



DOE Weatherization Program: A Review of Funding, Performance, and Cost-Effectiveness Studies

Fred Sissine
Specialist in Energy Policy

January 11, 2012

Congressional Research Service

7-5700

www.crs.gov

R42147

Summary

This report analyzes the Department of Energy's (DOE's) Weatherization Assistance Program (WAP, the "program"). It provides background—a brief history of funding, program evolution, and program activity—and a review of program assessments and benefit-cost evaluations.

Budget debate over the program is focused on a \$5 billion appropriation in the Recovery Act of 2009, a report that state and local governments have yet to commit about \$1.5 billion of that total, and concerns about the quality of weatherization projects implemented with Recovery Act funding. During the debate over FY2011 funding, the House Republican Study Committee called for program funding to be eliminated. In April 2011, Congress approved \$171 million for the program in the final continuing resolution for FY2011 (P.L. 112-10). For FY2012, DOE requested \$320 million, the House approved \$30 million, the Senate Appropriations Committee recommended \$171 million, and the Conference Committee approved \$68 million. The budget debate provides the context for this report, but details of the current debate are beyond the scope of this report, which is focused on evaluations of program cost-effectiveness.

WAP is a formula grant program: funding flows from DOE to state governments and then to local governments and weatherization agencies. Over the 32 years from the program's start-up in FY1977 through FY2008, Congress appropriated about \$8.7 billion (in constant FY2010 dollars). The \$5 billion provided by the Recovery Act added more than 50% to the previous spending total.

Over the program's history, DOE's Oak Ridge National Laboratory (ORNL) and the Office of Management and Budget have used process and impact evaluation research methods to assess WAP operations and estimate cost-effectiveness. Virtually all the studies conducted through 2005 showed that the program was moderately cost effective. The studies included measures of operational effectiveness, energy savings, and non-energy benefits. The timing of past studies was a bit sporadic, driven mainly by new statutory requirements and program audits. Performance assessments have alternately identified improvements in program operations or identified operational problems that subsequently stimulated program improvements. Only the intensive evaluation study of program year 1989 (published in 1993) was designed to directly draw a national sample to produce empirical data on program cost-effectiveness. Such in-depth evaluations are costly and time-consuming. Most other "metaevaluations" were much less costly, using available state-level evaluation studies as the basis to infer national-level program impacts.

The large infusion of Recovery Act funding—and attendant changes in program structure—heightened interest in conducting a fresh assessment of operations and new scientifically based evaluations of program impacts. Unforeseen recession-driven events delayed use of Recovery Act funding. For example, the DOE Inspector General (DOE IG) found that recession-driven budget shortfalls, state hiring freezes, and state-wide planned furloughs delayed program implementation—and created barriers to meeting spending and home weatherization targets. In December 2011, the Government Accountability Office (GAO) released a performance audit which found that the Recovery Act phase of the program was successfully addressing most goals and the challenges identified by the DOE IG. Recently-launched evaluation studies by DOE aim to determine whether Recovery Act funding was used cost-effectively and whether it fulfilled goals for job creation.

The report concludes by reviewing the debate over the use of "outside" contractors to improve the objectivity and independence of weatherization program evaluations.

Contents

Background.....	1
Statutory Authority	1
Program Structure.....	1
Types of Weatherization Services.....	1
Eligible Groups	2
Low-Income Population and Energy Cost Burden.....	2
Allocation Formula and Intergovernmental Administration.....	3
Funding Use Breakdown.....	3
Distribution Factors.....	4
Funding Allocation Priorities and Proportional Funding Cuts.....	4
Funding History: Selected Highlights.....	5
Funding Trend Highlights.....	5
Funding History Highlights, FY1977 Through FY2008.....	6
Presidents Ford, Carter, and Reagan Administrations.....	6
Presidents George H. W. Bush and Clinton Administrations	7
President George W. Bush Administration.....	7
President Barack Obama Administration: Recovery Act Expands Program	9
Program Appeared “Shovel Ready”	9
New Program Requirements.....	10
Prevailing Wage	10
Additional Training	10
Recession Impeded State Capacity.....	10
Major Implementation Delays	11
Performance Assessments and Evaluation Studies of Cost-Effectiveness.....	12
Background.....	12
GPRAs Performance Assessment Requirements	12
DOE’s Approach to Assessment and Evaluation.....	13
Choice of Studies for Review.....	13
Selected Performance Assessments.....	14
2003 IG Performance Audit	14
2003 OMB Assessment with PART Method	15
2011 GAO Performance Audit of Recovery Act Funding.....	19
Selected Evaluation Studies.....	21
1993 National Evaluation Report.....	21
Three Metaevaluations (1997, 1999, 2005).....	21
2010 Technical Memorandum (FY2010 Projected Data).....	26
Current Plans for Assessment and Evaluation Studies.....	28
Plan for a Strategic Assessment (Deloitte)	28
Evaluation Plan for Regular Program (APPRISE, 2007-2008 Data)	29
Design for the APPRISE Second National Retrospective (Impact) Evaluation.....	30
Design for APPRISE Process Evaluation.....	31
Evaluation Plan for Recovery Act (APPRISE, 2009-2011 Data).....	31
Impacts of Davis Bacon Requirement.....	32
Impacts on Administration, Training, and Employment	32
Other Expected Results	33

Issue of Evaluation Independence and Objectivity.....	33
Debate Over the Use of Evaluation Contractors.....	34
Metcalf Argues: Use of Contractors Improves Evaluation Independence	34
Reingold Counter-Argues: With Contractors, Threats to Independence Remain.....	36
Klerman Contributes to the Debate.....	38
The Limits of Independence: Is There an Optimum?.....	40
Implications of the Independence Debate for DOE Evaluations.....	41
DOE Addresses Threats to Independence	41
Further Considerations for DOE’s Active Evaluations	42
Some Policy Options That May Enhance Independence	43

Figures

Figure B-1. DOE Weatherization Program Funding History: Requests and Appropriations.....	47
--	----

Tables

Table 1. Funding Allocation Priorities and Proportional Funding Cuts.....	5
Table 2. PART Rating Scale.....	17
Table 3. Benefit-Cost Ratios for the 1993 National Evaluation, Three Metaevaluations (1997, 1999, and 2005), and the 2010 Technical Memorandum	24
Table 4. DOE Assessment and Evaluation Activities Underway.....	29
Table 5. Debate: Does Evaluation Contracting Promote Independence?	40
Table A-1. DOE Weatherization Funding, Units Weatherized, and Benefit-Cost Ratio	44

Appendixes

Appendix A. DOE Weatherization Program: Historical Data.....	44
Appendix B. DOE Weatherization Funding Chart.....	46

Contacts

Author Contact Information.....	48
---------------------------------	----

Background

The DOE Weatherization Assistance Program enables low-income families to permanently reduce their energy bills by making their households more energy efficient. DOE program guidelines specify that a variety of energy efficiency measures are eligible for support under the program. The measures include insulation, space-heating equipment, energy-efficient windows, water heaters, and efficient air conditioners.

Statutory Authority

The program was created under Title IV of the Energy Conservation and Production Act of 1976 (P.L. 94-385).¹ The statute specifies that the program's primary purpose is

to increase the energy efficiency of dwellings owned or occupied by low-income persons, reduce their total residential energy expenditures, and improve their health and safety, especially low-income persons who are particularly vulnerable such as the elderly, the handicapped, and children.²

The 1973 oil crisis caused rapid increases in energy prices, which caused major economic dislocations for the nation. As one result, the program was designed to save imported oil and cut heating bills for low-income households. This included senior citizens living on fixed incomes and Social Security, who were especially hard hit by rising energy bills.³

The Department of Health and Human Services (HHS) operates a Low-Income Home Energy Assistance Program (LIHEAP), which was designed to help pay energy bills for low-income households.⁴ Since the inception of LIHEAP in 1981, up to 15% of funds could be used for weatherization. In 1990, the statute was amended to allow states to use up to 25% of funds for weatherization, without requiring a waiver from HHS.⁵

Program Structure

Types of Weatherization Services

At first, weatherization providers emphasized low-cost measures, such as covering windows with plastic sheets and caulking and weatherstripping windows and doors. Many of these activities involved emergency and temporary measures. With the accumulation of experience over time, the emphasis shifted to more permanent and more cost-effective measures. The range of qualified measures expanded to include storm windows and doors, attic insulation, space heating and water

¹ More detailed background is available in DOE's *Weatherization Assistance Program Overview* at http://www.waptac.org/data/files/website_docs/briefing_book/wap_programoverview_final.pdf.

² Title 42 of the U.S. Code, Chapter 81, Subchapter III, Part A, 6861.

³ DOE, *History of the Weatherization Assistance Program*. http://www1.eere.energy.gov/wip/wap_history.html

⁴ In 1977, Congress first appropriated funds to help low-income households pay energy bills. In 1981, P.L. 97-35 gave this activity statutory authority as the LIHEAP program.

⁵ For further information on LIHEAP, see CRS Report RL31865, *The Low Income Home Energy Assistance Program (LIHEAP): Program and Funding*, by Libby Perl.

heating systems, furnace and boiler replacements, and cooling efficiency measures.⁶ The cooling efficiency measures include activities such as air conditioner replacements, ventilation equipment, and screening and shading devices.

DOE recounts that the use of home energy audits was key to adapting portfolio of measures:

In the 1990s, the trend toward more cost-effective measures continued with the development and widespread adoption of advanced home energy audits. This proved to be a key advance for weatherization service providers since it required every home to be comprehensively analyzed before work began in order to select the most cost-effective measures and the best approach. This custom analysis of every home has become the hallmark of weatherization and ensures each client receives the most cost-effective treatment.⁷

Eligible Groups

DOE's Weatherization Program is one of the largest energy efficiency programs in the nation. It is implemented in all 50 states, in the District of Columbia, in the U.S. trust territories, and by Native American Tribes. Vulnerable groups are targeted, including the elderly, people with disabilities, and families with children. A high priority is given to households with an elderly or disabled member. In FY2000, 49% of the weatherized households were occupied by an elderly resident or by a person with a disability.

Low-Income Population and Energy Cost Burden

The American Recovery and Reinvestment Act of 2009 (Recovery Act, P.L. 111-5; §407a) revised the program guidelines to raise the low-income eligibility ceiling from 150% to 200% of the poverty level.⁸ Low-income households have lower total energy use and smaller bills than the non-low-income population. However, those bills represent a higher proportion of total income. For 2009, Oak Ridge National Laboratory (ORNL) estimated the average energy cost burden at about 10% of income for low-income households compared to about 3.3% for non-low-income households.⁹ DOE elaborates further:

Low-income households have lower average residential energy usage and lower residential energy bills than the non-low-income population, but this difference is not in proportion to household income. The average income of low-income households as provided in the 2005 RECS and adjusted for inflation was estimated at \$18,624 compared to \$71,144 for non-low-income households. In 2009 the group energy burden of low-income households, defined as average residential energy expense divided by average income, was estimated to be 10 percent of income for low-income households compared to 3.3 percent for non-low-income

⁶ DOE, *History of the Weatherization Assistance Program*.

⁷ Ibid.

⁸ ORNL, *Weatherization Assistance Program Technical Memorandum Background Data and Statistics* (ORNL/TM-2010/66, prepared by Joel Eisenberg) March 2010, p. 3. The Program now defines eligibility as household income at or below 200% of the poverty income guidelines or of the HHS LIHEAP eligibility guidelines. Of the 111 million households nationwide, about 38.6 million are eligible for LIHEAP, and about 16.6 million have an income below the poverty level. The geographic distribution is roughly 21% in the Northeast, 24% in the Midwest, 36% in the South, and 19% in the West. http://weatherization.ornl.gov/pdfs/ORNL_TM-2010-66.pdf

⁹ Ibid, p. 5. The energy burden was defined as average residential energy expense divided by average income.

households.... Households that actually received energy payment assistance, estimated at just over 5 million in 2005, had an even higher energy burden of 11.5 percent of income.¹⁰

Allocation Formula and Intergovernmental Administration

DOE employs a formula to allocate funding to the states and territories.¹¹ Each state and territory, in turn, decides how to allocate its share of the funding to local governments and jurisdictions.¹² Funds made available to the states are allocated to local governments and nonprofit agencies for purchasing and installing energy efficiency materials, such as insulation, and for making energy-related repairs.¹³

Funding Use Breakdown

The law directs DOE to reserve funds for national training and technical assistance (T&TA) activities that benefit all states and Native American Tribes. DOE allocates funding for T&TA activities at both the state and local levels. The total funding for national, state, and local T&TA was originally limited to 10% of an annual congressional appropriation.¹⁴ The Recovery Act allowed the T&TA share to increase temporarily to 20%.

The remaining funds comprise the total allocation to state programs. The program allocation consists of two parts: the base allocation and the formula allocation. The base allocation for each state is fixed, but the amount differs for each state. The fixed base was computed so that a revised formula would not cause large swings from previous allocations, which could disrupt a state's program operations.¹⁵ **Appendix A** provides a history of total annual funding.

In FY2010, a new program account for "Innovations in Weatherization" was funded. The new activity was designed to demonstrate new ways to increase the number of low-income homes weatherized and lower the federal cost per home for residential retrofits, while also establishing a stable funding base. Partnerships with traditional weatherization providers such as non-profits,

¹⁰ DOE, *Weatherization Assistance Program Technical Memorandum Background Data and Statistics*, March 2010, p. 5. RECS stands for the Residential Energy Consumption Survey prepared by DOE's Energy Information Administration (EIA). http://weatherization.ornl.gov/pdfs/ORNL_TM-2010-66.pdf

¹¹ The Energy Independence and Security Act of 2007 (P.L. 110-140, §411c) added Puerto Rico and the territories of the U.S. to the definition of "State" for the purpose of funding allocations. Beginning with Program Year 2009, the territories of American Samoa, Guam, Commonwealth of the Northern Mariana Islands, Commonwealth of Puerto Rico and the U.S. Virgin Islands were added to the program.

¹² Administrative rules, eligibility standards, the types of aid, and benefit levels are primarily decided at the state level. Eligibility is automatically given to applicants receiving Temporary Assistance to Needy Families (TANF) or Supplemental Security Income (SSI). Also, if a state elects, program eligibility can be extended to a household that meets (LIHEAP) eligibility criteria.

¹³ Most of the grantees are state-designated community action agencies, which administer multiple types of social service grants for low-income persons. No more than 10% of grant funds allocated to states may be used for administration.

¹⁴ DOE, *Weatherization Assistance Program Funding*, p. II-12.

¹⁵ The sum of the base allocations for all states totals \$171,258,000. The total formula allocations equal the total program allocations minus the base allocations.

unions, and contractors is the focus. The partners are expected to leverage financial resources, with a goal of \$3 of non-federal contributions for each \$1 that DOE provides.¹⁶

Distribution Factors

The distribution of total formula allocations across the states is based on three factors: the relative size of the low-income population, climatic conditions, and residential energy expenditures. The low-income population factor is the share of the nation's low-income households in each state expressed as a percentage of all U.S. low-income households. The climatic conditions factor is obtained from the heating and cooling degrees for each state, treating the energy needed for heating and cooling proportionately. The residential energy expenditure factor is an approximation of the financial cost burden that energy use places on low-income households.¹⁷

Funding Allocation Priorities and Proportional Funding Cuts

In the event of funding cuts below a minimum threshold level, DOE program rules specify how cuts would be carried out.¹⁸ The rule provides funding according to four priority levels. In descending order, the priorities are: national training and technical assistance (TTA) activities, TTA for state and local levels, base allocation to states, and—if the funds remaining after TTA exceed a threshold of \$209.7 million—then a formula is used to spread the remainder among the states. **Table 1** illustrates these priorities. In the other case—if the funds remaining after the total TTA allocation are less than \$209.7 million—then there is no formula allocation and the base allocation is reduced proportionally.¹⁹

¹⁶ DOE, *FY2011 Budget Request*, vol. 3, p. 440.

¹⁷ DOE, *Weatherization Assistance Program Funding*, p. II-12. The approximation is necessary due to the lack of state-specific data on residential energy expenditures by low-income households.

¹⁸ The interim final rule on the revised allocation formula, which was issued on June 5, 1995, set the minimum threshold at \$209.7 million.

¹⁹ DOE, *Weatherization Assistance Program Funding*, p. II-12. The rules state that if the post-TTA allocation falls below the threshold of \$209.7 million set under P.L. 103-332, then each state's program allocation shall be reduced by the same percentage relative to its allocation under P.L. 103-332. For example, if total program allocations for a given year were 10% below the amount under P.L. 103-332, then each state's program allocation would be 10% less than under P.L. 103-332.

Table I. Funding Allocation Priorities and Proportional Funding Cuts

Priority	Program Component	Share of Annual Appropriation	After Total TTA, Remaining Funds greater than \$209.7 million	After Total TTA, Remaining Funds less than \$209.7 million
1	TTA—National	—	—	—
2	TTA—State & Local	—	—	—
	Total TTA	Maximum 20%	—	—
3	Base Allocation	Minimum 80%	yes	yes, reduced proportionally
4	Formula Allocation		yes, based on population, climate, and energy use	no

Source: DOE, *Overview of the Allocation Formula for the Weatherization Program*, <http://www.waptac.org/data/files/Website.../Allocation%20Funding.docx>; “Weatherization Assistance Program for Low-Income Persons,” *Federal Register*, March 25, 2009, <http://www.federalregister.gov/articles/2009/03/25/E9-6628/weatherization-assistance-program-for-low-income-persons>.

Notes: For further details, see “Weatherization Assistance Program for Low-Income Persons,” *Federal Register*, March 25, 2009, <http://www.federalregister.gov/articles/2009/03/25/E9-6628/weatherization-assistance-program-for-low-income-persons>.

Funding History: Selected Highlights²⁰

Funding continuity has been elusive for this program. Over its history, WAP program funding has followed an up-and-down pattern, framed by occasional funding spikes and proposals to eliminate program funding and operations.

Funding Trend Highlights

The chart in **Figure B-1** of Appendix B, and the data in **Table A-1** of Appendix A, show the variation in the program’s historical funding trend, calibrated in constant 2010 dollars.²¹ **Table A-1** shows that the accumulated congressional appropriations for the 32-year period from FY1977 through FY2008 reached a sum of nearly \$8.7 billion.²² For FY2009, the Recovery Act made a special one-time appropriation of \$5.0 billion, which added about 57% to the sum of all previous appropriations through the end of FY2008.

Figure B-1 shows the alternating pattern of support, characterized by occasional agreements—and periods of marked differences—between administration and congressional funding viewpoints. A few observations on this history:

²⁰ Unless otherwise noted, all funding figures refer to constant FY2010 dollars.

²¹ The program funding from 1977 through 2008 can also be found at http://www.waptac.org/data/files/website_docs/briefing_book/2_programfunding_final.pdf.

²² In constant 2010 dollars.

- **Funding Range.** Except for the FY2009 Recovery Act, single-year program funding has ranged between a high of about \$500 million in FY1979 to a low of about \$150 million in FY1996 (constant FY2010 dollars);
- **Funding Trend.** Except for the FY2009 Recovery Act appropriation, program funding has been on a long-term downtrend since FY1979 (constant FY2010 dollars);
- **Past Requests for Termination.** In 8 out of 34 years (from FY1978 through FY2011), the administration requested zero funding (program termination): FY1982, FY1983, FY1984, FY1987-FY1990, and FY2009;
- **Administration-Congress Agreements.** In 7 years of the 35-year funding history from FY1977 through FY2011, Congress approved nearly the same funding (less than 6% difference) level as the administration requested: FY1979, FY1980, FY1981, FY1985, FY2001, FY2006, and FY2010;
- **Request-Appropriation Percent Differences.** Including the 8 years of zero funding requests, in 20 out of 35 years, the final congressional appropriation differed (plus or minus) from the administration request by more than 20%;
- **Request-Appropriation Differences Larger than \$50 Million.** In 20 out of 35 years (since FY1977), the final congressional appropriation differed (plus or minus) from the administration request by more than \$50 million;
- **Appropriation Exceeded Request by \$200 Million.** In 10 out of 35 years (since FY1977), Congress appropriated a funding level that exceeded the administration request by more than \$200 million (in 2010 dollars): FY1982-FY1984, FY1987-FY1992, and FY2009.

Funding History Highlights, FY1977 Through FY2008

Presidents Ford, Carter, and Reagan Administrations

Congress started the Weatherization Program in FY1977, during the Ford Administration. Prodded by the second oil import crisis, funding was driven rapidly upward during the Carter Administration through FY1980. For FY1981, the Carter Administration requested \$428.2 million (in constant 2010 dollars), which the 96th Congress approved. In January 1981, the outgoing Carter Administration issued a DOE FY1982 budget request that sought \$402.8 million (in constant 2010 dollars) for the Weatherization Program.

Shortly after coming into office in January 1981, the incoming Reagan Administration reversed the previous trend by requesting that \$26.2 million be rescinded from the FY1981 appropriation. The 97th Congress approved the rescission, bringing program funding down to \$376.6.0 million (in constant 2010 dollars) In March 1981, the Reagan Administration issued an FY1982 request for zero funding, noting that it planned to restructure WAP as a block grant program under the Department of Housing and Urban Development. In DOE's FY1983 request, the Reagan Administration again proposed to terminate the Weatherization Program and other DOE grant programs:

The Energy Conservation Grants account consolidates financial and technical assistance programs carried out under the Energy Conservation appropriation which are proposed for

termination in FY1983 in support of the Administration's proposal to dismantle DOE. These programs include State and Local Assistance.... The budget reductions are in response to the fact that motivated by rising energy cost and federal tax policies, individuals, businesses and other institutions have undertaken major conservation efforts. The President's economic recovery program, in conjunction with oil price decontrol and increasing natural gas prices will accelerate this trend. Public awareness of energy conservation benefits and the high level of private investment in energy conservation clearly show that the State/Local grant programs do not warrant further federal support.²³

Despite the zero request, Congress approved nearly \$473 million in 2010 dollars for FY1983. The Reagan Administration sought zero funding again in FY1984, FY1987, FY1988, and FY1989. Following FY1983, appropriations trailed downward steadily—and stood at about \$258 million in 2010 dollars in FY1989.

Presidents George H. W. Bush and Clinton Administrations

In parallel to the Reagan Administration requests for zero funding in FY1987, FY1988, and FY1989, the George H.W. Bush Administration sought zero funding for FY1990 and only relatively small amounts for FY1991 (\$22 million, in 2010 dollars) and FY1992 (\$35 million, in 2010 dollars).

In its first request, the Clinton Administration sought a major increase for FY1994 (\$332 million, in 2010 dollars), and sustained the request at about that level for FY1995. Congress approved most of those two requests. For FY1996, the Administration came back with a request at nearly the same level (about \$306 million, in 2010 dollars), but Congress approved only about \$150 million (in 2010 dollars)—less than half the request. In real dollar terms, this was the lowest appropriation since FY1977. For FY1997 through FY2001, the Administration's requests hovered in a range from about \$190 million to \$200 million, in 2010 dollars. Congress generally approved \$20 million to \$40 million less than those requests, in 2010 dollars. For FY2001, Congress approved about \$187 million (in 2010 dollars), which was just about \$1 million less than was requested by the outgoing Clinton Administration.

President George W. Bush Administration

Bush Administration's 2001 Initiative to Increase Funding

In May 2001, President George W. Bush's National Energy Policy Development (NEPD) Group released the National Energy Policy Report. In regard to improving national energy security, that report stated:

Energy security also requires preparing our nation for supply emergencies, and assisting low-income Americans who are most vulnerable in times of supply disruption, price spikes, and extreme weather.²⁴

The report cited the cost-effectiveness of the program:

²³ DOE, *FY1983 Budget Request* (v. 3, Energy Conservation Grant Programs), February 1982.

²⁴ The White House, *National Energy Policy*, Report of the National Energy Policy Development Group, May 2001, p. xv, <http://georgewbush-whitehouse.archives.gov/energy/2001/index.html>.

Currently, each dollar spent on home weatherization generates \$2.10 worth of energy savings over the life of the home; with additional economic, environmental, health, and safety benefits associated with the installation and resulting home improvements. Typical savings in heating bills, for a natural gas heated home, grew from about 18 percent in 1989 to 33 percent today.²⁵

As one part of the plan to assist low-income people with high energy costs,²⁶ the report indicated that the Bush Administration was committed to increasing support for DOE's program over the long-term. Specifically, the NEPD Group recommended that:

... the President increase funding for the Weatherization Assistance Program by \$1.2 billion over ten years. This will roughly double the spending during that period on weatherization. Consistent with that commitment, the FY2002 Budget includes a \$120 million increase over 2001.²⁷

As **Table A-1** and **Figure B-1** show, the Congress responded to the President's funding initiative by approving a nearly 50% increase for FY2002—raising funding from about \$187 million to about \$277 million, in constant 2010 dollars. The requests—and appropriations—climbed steadily for the next four years, reaching \$261 million in FY2006. After that, the requests turned sharply downward and appropriations dropped slightly, standing at \$232 million in FY2008.

Bush Administration's FY2009 Request to Terminate Funding

The FY2009 request observed that a 2003 assessment had rated the program as “moderately effective,”²⁸ and that:

the program coordinates effectively with other related government programs in its efforts to meet interrelated Departmental goals and achieves its goals of a favorable benefit-cost ratio and other performance goals, based on internal programmatic assessments.²⁹

DOE also noted that a new evaluation of the program's benefits and costs was underway.

However, in a reversal of its previous requests, the George W. Bush Administration requested that funding for the Weatherization Program be terminated in FY2009. The main rationale given for program termination was that the energy efficiency technology programs operated by DOE's Office of Energy Efficiency and Renewable Energy (EERE) provided a much higher benefit-to-cost ratio than that for the Weatherization Program.³⁰ Specifically, the narrative for the EERE in DOE's FY2009 Congressional Budget Request stated that:

In FY2009, Weatherization Assistance Program funds are redirected to R&D programs which deliver greater benefits. EERE's Energy Efficiency portfolio has historically provided

²⁵ The White House, *National Energy Policy*, p. 2-3.

²⁶ The report also called for increases to funding for the Low Income Home Energy Assistance Program (LIHEAP) at the Department of Health and Human Services.

²⁷ The White House, *National Energy Policy*, p. 2-12.

²⁸ The assessment was obtained from use of the Program Assessment Rating Tool (PART), which was developed by the Office of Management and Budget to provide a standardized way to assess federal programs.

²⁹ DOE, *FY 2009 Congressional Budget Request*, vol. 3, p. 463.

³⁰ EERE's technology programs include Vehicle Technologies, Building Technologies, and Industrial Technologies.

approximately a 20 to 1 benefit to cost ratio. In comparison, Weatherization has a benefit cost ratio of 1.53 to 1.³¹

President Barack Obama Administration: Recovery Act Expands Program

Under the Obama Administration, the American Recovery and Reinvestment Act of 2009 (Recovery Act, P.L. 111-5) identified the program as one avenue to help address the recession and provided the program with a major one-time increase—\$5 billion to weatherize an estimated 600,000 homes.³² This objective aimed to help achieve the President's goal of weatherizing one million homes per year. In addition to the funding increase, some amendments to the original authorizing law were enacted. The amendments allowed more cost-effective measures to be installed in more homes. One amendment raised the ceiling on cost per dwelling from \$2,500 to \$6,500.³³

Program Appeared “Shovel Ready”

The Recovery Act aimed to stimulate the U.S. economy, create jobs, and make infrastructure investments in energy and other areas. During congressional consideration of the Recovery Act, and following its enactment, the conventional wisdom was that DOE's Weatherization Program was about as close to meeting the definition of "shovel ready" as virtually any program in DOE's portfolio. Specifically, the Recovery Act weatherization effort had the following attributes: an existing programmatic infrastructure, including processes and procedures, which had been in place for many years. DOE's Inspector General observed that:

The techniques for weatherization tasks were well known and comparatively uncomplicated, and the requisite skills were widely available; performance metrics were relatively easy to establish and understand; the potential benefits for low income citizens were easily recognized; and, the potential beneficial impact on energy conservation was obvious.³⁴

With that well-entrenched program structure, there was a strong expectation that the \$5 billion in Recovery Act funds would have a prompt and easily measurable impact on job creation and economic stimulation.³⁵

³¹ DOE, *FY 2009 Congressional Budget Request*, vol. 3, p. 44.

³² DOE, Office of the Inspector General, *Special Report: Progress in Implementing the Department of Energy's Weatherization Assistance Program Under the American Recovery and Reinvestment Act* (OAS-RA-10-04), February 2010, p. 1. This was a major increase over the \$450 million appropriated for the program in FY2009. <http://energy.gov/sites/prod/files/igprod/documents/OAS-RA-10-04.pdf>

³³ DOE, *History of the Weatherization Assistance Program*.

³⁴ DOE IG, *Special Report on the Weatherization Program*, p. 5.

³⁵ Ibid.

New Program Requirements

The Recovery Act modified the program to incorporate a new labor requirement for “prevailing wage” and to embrace new training requirements.

Prevailing Wage

The Recovery Act required that state and local recipients of weatherization funds ensure that laborers be paid at least the “prevailing wage” as determined under the Davis-Bacon Act.³⁶ This requirement had not been previously applied to weatherization program activities. As a result, grantees lacked information on which to base wage rates. Many grantees chose not to begin work until the prevailing wage rates were formally established by the Department of Labor (DOL). Even after DOL’s work was complete, additional delays occurred while grantees prepared guidance for sub-recipients on how to apply the wage rates. Thus, the delivery of Recovery Act-funded weatherization activities did not reach full momentum until after guidance was completed in October 2009.³⁷

Additional Training

To ensure successful implementation of program criteria under the Recovery Act, DOE required that program personnel at state and community action agencies receive additional training. The aim of the training was to ensure that recipients understood Davis-Bacon Act requirements, new income eligibility requirements, increased allowable costs per unit, and monitoring of work performed by sub-recipients. As with other state and local activities, recession-driven budget shortfalls and staff furloughs delayed many required state training initiatives.³⁸

Recession Impeded State Capacity

Recession-driven state budget shortfalls caused certain states to administer hiring freezes that applied to all employees regardless of the source of their funding, including those tasked with weatherization-related work. In some other states, progress was impacted because personnel involved with the program were subjected to significant state-wide furloughs. Further, the approval of state budgets was delayed in states such as Pennsylvania, as legislators deliberated over how to address overall budget shortfalls. Lacking staff, states were unable to perform required implementation tasks necessary to handle the large infusion of Recovery Act funding for DOE’s weatherization program. Without adopted budgets, states did not have any spending authority. As a result, many states were not able to obligate or expend any weatherization program funds.

³⁶ Davis-Bacon is the common name applied to legislation enacted in the 1930s that requires all federal construction projects to pay prevailing wages. For more background on the Davis Bacon Act, see CRS Report R40663, *The Davis-Bacon Act and Changes in Prevailing Wage Rates, 2000 to 2008*, by Gerald Mayer.

³⁷ DOE IG, *Special Report on the Weatherization Program*.

³⁸ *Ibid.* p. 4.

Major Implementation Delays

DOE had taken a number of proactive steps to foster timely implementation of the program. In spite of those efforts, grantees had made little progress in weatherizing homes by the beginning of 2010. As of February 2010, the one-year anniversary of the Recovery Act, only a small percentage of Recovery Act weatherization funds had been spent and few homes had actually been weatherized. Only \$368.2 million (less than 8%) of the total award of \$4.73 billion had been drawn by grantees. With the low spending rates, state and local grant recipients fell significantly short of goals to weatherize homes.

The lack of progress by state and local grantees to implement the weatherization program funding alarmed the DOE Inspector General (IG). In early 2010, the IG found that the nation had not realized the potential economic benefits of the \$5 billion in Recovery Act funds allocated to the program. In particular, the job creation impact of what many considered to be one of DOE's most "shovel ready" projects had not materialized. Further, the IG observed that modest income residents had not enjoyed the benefits of reduced energy use and better living conditions that had been promised as part of the Recovery Act weatherization effort.

The IG's report found:

Department officials worked aggressively with the states and other responsible agencies to mitigate these challenges.... However, as a practical matter, program challenges, such as those identified in this report, placed the Recovery Act-funded Weatherization Program "on hold" for up to nine months.³⁹

Thus, the Recovery Act goals proved to be much more difficult to achieve than originally envisioned. The IG report noted:

The results of our review confirmed that as straight forward as the program may have seemed, and despite the best efforts of the Department, any program with so many moving parts was extraordinarily difficult to synchronize. In this case, program execution depended on the ability of the Federal government (multiple agencies, in fact), state government, grant sub-recipients and weatherization contractors, working within the existing Federal and state regulatory guidelines, to respond to a rapid and overwhelming increase in funding. Further, anticipated stimulus impact was affected by certain conditions and events clearly outside of Departmental control including state budget difficulties; availability of trained and experienced program staff; and, meaningful changes in regulatory requirements.⁴⁰

Thus, despite the assumption that the program was "shovel ready," the IG uncovered several key administrative and intergovernmental barriers to delivery of Recovery Act funded weatherization services.

³⁹ DOE Inspector General, *Special Report*, p. 2.

⁴⁰ DOE Inspector General, *Special Report*, p. 5.

Performance Assessments and Evaluation Studies of Cost-Effectiveness

This part of the report reviews a selection of key studies that examined Weatherization Program management processes, performance, and economic (energy and non-energy) impacts. Performance assessments and evaluation studies are related activities within a family of research on performance measurement and program evaluation. Both activities support the Government Performance and Results Act (GPRA) and the Office of Management and Budget's (OMB's) concerns for federal program management. In general, a performance assessment employs anecdotal data that is collected over a shorter time-frame, with a focus on management responsibilities and budget processes. In contrast, an evaluation study employs survey research methods to gather comprehensive data that is collected over a longer time-frame, with a focus on revealing the net impacts and cost-effectiveness of program operations.⁴¹

Background

GPRA Performance Assessment Requirements

The Government Performance and Results Act of 1993 (GPRA, P.L. 103-62) established requirements for federal agencies to conduct regular performance assessments based on general methods of management by objectives and strategic planning. The law directs OMB to lead GPRA implementation,⁴² in cooperation with federal agencies. OMB describes this responsibility in Part 6 of its Circular A-11:

GPRA provides the foundation for performance planning, reporting, and budgeting for Federal agencies. GPRA requires agencies to prepare strategic plans, related performance budgets/annual performance plans, and annual performance reports (31 U.S.C. §1115). The legal requirements for an annual performance plan are met by a performance budget. The annual performance report requirement (APR) will be fulfilled by either the annual performance and accountability report (PAR) or by the congressional budget justification for agencies that choose to produce a separate annual financial report (AFR) and APR.⁴³

Among more specific GPRA purposes, Circular A-11 identifies a key GPRA purpose to:

Improve Congressional decision-making by providing more objective information on achieving statutory objectives, and on the relative effectiveness and efficiency of Federal programs and spending....⁴⁴

⁴¹ For more on the definitions—and complementary aspects—of performance assessment and program evaluation studies, see: GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, May 2011, <http://www.gao.gov/assets/80/77277.pdf>; and OMB, *Performance Measurement Challenges and Strategies*, June 18, 2003, http://www.whitehouse.gov/sites/default/files/omb/part/challenges_strategies.pdf.

⁴² OMB's GPRA guidance is available on its website at <http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m#h1>.

⁴³ OMB, *Circular No. A-11*, July 21, 2010, Section 200.2, http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/a_11_2010.pdf.

⁴⁴ OMB, *Circular No. A-11*, Section 200.2.

Over the years, DOE has integrated its responses to GPRA requirements into its strategic plan, annual performance report,⁴⁵ annual performance and accountability report,⁴⁶ and annual budget request. DOE notes that it has, since 2002, been working with OMB to assess its programs with the Performance Assessment Rating Tool (PART).⁴⁷ Evidently, PART had become the main vehicle for reporting of GPRA assessment requirements. PART is described in greater detail in the next section of the report.

Circular A-11 defines program assessment:

A determination, through objective measurement and systematic analysis, of the manner and extent to which Federal programs achieve intended objectives.⁴⁸

It also defines program evaluation:

Program and practice evaluation is an assessment, through objective measurement and systematic analysis, of the manner and extent to which programs or practices achieve intended results. Program and practice evaluations should be performed with sufficient scope, quality, and independence. Although agencies may cite rigorous evaluations commissioned independently by organizations such as the Government Accountability Office, Office of the Inspector General, or other groups, these evaluations should not completely supplant rigorous evaluations commissioned by the agencies themselves.⁴⁹

DOE's Approach to Assessment and Evaluation

DOE's Weatherization Program has been assessed through a combination of assessment and evaluation methods. DOE's GPRA-driven program assessment tools—strategic plan, annual performance plan, performance and accountability report, and performance budget—tend to be quite broad, focused on the larger agency missions of R&D, science, and defense. Those documents tend to mention the program incidentally, mainly citing achievements in terms of annual number of units weatherized. For a more in-depth examination of program operations, DOE has a history of directing the Oak Ridge National Laboratory (ORNL) to conduct evaluation impact studies.

Choice of Studies for Review

In 1993, DOE published a major evaluation report, the first “scientific” study of program performance, impact, and cost-effectiveness. That report established impact evaluation studies as a key feedback component that promoted program design changes and improvements. However, its cost, complexity, and lag time led DOE to subsequently rely on aggregations of available state-level evaluations. Also, in 1993, GPRA stimulated the emergence of a second track of less in-

⁴⁵ DOE, *FY2009 Annual Performance Report*. For the Weatherization Program, this report only cites one performance measure, the number of homes weatherized in FY2009, compared with the number weatherized in FY2006 through FY2008. p. 87. <http://www.cfo.doe.gov/cf1-2/2009APR.pdf>

⁴⁶ DOE's *FY2006 Performance and Accountability Report* appears to be the most recent one available. It reports a few selected findings about the Weatherization Program from OMB's PART assessment on pp. 11, 19, and 102. <http://www.cfo.doe.gov/progliaison/2006finalpar2.pdf>

⁴⁷ DOE, *Strategic Plan (2006)*, p. 27. <http://www.cfo.doe.gov/strategicplan/docs/2006StrategicPlan.pdf>

⁴⁸ OMB, *Circular No. A-11*, Section 200.4.

⁴⁹ *Ibid.*

depth performance assessments that had a shorter-term budget focus. Both tracks of analysis—performance assessments and evaluation studies—have the same general goals of providing feedback that can be used to improve program operations. So both types of studies were selected for review.

Selected Performance Assessments

2003 IG Performance Audit

A performance audit is an assessment tool used to examine the performance and management of a program against objective criteria. It is designed to facilitate oversight and may serve several objectives, including the assessment of program effectiveness and results.⁵⁰ From late 2002 through early 2003, DOE's Office of the Inspector General (IG) conducted a performance audit of the program. The audit period included Program Year 2000 (PY2000), PY2001, and planned activities for PY2002.⁵¹ The total DOE budget for the audit review period was \$518 million. The purpose of the audit was to “determine whether the program was properly administered and was achieving its goals.”⁵² The audit was conducted in accordance with Government Auditing Standards for performance audits.⁵³

The IG observed that the program has a long-established structure for transferring funds to state and local agencies and that improvements over the years had made use of the funds more efficient and effective.⁵⁴ As further background, the IG observed that, in addition to DOE funds, state and local agencies also obtain funding for weatherization from the Department of Health and Human Services' (HHS) Low-Income Home Energy Assistance Program (LIHEAP) and other programs funded by utilities, states, and other sources.

Local agencies report to DOE annually through state weatherization offices on expenditures, number of units completed, and other performance-related measures. For PY2001, the IG noted that “about 900 agencies received funding that ranged from as low as a few thousand dollars to

⁵⁰ Government Accountability Office (GAO), *Government Auditing Standards: 2003 Revision*, June 2003 (Section 2.09). GAO provides a more detailed definition, as follows: Performance audits entail an objective and systematic examination of evidence to provide an independent assessment of the performance and management of a program against objective criteria as well as assessments that provide a prospective focus or that synthesize information on best practices or crosscutting issues. Performance audits provide information to improve program operations and facilitate decision making by parties with responsibility to oversee or initiate corrective action, and improve public accountability. Performance audits encompass a wide variety of objectives, including objectives related to assessing program effectiveness and results; economy and efficiency; internal control; compliance with legal or other requirements; and objectives related to providing prospective analyses, guidance, or summary information. Performance audits may entail a broad or narrow scope of work and apply a variety of methodologies; involve various levels of analysis, research, or evaluation; generally provide findings, conclusions, and recommendations; and result in the issuance of a report. <http://www.gao.gov/govaud/yb2003.pdf#Page=27>

⁵¹ DOE generally defines a program year as the 12-month period starting on July 1 of the year after DOE receives the appropriation from Congress.

⁵² DOE, Office of the Inspector General, *Audit of the Weatherization Assistance Program*. (Audit Report No.: OAS-L-03-15) April 18, 2003, p. 1. (Hereinafter referred to as DOE IG.)

⁵³ GAO, *Government Auditing Standards* (July 2007 Revision). <http://www.gao.gov/govaud/govaudhtml/index.html>

⁵⁴ DOE IG, p. 1.

as high as \$4 million.”⁵⁵ The IG found two local government reporting issues: administrative costs and number of households weatherized.

Regarding the first concern, it identified instances where local agencies reported administrative-type expenses as program operating costs:

Specifically, we observed that certain organizations inappropriately charged expenses such as administrative staff, office rent, and administrative supplies as direct program costs and thus understated total administrative costs. Public laws and federal regulations limit the amount of weatherization grant funds that may be used for administrative purposes. If local agencies continue to under report administrative costs, states could ultimately exceed statutory limitations for administrative expenses over the life of the grant.⁵⁶

DOE advised the IG that it is often difficult for local agencies to operate within the limits for administrative costs. The IG concluded that the DOE Program Office should work more closely with states and local agencies to ensure that administrative costs are minimized and that the costs incurred are reported accurately and consistently.

Regarding the concern about local agency reporting on the number of households weatherized, the IG found that:

... data regarding the number of households weatherized was not reported on a consistent basis. For example, data reported by some states related strictly to the number of homes weatherized using Department funds. Other states, however, combined the results of weatherization efforts funded by the HHS LIHEAP with those completed with Departmental funds. Merging performance data for such states distorts program results and could make it appear program efficiencies and energy savings are greater than that actually achieved with available funding.⁵⁷

A recent OMB program assessment found that DOE had addressed the IG’s concern about distorted efficiency (benefit-cost) measures:

Average cost per home employed in the calculation of benefit/cost ratios now reflects total costs (including non-DOE funds) expended per unit in the states whose evaluations were used in the energy savings estimates. This substantially raises the per-unit investment from a DOE-only level making the benefit/cost ratio more conservative.⁵⁸

2003 OMB Assessment with PART Method

PART Background

The Performance Assessment Rating Tool (PART) was developed by the Office of Management and Budget (OMB) in 2002 as a key component for implementing GPRA and the President’s

⁵⁵ Ibid.

⁵⁶ Ibid. p. 2.

⁵⁷ Ibid. p. 2.

⁵⁸ Office of Management and Budget (OMB). *Detailed Information on the Weatherization Assistance Assessment*. Weatherization Assistance. Program Assessment Summary. 2003. (Hereinafter referred to as “OMB, *Detailed Assessment*.”) <http://www.whitehouse.gov/omb/expectmore/summary/10000128.2003.html>

Management Agenda (PMA),⁵⁹ particularly the Budget and Performance Integration initiative.⁶⁰ PART was designed to assess program planning, management, and performance against quantitative, outcome-oriented goals. It can inform funding and management decisions aimed at improving program effectiveness.⁶¹ As a diagnostic instrument for evaluating efficiency and effectiveness, PART aims to help managers identify, anticipate, and rectify performance problems.

PART was designed to provide a basis for DOE and OMB to agree upon meaningful annual and long-term targets. The PART assessment process aims to unify annual performance targets and goals with long-term goals. PART was designed to be an iterative process, capable of tracking the evolution of program performance over time through periodic reassessments. OMB's recommendations to foster program improvement are central to the PART process. Program offices track actions taken to implement PART recommendations and report those actions to OMB semi-annually.

According to OMB, the ongoing cycle of reviewing and implementing PART recommendations, coupled with the use of performance data from assessments and periodic reassessments, signify the perception of PART as an integral process for planning and budget decision-making, as distinguished from a set of one-time program evaluations.⁶²

PART assessments help inform budget decisions and identify actions to improve results. Agencies are held accountable for implementing PART follow-up actions and working toward continual improvements in performance.

PART asks a series of questions that cover four areas:

- program purpose and design,
- performance measurement, evaluation, and strategic planning,
- program management, and
- program results.

To earn a high PART rating, a program must use performance assessment to manage, justify its resource requests on expected (projected) performance, and continually improve efficiency. All of those activities are goals of the Budget and Performance Integration Initiative. The PART performance rating scale is shown below:

⁵⁹ This was an initiative of the George H. W. Bush Administration.

⁶⁰ DOE, *Performance and Accountability Report, Fiscal Year 2006*. (DOE/CF-0012) p. 11. <http://www.cfo.doe.gov/progliaison/2006finalpar2.pdf>

⁶¹ OMB, *What Constitutes Strong Evidence of a Program's Effectiveness?* 2004. http://www.whitehouse.gov/sites/default/files/omb/part/2004_program_eval.pdf

⁶² DOE, *Performance Report Highlights FY2006*, p. 11.

Table 2. PART Rating Scale

(the numerical rating represents the total point score on the PART questionnaire)

Rating	Range
Effective	85 - 100
Moderately Effective	70 - 84
Adequate	50 - 69
Ineffective	0 - 49

Source: National Institutes of Health. GPRA and PART.
<http://sourcebook.od.nih.gov/SD%20Orientation/PART.pdf>

PART Report on DOE Weatherization Program

The summary of a PART report on the DOE Weatherization Program found that it was performing at the “moderately effective” level.⁶³ The assessment defines the moderately effective rating as follows:

In general, a program rated Moderately Effective has set ambitious goals and is well-managed. Moderately Effective programs likely need to improve their efficiency or address other problems in the programs’ design or management in order to achieve better results.⁶⁴

The PART assessment cited three findings to explain the “moderately effective” rating. First, it noted that the program met its annual performance target for the number of homes weatherized. Second, OMB stated that the program lacked an assessment of performance that was current, comprehensive, and independent. It noted that the program reported on “internal assessments” that showed a favorable benefit-cost ratio. However, OMB found that those assessments relied in part on old data and were conducted by Oak Ridge National Lab (ORNL), which it says is not an “independent” source.⁶⁵ Third, OMB noted that the DOE Inspector General identified issues with the way state and local agencies report on program management.⁶⁶ DOE responded that it was conducting an independent analysis of the cost-effectiveness of the program and addressing the IG’s audit recommendations.

The detailed PART assessment report presented trend data for key performance measures, including number of homes weatherized, cost per home, and the overall benefit-cost ratio.⁶⁷ Also, it described and rated strategic program features, such as goals, objectives, design, targets, and budgeting.

The PART Assessment defined the benefit-cost ratio as a long-term measure of program efficiency.⁶⁸ OMB stated that the benefit-cost ratio represents the value of energy saved,⁶⁹ divided

⁶³ OMB, *Detailed Assessment*.

⁶⁴ Weatherization Assistance, Program Assessment Summary, Tab for “RATING: What This Rating Means.” <http://www.whitehouse.gov/omb/expectmore/rating.html>

⁶⁵ OMB described each evaluation performed by ORNL as an “internal program assessment,” because it was not conducted by an independent third party. The issue of independence is treated in more detail later, in another section.

⁶⁶ These issues were raised in the 2003 IG performance audit, which is reviewed in the previous section.

⁶⁷ OMB, *Detailed Assessment*.

⁶⁸ OMB cited figures that exclude values for non-energy benefits.

by program costs.⁷⁰ The ratio depended in part on estimated long-term energy prices at the time of the assessment and on an average energy savings per household of 29.1 million British thermal units (MBtu).⁷¹ PART reported that estimates of the benefit-cost ratio for energy savings under various price scenarios ranged from 1.19 to values greater than 2.00,⁷² never measuring less than 1.00.⁷³

PART assessment question number 4.3 asked:

Does the program demonstrate improved efficiencies or cost effectiveness in achieving program performance goals each year?

The response was:

Benefit-cost ratio rose from 1.06 in 1989 to 1.79 in 1996, and then declined to 1.51 and 1.30 in 1999 and 2002, respectively. These estimates depend largely on EIA estimated long-term energy prices.

PART question number 2.6, on strategic planning, asked:

Are independent and quality evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?⁷⁴

The response touched upon several aspects of the evaluation strategy, including frequency of studies, measurement of non-energy benefits, and independence of the evaluator. It stated:

The program does not conduct annual evaluations on a national basis because of the high cost of such evaluation and the limited amount of change that occurs in program activities from year to year. The program has contracted with Oak Ridge National Laboratory (ORNL) to devise evaluation methodologies and report periodically on program results based on state grantee-level performance evaluation. ORNL has also conducted selective evaluation activities designed to inform program management of performance characteristics in areas in which the program performance has been below average (hot climate zones) or in areas in which there has been growing strategic program interest but little evaluation data. The latter includes base load electric measures as well as non-energy benefits. To assure independence, the program should consider using an alternative contractor in future assessments, or at least having future ORNL reports assessed by a third party.⁷⁵

(...continued)

⁶⁹ The value of energy was discounted over future years at a rate of 3.2%.

⁷⁰ The program costs excluded some management costs.

⁷¹ The price data were obtained from DOE's Energy Information Administration (EIA).

⁷² The range for the benefit-cost ratio was tested for the 90% confidence range.

⁷³ ORNL benefit-cost estimates are discussed in the next section and are shown in **Appendix A**.

⁷⁴ OMB, *Weatherization Assessment*, Question # 2.6.

⁷⁵ Peer reviewers of ORNL weatherization reports are generally employees of ORNL or DOE. Ibid.

As evidence of its evaluation activities, DOE's response included references to several ORNL program metaevaluations and other evaluation studies.⁷⁶ Some key ORNL studies are described in the next section.

2011 GAO Performance Audit of Recovery Act Funding

The Government Accountability Office (GAO) defines a performance audit broadly:

Performance audits provide objective analysis so that management and those charged with governance and oversight can use the information to improve program performance and operations, reduce costs, facilitate decision making by parties with responsibility to oversee or initiate corrective action, and contribute to public accountability.⁷⁷

According to GAO, a performance audit may embrace a wide variety of objectives, including:

- an analysis of the “relative cost-effectiveness of a program or activity,” and/or
- a determination of the “current status or condition of program operations or progress in implementing legislative requirements.”⁷⁸

Thus, by GAO's definition, a performance audit may include activities with the same purposes as those of an impact and/or process evaluation study.

The Recovery Act directed GAO to conduct bimonthly reviews and reporting on selected states and localities' use of the funds it provided, including the funds for the DOE Weatherization Program.⁷⁹ GAO describes these bimonthly review as performance audits. In addition to the bimonthly audits, GAO's response to the directive has included one performance audit that covered only the Weatherization Program.

For example, in December 2009, GAO published a performance audit report on Recovery Act funds used by states and localities.⁸⁰ The DOE Weatherization Program was one of many programs covered. For the weatherization component, GAO covered 16 states and 24 localities. GAO found that, as of the end of FY2009, states had outlaid about 2% (\$113 million) of weatherization funds and had completed about 1% (7,300) of the targeted number of homes to be

⁷⁶ The cited evidence was: State-level Evaluation of WAP 1990-1996: A Meta Evaluation of 17 States Evaluations - 1997 (ORNL/CON-435); Metaevaluation of National Weatherization Assistance program Based on State Studies, 1996-1998 (ORNL/CON-467); Metaevaluation of National Weatherization Assistance Program Based on State Studies, 1993-2002 (ORNL/CO-488); Nonenergy Benefits from the Weatherization Assistance Program: A Summary of Findings from the Recent Literature, 2002 (ORNL/CON-484).

⁷⁷ GAO, *Government Auditing Standards* (Definition of Performance Audits in Chapter 1: Use and Application of GAGAS), July 2007 Revision (GAO-07-731G. <http://www.gao.gov/govaud/govaudhtml/d07731g-3.html>

⁷⁸ Ibid.

⁷⁹ The American Reinvestment and Recovery Act of 2009 (P.L. 111-5), Sections 901 and 902, H.Rept. 111-16, p. 77-78 and 463.

⁸⁰ GAO reported that the audit was conducted from September 18 through December 4, 2009. The review covered about two-thirds of federal assistance funds and about 65% of the national population. GAO, *Recovery Act: Status of States' and Localities' Use of Funds and Efforts to Ensure Accountability*, (GAO-10-231) December 2009, <http://www.recovery.gov/accountability/documents/d10231.pdf>.

weatherized.⁸¹ Many contracts between states and local agencies had been delayed due to concerns about staff capacity and new labor requirements.⁸²

As another example, in May 2010, GAO published another audit report on Recovery Act funds that covered DOE Weatherization Program funds.⁸³ As of the mid-point in FY2010, states had outlaid about 13% (\$659 million) of weatherization funds and had completed about 13% of the targeted number of homes to be weatherized. GAO made several observations about the status of program implementation and offered some recommendations for program improvement.⁸⁴

In January 2011, the GAO initiated a performance audit focused solely on Recovery Act funding for the DOE Weatherization Program.⁸⁵ The report was released in late December 2011.⁸⁶ The study aimed to examine: (1) status of funds, (2) implementation challenges facing recipients, (3) achievement of energy and cost savings, and (4) changes in the quality of employment data reported by recipients. To address the four objectives, GAO conducted a web survey of all 58 recipients and received 55 responses.⁸⁷ GAO also interviewed officials from DOE and selected national associations, conducted site visits in seven states, and conducted telephone interviews with two other state agencies.⁸⁸ Selected findings, as of the end of FY2011, included: (1) recipients had spent about \$3.46 billion (73%) of the \$4.75 billion allocated,⁸⁹ (2) recipients reported that implementation challenges were declining,⁹⁰ (3) the average cost per home was about \$4,900,⁹¹ about 563,000 homes had been weatherized, and DOE projected it would exceed the original target of 607,000 homes by the deadline set for the end of March 2012;⁹² and (4) GAO found that the quality of employed data reported by recipients had improved.⁹³

⁸¹ Ibid. p. 3-4.

⁸² Ibid. p. 91-93.

⁸³ For this phase, GAO collected updated state interview data from nine states and the District of Columbia, interviewed 31 local service providers, and conducted several site visits. GAO, *Recovery Act: States' and Localities' Uses of Funds and Actions Needed to Address Implementation Challenges and Bolster Accountability* (GAO-10-604) May 2010, <http://www.gao.gov/new.items/d10604.pdf>.

⁸⁴ GAO recommended that DOE clarify program guidance to address issues involving income eligibility, training standards, state monitoring, multi-family units, local agency internal controls, and methods for assessing costs and cost-effectiveness. Ibid. p. 103-125.

⁸⁵ Personal communication with Arvin Wu at GAO. January 2011.

⁸⁶ GAO, *Recovery Act: Progress and Challenges in Spending Weatherization Funds*, (GAO-12-195) December 2011, <http://www.gao.gov/products/GAO-12-195>.

⁸⁷ The 58 recipients include the 50 states, seven territories, and District of Columbia.

⁸⁸ Ibid. p. 2-4.

⁸⁹ Ibid. p. 9.

⁹⁰ Ibid. p. 19-22.

⁹¹ Ibid. p. 17 and 31-34. GAO does not make an independent estimate of cost-effectiveness. Instead, it refers to an ORNL Technical Memorandum issued March 2010 (reviewed in a later section of this report, on p. 26) that provided a preliminary estimate of cost-effectiveness, including a benefit-cost ratio of 1.8 for energy savings, rising to 2.5 with non-energy benefits included.

⁹² Ibid. p. 10 and 14.

⁹³ Ibid. p. 35-36.

Selected Evaluation Studies

Several DOE reports, prepared primarily by staff of the Oak Ridge National Laboratory (ORNL), have assessed the energy cost savings and non-energy benefits of the DOE Weatherization Program.⁹⁴ In particular, a comprehensive evaluation of the PY1989 program was published in 1993.⁹⁵

1993 National Evaluation Report

In 1993, DOE issued a report entitled *National Impacts of the Weatherization Assistance Program in Single Family and Small Multifamily Dwellings*.⁹⁶ The reporting project was an exhaustive five-year, first-of-its-kind, comprehensive evaluation study of the energy savings and cost-effectiveness of the program during PY1989.⁹⁷ A letter accompanying the report noted that the program is “the nation’s largest residential conservation program, and one of its oldest.” The letter captures the highlights:

The report evaluates the national and regional energy savings and cost effectiveness of the program in single-family dwellings, mobile homes, and small (2-unit to 4-unit) multifamily dwellings. It is based upon a representative national sample of nearly 15,000 dwellings weatherized by 368 local weatherization agencies, and a control group of homes waiting to be weatherized by the same agencies.⁹⁸

The report estimated an overall program benefit-cost ratio of 1.09 for the entire program.⁹⁹ Adding estimates of non-energy benefits—such as comfort, employment, reduced environmental impacts, and preservation of affordable housing—yielded a societal benefit-cost ratio of 1.72.

Three Metaevaluations (1997, 1999, 2005)

After the 1993 National Evaluation report, DOE relied upon “metaevaluations”¹⁰⁰ of state evaluation studies as the main tool to assess program energy savings and cost effectiveness. All of

⁹⁴ Many of ORNL’s evaluation reports on the Weatherization Program are available at <http://weatherization.ornl.gov/publications.shtml>.

⁹⁵ Subsequently, ORNL conducted four “metaevaluations” of the Program’s energy savings, which used studies conducted by individual states between the years 1990–1996, 1996–1998, 1993–2002, and 1993–2005. For more details about those studies, see: ORNL, *National Evaluation of the Weatherization Assistance Program: Preliminary Evaluation Plan for Program Year 2006*, February 2007, p. 1. http://weatherization.ornl.gov/pdfs/ORNL_CON-498.pdf

⁹⁶ ORNL managed the five-part study, which was based mainly on data from PY1989, supplemented with data from 1991 and 1992. The 400-page report is available at http://weatherization.ornl.gov/pdfs/ORNL_CON-326.pdf.

⁹⁷ Home Energy. *Evaluating DOE’s Weatherization Assistance Program*. July/August 2010. At that time, the program weatherized about 250,000 dwellings per year through approximately 1,110 local weatherization agencies. The evaluation consisted of five parts: a network study, a resources and population study, a multifamily study, a single-family study, and a fuel oil study. The average savings for all single-family and small multifamily dwellings in 1989 was estimated to be 17.6 million Btu per year, 18.2% of the energy used for space heating and 13.5% of total energy use. It was estimated that over a 20-year lifetime, the program would save the equivalent of 12 million barrels of oil. http://www.homeenergy.org/article_full.php?id=725

⁹⁸ ORNL, *Dear Colleague Letter from Marilyn A. Brown* [To accompany National Weatherization Evaluation Report], October 1993.

⁹⁹ More recent DOE reports appear to suggest that the benefit-cost figure was later revised to a value of 1.06.

¹⁰⁰ A metaevaluation involves synthesizing results from a number of individual studies of state weatherization efforts in order to estimate total national impacts.

the metaevaluations were conducted by ORNL. Compared with the 1993 National Evaluation, metaevaluations became a low-cost way for DOE to update energy savings and cost-effectiveness. This section reviews three key metaevaluations, covering program operations from 1990 through 2005.

1997 Metaevaluation (Covers 1990-1996)

DOE asked ORNL to prepare the 1997 metaevaluation.¹⁰¹ The study involved locating, assembling, and summarizing the results of all state-level evaluations that had become available since 1990. ORNL found 17 state-level evaluation studies from 15 states that it used to cover program operations during the period from 1990 through 1995.¹⁰² Energy savings were expected to be higher than those found in the 1993 National Evaluation, which was based on 1989 data. Typical savings ranged from 18% to 24%, which was greater than the range from 12% to 16% found in the 1993 National Evaluation.¹⁰³ Key reasons for the expected improvement included some advances in weatherization procedures, such as the use of advanced audits and blower-door directed air sealing.¹⁰⁴ The study concluded that program performance had improved significantly over that seven-year period.¹⁰⁵

The 1997 metaevaluation study used three perspectives to estimate cost-effectiveness. The program perspective compared energy benefits to total program costs. The installation perspective compared energy benefits to installation costs. The societal perspective compared the summed value of energy and non-energy benefits to total costs.¹⁰⁶ ORNL reported that the metaevaluation calculations employed the same procedures and assumptions used in the 1993 National Evaluation. The values of 1.79 (program perspective), 2.39 (installation perspective), and 2.40 (societal perspective), respectively, are shown in **Table 3**.¹⁰⁷

The study found that the synthesis of state-level evaluations offered a reasonable characterization of national program performance:

Although national level evaluation efforts are sometimes needed to definitively demonstrate program performance, reviews of state-level evaluations provide useful, and inexpensive, benchmarks of progress during the years between such large-scale national assessments.¹⁰⁸

¹⁰¹ ORNL, *State-Level Evaluations of the Weatherization Assistance Program in 1990-1996: A Metaevaluation that Estimates National Savings*, (ORNL/CON-435) January 1997. (http://weatherization.ornl.gov/pdfs/ORNL_CON-435.pdf)

¹⁰² The 15 states included 9 with published results—Colorado, Indiana, Iowa, New York, North Carolina, North Dakota, Ohio (2 studies), Texas, and Vermont (2 studies)—and 6 with unpublished results, including Delaware, Kansas, Minnesota, Nebraska, Wisconsin, and Wyoming.

¹⁰³ *Ibid.* p. xii.

¹⁰⁴ *Ibid.* p. xi.

¹⁰⁵ *Ibid.* p. 11. Three types of evidence supported the finding that savings were increasing: a literature review, within-state comparisons of savings over time, and regression results.

¹⁰⁶ *Ibid.* p. 16.

¹⁰⁷ *Ibid.* p. 17.

¹⁰⁸ *Ibid.* p. 18.

1999 Metaevaluation (Covers 1996-1998)

As a follow-up to the 1996 metaevaluation, ORNL performed another metaevaluation, which was published in 1999. The report synthesized results from 10 individual evaluation studies of state weatherization efforts that were completed between April 1996 and September 1998.¹⁰⁹

The stated objectives of the 1999 study were (1) to identify average energy savings of households in the states that provided information; (2) to identify key variables that explain the magnitude of those weatherization-induced savings; and (3) to use the findings from the state studies to estimate average household energy savings that could be expected nationwide.¹¹⁰

The study reported that program-induced savings continued at a higher rate than those found in the 1993 national evaluation.¹¹¹ However, the study found no dramatic changes were made in the structure or practices of the program during the period after the 1996 study.¹¹² Thus, no major increases in savings were expected above those found in the 1996 study.

To make the benefit-cost ratios in the 1999 study comparable to those in the 1996 metaevaluation and the 1993 national evaluation, ORNL used the same assumptions and procedures. In particular, the average measure lifetime was assumed to be 20 years and the discount rate used was 4.7%. Cost-effectiveness was calculated for the weatherization program nationwide. As in past evaluations, three different benefit-cost perspectives (program, installation, and societal) were examined.

Because higher average national energy savings were estimated in the 1996 and 1999 metaevaluations, the benefit-cost ratios for those years were higher than the ones reported in the 1993 National Evaluation. The benefit-cost ratios for the three perspectives are shown in **Table 3**.

¹⁰⁹ ORNL, *Metaevaluation of National Weatherization Assistance Program Based on State Studies, 1996-1998*, (ORNL/CON-467) 1999. ORNL's review included studies of Colorado, Delaware, District of Columbia, Indiana, Iowa (2), Minnesota (2), Ohio, and Vermont. Of the 10 state studies, 3 used control groups and 7 did not. pp. 4-5.

¹¹⁰ ORNL, *Metaevaluation 1996-1998*, p. x. The key variable(s) associated with energy savings were identified by running a regression analysis using energy savings as the dependent variable and a number of potentially related factors as independent variables. Estimates of average household energy savings nationwide were accomplished by taking the regression equation from the model with the best predictive ability and inserting the average national values for the independent variable(s).

¹¹¹ *Ibid.* p. 14.

¹¹² ORNL, *Metaevaluation 1996-1998*, p. 1.

Table 3. Benefit-Cost Ratios for the 1993 National Evaluation, Three Metaevaluations (1997, 1999, and 2005), and the 2010 Technical Memorandum
(Three benefit-cost perspectives shown)

Study	Program Perspective	Installation Perspective	Societal Perspective
National Evaluation (1989)	1.06	1.58	1.61
Metaevaluation (1990-1996)	1.79	2.39	2.40
Metaevaluation (1996-1998)	1.51	2.02	2.12
Metaevaluation (1993-2005)	1.34	—	2.53
Technical Memorandum	1.80	---	2.51

Sources: ORNL/CON-326; ORNL/CON-435; ORNL/CON-467; RNL/CON-493; ORNL/TM-2010/66.

2005 Metaevaluation (Covers 1993-2005)

The most recent metaevaluation, published in 2005, estimated average household energy savings at 30.5 million British thermal units (MBtu) per year with a benefit/cost ratio from the program perspective estimated at 1.34 with 2003 prices. Costs were based on those derived from the state evaluations themselves.¹¹³

Like the previous metaevaluations, ORNL undertook this one to update the findings from the 1993 national evaluation.¹¹⁴ This metaevaluation was based on data from 38 evaluation studies of weatherization programs in 19 states,¹¹⁵ published between 1993 and 2005.¹¹⁶

The overall method of the 2005 metaevaluation paralleled that of the previous metaevaluations, but ORNL describes some improvements:

1. Expanded Range of Time and Geography. The overall similarity of results across the various studies led ORNL to include findings from a 2003 study that covered all post-1992 state-level studies. ORNL found that this approach increased sample size, improved the ability to cover all major U.S. climate regions, and added to the statistical rigor of the results.
2. States as the Unit of Analysis. ORNL aggregated data from the 38 studies by state. This aspect contrasts with previous metaevaluations, which had treated multiple studies of the same state as separate empirical observations.¹¹⁷

¹¹³ Personal communication with Robert Adams, Supervisor of the DOE Weatherization Program, January 30, 2011. The 2005 Metaevaluation is at http://weatherization.ornl.gov/pdfs/ORNL_CON-493.pdf.

¹¹⁴ ORNL, *Estimating the National Effects of the U.S. Department of Energy's Weatherization Assistance Program With State-Level Data: A Metaevaluation Using Studies From 1993 to 2005*, (ORNL/CON-493), September 2005. http://weatherization.ornl.gov/pdfs/ORNL_CON-493.pdf

¹¹⁵ The 19 states were: Colorado, Delaware, District of Columbia, Georgia, Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, New York, North Carolina, Ohio, Texas, Vermont, Washington, West Virginia, Wisconsin, and Wyoming.

¹¹⁶ ORNL, 2005, p. xi and 4.

¹¹⁷ Ibid. p. 3. This metaevaluation included new data on measured energy savings from six states: Delaware, District of Columbia, Iowa, Illinois, Texas, and Wisconsin.

3. Studies Combined into a Single State Value. For each relevant variable (e.g., energy savings per gas-heated household), a single average value was used to represent the entire state. ORNL says this procedure prevents overall study results from being skewed due to the presence of multiple studies with similar results in a single state.
4. Weighted Average Used to Account for Multiple Studies. Where there were multiple studies, the average values computed in each study were used to determine a weighted average for each variable, with the weighting done by sample size.¹¹⁸

Under this revised approach, those states that represent a larger part of the national program contribute more heavily to the analysis and findings. This is appropriate because the purpose of the study is to estimate energy savings nationwide.¹¹⁹

Table 3 shows the benefit-cost values in the various ORNL metaevaluations. The similarity among the values is not surprising because those meta analyses used data from many of the same state-level studies.¹²⁰

ORNL states that all three of the metaevaluations, including the 2005 study, focused primarily on energy savings in homes heated by natural gas because the large majority of state-level studies addressed that fuel.¹²¹ ORNL found a statistically significant difference between the natural gas savings per household in the 2005 metaevaluation and the savings reported in the 1993 national evaluation.¹²² However, the 2005 study noted that the other two-thirds of the states were not included in the metaevaluation, and thus did not provide data for the study, which raises the possibility that the sample examined may not fully represent the nation as a whole.¹²³

ORNL found that actual benefit-cost values are likely to be higher than those reported for natural gas savings, as shown in **Table 3**. This was due to the fact that reported expenditures also included the cost of installing measures to reduce baseload electricity use, but only the benefits of natural gas savings were used in those benefit-cost calculations.

Overall, the study found significant program improvements since 1989 (as reported in the 1993 national evaluation) and a generally higher benefit-cost ratio. However, it also acknowledged the shortcomings of the metaevaluation method and called for a new national evaluation:

While the metaevaluations performed over the last decade have consistently shown higher natural gas savings per household than those reported in the national evaluation, there is a need to corroborate those findings through a rigorous examination of Weatherization Program efforts nationwide. Even the current metaevaluation is based on studies performed in only a third of the states, and those may not be fully representative of the entire Weatherization Program. Also, the value for preweatherization energy consumption, which is a major input for the calculation of national savings, is based on 1989 data. In addition, while

¹¹⁸ Ibid. p. 3-4.

¹¹⁹ Ibid. p. xi.

¹²⁰ Ibid. p. xii.

¹²¹ Within the sample of 19 states, studies in 17 states addressed savings in dwellings heated by natural gas.

¹²² Ibid. p. 15.

¹²³ Ibid. p. 5.

state-level evaluations have put a strong emphasis on gas-heated houses, few studies have been conducted on electrically heated dwellings. And it is important to note that the biggest recent change to the Weatherization Program – the addition of baseload measures such as highly efficient refrigerators, water heaters, and light bulbs – has barely been addressed by state-level studies. For all these reasons, there is a strong need for a new national evaluation to thoroughly explore the current operations and achievements of the Weatherization Program across the entire nation.¹²⁴

DOE has engaged APPRISE, a nonprofit analytical organization, to conduct the second national evaluation. The status of plans for that evaluation are described in the section below, entitled “Evaluation Plan for the Regular Program.”

2010 Technical Memorandum (FY2010 Projected Data)

The Recovery Act provided a large infusion of funding and prescribed major new requirements for the program. In response to concerns about these major changes, ORNL prepared a technical memorandum to update estimates from the metaevaluations.¹²⁵ The memo only attempted to make preliminary projections based on an audit; it did not attempt to generate empirical data from an impact evaluation study. Specifically, the memo recounted results from an engineering modeling approach that used the National Energy Audit Tool (NEAT) to estimate the performance of the program by using typical homes in each state.¹²⁶ These homes varied by building type, primary heating fuel and prices, and local weather conditions.

ORNL employed the Weatherization Assistant residential audit package to estimate annual energy savings under new parameters set by the Recovery Act.¹²⁷ The estimated annual savings for heating and cooling projected for FY2010 were 29.0 MBtu per household.¹²⁸ In comparison, the 2005 metaevaluation estimated savings at 30.5 MBtu over the period from 1993 through early 2005.¹²⁹

The Recovery Act raised the average per-unit investment cost ceiling from \$2,500 to \$6,500. The likely increase in investment per unit was expected to increase total energy savings. However, the higher funding level was accompanied by a shift in formula allocation that allowed proportionally more funding for states in warmer climates.¹³⁰

ORNL projected a 2010 benefit-cost ratio of 1.80 (from the program perspective). Most of the increase relative to the 2005 benefit-cost ratio of 1.54 was attributed to the change in energy prices from 2003 to 2010. Costs were based on the Recovery Act ceiling of \$6,500 for average costs. However, the ceiling for cost-effective expenditures varied from state to state, based on the

¹²⁴ Ibid. p. xiv.

¹²⁵ ORNL, *Weatherization Assistance Program Technical Memorandum Background Data and Statistics*, (ORNL/TM-2010/66), Prepared by Joel Eisenberg, March 2010, p. 1. http://weatherization.ornl.gov/pdfs/ORNL_TM-2010-66.pdf

¹²⁶ NEAT was developed by ORNL. It is part of “The Weatherization Assistant,” a family of computer audit software programs that DOE designed to help state and local agencies implement the program. The Weatherization Assistant serves as the umbrella program for NEAT and for DOE’s Manufactured Home Energy Audit (MHEA). *The Weatherization Assistant Features*, http://weatherization.ornl.gov/assistant_features.shtml.

¹²⁷ Ibid. p. 7.

¹²⁸ ORNL, *Technical Memorandum*, p. 7.

¹²⁹ ORNL, *Metaevaluation 2005*.

¹³⁰ ORNL, *Technical Memorandum*, p. 7.

requirement that the savings-to-investment ratio (SIR)¹³¹ reach one (1.0) or greater for each of three factors: the last measure installed, energy intensity, and local energy prices.¹³²

Traditionally, the program benefit-cost ratio has been expressed in terms of estimated national energy cost savings per dollar of investment for homes heated with natural gas, with cost savings based on national average residential natural gas prices. ORNL's estimates in the memo differ somewhat, due to reliance on state-by-state energy savings and investment numbers used to generate the energy savings estimates. This aspect allowed the dollar savings estimates to vary by region, by fuel type, by housing type, and with the adjusted average cost ceiling provided under the Recovery Act. The projected 2010 societal benefit/cost ratio (all benefits divided by all costs) is 2.51. This is almost identical to the societal benefit/cost ratio of 2.53 computed in the 2005 metaevaluation.¹³³

ONRL reflected on the limits of the data used to make estimates, noting again the need for a new national evaluation:

The methods used here are more complex and hopefully better at reflecting current reality than those used in previous years. Nonetheless, one needs to keep in mind the limitations of available input data that constrain an analysis such as this. There is no nationally available data on energy-related housing characteristics of weatherized households, nor is there data on the variations in the cost of measures installed from one locale to another. There is also no way of knowing at this point the exact characteristics of the housing stock and housing types that will actually get weatherized with the greatly expanded revenues available. Making these estimates more precise requires populating the methodology with more accurate data that will flow from the National Evaluation effort.¹³⁴

ONRL noted that the memo's estimates were derived from a different methodology than that used in the metaevaluations. There are several reasons that a change in method was needed. First, the metaevaluation estimates are dated and do not reflect recent changes in program operations that materially impact household savings. These include a major change in the program's average cost ceiling, from \$2,500 to \$6,500, an expansion in allowable measures to include electricity measures such as refrigerator replacement and lighting changeout, and a major increase in program funding impacting the allocation of resources among different regions and climate zones.

Second, the 2005 metaevaluation results describe only homes heated with natural gas and do not reflect the diversity of heating fuels used in treated homes nor do the results reflect potential cooling savings. The method used for the metaevaluations reflected only savings in single-family homes and those savings were never adjusted for variations in the treated housing stock. Finally, the estimates in the metaevaluations were based on national average energy prices and were not

¹³¹ DOE, WAPTAC, *Simple Savings to Investment Ratio (SIR) Comparison*. Actual SIR calculations supported by NEAT, MHEA, and other approved audit tools account for the present value (PV) of money and fuel escalation rates over the lifetime of the measures to arrive at more accurate savings numbers. DOE describes the SIR in a worksheet available online at http://www.google.com/search?rlz=1C1CHFX_enUS389US389&aq=f&sourceid=chrome&ie=UTF-8&q=savings+investment+ratio.

¹³² ORNL, *Technical Memorandum*, p. 7.

¹³³ *Ibid.* p. 9

¹³⁴ *Ibid.* p. 9.

varied to reflect the diversity of energy prices weighted by the location of the weatherization work being performed.¹³⁵

ONRL states that the methodology used to prepare statistics for the memo corrects for many of the above-noted deficiencies. However, it acknowledges that the findings do not present a statistically valid representation of the program's performance and results:

There are too many assumptions and uncertainties incorporated in it to allow that to be the case. Much of this is the result of a lack of up-to-date information on program operations, particularly regarding measures installed and their cost as well as the energy-related characteristics of the homes weatherized. Rather, the results should be treated as the best currently available estimate that can serve until more rigorous results are provided by the new National Evaluation.¹³⁶

Current Plans for Assessment and Evaluation Studies

Three activities are underway to assess and evaluate DOE's Weatherization program: a DOE (Deloitte) strategic assessment, a DOE (APPRISE) impact study for PY2007 through PY2008, and a DOE (APPRISE) impact evaluation for the Recovery Act period of PY2009 through PY2011. **Table 4** provides a brief overview of the goals and target dates for the three activities. Descriptions of the activities follow.

Plan for a Strategic Assessment (Deloitte)

In November 2008, DOE's Weatherization Assistance Program Office announced that it had contracted with Deloitte Limited Liability Partnership (LLP) to perform a strategic assessment of the program.¹³⁷ Deloitte is tasked with conducting a fundamental analysis of the program's objectives, impact metrics, market delivery vehicles, and finance mechanisms. The study aims to identify fundamental improvements in program design and delivery, in contrast to the more traditional evaluation of program benefits and costs that ORNL has conducted in the past.¹³⁸

¹³⁵ Ibid. p.13.

¹³⁶ Ibid. p. 13.

¹³⁷ Deloitte Touche Tohmatsu International is a major, international accounting and auditing firm. Deloitte LLP is the U.S. subsidiary of the international firm.

¹³⁸ DOE, *Weatherization Program Notice 09-1*, November 17, 2008.

Table 4. DOE Assessment and Evaluation Activities Underway
(As of December 2011)

	Strategic Assessment	Second National Retrospective Evaluation	Recovery Act Retrospective Evaluation
Organization	Deloitte	ORNL-APPRISE	ORNL-APPRISE
Process Evaluation	yes	yes	yes
Impact Evaluation	no	yes	yes
Period Covered	—	2007 through 2008	2009 through 2011
Focus	program design and measurement tools	comprehensive energy and non-energy impacts, benefit-cost	comprehensive energy and non-energy impacts, benefit-cost
Target Date for Completion/Report	No date published	Fall 2012	Early 2014

Source: DOE, various references.

Notes: The only published target dates for completion were obtained from ORNL, *Evaluation of the National Weatherization Assistance Program during Program Years 2009-2011 (American Reinvestment and Recovery Act Period)*, ORNL/TM-2011/87, May 20, 2011 (draft date), pp. 1 and 123.

Evaluation Plan for Regular Program (APPRISE, 2007-2008 Data)

DOE acknowledges that OMB's PART review and the IG's report have highlighted the need for an updated comprehensive evaluation of the program to ensure that objectives are being met and that estimates of energy savings, bill reductions, program costs, and program benefits are valid. Since the first national evaluation was published in 1993, there have been several changes made to program policies, procedures, technologies, and techniques. DOE notes that it is important to assess how well these changes have worked. For example, there is a need to assess how well new policies—such as whole-house weatherization—are impacting the program.¹³⁹

In response to the growing need for updated and valid program data, the program moved forward in FY2005 to begin planning another national evaluation. With a committee comprised of state and local weatherization staff, the program office started planning the evaluation and identifying areas where the program might be improved. DOE states that the 1993 national evaluation brought about major program changes such as computerized audits, less use of door and window measures, and increased emphasis on hot climate opportunities. DOE anticipates that the second national evaluation will have a similar effect on future program policy and performance.¹⁴⁰

In February 2007, ORNL released a plan for a second national retrospective evaluation, which would use data collected during PY2006.¹⁴¹ ORNL says that the program has been changed by

¹³⁹ DOE, *Weatherization Program Briefing Book*. National Evaluation. 2006.
<http://www.nyswda.org/DirManual/Documents/06DOEBriefingBook/Issues.pdf>

¹⁴⁰ Ibid.

¹⁴¹ ORNL, *National Evaluation of the Weatherization Assistance Program*, 2007. p. 1. The report says that DOE undertook the new evaluation because the program is now “vastly different” from the PY1989 program that was examined in the 1993 national evaluation.

findings from the 1993 study, a strategic planning process, and other initiatives that had caused the program to change significantly. The program has incorporated new funding sources, management principles, audit procedures, and energy efficiency measures.¹⁴²

ORNL is supervising a competitively selected independent contractor team to conduct the evaluation. The team is led by the Applied Public Policy Research Institute for Study and Evaluation (APPRISE).¹⁴³ The study will assess program performance for PY2007 and PY2008, the period immediately preceding the Recovery Act. Performance measures will include program costs and benefits. In order to conduct a comprehensive savings analysis, the design includes collection of a full year of post-weatherization billing data.¹⁴⁴ The evaluation is expected to be completed by Fall 2012.¹⁴⁵

Design for the APPRISE Second National Retrospective (Impact) Evaluation

DOE's central evaluation question is: How much energy did the weatherization program save in 2007 and 2008? DOE plans to use well-known analytical approaches to answer this question. Electricity and natural gas billing histories will be collected pre- and post-weatherization for a sample of weatherized homes (the treatment group) and a sample of comparable homes that were not weatherized (the comparison group).¹⁴⁶ The comparison group for the 2007 treatment group will be homes weatherized in 2008. The comparison group for the 2008 treatment group will be homes weatherized in 2009. A national sample of 400 local weatherization agencies will be selected to provide information on about one-third of their weatherized homes. Billing history data will be normalized using three different analytic methods, including the Princeton Scorekeeping Method (PRISM).¹⁴⁷

DOE will also address an important complementary evaluation question: How cost-effective were those energy savings? In principle, all measures should be cost-effective, because each one is required to meet the savings-investment ratio (SIR) test. Evaluations are conducted to compare expectations with measured values. DOE says it will collect cost information associated with each weatherized home included in the treatment sample. Total projected energy cost savings over the lifetime of the measures will be divided by total costs to estimate a benefit-cost ratio for the program.

¹⁴² Ibid.

¹⁴³ Home Energy, *Evaluating DOE's Weatherization Assistance Program*, June 30, 2010. APPRISE is a nonprofit energy research company. Key team members include the Energy Center of Wisconsin, Michael Blasnik & Associates, and Dalhoff Associates, LLC.

¹⁴⁴ As of early 2011, there was an indication that the study might produce some early results for PY2007 and PY2008. As of December 2011, such results had not been published. There was also an early 2011 expectation that the study would include some PY 2009 and PY 2010 results, which would become available in following years. Personal communication with Robert Adams, DOE Weatherization Program Office, January 2011.

¹⁴⁵ ORNL. *Evaluation of the National Weatherization Assistance Program during Program Years 2009-2011 (American Reinvestment and Recovery Act Period)*, ORNL/TM-2011/87, May 20, 2011 (draft date), pp. 1 and 123.

¹⁴⁶ Home Energy, *Evaluating DOE's Weatherization Assistance Program*. DOE reports that it is difficult to measure the impact of energy efficiency programs on homes that use bulk fuels, such as propane and fuel oil. This is because residents use different suppliers and sometimes use additional fuels, and because they do not always fill up the tank completely. To meet this challenge, the evaluation team will install submeters to measure energy savings in homes heated with bulk fuels.

¹⁴⁷ Ibid.

The study will attempt to quantify nonenergy benefits, which will be sorted into three categories—utility benefits, occupant benefits, and societal benefits. Utility providers benefit because weatherization reduces arrearages and service shutoffs. Occupants benefit because weatherization makes homes more comfortable, healthier, safer, and more valuable. Society benefits because weatherization curbs greenhouse gas emissions, reduces other forms of pollution, and conserves water. Further, it increases local employment and it increases local economic activity caused by the multiplier effect.¹⁴⁸

Design for APPRISE Process Evaluation

DOE intends that the Second National Retrospective Evaluation also examine program processes. To do that, it plans to survey all program grantees and subgrantees to collect data about program operations, training and client education activities, and quality assurance procedures. That survey data will be used to provide a program snapshot for PY2008. Case studies and a field study will be incorporated into the process evaluation design to furnish additional insights on program implementation.¹⁴⁹

Evaluation Plan for Recovery Act (APPRISE, 2009-2011 Data)

The Recovery Act committed \$5 billion over two years to an expanded weatherization program. This large sum greatly heightened interest in the program, the population served, its energy and cost savings, and its cost-effectiveness. Solid answers to many of the questions about the program and its performance require a comprehensive evaluation. DOE is funding an evaluation by APPRISE¹⁵⁰ in order to ascertain program effectiveness and to improve program performance.¹⁵¹ The evaluation project has a budget of \$19 million. GAO reviewed the plan's design for evaluating energy cost-effectiveness and found it methodically sound.¹⁵²

DOE observes that the weatherization program has changed considerably. In particular, it notes that there are four key differences between the pre-Recovery Act program and the post-Recovery Act program.¹⁵³ First, to expend \$5 billion and to increase weatherization activity by 300%, a greatly expanded weatherization workforce was recruited, trained, organized, and sent into the field. To support that expansion, the percentage of program funds allowed for training and technical assistance was raised from 10% to 20%. Second, all states and U.S. territories received large funding increases, and some states and local agencies struggled with weatherization budgets several times larger than previous ones. Faced with that major program expansion, many states and local agencies chose to implement innovations in program delivery and management. Third, DOE also has made provisions to set aside substantial sums to support innovations in program funding and design.

Fourth, to accommodate the Recovery Act funding expansion, major changes were incorporated into the program:

¹⁴⁸ Ibid.

¹⁴⁹ Ibid.

¹⁵⁰ ORNL is developing an evaluation plan for the Recovery Act funding period that it is coordinating with APPRISE.

¹⁵¹ ORNL, *Technical Memorandum*, p. 1. Partial results were expected to emerge as early as late 2011.

¹⁵² GAO, *Performance Audit for Recovery Act Spending on DOE Weatherization Program*, p. 34.

¹⁵³ Home Energy, *Evaluating DOE's Weatherization Assistance Program*.

- The definition of the threshold for a low-income household was raised from 150% to 200% of the federal poverty income guidelines;
- The ceiling on average cost per unit—that is, the average amount of money that grantees can spend to weatherize their homes—was increased from \$2,500 to \$6,500; and
- Wages for weatherization workers were subjected to Davis-Bacon Act prevailing wage requirements for the first time.

DOE/ORNL says that many features of the 2009-2011 evaluation will parallel those of the retrospective evaluation for 2007-2008, as described in the previous section. Analyses of energy savings and cost-effectiveness will be similar, although billing histories may be collected to also assess the effects of certain program innovations.

The main difference between the retrospective evaluation of 2007-2008 and the evaluation of 2009-2011 will be the process evaluation component. DOE says that funding and associated provisions have had a significant impact on the weatherization community and program operations. Thus, it plans two special process evaluation studies as part of the Recovery Act evaluation project. One would focus on Davis-Bacon requirements and the other would address changes in the composition of the national weatherization community.

In May 2011, ORNL published a plan for the Recovery Act evaluation that includes details about both the impact and the process components of the assessment.¹⁵⁴ The schedule for the evaluation activities covers the period from the second quarter of calendar year (CY) 2011 through the first quarter of CY2014.¹⁵⁵

Impacts of Davis Bacon Requirement

Congress directed that weatherization projects funded with Recovery Act appropriations must follow Davis-Bacon labor rules.¹⁵⁶ The Department of Labor is responsible for identifying prevailing wages in the construction industry. These wages are identified for a set of construction industry jobs and are estimated for each county in the United States. Two possible questions DOE has identified for this part of the evaluation research are:

- (1) What were the actual, monetary administrative costs for complying with Davis-Bacon?
- (2) Overall, how did application of Davis-Bacon influence the cost-effectiveness of WAP?

Impacts on Administration, Training, and Employment

DOE observes that the large flow of Recovery Act funds into low-income weatherization changed the national weatherization network. The weatherization labor force grew larger and new

¹⁵⁴ ORNL, *Evaluation of the National Weatherization Assistance Program during Program Years 2009-2011 (American Reinvestment and Recovery Act Period)*, ORNL/TM-2011/87, May 20, 2011 (draft date), pp.1 and 123.

¹⁵⁵ ORNL notes that Program Year (PY) 2009 started in April 2009 and PY2011 ends in June 2012. Ibid. pp.122-123.

¹⁵⁶ The weatherization program was not subject to Davis-Bacon rules before the Recovery Act.

stakeholders were drawn into the network. The increased funding had both positive and negative effects on long-standing leveraging relationships. DOE is considering several possible research questions. Four of the questions are:

- (1) How have relationships between state weatherization offices and local weatherization agencies changed?
- (2) Did the Recovery Act change the way that local agencies use contracts to procure weatherization services?¹⁵⁷
- (3) What approaches did local agencies and/or contractors use to recruit and train new qualified, reliable, and trustworthy weatherization crew members, and how effective were these approaches?
- (4) Did the expanded weatherization workforce find work opportunities in the energy efficiency field outside of the DOE weatherization program?

Other Expected Results

The evaluation design is expected to produce a variety of new information and findings about the program. Basic results are expected to include estimated energy savings for 2007 through 2011, along with estimates of cost-effectiveness and non-energy benefits. The design aims to uncover insights into energy savings attributable to particular measures and the strengths and weaknesses of computer audits, versus priority lists, as a way to select appropriate weatherization measures. Also, the design is expected to produce program operational data for two levels of average home investments (\$2,500 and \$6,500) and two levels of home eligibility (150% and 200% of the poverty level). Findings are expected to yield some indication of the impact on program benefits attributable to changing the federal grant formula in a way that distributes more funds to states in warmer climates.¹⁵⁸

Issue of Evaluation Independence and Objectivity

The OMB Inspector General has criticized DOE for depending almost exclusively on “in-house” ORNL staff to conduct evaluation studies. The core issue is whether an “in-house” evaluation effort can be sufficiently objective or neutral in its conduct of research. This issue is part of a quite spirited debate over federal program evaluation that goes far beyond the DOE weatherization program.¹⁵⁹ A key assumption of the debate is that a separation of evaluation duties from program implementation duties would increase the “independence” of the evaluating organization. In response to criticism regarding ORNL, DOE engaged a competitive process to retain an outside contractor to conduct the second national evaluation (PY2007-PY2008) and the evaluation of the Recovery Act period (PY2009-PY2011). As noted previously, the competition led to the selection of the APPRISE contracting firm.

¹⁵⁷ For example, did the Recovery Act funding change the use of requests for proposals (RFPs) versus bids?

¹⁵⁸ Home Energy, *Evaluating DOE’s Weatherization Assistance Program*.

¹⁵⁹ The current debate over the use of outside evaluation contractors is part of a long-term historical debate over the relative merits of having the evaluation activity conducted by an organization located inside (internal evaluator)—or outside (external evaluator)—of the institution that operates the program. The internal versus external issue, and the debate over evaluation contractors, continues to be a concern for many federal programs.

Debate Over the Use of Evaluation Contractors

A recent debate between professors Charles Metcalf and David Reingold focused on the extent to which the use of “outside” contractors to conduct federal program evaluations can significantly improve research independence and objectivity.¹⁶⁰ In a separate study, Professor Jacob Klerman continues the dialogue, adding further perspectives to the debate. All three professors suggest ways to strengthen the independence of evaluations conducted with contractors.

Metcalfe Argues: Use of Contractors Improves Evaluation Independence

Charles Metcalf speaks from experience as both a professor and an evaluation contractor. He says the central issue for evaluation independence is control over research processes and reporting. Metcalf acknowledges that outside pressures can affect a contractor’s objectivity. In particular, he contends that client (agency/program staff) authority to accept or reject the evaluation findings and report—and to time its release—can give it significant control over evaluation content.¹⁶¹ Metcalf argues that the independence and objectivity of contractor evaluations are promoted by four main conditions: maturity of the contractor organization, contractor reputation and longevity, formal agency-contractor agreements, and research transparency.¹⁶²

First, he claims that, over the past 40 years, the methods and practices of federal program evaluation have matured to the point where major evaluation contractor organizations have become generally objective and independent:

Mathematica Policy Research (MPR), and others such as Abt Associates, MDRC, and RAND ... have established objectivity and independence as core values, and by and large have successfully defended these values in their conduct with funding organizations. Our track record establishes that government supported policy research can be and is independent, with proper safeguards.¹⁶³

Second, Metcalf claims that the ability of a contractor organization to establish a reputation for objectivity enables it to enjoy longevity. In turn, this longevity helps the contractor to be independent from any single administration:

If we define the client to be the current administration, the contractor and often the contract outlive the current client. For contract research institutions that regard themselves as doing policy research for a sequence of clients with differing policy agendas, the institutions’ reputation for objectivity (as well as the quality of their work) is their primary currency in maintaining their longevity.¹⁶⁴

¹⁶⁰ Richard P. Nathan, *Journal of Policy Analysis and Management*, (Point/Counterpoint) *Can Government-Sponsored Evaluations Be Independent?* v. 27, no. 4, pp. 926-944, 2008.

¹⁶¹ Metcalf says this “ownership power” is the client’s primary way to influence content. *Ibid.* p. 930.

¹⁶² Metcalf, Charles E. *Threats to Independence and Objectivity of Government-Sponsored Evaluation and Policy Research*. *Journal of Policy Analysis and Management*. vol. 27, 2008. p. 927-934.

¹⁶³ RAND and MDRC are both non-profit, non-partisan policy research organizations. The RAND acronym was originally formed by a contraction of the term “research and development.” The MDRC acronym was originally formed from the term, Manpower Demonstration Research Corporation. *Ibid.* p. 927.

¹⁶⁴ *Ibid.* p. 932.

Third, Metcalf argues that the agency's ability to influence content is largely constrained by the evaluation contract:

For contract research, the researcher and his or her employing institution are collecting data and conducting research on an issue defined and paid for by a client. In this environment, two questions must be answered: (1) Who owns the research product? and (2) Who controls the content and the dissemination of the research product? These questions are typically defined by the contract, and the rights no longer lie intrinsically with the institution nor, by extension, with the researchers it employs to conduct the research.¹⁶⁵

Further, he claims that there is an increased tendency for researchers and clients to agree explicitly, in advance, on the research methodology and the scope of reports. This process tends to protect the client (agency) from unexpected surprises, and protects the researcher (contractor) by allowing less "wobble room" for the client to reinterpret results.¹⁶⁶

Fourth, Metcalf contends that research transparency is a key factor for contract evaluation independence. In this regard, he draws a parallel to academic research:

[T]he basic guardian of both quality and objectivity in the conventional world of scholarly research is the required openness to the examination of others. If you don't adhere to this rule, your research is presumptively not credible.¹⁶⁷

Metcalf argues that openness in the various phases of federal program evaluations is guaranteed by the government's competitive procurement process and by the Freedom of Information Act (FOIA). In procurement, the agency makes the request for proposals (RFP) publicly available. After the contract is awarded, the public can submit a FOIA request to obtain both the contract document (including the statement of the work) and the winning proposal:

Regardless of internal contract restrictions, the contractor has the implicit nonexclusive right to request its own report under the FOIA, and to release it freely without restriction. Thus, while the government owns the research products it purchases through contract, the FOIA protects the researcher's right to disseminate his or her results—and, indeed, the right of anyone else to disseminate the results.¹⁶⁸

Metcalf further claims that, in practice, the right of evaluation contractors to disseminate research results through conference participation and publication parallels the rights of researchers in academic institutions.¹⁶⁹ Also, in parallel to academic peer review, he notes that third parties often have a role in evaluations:

First, we have already stressed the importance of the ultimate availability of research reports to the public, time delays notwithstanding. Second, most large-scale studies engage advisory panels and other forms of peer review at both the design and report stages.¹⁷⁰

¹⁶⁵ *Ibid.*, p. 928.

¹⁶⁶ Also, he says this environment has promoted greater use of experimental research methods, which leave less room for subjective judgment than non-experimental methods. *Ibid.*, p. 932.

¹⁶⁷ *Ibid.*, p. 928.

¹⁶⁸ Since federal contracts currently tend to restrict automatic dissemination rights before the contract ends, Metcalf finds FOIA-guaranteed access very important. *Ibid.*, p. 930.

¹⁶⁹ *Ibid.*, p. 933.

¹⁷⁰ The use of advisory panels and related mechanisms can contribute a form of "third party" balance. *Ibid.*, p. 931.

Thus, he observes that third party involvement is an integral aspect of transparency.

While acknowledging that threats to independence persist, Metcalf contends that the contracting environment has had a positive effect on evaluation. He concludes that the practice of using external evaluation contractors has shown an ability to enhance research credibility and independence.¹⁷¹

Reingold Counter-Argues: With Contractors, Threats to Independence Remain

Professor Reingold starts from the same assumptions as Professor Metcalf. Reingold says separation of the evaluation activity is assumed to establish evaluator independence which—along with quality research techniques—allows program evaluation to serve an accountability function.¹⁷²

[U]nlike other services frequently contracted out by governmental and nongovernmental organizations, the rationale for contracting out program evaluation rests largely in its ability to establish independence by creating a clear division of labor between those conducting evaluations and those responsible for managing programs being evaluated.¹⁷³

He elaborates further that the process for selecting a contractor is vital to independence:

[T]he perception of independence is typically established via the selection process used to identify and hire the person or organization responsible for designing and implementing the evaluation. For many organizations that purchase evaluation services, independence is established by selecting someone who is external to the organization or program being evaluated and who is free from financial self-interest or ideology that may bias the evaluation to produce a desired outcome, and where the evaluator has control (or ownership) over the evaluation design and its intellectual property (data and findings).¹⁷⁴

Reingold argues that all of the “safeguards” to independence that Metcalf identified are overwhelmed by three main factors: (1) agency funding control gives it influence over evaluation content, (2) contractor organizations have financial self-interests in accommodating agency wishes, and (3) their mutual self-interests undermine the competitiveness of the selection process.

First, Reingold argues that an agency’s control over evaluation funding allows it to retain a strong influence over evaluation content:

[W]hen purchasing evaluation services via contracts, the purchasing agency has total control over all aspects of the services to be provided ... [thus] ... [i]t is inconsistent to argue that contracted program evaluation can be independent when the agency purchasing the evaluation owns (or controls) the research question and design, the method of data collection, the strategy of analysis, the data, the final report, and the rules governing the dissemination of results.¹⁷⁵

¹⁷¹ Ibid, p. 933-934.

¹⁷² Ibid, p. 934.

¹⁷³ Ibid, p. 935.

¹⁷⁴ Ibid, p. 935.

¹⁷⁵ Journal of Policy Analysis and Management. *David A Reingold’s Response to Charles E. Metcalf*, Fall 2008, pp. 937 and 942. Reingold notes that program evaluation is not the only realm of scientific analysis that is subject to the (continued...)

Second, Reingold contends that contractor organizations have a financial self-interest in accommodating agency wishes:

[E]valuation contractors view their relationships with those responsible for running programs and funding evaluations as clients where satisfaction with the deliverable will influence the likelihood of securing additional business in the form of more evaluation contracts.¹⁷⁶

In other words, he argues, independence can be lost when experts that are expected to produce independent facts also have a substantial incentive to acquiesce to a client's interests.

Third, Reingold claims that self-interests may begin to erode the contractor selection process, which underpins the independence sought by separating evaluation activity from program management:

When program staff are allowed to strongly influence the selection of a firm to conduct an evaluation of their program, they frequently select firms that will “work with them” to produce a “helpful” evaluation. This is code for selecting a firm that will tailor and implement an evaluation design that will likely provide positive findings for the program under study.¹⁷⁷

Over time, this threat to independence can deepen. Ultimately, he says, the complementary self-interests of agency and contractor may undermine the competitive foundation of the selection process:

When a firm has established itself as a reliable partner that will cooperate with the demands of a particular client, the selection process has virtually eliminated any real competition that would allow for a truly independent evaluation.¹⁷⁸

In sum, Reingold argues that threats to independence and objectivity can enter from both sides—a program's interest in self-protective control and a contractor's interest in preserving a business relationship. He concludes that:

Unfortunately, there are signs this arrangement [the effort to achieve greater independence through external contracting] is not working.¹⁷⁹

To address the problem, Reingold suggests two policy options to improve the independence of evaluation contracting. One option is to substitute grants for contracts, as the grants confer greater independence to the recipient. However, grants typically allow less oversight than contracts, which means that quality could suffer. Another option is to expand third party involvement. This option is described in the last section of the report.

(...continued)

problem of contracting bias.

¹⁷⁶ Ibid, p. 937.

¹⁷⁷ Ibid, p. 939.

¹⁷⁸ Ibid, p. 937.

¹⁷⁹ Ibid, p. 935.

Klerman Contributes to the Debate

Since the Metcalf-Reingold debate, there has continued to be a focus on means to reduce threats to evaluation independence. For example, Professor Jacob Klerman echoes some points from both Metcalf and Reingold in describing his view of the challenges to contractor evaluation independence. Klerman stresses two key threats to independence: the mutually reinforcing self-interests of funder and contractor and the self-interests of third parties. Also, he offers two ways to improve independence by modifying formal contract agreements.

First, Klerman characterizes the main threat to independence as an “inherent tension” between funder and contractor.¹⁸⁰ Beyond stated goals for quality and objectivity,¹⁸¹ he claims that both parties bring “unstated goals” involving pre-set policy opinions and organizational survival interests that can color an evaluation project’s process and findings. He echoes Reingold’s argument that self-interests pose a key threat to independence:

[E]valuation is sometimes politics (or business) by other means. For the funder, showing that the current administration’s programs “work” (and those of the previous administration did not “work”) yields political benefits. Sometimes the concerns are quite direct. A negative evaluation result might cause shrinking of a program, loss of budget for the funder’s organization, in the extreme even loss of jobs for the funder’s lead evaluation staff. For the contractor, this is a business. The “right answer”—that is, the answer that the funder “wants to hear”—now is likely to lead to more evaluation business in the future; the “wrong answer” now is likely to lead to less evaluation business in the future.¹⁸²

Second, Klerman proposes that when seeking a balance point between funder and contractor over evaluation content, one option is to involve a third party. However, “[e]ven nominating some third party leaves open the question of the third party’s policy and organizational goals.”¹⁸³ From his perspective, all three parties to an evaluation—funder, contractor, and third party—are propelled by self-interests that pose threats to independence. However, Klerman raises the question of whether there may be an “appropriate balance” of these interests that could actually reinforce independence.¹⁸⁴

In conclusion, Klerman—like Metcalf—claims that independence may be improved by augmenting, or otherwise improving, formal contract agreements. He describes one strategy for improvement:

The challenge is to devise contractual mechanisms that assure satisfaction of the stated goal—a high-quality and independent evaluation. Ideally, a contract would give the funder

¹⁸⁰ He points out that this unconventional nature is similar to that for certain other products. Auditing is cited as the leading example, where the tension arises because management is the direct client, but the audit will also be used by the broader investing community. Reputation and financial liability are noted as key incentives for auditing firms to show objectivity and independence. He cites the case of Enron and Arthur Andersen as an example where organization interests overwhelmed those incentives. Other such “gatekeepers” are observed, including lawyers in merger negotiations, bond raters, and securities analysts. Klerman, Jacob Alex, *Contracting for Independent Evaluation: Approaches to an Inherent Tension*, Evaluation Review, July 20, 2010, pp. 299-333.

¹⁸¹ Klerman defines objectivity in the methodological sense, to mean “conclusions following from the analysis, no more and no less.” Ibid. p. 329.

¹⁸² Ibid. p. 302.

¹⁸³ Ibid. p. 304.

¹⁸⁴ Ibid. p. 310.

enough authority to handle standard contractual issues while preventing either side from deviating from the stated goals to further policy or organizational goals.¹⁸⁵

As a starting point for this strategy, Klerman suggests that a clear separation be established between an official evaluation report and a contractor's own publication of analysis from the underlying evaluation. He recommends that:

The funder would retain almost unfettered rights to the official contract report. Those rights would include the right never to officially release the document but not the right to change the contractor's text while continuing to list the contractor and its individual staff members as authors. The contractor would retain clearly defined rights to publish any findings from the evaluation.¹⁸⁶

Klerman's main strategy for improving independence encompasses 19 approaches to modify evaluation contracts, which are grouped into three categories: (1) approaches applied when a funder retains the right to specify the final report text, specify the report release, and control discussion of the process; (2) approaches that limit the funder's right to specify the final text, and (3) approaches that limit the funder's right to specify report release.¹⁸⁷

Table 5 summarizes the main points of the Metcalf-Reingold debate, including the points contributed by Klerman.

¹⁸⁵ Ibid, p. 304.

¹⁸⁶ Ibid, p. 330.

¹⁸⁷ Ibid. pp. 307-331. A review of Klerman's 19 approaches is beyond the scope of this report.

Table 5. Debate: Does Evaluation Contracting Promote Independence?

(Metcalf – Reingold Arguments)

Issue Premise: Pressures exist that threaten the independence and objectivity of evaluation studies. In order to fulfill an accountability function, evaluation studies must be conducted independently of agency and program influences.		
Dimension/Debate Point	Contracting promotes evaluation independence. (Professor Metcalf)	Contracting does not guarantee evaluation independence. (Professor Reingold)
separate evaluation activity	The division of labor promotes independence.	Contractors have a self-interest in preserving their business/organization, which makes them “bend” to program/agency influence.
select contractor competitively	Competitive procurement of contractors promotes independence.	The contractor selection process determines independence. Funding control gives the program/agency influence over contractor selection. Also, the competitiveness of the selection process may be undermined by organizational self-interests on both sides.
make evaluation transparent	Research process transparency promotes independence. Prior agreements over research design and methods helps ensure independence and objectivity. Responses to RFPs are publicly available. FOIA allows contractors and others to make research results publicly available.	Funding control gives influence over research design, data collection, analysis, reporting, and dissemination.
involve third party	Third party/peer review activity promotes independence.	Third parties also have organizational self-interests.

Source: Richard P. Nathan, Journal of Policy Analysis and Management, (Point/Counterpoint) *Can Government-Sponsored Evaluations Be Independent?* v. 27, no. 4, pp. 926-944, 2008. Also see Klerman, Jacob Alex, *Contracting for Independent Evaluation: Approaches to an Inherent Tension*. *Evaluation Review*, July 20, 2010, pp. 299-333.

Notes: Although the original debate was framed by a 2008 point-counterpoint dialogue between Metcalf and Reingold; in 2010, Klerman made substantial contributions to the debate, which are incorporated into the table.

The Limits of Independence: Is There an Optimum?

Agencies of other national governments are also concerned about program evaluation independence. An example is the United Kingdom’s Department for International Development (DFID). An assessment of DFID’s evaluation function addressed many of the same issues of control over the research process and findings that were the focus of the Metcalf-Reingold debate. The study observed that independence is important, a critical aspect of evaluation credibility and reliability:

To be sure, independence on its own does not guarantee quality (relevant skills, sound methods, adequate resources and transparency are also required) but there is no necessary trade-off between independence, quality or credibility. Indeed, evaluation quality without independence does not assure credibility. Furthermore, in open and accountable working environments, evaluation independence induces credibility, protects the learning process and induces program managers and stakeholders to focus on results. Thus, evaluation independence, quality and credibility are complementary characteristics that together contribute to evaluation excellence.¹⁸⁸

¹⁸⁸ Picciotto, Robert. *Evaluation Independence at DFID: An Independent Assessment Prepared for the Independent Advisory Committee for Development Impact*. August 29, 2008, p. 5, <http://www.wipo.int/about-> (continued...)

The DFID assessment of evaluation function suggests that the quest for independence—if taken too far—could lead to dysfunction. The report suggests that there may be limits to evaluation independence:

Optimum independence is not maximum independence. Accurate and fair evaluations combine intellectual detachment with empathy and understanding. The ability to engage with diverse stakeholders and secure their trust while maintaining the integrity of the evaluation process is the acid test of evaluation professionalism. This is why diminishing returns set in when evaluation independence assumes extreme forms of disengagement and distance. Independence combined with disengagement increases information asymmetry, ruptures contacts with decision makers and restricts access to relevant sources of information. It leads to isolation, a lack of leverage over operational decision making and a chilling effect on learning. Thus, the basic challenge of evaluation governance design consists in sustaining full independence without incurring isolation.¹⁸⁹

That finding appears to be a more elaborate expression of a key conclusion that Klerman makes in regard to achieving an “appropriate balance” of the three potential parties to a contract evaluation:

[N]o single party—not the funder, not the contractor, and not some third parties—will always be without bias. Different approaches shift the relative power of the three groups. Furthermore, different approaches have different costs—in dollars, in the formality of the funder–contractor relationship, and in time to complete the study. The appropriate balance will vary with the particular situation, and, in particular, with the prevalence of problematic behaviors on all sides—funder, contractor, and third parties.¹⁹⁰

As noted previously, Klerman proposes several means to modify the client-contractor relationship to address the inherent tension and reduce—but not eliminate—the threats to objectivity and independence.¹⁹¹

Implications of the Independence Debate for DOE Evaluations

DOE Addresses Threats to Independence

DOE is aware of the concern about contractor independence for impact evaluations. It has published guidance for program staff that addresses the issue:

Evaluation should be conducted by outside independent experts. This means that, even if staff commission a study (fund an evaluation contractor) that contractor should have some degree of independence from the program office that is being evaluated. Also, the contractor should have no real or perceived conflict of interest. Although the program staff person may

(...continued)

[wipo/en/oversight/evaluation/pdf/iadci_dfid.pdf](https://www.wipo/en/oversight/evaluation/pdf/iadci_dfid.pdf).

¹⁸⁹ Ibid.

¹⁹⁰ Ibid, p. 310.

¹⁹¹ His 19 suggestions are grouped into three categories: (1) approaches when a funder retains the rights to specify the final text, specify the report release, and control discussion of the process; (2) approaches that limit the right to specify the final text; and (3) approaches that limit the right to specify report release. A review of Klerman’s suggestions is beyond the scope of this report.

work with the contractor during the consultation phase to clarify information needs and discuss potential evaluation criteria and questions, the staff should establish some line of separation from the contractor for much of the remainder of the evaluation study—i.e., put up a firewall after the initial consultation period is concluded.¹⁹²

Also, in order to address remaining concerns about outside contractor independence, DOE guidance further recommends the establishment of a third party group of experts to critique the contractor's work:

It is further highly recommended that a panel of external evaluation experts who are not part of the contractor's team be assembled to serve as reviewers of the contractor's work. This would include the Evaluation Plan, and the draft and final reports. Having the work of the contractor's team itself evaluated helps ensure the evaluation methodology is sound and the study is carried out efficiently. It also sets up a second firewall that raises the credibility of the study to even a higher level (important for those who remain skeptical of evaluation studies commissioned by the program being evaluated).¹⁹³

Further Considerations for DOE's Active Evaluations

DOE has addressed the first point of the independence debate. It separated the evaluation activity from the program activity by choosing outside contractors to perform a strategic assessment and two major evaluation studies. Additional comments and questions may focus on the other three points of the independence debate:

- **Contractor Selection.** All three debaters stressed that the selection process is critical to evaluation independence. Was the selection of Deloitte and APPRISE competitive? What objective indicators should be used to gauge the competitiveness of the selection process? Is there an optimum degree of selection process competitiveness?
- **Transparency.** Metcalf argued that transparency helps guarantee evaluation independence. Will this dimension be fulfilled by the public release of the three final reports? Otherwise, what objective indicators should be used to gauge transparency for each evaluation? Is there an optimum degree of transparency?
- **Third Party Involvement.** Klerman described the possibility of an “appropriate balance” that included a third party. Related to the ongoing evaluation of the Recovery Act period (2009-2011), the DOE IG issued a progress report and GAO has contributed a major performance audit. Is that sufficient third party involvement? Otherwise, what objective indicators should be used to gauge the sufficiency of third party involvement for each evaluation? Is there an optimum degree of third party involvement?

¹⁹² DOE EERE, *Program Evaluation: Lessons Learned* (Lessons Learned from Outcome/Impact Evaluation Studies, Item #4), http://www1.eere.energy.gov/ba/pba/program_evaluation/m/lessons_learned.html.

¹⁹³ Ibid.

Some Policy Options That May Enhance Independence

The debate between Metcalf and Reingold produced some other ideas for establishing independence of the evaluation function.¹⁹⁴ One notable strategy suggests using other (third party) organizations to play a more central role in managing the program evaluation process. Reingold suggests:

Ideally, Congress, the GAO, and perhaps the National Academy of Sciences [National Research Council] could play a more active role in screening and selecting evaluation plans and firms.¹⁹⁵

Three variations on this strategy, in increasing order of dependence on the outside agency are:

- (1) Have the DOE Office of the Inspector General (DOE IG) or GAO or Congressional Budget Office (CBO) ensure that the third party contractor evaluation work is performed in an independent manner.
- (2) Move the contractor (evaluator) selection process from the agency to the Office of the Inspector General, GAO, CBO, or the National Research Council (NRC).
- (3) Transfer full responsibility for conduct of the evaluation to the DOE IG, GAO, CBO, or NRC.

¹⁹⁴ Reingold, JPAM, p. 938-940.

¹⁹⁵ Ibid. p. 940.

Appendix A. DOE Weatherization Program: Historical Data

Table A-1. DOE Weatherization Funding, Units Weatherized, and Benefit-Cost Ratio
(\$ millions)

Fiscal Year	Request \$Current	Approp. \$Current	Deflator \$2005	Deflator \$2010	Request \$ 2010	Approp. \$2010	Annual Units	Cumulative Units	Benefit-Cost Ratio
1977	0.0	27.5	0.3750	0.3362	0.0	81.6	1,622	1,622	
1978	54.1	65.0	0.4003	0.3589	150.4	180.7	6,742	8,364	
1979	198.5	199.0	0.4325	0.3878	510.7	511.8	15,387	23,751	
1980	199.0	199.0	0.4707	0.4220	470.4	470.3	232,751	256,502	
1981	175.0	175.0	0.5171	0.4636	376.6	376.5	352,906	609,408	
1982	0.0	144.0	0.5525	0.4954	0.0	290.0	122,992	732,400	
1983	0.0	245.0	0.5768	0.5172	0.0	472.6	156,629	889,029	
1984	0.0	190.0	0.5981	0.5363	0.0	353.5	209,261	1,098,290	
1985	190.0	191.1	0.6175	0.5537	342.4	344.4	163,860	1,262,150	
1986	152.9	182.1	0.6318	0.5665	269.3	320.7	149,047	1,411,197	
1987	0.0	161.3	0.6486	0.5815	0.0	276.7	105,440	1,516,637	
1988	0.0	161.3	0.6694	0.6002	0.0	268.1	105,465	1,622,102	
1989	0.0	161.3	0.6954	0.6235	0.0	258.1	85,115	1,707,217	1.06
1990	0.0	162.0	0.7210	0.6465	0.0	250.0	84,441	1,791,658	
1991	15.0	198.9	0.7483	0.6709	22.3	295.8	105,769	1,897,427	
1992	24.0	194.0	0.7678	0.6884	34.8	281.1	99,587	1,997,014	
1993	80.0	185.4	0.7848	0.7037	113.4	262.9	103,394	2,100,408	
1994	239.4	214.8	0.8014	0.7186	332.4	298.2	114,904	2,215,312	
1995	249.8	226.4	0.8184	0.7338	339.6	307.8	102,981	2,318,293	
1996	229.0	111.7	0.8342	0.7480	305.5	149.0	76,393	2,394,686	1.79
1997	155.5	120.8	0.8495	0.7617	203.7	158.3	71,597	2,466,283	
1998	154.1	124.8	0.8603	0.7714	199.3	161.4	68,470	2,534,753	
1999	154.1	133.0	0.8717	0.7816	196.7	169.8	71,984	2,606,737	1.51
2000	154.0	135.0	0.8889	0.7970	192.8	169.0	74,316	2,681,053	
2001	154.0	153.0	0.9099	0.8158	188.3	187.1	77,709	2,758,762	
2002	273.0	230.0	0.9249	0.8293	328.4	276.7	104,860	2,863,622	1.30
2003	277.1	223.5	0.9442	0.8466	326.6	263.4	100,428	2,964,050	
2004	288.0	227.2	0.9684	0.8683	330.9	261.0	99,593	3,063,643	
2005	291.2	228.2	1.0000	0.8966	324.0	253.9	97,500	3,161,143	1.34
2006	230.0	242.6	1.0342	0.9273	247.5	261.0	104,283	3,265,426	1.54

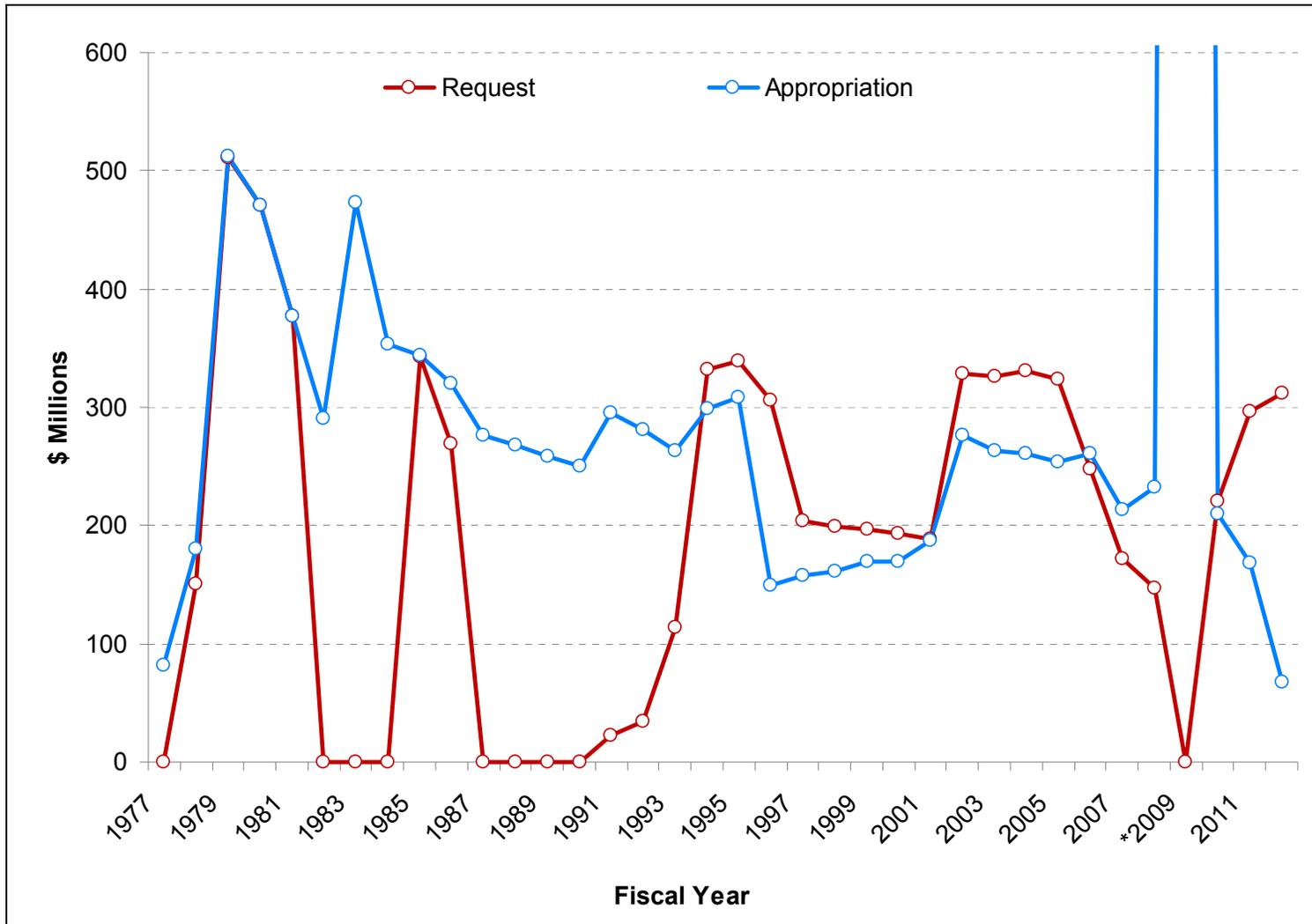
Fiscal Year	Request \$Current	Approp. \$Current	Deflator \$2005	Deflator \$2010	Request \$ 2010	Approp. \$2010	Annual Units	Cumulative Units	Benefit-Cost Ratio
2007	164.2	204.6	1.0643	0.9543	171.5	213.6	89,772	3,355,198	1.54
2008	144.0	227.2	1.0890	0.9764	147.0	232.0	6,116	3,361,314	1.65
2009	0.0	5,427.5	1.1054	0.9911	0.0	5,468.8			1.67
2010	220.0	210.0	1.1153	1.0000	220.0	210.0			1.80
2011	300.0	171.0			296.1	168.8			
2012	320.0				311.5				
Total		11,453.0				14,504.5			
Subtotal, 1977- 2008		5,644.5				8,657.0			

Source: DOE Weatherization Budget History; ExpectMore PART Results; and various evaluation studies.

Appendix B. DOE Weatherization Funding Chart

Figure B-1, below, shows the entire history of annual requests and appropriations for the DOE Weatherization Assistance Program (WAP). Requested amounts for FY1977, FY1978, and FY1979 were not available. A final FY2012 appropriation figure had not yet been determined at the time this report was prepared.

Figure B-I. DOE Weatherization Program Funding History: Requests and Appropriations
 (\$ millions, FY2010 constant dollars)



Source: DOE budget requests (various years); DOE Weatherization Budget History; and ExpectMore PART Results.

Notes: For FY2009, WAP received an appropriation of \$427.5 million (\$ current) plus \$5 billion (\$ current) from the Recovery Act (P.L. 111-5). The total amount is too large to be shown on this chart, but the funding amounts appear in Appendix A, **Table A-I**.

Author Contact Information

Fred Sissine
Specialist in Energy Policy
fsissine@crs.loc.gov, 7-7039