



214 Massachusetts Avenue, NE • Washington DC 20002 • (202) 546-4400 • [heritage.org](http://heritage.org)

*CONGRESSIONAL TESTIMONY*

---

**Do Federal Social Programs for  
Children Work?**

**Testimony before  
the Committee on the Budget  
United States Senate**

**June 26, 2013**

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

My name is David Muhlhausen. I am Research Fellow in Empirical Policy Analysis in the Center for Data Analysis at The Heritage Foundation. I thank Chairwoman Patty Murray, Ranking Member Jeff Sessions, and the rest of the committee for the opportunity to testify today on the effect of sequestration on children. 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. To answer in the affirmative, federal social programs must ameliorate the social problems they target. In essence, social programs seek to improve human behavior in ways that will make people better off.

Two types of federal social programs—early childhood education and youth job-training programs—are the focus of my testimony.<sup>2</sup> Determining the effectiveness of these social programs is particularly relevant given the current political debate over the federal government’s persistent deficits and debt. For example, President Barack Obama has claimed that “70,000 young children would be kicked off Head Start” due to sequestration.<sup>3</sup> The clear implication is that 70,000 children will somehow be harmed by not attending Head Start. This would be true only if Head Start is an effective program that actually benefits the children it serves.

Before I review evaluations of federal early childhood education and youth job-training programs, the standards Congress should use in judging the effectiveness of social programs are discussed.

### **Standards for Assessing the Effectiveness of Federal Social Programs**

Given the fiscal crises that the federal government is facing, holding federal social programs accountable for their performance is necessary to regain control over excessive spending. Operating with increasingly scarce resources, federal policymakers need to start denying funds to ineffective programs, even if calls for funding these programs *seem* morally compelling. Calling for more spending on social programs may seem morally compelling, but continuing to spend taxpayer dollars on programs that do not produce their intended results is morally indefensible. Americans, especially income tax-payers, deserve better.

Social programs should be carefully evaluated to determine whether they do, in fact, work. Determining whether these programs work requires reliably sorting out the effect of a social program from confounding factors, which is a difficult task. Unfortunately, Congress too often relies on self-serving anecdotal observations offered by individuals and organizations dependent on federal government funding.

*Science Versus Anecdotal Observations.* There are numerous methods of making sense of the world around us. We frequently make personal observations of events around us to bring order to our lives. We often assign cause-and-effect relationships to events we personally experience. For instance, learning that touching a hot stove will burn one’s

hand is an easy cause-and-effect association that does not need to be tested more than once. We can easily correlate the act of touching the stove with the pain felt. Firsthand experience is often instrumental to developing knowledge. Every day, we make personal observations that guide us in our activities. We often seek the advice of others based on their personal experiences.

Congress frequently seeks policy advice through hearings. At congressional hearings, congressional committees seek the testimony of experts. On many occasions, these committees are collecting advice on the merits of social programs. As is often the case at these hearings, the invited panelists offer their opinions of the pros and cons of the social program of interest. A frequent type of panelist is an administrator of a social program that is financially dependent on continued federal funding.

Members of Congress should take any claim of effectiveness from individuals dependent on federal funding with a healthy dose of skepticism. No one who comes before Congress with hat in hand seeking federal funding is going to admit that they do not know if their program works or that their program is ineffective. The same holds true for claims of impending doom if budget cuts occur, no matter how small or large. With the federal government spending hundreds of billions of dollars per year on social programs, we should expect Congress to not rely on personal opinions that are too often self-serving.

Further, the usefulness of personal observations or experiences can be suspect when assessing complex social interactions that can have multiple causes. This problem is particularly acute when assessing the effectiveness of social programs where multiple factors can cause the outcomes of interest.

Assessing the effectiveness of federal social programs should be based on evaluations with two important characteristics. First, policymakers should rely on experimental designs that use random assignment. Second, policymakers should rely on large-scale evaluations that assess the effectiveness of federal social programs in multiple settings.

*Experimental designs.* Impact evaluations often assess impacts by comparing treatment or intervention groups to control or comparison groups. Determining the impact of social programs requires comparing the conditions of those who received assistance with the conditions of an equivalent group that did not experience the intervention. However, evaluations differ by the quality of methodology used to separate the net impact of programs from other factors that may explain differences in outcomes between comparison and intervention groups.<sup>4</sup>

Experimental evaluations are the “gold standard” of evaluation designs. Randomized experiments attempt to demonstrate causality by (1) holding all other possible causes of the outcome constant, (2) deliberately altering only the possible cause of interest, and (3) observing whether the outcome differs between the intervention and control groups.

When conducting an impact evaluation of a social program, identifying and controlling for all the possible factors that influence the outcomes of interest is impossible. We

simply do not have enough knowledge to accomplish this task. Even if we could identify all possible causal factors, collecting complete and reliable data on all of these factors would likely still be beyond our abilities. For example, it is impossible to isolate a person participating in a social program from his family in order to “remove” the influences of family. This is where the benefits of random assignment become clear.

Because we do not know enough about all possible causal factors to identify and hold them constant, randomly assigning test subjects to intervention and control groups allows us to have a high degree of confidence that these unidentified factors will not confound our estimate of the intervention’s impact. Random assignments should evenly distribute these unidentified factors between the intervention and control groups of an experimental evaluation.

However, the benefits of random assignment are most likely to occur when large sample sizes are used. Randomized evaluations using small sample sizes do not have the same scientific rigor as randomized evaluations using large sample sizes. Random assignment helps to ensure that the control group is equivalent to the intervention group in composition, predispositions, and experiences. Randomization is supposed to result in the intervention and control groups having an identical composition. The groups are composed of the same types of individuals in terms of their program-related and outcome-related characteristics. In addition, the intervention and control groups should have identical predispositions. Members of both groups are similarly disposed towards the program. Further, the intervention and control groups should have identical experiences with regards to time-related internal validity processes, such as maturation, and history.<sup>5</sup>

Randomized experiments have the highest internal validity when sample sizes are large enough to ensure that idiosyncrasies that can affect outcomes are evenly distributed between the program and control groups. With small sample sizes, disparities in the program and control groups can influence the findings. For this reason, evaluations with large samples are more likely to yield scientifically valid impact estimates.

*Multi-site designs.* Congress can take several steps to ensure that federal social programs are properly assessed using experimental evaluations. These experimental evaluations should be large in scale and based on multiple sites to avoid the problems of simplistic generalizations. A multitude of confounding factors influences the performance of social program. Thus, the larger the size of the evaluation (e.g., sample size and number of sites), the more likely the federal social program will be assessed under all of the conditions under which it operates.

When Congress creates social programs, the funded activities are implemented in multiple cities or towns. While individual social programs operating in a single location and funded by the federal government may undergo experimental evaluations, these small-scale, single-site evaluations do not inform policymakers of the general effectiveness of national social programs. Small-scale evaluations assess only the impact on a small fraction of the people served by federal social 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. Simply, small-scale evaluations are poor substitutes for large-scale, multisite evaluations. As will be detailed later in my testimony, Congress created the national Early Head Start program based upon the findings of the small-scale Carolina Abecedarian evaluation. After undergoing a multisite experimental evaluation, the federal government failed to replicate original effects of the Abecedarian Project on a national scale.

Thus, federal social programs should be evaluated in multiple sites so that social programs can be tested in the various conditions in which they operate and in the numerous types of populations that they serve. 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.<sup>6</sup>

The case of police departments performing mandatory arrests in domestic violence incidents is a poignant example of why caution should be exercised when generalizing findings from a single evaluation. During the 1980s, criminologists Lawrence W. Sherman and Richard A. Berk analyzed the impact of mandatory arrests for domestic violence incidents on future domestic violence incidents in Minneapolis, Minnesota.<sup>7</sup> Compared to less severe police responses, the Minneapolis experiment found that mandatory arrests lead to significantly lower rates of domestic violence. Sherman and Berk urged caution, but police departments across the nation adopted the mandatory arrest policy based on the results of one evaluation conducted in one city.

However, what worked in Minneapolis did not always work in other locations. Experiments conducted by Sherman and others in Omaha, Nebraska; Milwaukee, Wisconsin; Charlotte, North Carolina; Colorado Springs, Colorado; and Dade County, Florida, found mixed results.<sup>8</sup> Experiments in Omaha, Milwaukee, and Charlotte found that mandatory arrests lead to long-term *increases* in domestic violence. Apparently, knowing that they would automatically be arrested prompted repeat offenders to become more abusive. It seems that the following sick logic occurred: If the offender is going to automatically spend the night in jail, then he might as well beat his wife or girlfriend extra good. In a subsequent analysis of the disparate findings, Sherman postulated that arrested individuals who lacked a stake in conformity within their communities were significantly more likely to engage in domestic violence after arrest, while married and employed arrested individuals were significantly less likely to commit further domestic violence infractions.<sup>9</sup> Thus, mandatory arrest policies may be more likely to work in communities with high rates of marriage and employment, than communities with lower rates of marriage and employment.

Contradictory results from evaluations of similar social programs implemented in different settings are a product not only of implementation fidelity (the degree to which social programs are implemented as originally intended), 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>10</sup> “Causal density,” a term coined by Manzi, is “the number and complexity of potential causes of the outcomes of interest.”<sup>11</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 likely contributed to contradictory effects of mandatory arrest policies 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. Combined with random assignment, this approach is the best method for assessing the effectiveness of federal social programs.

Using evidence from scientifically rigorous multisite experimental evaluations of national programs, my testimony makes the case that real reductions in spending or slowing the rate in increase in spending on early childhood education and youth job-training programs will not harm children and youth. The reason for my conclusion is that the best research available finds that these social programs are highly ineffective. With the federal government’s debt approaching \$17 trillion, the American public has nothing to fear from reduced funding for ineffective social programs.

### **Early Childhood Education Programs**

Proponents 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—that are nowhere near being the definitive studies on the subject.<sup>12</sup> Policymakers should be very skeptical about speculated payoffs to society based upon two small-scale evaluations of early childhood education programs.<sup>13</sup> For example, James Heckman of the University of Chicago and his coauthors estimate that the Perry program, an early childhood education program that primarily targeted black children, produced \$7 to \$12 in societal benefits for every dollar invested.<sup>14</sup> The major benefit of the program is derived from reduced crime.<sup>15</sup>

Based on Heckman’s research, President Barack Obama during his 2013 State of the Union Address made the broad generalization that “Every dollar we invest in high-quality early childhood education can save more than seven dollars later on—by boosting graduation rates, reducing teen pregnancy, even reducing violent crime.”<sup>16</sup> President Obama is making a narrow-to-broad generalization, where he assumes that a program implemented in Ypsilanti, Michigan will have the same effect everywhere else in the nation. There are several problems with making broad policy generalizations based upon the Perry and Abecedarian evaluation findings.

First, the results of these outdated evaluations have never been replicated. 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.

Second, as Amy E. Lowenstein of New York University points out, the Perry and Abecedarian findings are based on very small samples of children (123 and 111, respectively).<sup>17</sup> The small sample sizes pose serious drawbacks to making assertions about effectiveness.

Commenting on the Perry and Abecedarian evaluations, Charles Murray of the American Enterprise Institute correctly observes,

The main problem is the small size of the samples. Treatment and control groups work best when the numbers are large enough that idiosyncrasies in the randomization process even out. When you're dealing with small samples, even small disparities in the treatment and control groups can have large effects on the results. There are reasons to worry that such disparities existed in both programs.<sup>18</sup>

Third, the sample children for the Perry and Abecedarian evaluations consisted almost entirely of low-income blacks.<sup>19</sup> Can these programs have the same effect on whites and Hispanics? There is virtually no evidence that the results of the Perry and Abecedarian evaluations can be generalized to other populations.

Fourth, the beneficial impacts of these programs appear to be restricted to females in the treatment group.<sup>20</sup> According to Lowenstein, "treated females showed sharp increases in years of schooling, improved economic outcomes, reductions in criminal behavior and drug use, and increased marriage rates, but there were no significant long-term effects for males."

Fifth, the findings cannot be generalized to other locations.<sup>21</sup> Lowenstein warns that "we must be cautious in drawing conclusions about crime effects based on the reductions in crime found in the Perry Preschool study, because there is no way to know if these effects were specific to Ypsilanti, Michigan, where the Perry Preschool was located, or if they would have emerged regardless of where the study took place."<sup>22</sup>

Sixth, Robinson G. Hollister of Swarthmore College has pointed out that while the Perry evaluation was initially supposed to be based on random assignment, "the researchers made several nonrandom adjustments to the assignment, for instance, moving siblings so that they would be together in the treatment or control group, or moving all children of working mothers to the control group."<sup>23</sup> As a result, 20 percent of the sample used to make inferences about the effectiveness of the programs was not randomly assigned. For Hollister, the failure to carry out the experimental design "greatly undermine[s] one's ability to take estimates of the 'impacts' as sound."<sup>24</sup> The bottom line is that the Perry evaluation is not really based on a true experimental design, and, thus, it does not benefit from the strong internal validity of true experimental designs.

Seventh, the impacts of the Perry program seesaw over time. According to Hollister,

Further doubts about the reliability of the estimates arise from the fact that the estimated impacts in given areas, for example, academic achievement test scores, vary sharply over time (age of the child). For instance, the crime data suddenly show big differences in favor of the program in the age 27 data. The estimated impacts on crime play a large role in the overall high benefit-cost ratios that have been highly touted.<sup>25</sup>

Suddenly, the benefits of the program are prevalent long after the individuals participated in the program.

Last, the Perry and Abecedarian programs are not representative of the vast majority of early childhood education programs operating today. These programs were “carefully constructed, high quality, expensive programs” that “do not reflect the assortment of scaled-up [early childhood education] programs available to most low-income families with young children today.”<sup>26</sup> The Perry and Abecedarian programs “represent the exception rather than the rule.”<sup>27</sup> Thus, Lowenstein concludes that the claims of advocates are “somewhat misleading.”<sup>28</sup>

The Perry and Abecedarian programs are not realistic models to draw conclusions about the effectiveness of federal early childhood education programs. Fortunately, we have ample evidence based upon multisite experimental evaluations.<sup>29</sup>

*Early Head Start.* Early Head Start, created during the 1990s, is a federally funded community-based program that serves low-income families with pregnant women, infants, and toddlers up to age three. The results of the multisite experimental evaluation of Early Head Start are particularly important because the program was inspired by the findings of the Abecedarian Project.<sup>30</sup> By the time participants reached age three, Early Head Start had beneficial impacts on two out of six outcome measures for child cognitive and language development, while the program had beneficial effects on four out of nine measures of child-social-emotional development.<sup>31</sup> While the short-term (age three) findings indicated modest positive impacts, almost all of the positive findings for all Early Head Start participants were driven by the positive findings for black children. The program had little to no effect on white and Hispanic participants, who are the majority of program participants. For Hispanic children, the program failed to have a short-term impact on all six measures of child cognitive and language development, while the program had a beneficial effect on only one of nine measures of child-social-emotional development. For white children, the program failed to produce any beneficial impacts on these outcome measures.

For the long-term findings, the overall initial effects of Early Head Start at age three clearly faded away by the fifth grade.<sup>32</sup> For the 11 child-social-emotional outcomes, none of the results were found to have statistically meaningful impacts.<sup>33</sup> Further, Early Head Start failed to have statistically measurable effects on the 10 measures of child academic



outcomes, including reading, vocabulary, and math skills.

What happened when the long-term results were analyzed by race and ethnicity? There were only two beneficial impacts for black children on 11 of the child-social-emotional outcomes. For Hispanic and white children, there was no beneficial effects for all these outcomes.

For child academic outcomes, the long-term findings by race and ethnicity were consistent. Early Head Start failed to affect all 10 academic outcomes for each of the subgroups.

*Head Start.* Created as part of the War on Poverty in 1965, Head Start is a preschool community-based program intended to help disadvantaged children catch up to children living in more fortunate circumstances. Despite Head Start's long life, the program never underwent a thorough, scientifically rigorous evaluation of its effectiveness until Congress mandated an evaluation in 1998. The Head Start Impact Study began in 2002, and the immediate-term, short-term, and long-term results released in 2005, 2010, and 2012, respectively, are disappointing.<sup>34</sup> According to *CQ News*, the 2012 study "revealed that children who attended Head Start had lost most of its benefits by the time they reached third grade."<sup>35</sup> This assessment is entirely wrong. Almost all of the benefits of participating in Head Start disappeared by kindergarten.

Overall, the evaluation found that the program largely failed to improve the cognitive, socio-emotional, health, and parenting outcomes of children in kindergarten and first grade who participated compared with the outcomes of similar children who did not participate. According to the report, "[T]he benefits of access to Head Start at age four are largely absent by 1st grade for the program population as a whole."<sup>36</sup> Alarmingly, Head Start actually had a harmful effect on three-year-old participants once they entered kindergarten. Teachers reported that non-participating children were more prepared in math skills than the children who participated in Head Start.

The third-grade follow-up to the Head Start Impact Study followed students' performance through the end of third grade.<sup>37</sup> The results shed further light on the ineffectiveness of Head Start. By third grade, Head Start had little to no effect on cognitive, social-emotional, health, or parenting outcomes of participating children.

In addition to the failures of Early Head Start and Head Start, multisite experimental evaluations of the Enhanced Early Head Start with Employment Services, which provides early childhood care and employment training services to families, and the now-defunct Even Start Family Literacy Program, which was intended to meet the basic educational needs of parents and children, failed to produce beneficial impacts.<sup>38</sup> The scientific rigor of these evaluations clearly demonstrates that the federal government has serious trouble operating early childhood education programs. These programs have done a poor job of improving the cognitive abilities and socio-emotional development of children.

## Youth Job-Training Programs

The federal government has spent decades trying to improve the earnings of disadvantaged youth through various employment and training programs, but the Government Accountability Office has concluded that little evidence shows that youth and adult training programs are effective.<sup>39</sup>

*Job Training Partnership Act (JTPA)*. Conducted in 16 sites across the nation during the late 1980s and early 1990s, the JTPA evaluation tracked program effects for more than 20,000 adult men, adult women, and out-of-school youths over the course of 30 months.<sup>40</sup> The performance of JTPA programs is widely considered to be a failure, especially for youth.

Overall, JTPA programs failed to raise the incomes of female youth and male youth without an arrest record prior to random assignment. However, JTPA programs had a harmful impact on the incomes of male youth with prior arrest histories. Even more alarming, male youth nonarrestees were more likely to be arrested for crimes after participating in training, compared to similar counterparts not given access to training.

*Job Corps*. 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. Before the U.S. Senate Committee on Appropriations, Subcommittee on Labor, Health and Human Services, Education, and Related Agencies in 2011, Secretary of Labor Hilda L. Solis testified that the “Job Corps program has a long history of preparing disadvantaged youth for a successful transition into the workforce.”<sup>41</sup> Is Job Corps an effective program? Its primary hypothesis relating to employment and earnings is that “youth who obtain Job Corps education and training will become more productive and, hence, will have greater employment opportunities and higher earnings than those who do not.”<sup>42</sup> Fortunately, we have a multisite experimental impact evaluation of Job Corps (“2008 outcome study”) to assess the program’s effectiveness.<sup>43</sup>

The 2008 outcome study found:

- Compared to non-participants, Job Corp participants were less likely to earn a high school diploma (7.5 percent versus 5.3 percent);<sup>44</sup>
- Compared to non-participants, Job Corp participants were no more likely to attend or complete college;<sup>45</sup>
- Four years after participating in the evaluation, the average weekly earnings of Job Corps participants was \$22 more than the average weekly earnings of the control group; and<sup>46</sup>
- Employed Job Corps participants earned \$0.22 more in hourly wages compared to employed control group members.<sup>47</sup>

If the Job Corps actually improves the skills of its participants, then it should have substantially raised their hourly wages. However, \$0.22 increase in hourly wages suggests that Job Corps does little to boost the job skills of participants.

Other impact evaluations of Job Corps have found similar results. In 2001, the *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes* (“2001 outcome study”) measured the impact of the Job Corps on participants’ employment and earnings.<sup>48</sup> While the 2001 outcome study found some increases in the incomes of participants, the gains were trivial. For example, compared to non-participants, the estimated average increase in the weekly incomes of all participants over four years was never more than \$25.20.<sup>49</sup>

Another evaluation, the *National Job Corps Study: Findings Using Administrative Earnings Records Data* (“2003 study”), was published in 2003, but the Labor Department withheld it from the general public until 2006.<sup>50</sup> The 2003 study found that Job Corps participation did not increase employment and earnings. Searching for something positive to report, the 2003 study concludes that “There is some evidence, however, of positive earnings gains for those ages 20 to 24.”<sup>51</sup>

*Why Withhold the 2003 Study?* Based on survey data, the 2001 cost-benefit study assumed that the gains in income for participants will last indefinitely, a notion unsupported by the literature on job training.<sup>52</sup> But included in the 2003 study is a cost-benefit analysis that directly contradicts the positive findings of the 2001 cost-benefit study.

The 2003 study used official government data, instead of self-reported data, and used the more reasonable assumption that benefits decay, rather than last indefinitely.<sup>53</sup> Contradicting the 2001 cost-benefit study, the 2003 study’s analysis of official government data found that the benefits of the Job Corps do not outweigh the cost of the program. Even more damaging, the 2003 study re-estimated the 2001 cost-benefit study with the original survey data using the realistic assumption that benefits decay over time. According to this analysis, the program’s costs again outweighed its benefits.

*Is Job Corps Worth \$1.7 Billion Per Year?* According to Job Corps, the cost of the program per participant in program year 2009 was \$26,551.<sup>54</sup> This estimate excludes program administration expenses, so it undercounts the true cost of the program on a per participant basis. The Office of Inspector General estimates that the actual cost per participant is \$37,880—a difference of \$11,329.<sup>55</sup> Perhaps a more important performance metric is the cost per successful job placement. For this measure, the OIG estimates that each Job Corps participant who is successfully placed into any job costs taxpayers \$76,574.<sup>56</sup>

If Job Corps actually improves the skills of its participants, then it should have substantially raised their hourly wages. The 2001 study found participants earned \$0.24 more per hour than nonparticipants.<sup>57</sup> Six months later, this difference had decreased to \$0.22 per hour.<sup>58</sup> Job Corps does not provide the skills and training necessary to substantially raise the wages of participants. One is certainly within reason to question whether the program is a waste of taxpayers’ dollars as it costs \$76,574 per participant placed in any job with an average participation period of eight months,.

*JOBSTART.* The JOBSTART Demonstration evaluated the impact of 13 job-training programs that were offered by community-based organizations, schools, and the Job Corps across the nation.<sup>59</sup> The targets of the training programs were 17- to 21-year-old “economically disadvantaged” school dropouts with poor reading skills. Overall, the programs failed to increase the earnings of participants. Of the 13 sites, 12 were found to be ineffective at raising the incomes of participants.<sup>60</sup> However, one site—the Center for Employment Training (CET) in San Jose, California—had a positive impact on earnings. For policymakers, the important question is whether the CET results can be replicated at different sites and for different populations.

*CET Replication.* Based on the JOBSTART evaluation results for the CET program in San Jose, California, the U.S. Department of Labor, in 1992, sought to replicate the program at 16 other sites across the nation. Twelve of the sites were evaluated.<sup>61</sup> The key elements of the CET model include a full-time commitment to participate in employment and training services in work-like settings.<sup>62</sup> In addition, employers were involved in designing and delivering services.<sup>63</sup>

In a classic example of not being able to replicate the results of a “proven” social program, CET Replication job-training programs failed to increase the employment and earnings of participants. Over more than a five-year follow-up period, the CET model had little to no effect on the employment and earnings outcomes at these 12 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>64</sup>

However, CET participation was associated with some harmful outcomes. Male youth experienced 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 declines in the number of months worked and earnings.

*Quantum Opportunity Program (QOP).* 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.<sup>65</sup> As an afterschool program, QOP provided case management and mentoring, additional education, developmental and community service activities, supportive services, and financial incentives.<sup>66</sup> QOP provided services to participants year-round for five years. The results of the QOP demonstration are particularly important because the program included several features of Workforce Investment Act’s (WIA) youth programs’ funding stream.<sup>67</sup>

QOP has many similarities with WIA youth programs, including:

- Case management and mentoring by adult staff;
- Basic education and study skills tutoring;
- Community service training;
- Year-round services, including summer jobs;
- An assortment of support services, including transportation, childcare, food, and emergency financial assistance; and
- Technical assistance to local service providers.<sup>68</sup>

According to the authors of the QOP evaluation:

These similarities between QOP and WIA youth programs suggest that the findings from the evaluation of the QOP demonstration might reveal some of the implementation challenges that WIA youth programs might encounter and *indicate whether WIA youth programs are likely to be effective* [Emphasis added].<sup>69</sup>

Thus, the findings from the QOP experimental evaluation, according to its authors, provide some insight about the effectiveness of WIA youth programs.

The QOP demonstration was implemented at seven sites across the nation. Five sites were funded by the Department of Labor, while the remaining two sites were funded by the Ford Foundation.<sup>70</sup> The total cost per participant for the Labor-funded sites was \$18,000 to \$22,000, while the cost per participant in the Ford-funded sites ranged from \$23,000 to \$49,000.<sup>71</sup>

At the initial and six-year follow-up periods, participation in QOP failed to have beneficial impacts on the employment and earnings of participants.<sup>72</sup> The job skills learned from QOP apparently had no effect on earnings. However, youth participating in QOP were more likely to be arrested by the six-year follow-up period. In addition, these youth were less likely to find jobs that provided health insurance benefits.

### **Conclusion**

Do federal early childhood education and youth job-training programs work? Based on the scientifically rigorous multisite experimental evaluations, the answer certainly cannot be in the affirmative. Despite the best social engineering efforts, overwhelming evidence points to the conclusion that these social programs are ineffective.

It cannot be just a coincidence that these multisite experimental evaluations overwhelmingly find failure. While we all agree on the importance of children having a solid foundation when entering school, this belief, no matter how noble, does not change the fact that federal early childhood education programs are ineffective. The same holds true for youth job-training programs.

Concerns over effects of sequestration on children and youth are unwarranted. Reduced funding for ineffective programs will not harm children and youth, because these programs largely do not work in the first place. Private companies are not hurt by eliminating inefficient divisions and neither are people when ineffective government programs are cut. In fact, reduced government spending will likely help children face a smaller financial burden of enormous debt that Congress's overspending has already imposed upon them.

Our nation faces a severe debt crisis that threatens our very future. Americans should not fear reductions in funding for these social programs. Now is the time for deep budget cuts to 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> While my testimony focuses on early childhood education and youth job-training programs, *Do Federal Social Programs Work?* reviews the failure of other federal social programs that serve children and youth, including the 21st Century Community Learning Centers, Upward Bound, and sexual abstinence education programs.

<sup>3</sup> The White House, Office of the Press Secretary, “Fact Sheet: Examples of How the Sequester Would Impact Middle Class Families, Jobs and Economic Security,” February 8, 2013, at <http://www.whitehouse.gov/the-press-office/2013/02/08/fact-sheet-examples-how-sequester-would-impact-middle-class-families-job> (accessed June 19, 2013).

<sup>4</sup> For a detailed discussion of evaluation methodology, see *ibid.*

<sup>5</sup> The internal validity threat of history occurs when events taking place concurrently with the intervention could cause the observed effect, while maturation occurs when natural changes in participants that occur over time could be confused with an observed outcome. For a more detailed discussion of threats to internal validity, see Muhlhausen, *Do Federal Social Programs Work?*

<sup>6</sup> Erica B. Baum, “When the Witch Doctors Agree: The Family Support Act and Social Science Research,” *Journal of Policy Analysis and Management*, Vol. 10, No. 4 (Autumn 1991), pp. 603–615, and Judith M. Gueron, “The Politics of Random Assignment: Implementing Studies and Affecting Policy,” in Frederick Mosteller and Robert Boruch, eds., *Evidence Matters: Randomized Trials in Education Research* (Washington, DC: Brookings Institution, 2002), pp. 15–49.

<sup>7</sup> Lawrence W. Sherman and Richard A. Berk, “The Specific Deterrent Effects of Arrest for Domestic Assault,” *American Sociological Review* 49, No. 2 (April 1984), pp. 261–272.

<sup>8</sup> Lawrence W. Sherman, *Domestic Violence: Experiments and Dilemmas* (New York: Free Press, 1992); Lawrence W. Sherman, Douglas A. Smith, Janell D. Schmidt, and Dennis Rogan, “Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence,” *American Sociological Review* Vol. 57 (October 1992), pp. 680–690; Lawrence W. Sherman, Janell D. Schmidt, Dennis Rogan, Douglas A. Smith, Patrick R. Gartin, Ellen G. Cohn, Dean J. Collins, and Anthony R. Bacih, “The Variable Effects of Arrest on Criminal Careers: The Milwaukee Domestic Violence Experiment,” *Journal of Criminal Law & Criminology* 83, No. 1 (1992), pp. 137–169.

<sup>9</sup> Sherman, *Domestic Violence*.

<sup>10</sup> Jim Manzi, “What Social Science Does—and Doesn’t—Know,” *City Journal* 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 June 21, 2013).

<sup>11</sup> *Ibid.*

<sup>12</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>13</sup> James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savellyev, Adam Yavitz, “The Rate of Return to the HighScope Perry Preschool Program,” *Journal of Public Economics*, Vol. 94 (2010), pp. 114–128; Art Rolnick and Rob Grunewald, “Early Childhood Development: Economic Development with a High Public Return,” *The Region*, December 2003, pp. 6–12.

<sup>14</sup> Heckman et al., “The Rate of Return to the HighScope Perry Preschool Program,” pp. 115–116.

<sup>15</sup> *Ibid.*, p. 119.

<sup>16</sup> Barack Obama, “Remarks by the President in the State of the Union Address,” The White House, Office of the Press Secretary, February 12, 2013, at <http://www.whitehouse.gov/the-press-office/2013/02/12/remarks-president-state-union-address> (accessed June 18, 2013).

<sup>17</sup> Amy E. Lowenstein, “Early Care and Education as Educational Panacea: What Do We Really Know About Its Effectiveness,” *Educational Policy*, Vol. 25, No. 1 (2011), p. 102.

<sup>18</sup> Charles Murray, “The Shaky Science Behind Obama’s Universal Pre-K,” Bloomberg.com, February 20, 2013, <http://www.bloomberg.com/news/2013-02-21/the-shaky-science-behind-obama-s-universal-pre-k.html> (accessed June 19, 2013).

<sup>19</sup> Lowenstein, “Early Care and Education as Educational Panacea,” p. 102.

<sup>20</sup> *Ibid.*

<sup>21</sup> *Ibid.*

<sup>22</sup> *Ibid.*

---

<sup>23</sup>Robinson G. Hollister, “Opening Statement,” in *Social Experimentation, Program Evaluation, and Public Policy*, ed. Maureen A. Pirog (Wiley-Blackwell, 2008), pp. 19–20.

<sup>24</sup>Ibid.

<sup>25</sup>Ibid.

<sup>26</sup>Lowenstein, “Early Care and Education as Educational Panacea,” p. 102.

<sup>27</sup>Ibid.

<sup>28</sup>Ibid.

<sup>29</sup>The results of the evaluations for early childhood education and job-training programs reported in my testimony are based upon the 5 percent level of statistical significance.

<sup>30</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>31</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 Fulingi, 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

<sup>32</sup>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>33</sup>Vogel et al., *Early Head Start Children in Grade 5*, Table III.2, pp. 24–25.

<sup>34</sup>U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *Head Start Impact Study: First Year Findings*, June 2005, and U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *Head Start Impact Study: Final Report*, January 2010; and Mike 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,” U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, *OPRE Report 2012-45*, October 2012.

<sup>35</sup>*CQ News*, “Obama Details Plan for Expanding Preschool Education,” February 14, 2013.

<sup>36</sup>U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *Head Start Impact Study: Final Report*, January 2010p. xxxviii.

<sup>37</sup>Puma et al., “Third Grade Follow-Up to the Head Start Impact Study: Final Report.”

<sup>38</sup>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*, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, *OPRE Report No. 2012-05*, February 2012; U.S. Department of Education, *Third National Even Start Evaluation: Program Impacts and Implications for Improvement*, 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*, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, December 2004.

<sup>39</sup>U.S. Government Accountability Office, *Multiple Employment and Training Programs: Providing Information on Collocating Services and Consolidating Administrative Structures Could Promote Efficiencies*, GAO-11-92, January 2011.

<sup>40</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: The Urban Institute Press, 1996).

<sup>41</sup>Hilda L. Solis, “Statement of Hilda L. Solis, Secretary of Labor, before the Subcommittee of Labor, Health and Human Services, Education, and Related Agencies, Committee on Appropriations, United States Senate,” May 4, 2011, [http://www.dol.gov/\\_sec/media/speeches/20110504\\_subcom\\_labor.htm](http://www.dol.gov/_sec/media/speeches/20110504_subcom_labor.htm) (accessed June 21, 2013).

<sup>42</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), p. 30.



- 
- <sup>43</sup> Peter Z. Schochet, John Burghardt, and Sheena McConnell, “Does Job Corps Work? Impact Findings from the National Job Corps Study,” *American Economic Review*, Vol. 98, No. 5 (December 2008), pp. 1864-1886.
- <sup>44</sup> *Ibid.*, p. 1871.
- <sup>45</sup> *Ibid.*
- <sup>46</sup> *Ibid.*, p. 1872.
- <sup>47</sup> *Ibid.*
- <sup>48</sup> Schochet et al., *National Job Corps Study: The Impacts of Job Corps on Participants’ Employment and Related Outcomes*.
- <sup>49</sup> *Ibid.*, p. 130.
- <sup>50</sup> Erik Eckholm, “Job Corps Plans Makeover for a Changed Economy,” *New York Times*, February 20, 2007, at [www.nytimes.com/2007/02/20/washington/20jobcorps.html](http://www.nytimes.com/2007/02/20/washington/20jobcorps.html) (June 21, 2013) and Peter Z. Schochet, Sheena McConnell, and John Burghardt, *National Job Corps Study: Findings Using Administrative Earnings Records Data: Final Report* (Princeton, N.J.: Mathematica Policy Research, Inc., October 2003).
- <sup>51</sup> Schochet et al., *National Job Corps Study: Findings Using Administrative Earnings Records Data: Final Report*, p. 70.
- <sup>52</sup> Pedro Carneiro and James Heckman, “Human Capital Policy,” NBER *Working Paper* No. 39495, February 2003.
- <sup>53</sup> Schochet et al., *National Job Corps Study: Findings Using Administrative Earnings Records Data: Final Report*.
- <sup>54</sup> U.S. Department of Labor, Office of the Inspector General, *Job Corps Needs to Improve Reliability of Performance Metrics and Results*, September 30, 2011, <http://www.oig.dol.gov/public/reports/oa/2011/26-11-004-03-370.pdf> (June 21, 2013).
- <sup>55</sup> *Ibid.*
- <sup>56</sup> *Ibid.*
- <sup>57</sup> *Ibid.*, p. 139.
- <sup>58</sup> Schochet et al., *National Job Corps Study: The Impacts of Job Corps on Participants’ Employment and Related Outcomes*.
- <sup>59</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, p. xvii.
- <sup>60</sup> *Ibid.*, p. 158, Table 5.13.
- <sup>61</sup> Cynthia Miller, Johannes M. Ros, 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), p. 1.
- <sup>62</sup> *Ibid.*, p. 7–8.
- <sup>63</sup> *Ibid.*, p. 9.
- <sup>64</sup> *Ibid.*, p. xi.
- <sup>65</sup> Allen Schirm and Nuria Rodriguez, *The Quantum Opportunity Program Demonstration: Initial Post Intervention Impacts* (Mathematica Policy Research, June 2004), p. 1.
- <sup>66</sup> *Ibid.*, p. v.
- <sup>67</sup> *Ibid.*
- <sup>68</sup> *Ibid.*, p. 7.
- <sup>69</sup> *Ibid.*, p. 7. Emphasis added.
- <sup>70</sup> *Ibid.*
- <sup>71</sup> *Ibid.*, p. vii.
- <sup>72</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).