not make couples more likely to stay together or get married."<sup>32</sup> In addition at the 36-month follow-up period, "BSF had no effect on couples' co-parenting relationship; it had small negative effects on some aspects of father involvement."<sup>33</sup> Not to be dismissed, the long-term follow-up did find a

beneficial impact of increased socio-emotional development for children in the intervention group, compared to children in the control group.

While the evaluation of the eight demonstration sites found federally funded marriage promotion programs to be ineffective overall, the results from Atlanta, Baltimore, Oklahoma City, and the Florida counties were contradictory.<sup>34</sup> In Atlanta, BSF led to a long-term decrease in the ability of participants to avoid destructive conflict behaviors. In Baltimore, unmarried couples participating in the program were less likely to be still romantically involved at the time of the 15-month follow-up. In addition, couples in the Baltimore program reported less support and affection in their relationships, and fathers were less likely to provide financial support for their children and less likely to engage in cognitive and social play with their children. By the time of the 36-month follow-up, these harmful impacts in Baltimore faded away.

While the short-term findings for the Florida counties indicated that the BSF yielded no beneficial or harmful impacts on participants, the long-term findings indicate the presence of several harmful impacts. For the relationship status of the couples, intervention group couples were less likely to be romantically involved and living together (married or unmarried), compared to their counterparts in the control group. In addition, fathers in the intervention group were less likely to live with and regularly spend time with their child.

In Oklahoma City, the opposite occurred. While unmarried couples in the program were no more likely to marry than were the control group couples at the time of the 15-month follow-up, Oklahoma participants reported improvements in relationship happiness, support and affection, use of constructive conflict behaviors, and avoidance of destructive conflict behaviors. Additionally, fathers participating in the program were more likely to provide financial support for their children than were their counterparts in the control group. While BSF still had no effect on marriage rates at the time of the 36-month follow-up, couples in the intervention group were more likely to report that neither partner had been unfaithful since random assignment, compared to control group couples.

If the Atlanta, Baltimore, and Florida counties sites were the only sites evaluated, then the results would indicate that federally sponsored marriage counseling for unmarried couples with children has harmful effects. Relying only on the more positive Oklahoma City results would have led to the opposite conclusion.

Contradictory results from evaluations of similar social programs implemented in different settings are a product not only of implementation fidelity,<sup>35</sup> but also of the enormous complexity of the social context in which these programs are implemented. Jim Manzi, a senior fellow at the Manhattan Institute, uses the conflicting results of experimental evaluations to explain the influence of “causal density” on the social sciences.<sup>36</sup> “Causal density,” a term coined by Manzi, is “the number and complexity of potential causes of the outcomes of interest.”<sup>37</sup> Manzi postulates that as causal density rises, social scientists will find greater difficulty in identifying all of the factors that cause the outcome of interest.

The confounding influence of causal density, in addition to implementation fidelity, likely contributed to contradictory effects of federal marriage promotion programs by location. To address causal density, experimental impact evaluations of federal social programs should be conducted using multiple sites. In fact, the total sum of the multiple sites should be nationally representative of the populations served by the social program being evaluated.

## Causing Harm

The results of the 20 multisite experimental evaluations of 21 federal social programs published since 1990 generally find that these programs are ineffective.<sup>38</sup> However, social program advocates too frequently concentrate on any beneficial, even if only modest, impacts that have been identified. Nevertheless, politicians and policy experts also need to recognize that federal social programs can produce harmful impacts too. These harmful effects rarely get mentioned in government press releases announcing the findings of evaluations. In addition to the BSF findings, the following is a brief summary of the harmful impacts found in multisite experimental evaluations of federal social programs published since 1990.<sup>39</sup>

For Early Head Start, white parents in the intervention group displayed higher dysfunctional parent-child interactions than their counterparts in the control group.<sup>40</sup> Further, participation in Early Head Start appears to have increased welfare dependency for Hispanics.

Enhanced Early Head Start with Employment Services is a demonstration program that involves regular Early Head Start services with the addition of employment and training services for parents. An experimental evaluation of the program based on two sites in Kansas and Missouri was performed. At the time of the 48-month follow-up, the longest job spells of mothers participating in the program were significantly shorter than the job spells of mothers in the control group.<sup>41</sup>

For the three-year-old cohort of the Head Start Impact Study, kindergarten teachers reported that math abilities were worse than for similar children not given access to the program.<sup>42</sup> For the four-year-old cohort, teachers reported that Head Start children in the first grade were more likely to be shy or socially reticent than their peers. By the third grade, teachers reported that the four-year-old cohort with access to Head Start displayed a higher degree of unfavorable emotional symptoms than similar children without access to the program.<sup>43</sup> Further, children in the four-year-old cohort self-reported poorer peer relations with fellow children than their counterparts in the control group.<sup>44</sup>

The role of the federal government in funding after-school programs increased substantially after passage of the Improving America's School Act of 1994, which created the 21st Century Community Learning Centers program. A multisite experimental impact evaluation of the 21st Century Community Learning Centers program found a whole host of harmful effects.<sup>45</sup> Overall, teachers found participating students to have disciplinary problems that were confirmed by student-reported data. According to their teachers, participating students were less likely to achieve at above average or high levels in class and were less likely to put effort into reading or English classes. These students were also more likely to have behavior problems in school than their counterparts. Teachers were more likely to have to call the parents of participating students about misbehavior. Participating students were also more likely to miss recess or be placed in the hall for disciplinary reasons, while also having parents come to school more often to address behavior problems. 21st Century students were also more likely to be suspended from school than similar students.

Upward Bound was created in 1965 and is an original War on Poverty social program. Through the provision of supplemental academic and support services and activities, Upward Bound is intended to help economically disadvantaged high school students successfully complete high school and attend college. Despite the program's lofty goal, Upward Bound participants with high expectations to earn a college degree were less likely than their

counterparts to earn associate's degrees, while being no more or less likely to attain any other college degree.<sup>46</sup>

The Department of Health and Human Services and Department of Labor funded the Employment Retention and Advancement (ERA) project, initiated in 1998, to assess the effectiveness of 12 different employment retention and advancement programs across the nation.<sup>47</sup> Participation in ERA programs targeting unemployed Temporary Assistance for Needy Families (TANF) recipients in Houston, Texas, and Salem, Oregon, was associated with increased dependence on the receipt of TANF benefits, while participation in the program in Fort Worth, Texas, was associated with increased dependence on food stamps. The Chicago ERA program targeting employed TANF recipients was associated with increased dependence on food stamps, while the Medford, Oregon, ERA program targeting employed individuals not on TANF was associated with decreased employment.

Conducted in five cities, the Moving to Opportunity (MTO) demonstration assessed the impact of offering families with children under 18 living in public housing developments or concentrated poverty areas the opportunity to move out of their neighborhoods. The evaluation consisted of two intervention groups, MTO voucher recipients and Section 8 voucher recipients, compared to a control group that did not receive MTO or Section 8 vouchers but was eligible to receive public housing assistance. For adults and children with access to MTO or Section 8 vouchers, several harmful impacts were produced.<sup>48</sup> Access to a MTO voucher was associated with increased dependence on drugs and alcohol for adults. Also, MTO adults had higher participation rates in food stamps and received more food stamp benefits than their similar counterparts not given access to MTO or Section 8 vouchers. Youth from families given access to MTO vouchers were less likely to be employed and more likely to have smoked than their peers. These youth were also more likely to be arrested for property crimes. As for Section 8, adults offered access were more likely to be currently unemployed and less likely to have employment spells with the same job for at least a year. In addition, Section 8 adults were less likely to be currently working and not receiving TANF than their counterparts. Section 8 youth were more likely to have smoked than their peers in the control group.

Adult men participating in JTPA programs were more likely to be dependent on AFDC benefits than similar men not given access to the training.<sup>49</sup> Male youths with no criminal arrest record at the time of random assignment were more likely to be arrested after participating in federal job-training programs, while male youth with histories of arrest experienced long-term declines in income.

In an attempt to help Americans start businesses, the Department of Labor teamed with the Small Business Administration to create an employment program to assist people in creating or expanding their own business enterprises.<sup>50</sup> After receiving entrepreneurship training, Project GATE participants spent more time collecting Unemployment Insurance benefits than their counterparts that were not taught how to be entrepreneurs. While Project GATE had no effect on the self-employment income of participants, participants experienced initial periods of decreased wages and salaries earned from overall employment.

The Quantum Opportunity Program (QOP) demonstration, operated by the U.S. Department of Labor and the Ford Foundation from 1995 to 2001, offered intensive and comprehensive services with the intention of helping at-risk youth graduate from high school and enroll in postsecondary education or training. QOP provided services to participants year-round

for five years. The findings from the QOP experimental evaluation, according to its authors, provide some insight about the effectiveness of WIA youth programs. For the initial post-intervention impacts, youth participating in QOP were less likely to find jobs that provided health insurance benefits.<sup>51</sup> At the six-year follow-up period, youth participating in QOP were more likely to be arrested.<sup>52</sup> Increasing criminality appears to be a common effect of federal job-training programs supposedly benefiting youth.

The previously discussed CET Replication job-training programs were associated with several harmful outcomes.<sup>53</sup> Men experienced periods of declines in employment, earnings, and number of months worked. Individual participants who possessed a high school diploma or GED at the time of random assignment experienced periods of declines in the number of months worked and earnings. In addition, participants in the high-fidelity sites were less likely to find jobs that provided health insurance. Also, those older than 18 and those with high school degrees or GEDs at the time of random assignment were less likely to have jobs that provided health insurance.

Job Corps is another federal training program that has negative effects. Created in 1964, Job Corps is a residential job-training program that serves disadvantaged youths aged 16 to 24 in 125 sites across the nation. A multisite experimental evaluation of Job Corps found, compared to non-participants, Job Corp participants were less likely to earn a high school diploma.<sup>54</sup> In addition, youth participating in the program worked fewer weeks and worked fewer hours per week than similar youth in the control group.<sup>55</sup>

In sum, federal social programs that harm their participants are not uncommon. This fact is all too often ignored by advocates of these social programs.

## **Conclusion**

With the federal debt reaching staggering heights, Congress needs to ensure that it is spending taxpayer dollars wisely. Multisite experimental evaluations are the best method for assessing the effectiveness of federal social programs. Yet to date, this method has been used on only a handful of federal social programs. While previous results have been disappointing, Congress needs to reverse the trend of not rigorously evaluating federal social programs.

\*\*\*\*\*

The Heritage Foundation is a public policy, research, and educational organization recognized as exempt under section 501(c)(3) of the Internal Revenue Code. It is privately supported and receives no funds from any government at any level, nor does it perform any government or other contract work.

The Heritage Foundation is the most broadly supported think tank in the United States. During 2012, it had nearly 700,000 individual, foundation, and corporate supporters representing every state in the U.S. Its 2012 income came from the following sources:

Individuals	81%
Foundations	14%
Corporations	5%

The top five corporate givers provided The Heritage Foundation with 2% of its 2012 income. The Heritage Foundation's books are audited annually by the national accounting firm of

McGladrey & Pullen. A list of major donors is available from The Heritage Foundation upon request.

Members of The Heritage Foundation staff testify as individuals discussing their own independent research. The views expressed are their own and do not reflect an institutional position for The Heritage Foundation or its board of trustees.

---

<sup>1</sup> David B. Muhlhausen, *Do Federal Social Programs Work?* (Santa Barbara, CA: Praeger, 2013).

<sup>2</sup> John M. Love, Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gun, Diane Paulsell, Kimberly Boller, Jill Constantine, Cheri Vogel, Allison Sidle Fulongi, and Christi Brady-Smith, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start, Volume 1: Final Technical Report*, Princeton, NJ: Mathematica Policy Research, June 2002, and Cheri A. Vogel, Yange Xue, Emily M. Moiduddin, Barbara Lepidus Carlson, and Ellen Eliason Kisker, *Early Head Start Children in Grade 5: Long-Term Follow-Up of the Early Head Start Research Evaluation Project Study Sample: Final Report*, OPRE Report # 2011-8 (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, December 2010).

<sup>3</sup> JoAnn Hsueh, Erin Jacobs, and Mary Farrell, *A Two Generational Child-Focused Program Enhanced with Employment Services: Eighteen-Month Impacts from the Kansas and Missouri Sites of the Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project* (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, March 2011), and JoAnn Hsueh and Mary E. Farrell, *Enhanced Early Head Start with Employment Services: 42-Month Impacts from the Kansas and Missouri Sites of the Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project*, OPRE Report # 2012-05 (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, February 2012).

<sup>4</sup> U.S. Department of Health and Human Services, Administration for Children and Families, *Head Start Impact Study: First Year Findings* (Washington, DC, June 2005); U.S. Department of Health and Human Services, Administration for Children and Families, *Head Start Impact Study: Final Report* (Washington, DC, January 2010); and Michael Puma, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer, *Third Grade Follow-up to the Head Start Impact Study Final Report* (Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, October 2012).

<sup>5</sup> Robert St. Pierre, Anne Ricciuti, Fumiyo Tao, Cindy Creps, Takeko Kumagawa, and William Ross, *Third National Even Start Evaluation: Description of Projects and Participants* (Abt Associates Inc., 2001); Robert St. Pierre, Anne Ricciuti, Fumiyo Tao, Cindy Creps, Janet Swartz, Wang Lee, Amanda Parsad, and Tracy Rimdzius, *Third National Even Start Evaluation: Program Impacts and Implications for Improvement* (Cambridge, MA: Abt Associates Inc., 2003); and Anna E. Ricciuti, Robert G. St. Pierre, Wang Lee, Amanda Parsad, and Tracy Rimdzius, *Third National Even Start Evaluation: Follow-Up Findings from the Experimental Design Study* (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance 2004).

<sup>6</sup> Susanne James-Burdumy, Mark Dynarski, and John Deke, "When Elementary Schools Stay Open Late: Results from the National Evaluation of the 21st Century Community Learning Centers Program," *Educational Evaluation and Policy Analysis* Vol. 29, No. 4 (December 2007), pp. 296-318 and U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *21st Century Community Learning Centers Descriptive Study of Program Practices* (Washington, DC, U.S. Department of Education, 2010).

<sup>7</sup> Barbara Devaney, Amy Johnson, Rebecca Maynard, and Chris Trenholm, *The Evaluation of Abstinence Education Programs Funded under Title V Section 510: Interim Report*, Princeton, N.J. Mathematica Policy Research, April 2002; Rebecca A. Maynard, Christopher Trenholm, Barbara Devaney, Amy Johnson, Melissa A. Clark, John Homrighausen, and Ece Kalay, *First-Year Impacts of Four Title V, Section 510 Abstinence Education Programs*, Princeton, NJ: Mathematica Policy Research, June 2005; and Christopher Trenholm, Barbara Devaney, Ken Fortson, Lisa Quay, Justin Wheeler, and Melissa Clark, *Impacts of Four Title V, Section 510 Abstinence Education Programs: Final Report*, Princeton, NJ: Mathematica Policy Research, April 2007.

<sup>8</sup> David Myers and Allen Schirm, *The Short-Term Impacts of Upward Bound: An Interim Report*. Princeton, NJ: Mathematica Policy Research, May 1997; U.S. Department of Education, Office of the Under Secretary, Policy and Program Studies Service, *The Impacts of Regular Upward Bound: Results from the Third-Follow-Up Data Collection* (Washington, DC: U.S. Department of Education, April 2004); and Neil S. Seftor, Arif Mamun, and



---

Allen Schirm, *The Impacts of Regular Upward Bound on Postsecondary Outcomes 7–9 Years after Scheduled High School Graduation: Final Report*, Princeton, NJ: Mathematica Policy Research, January 2009.

<sup>9</sup> Michael J. Puma and Nancy R. Burstein, “The National Evaluation of the Food Stamp Employment and Training Program,” *Journal of Policy Analysis and Management* Vol. 13, No. 2 (1994), pp. 311–330.

<sup>10</sup> Gayle Hamilton, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines, Sharon McGroder, Martha Zaslow, Jennifer Brooks, Surjeet Ahluwalia, Electra Small, and Bryan Ricchetti, *National Evaluation of Welfare-to-Work Strategies: How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs* (Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families and Office of the Assistant Secretary for Planning and Evaluation; and U.S. Department of Education, 2001)

<sup>11</sup> Richard Hendra, Keri-Nicole Dillman, Gayle Hamilton, Erik Lundquist, Karin Martinson, Melissa Wavelet, Aaron Hill, and Sonya Williams, *How Effective Are Different Approaches Aiming to Increase Employment Retention and Advancement? Final Impacts for Twelve Models*, MDRC, April 2010.

<sup>12</sup> Robert G. Wood, Sheena McConnell, Quinn Moore, Andrew Clarkwest, and JoAnn Hsueh, *Strengthening Unmarried Parents’ Relationships: The Early Impacts of Building Strong Families*, Princeton, NJ: Mathematica Policy Research, May 2010, and Robert G. Wood, Quinn Moore, Andrew Clarkwest, Alexandra Killewald, and Shannon Monahan, *The Long-Term Effects of Building Strong Families: A Relationship Skills Education Program for Unmarried Parents: Final Report*, (Princeton, NJ: Mathematica Policy Research, November 2012).

<sup>13</sup> JoAnn Hsueh, Desiree Principe Alderson, Erika Lundquist, Charles Michalopoulos, Daniel Gubits, David Fein, and Virginia Knox, *The Supporting Healthy Marriage Evaluation: Early Impacts on Low-Income Families* (Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2012).

<sup>14</sup> Larry Orr, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling, *Moving to Opportunity Interim Impacts Evaluation: Final Report* (Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, June 2003) and Lisa Sanbonmatsu, Jens Ludwig, Lawrence F. Katz, Lisa Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, Stacy Tessler Lindau, Matthew Sciandra, Fanghua Yang, Ijun Lai, William Congdon, Joe Amick, Ryan Gillette, Michael A. Zabeck, Jordon Marvakov, Sabrina Yusuf, and Nicholas A. Potter, *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation* (Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, November 2011).

<sup>15</sup> Larry L. Orr, Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave, *Does Training for the Disadvantaged Work?* (Washington, DC: Urban Institute Press, 1996).

<sup>16</sup> Jacob M. Benus, Terry R. Johnson, Michelle Wood, Neelima Grover, and Theodore Shen, “Self-Employment Programs: A New Reemployment Strategy: Final Impact Analysis of the Washington and Massachusetts Self-Employment Demonstrations,” *Unemployment Insurance Occasional Paper No. 95–4*. Washington, DC: U.S. Department of Labor, December 1995.

<sup>17</sup> Jeanne Bellotti, Sheena McConnell, and Jacob Benus, *Growing America through Entrepreneurship: Interim Report*, Impaq International, August 2006 and Jacob Benus, Theodore Shen, Sisi Zhang, Marc Chan, and Benjamin Hansen, *Growing America through Entrepreneurship: Final Evaluation of Project GATE*, Columbia, MD: Impaq International, December 2009.

<sup>18</sup> Peter Z. Schochet, John Burghardt, and Steven Glazerman, *National Job Corps Study: The Impacts of Job Corps on Participants’ Employment and Related Outcomes* (Princeton, NJ: Mathematica Policy Research, Inc., June 2001); Sheena McConnell and Steven Glazerman, *National Job Corps Study: The Benefits and Costs of Job Corps* (Princeton, NJ: Mathematica Policy Research, Inc., June 2001); and Peter Z. Schochet, Sheena McConnell, and John Burghardt, *National Job Corps Study: Findings Using Administrative Earnings Records Data: Final Report* (Princeton, NJ: Mathematica Policy Research, Inc., October 2003).

<sup>19</sup> George Cave, Hans Bos, Fred Doolittle, and Cyril Toussaint, *JOBSTART: Final Report on a Program for School Dropouts* (Manpower Demonstration Research Corporation, October 1993).

<sup>20</sup> Cynthia Miller, Johannes M. Bos, Kristen E. Porter, Fannie M. Tseng, and Yasuyo Abe, *The Challenge of Replicating Success in a Changing World: Final Report on the Center for Employment Training Replication Cites* (Manpower Demonstration Research Corporation, September 2005).

<sup>21</sup> Allen Schirm and Nuria Rodriguez, *The Quantum Opportunity Program Demonstration: Initial Post Intervention Impacts* (Mathematica Policy Research, June 2004) and Allen Schirm, Elizabeth Stuart, and Allison McKie, *The Quantum Opportunity Program Demonstration: Final Impacts*, (Mathematica Policy Research, July 2006).

<sup>22</sup> See Muhlhausen, *Do Federal Social Programs Work?*

- 
- <sup>23</sup> Lawrence J. Schweinhart, Helen V. Barnes, and David P. Wiekart, *Significant Benefits: The High/Scope Perry Preschool Study through Age 27* (Ypsilanti, Mich.: The High/Scope Press, 1993), and Frances A. Campbell and Craig T. Ramey, “Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families,” *Child Development*, Vol. 65 (1994), pp. 684-698.
- <sup>24</sup> See Muhlhausen, *Do Federal Social Programs Work?*
- <sup>25</sup> Geoffrey D. Borman, “National Efforts to Bring Reform to Scale in High-Poverty Schools: Outcomes and Implications” in *Scaled-Up in Education: Issues in Practice, Vol. II*, eds. Barbara Scheider and Sarah-Kathryn McDonald (Lanham, Md.: Rowman & Littlefield, Inc., 2007), pp. 41-67.
- <sup>26</sup> See Muhlhausen, *Do Federal Social Programs Work?*, pp. 80-98.
- <sup>27</sup> Miller et al., *The Challenge of Replicating Success in a Changing World*.
- <sup>28</sup> George Cave et al., *JOBSTART*.
- <sup>29</sup> Cynthia Miller et al., *The Challenge of Replicating Success in a Changing World*.
- <sup>30</sup> *Ibid.*, p. xi.
- <sup>31</sup> Fred Doolittle and Linda Traeger, *Implementing the National JTPA Study* (New York: Manpower Demonstration Research Corporation, 1990); Judith M. Gueron, “The Politics of Random Assignment: Implementing Studies and Affecting Policy,” 15-49 in *Evidence Matters: Randomized Trials in Education Research*, edited by Frederick Mosteller and Robert Boruch, (Washington, D.C.: Brookings Institution, 2002).
- <sup>32</sup> Wood et al., *The Long-Term Effects of Building Strong Families*, p. xiii.
- <sup>33</sup> *Ibid.*, p. xiii.
- <sup>34</sup> For the results of the individual sites, see Wood et al., *The Long-Term Effects of Building Strong Families*, Tables A.2a–A.9b, pp. A.5–A.19.
- <sup>35</sup> Implementation fidelity is the degree to which programs follow the theory underpinning the program and how correctly the program components are put into practice.
- <sup>36</sup> Jim Manzi, “What Social Science Does—and Doesn’t—Know,” *City Journal*, Vol. 20, No. 3 (Summer 2010), pp. 14–23, [http://www.city-journal.org/2010/20\\_3\\_social-science.html](http://www.city-journal.org/2010/20_3_social-science.html) (Accessed July 9, 2013).
- <sup>37</sup> *Ibid.*
- <sup>38</sup> Muhlhausen, *Do Federal Social Programs Work?*
- <sup>39</sup> For the full results of these evaluations, including the beneficial, harmful, and no impact findings, see Muhlhausen, *Do Federal Social Programs Work?*
- <sup>40</sup> Love et al., *Making a Difference in the Lives of Infants and Toddlers and Their Families*, Table VII.11, pp. 381–385.
- <sup>41</sup> Hsueh and Farrell, *Enhanced Early Head Start with Employment Services*, Table 3.2, pp. 36–37.
- <sup>42</sup> U.S. Department of Health and Human Services, *Head Start Impact Study: Final Report*, Exhibit 4.5, pp. 4-21–4-25.
- <sup>43</sup> Puma et al., *Third Grade Follow-up to the Head Start Impact Study Final Report*, p. 84.
- <sup>44</sup> *Ibid.*, Exhibit 4.3, pp. 81-82.
- <sup>45</sup> Burdumy et al., “When Elementary Schools Stay Open Late.”
- <sup>46</sup> Seftor et al., *The Impacts of Regular Upward Bound on Postsecondary Outcomes*, Table IV.2, p. 59.
- <sup>47</sup> Hendra et al., *How Effective Are Different Approaches Aiming to Increase Employment Retention and Advancement?*
- <sup>48</sup> Sanbonmatsu et al., *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*.
- <sup>49</sup> Orr et al., *Does Training for the Disadvantaged Work?*
- <sup>50</sup> Bellotti et al., *Growing America through Entrepreneurship: Interim Report*.
- <sup>51</sup> Schirm and Rodriguez, *The Quantum Opportunity Program Demonstration: Initial Post Intervention Impacts*.
- <sup>52</sup> Schirm et al., *The Quantum Opportunity Program Demonstration: Final Impacts*.
- <sup>53</sup> Miller et al., *The Challenge of Replicating Success in a Changing World: Final Report on the Center for Employment Training Replication Cites*.
- <sup>54</sup> Schochet, Burghardt, and McConnell, “Does Job Corps Work? Impact Findings from the National Job Corps Study.”
- <sup>55</sup> Schochet et al., *National Job Corps Study: The Impacts of Job Corps on Participants’ Employment and Related Outcomes*.