

Quantitative Findings from On-Site Evaluation of Energy Efficiency Program Service Delivery

Jacqueline Berger, APPRISE

ABSTRACT

This paper reports on an observational study of energy efficiency service delivery. The research documents how services are delivered to clients, how weatherization staff communicate with clients, and the consistency of implemented procedures with best practices. The work is unique because systematic techniques for observing and recording information, rather than descriptive case studies, were used, resulting in quantitative findings on service delivery strengths and areas for improved performance.

The methodical information was collected by providing observers with forms and rating scales, and a database in which to record the information. The forms listed the expected steps for each stage of the process – the audit, the measure installation, and the final inspection. Rating scales were developed by tallying the number and percent of steps completed correctly, with key steps required for a top rating to be assigned.

The observations quantitatively assessed the quality of work done and provided detailed data on where the program is succeeding, where procedures could be refined, and where additional training is needed. Additionally, some exploratory research was conducted to assess characteristics of delivery agencies that were associated with higher performance, and whether observed performance was related to actual energy savings achieved by the agencies.

The research found that there were the most opportunities for improvement with respect to client education provided during the audit, measure installation, and final inspection visits. While the education was rated low, the technical aspects of the work were rated in the mid-range, and the professional aspects of the work were rated highly.

Objectives

The objectives of the study were to document the following.

- How services are delivered to clients.
- The consistency of implementation procedures with best practices.
- The training needs of auditors, crews, and contractors.
- How program procedures could be modified to improve service delivery.

Study Design

The overall goal of the study was to accurately document how program services are delivered. We aimed to collect information in a systematic and quantitative manner to allow for collection of data that provides a rigorous assessment of service delivery and concrete information on implementation challenges and solutions. This approach required purposive selection of agencies across the country, development and extensive testing of detailed data collection instruments and quality rating scales, and informed analysis of the resulting data.

Weatherization Experts were recruited from the evaluation team's professional contacts across the country. All of the Weatherization Experts had extensive experience managing, implementing, and/or training on weatherization procedures, service delivery, and quality assurance. The Weatherization Experts were responsible for documenting the characteristics and quality of service delivery using a detailed set of data collection forms.

Social Scientists from various disciplines (urban planning, social anthropology, psychology, sociology, etc.) were also engaged for this study, to obtain a variety of perspectives on the interactions between clients and providers, and among providers. Social Scientists were responsible for documenting and assessing the interactions between weatherization staff and clients, the interactions among weatherization staff, and the characteristics of agencies, staff, training, and clients that influence service delivery. They provided context around the quantitative data, and findings from this research are largely excluded from this paper.

Sample Selection

Because of the small scale of the study relative to the wide variety of characteristics that describe the agencies and how they deliver program services, purposive sampling was the most appropriate method. Under the purposive method, the sample was selected in steps, to round out the sample and represent the important characteristics of the delivery agencies and programs. The purposive sampling differed from probability sampling in that random selection techniques were not used. As a result, confidence intervals cannot be developed for statistical analysis. However, the sample design allows for an assessment of overall program performance, with greater applicability than anecdotal results would allow.

When selecting the sample, we aimed to take the factors into account that were likely to be related to program performance – provider characteristics, service delivery characteristics, and county characteristics. Twenty local service delivery agencies were sampled from the 400 agencies that comprised the agency billing data sample (out of the nearly 1,000 agencies that deliver services across the country.) The sample was selected in the following manner.

1. The number of states per region was determined based on program funding.
2. States were selected within region based on the part of the region and the program size.
3. Agencies were selected within selected states to represent urban/rural, measure selection technique, and self-reported levels of client education, staff training, and quality control.
4. One agency was selected in each of the 20 selected states.

Data Collection Instruments

Design of the data collection instruments was a critical part of the research. These forms were planned to ensure that systematic data were collected about the work that was done and that the quality of the work was rated equivalently by the ten Weatherization Experts.

One of the key challenges in the form development was the fact that while the program has a set of general guidelines, it is implemented differently in every state, and differently by agencies within each state. Important variations include how the audit is performed, the diagnostic testing procedures that are required, what measures are eligible for selection, and the type of education that is provided. This variation made it challenging to develop general forms

and procedures that still collected the detailed quantitative data that were needed to assess the program.

The evaluation team decided that all agencies would be evaluated on the same scale, using the same forms, according to a set of best practices agreed to by the weatherization professionals who consulted on the form development. For example, even if an agency did not include a particular diagnostic test in their audit procedures, their work would be assessed as being less than complete. However, the Weatherization Experts would debrief the weatherization staff following the audit and record whether or not that test was part of the agency standards. This would allow an understanding of whether the standards needed to be addressed, or whether staff needed to be better trained to follow agency procedures.

A rating scale was developed for each aspect of the weatherization work that was rated for quality. The rating scale asked the observer to record whether or not each of several tasks were performed correctly, and then the score was based on the percentage of required tasks that were completed. Additionally, several of the items rated required the staff to have a “yes” for some specific tasks to receive one of the top two ratings (4/5 or 5/5 on the scale.) If an item was not applicable in the particular case, it was counted as a “yes” so that the exclusion of that activity would not count against the item being scored. A sample of the detailed rating scales that were used is furnished in this report to allow the reader to understand the assessment method.

Table 1 furnishes the general guidelines for how the rating scales were designed. The table shows the number of points needed, based on the total number of items rated, to score each rating.

- All of the items must have been done for the rating to be a 5.
- Between 90 and 99 percent of the items must have been done for the rating to be a 4.
- Between 75 and 89 percent of the items must have been done for the rating to be a 3.
- Between 50 and 74 percent of the items needed to be done for the rating to be a 2.
- The rating was 1 if less than half of the required items were done.

Table 1. Rating Design

Rating	1	2	3	4*	5*
% of Points Needed	0%-49%	50%-74%	75%-89%	90%-99%	100%
Total points	Number of Points needed for each rating				
4	0-1	2	3	--	4
5	0-2	3	4	--	5
6	0-2	3-4	--	5	6
Ratings for scales with total number of points between 6 and 13 are calculated in the same way, based on the % of points needed for each score.					
13	0-6	7-9	10-11	12	13
14	0-6	7-9	10-12	13	14
15	0-7	8-10	11-13	14	15

*In several cases, bolded items must be checked to receive a rating of 4 or 5. Examples of bold, required items are shown in later rating tables.

The data collection forms included several notes fields to allow the observers to record additional information that affected the quality of the work or specific challenges faced on site. For example, in some cases, the quality of work was rated low primarily because important safety precautions were not followed. In these cases, observers would document this information in the notes field.

Data collection forms were created for the three types of visits observed by the Weatherization Experts and for an evaluation of the audit write-up. The following forms were used.

1. Audit observation data collection
2. Audit write-up assessment
3. Measure installation observation data collection
4. Agency final inspection data collection

Observations Conducted

Weatherization Experts and Social Scientists visited 19 agencies in total because one of the 20 selected agencies was not performing weatherization work due to exhaustion of weatherization funding at the time that the visit was planned. Table 2 details the number of observations conducted of each type of visit by the Social Scientists and Weatherization Experts. In total, 155 audits, 159 installations, and 128 final inspections were observed.

Table 2. Observations Conducted

	Number of Observations		
	Social Scientist	Weatherization Expert	Total
Audit	43	112	155
Measure Installation	45	114	159
Final Inspection	37	91	128
TOTAL	125	317	442

Audits

Weatherization Experts observed 112 audits during the agency visits. The full evaluation report covers findings in the following areas.

- Audit Introduction and Home Walkthrough
- Heating, Ventilation, and Air Conditioning
- Water Heating
- Diagnostic Testing
- Refrigerators and Lighting
- Client Interaction and Education
- Exit Interview

- Audit Summary Ratings
- Auditor Professionalism
- Audit Write-up

This report provides a sample of the findings in two different categories.

Table 3 displays information from observations of air leakage and insulation diagnostics conducted during the audits. While the basic analyses, including measurements, inspections of windows and attics, and visual inspections were conducted in the majority of applicable homes observed, the more in-depth analyses were less commonly performed. For example, auditors only created access to inaccessible attics in ten percent of applicable cases, they measured insulation in exterior walls in 49 percent of applicable cases, and they inspected for all typical bypasses in 62 percent of the homes.

Table 3. Air Leakage and Insulation Diagnostics

	Applicable Observations	Action Taken	
		#	%
Measured surfaces	100	94	94%
Inspected windows	100	96	96%
Inspected all accessible attics	78	69	88%
Measured insulation in all accessible attics	79	70	89%
Created access to inaccessible attics	33	3	10%
Measured insulation in exterior walls	95	47	49%
Measured insulation in basement/crawlspace	74	55	74%
Inspected for all typical bypasses	100	62	62%
Visual inspection for air sealing opportunities	100	83	83%
Used blower door while inspecting for leaks	96	64	67%

Weatherization Experts were asked to rate the completeness of the observed audits, the completeness of testing performed, and the quality of testing performed. Table 4 displays the factors that were used to rate the completeness of the audit. There were nine elements included in the completeness rating. The completeness of the audit depended upon inspection of all areas, discussion with the client about home and energy issues, and use of a blower door test.

Table 4. Completeness of Audit Rating Design

Rating	Completeness of Inspection
1	Auditor walked through all rooms in the house
2	Auditor measured and recorded dimensions of building
3	Auditor physically checked existing insulation levels in the attic(s), walls, and basement/crawlspace
4	Auditor talked with client about comfort issues
5	Auditor talked with client about energy issues
6	Auditor talked with client about health and safety issues
7	Auditor assessed all accessible attics
8	Auditor assessed all accessible basement/crawlspace areas
9	Auditor conducted a blower door test

Table 5 displays the results from the audit completeness rating. The audits received a mid-level rating of 3.0. Many of the auditors did not receive higher scores because they did not discuss comfort, energy issues, and/or health and safety issues with the clients.

Table 5. Completeness of Inspection Rating

	Number Rated	Percent With Each Rating					Mean Rating
		Needs Improvement	—————→			Excellent	
		1	2	3	4	5	
Completeness of inspection	103	8%	32%	28%	17%	15%	3.0

The audit observations and data analysis led to the following conclusions about the strengths of the audit process.

- Weatherization staff demonstrated extraordinary respect and concern for clients served.
- The explanation of the program and process was usually comprehensive.
- Staff made good use of data collection forms.
- Auditors often met with installers to explain the audit findings and work scope.

The research led to the following conclusions about opportunities for improved audit performance.

- Use of utility bills could improve understanding of the home and services needed.
- Increased client engagement in the audit could improve program outcomes.
- Auditors should have increased understanding of pressure boundaries.
- Auditors should have increased understanding of testing purpose and procedures.
- Worst case draft testing should be implemented in all appropriate cases.
- Improved audit write-ups and work scope descriptions would increase installers' understanding of expected work.

Measure Installation

Weatherization Experts observed 114 installations during the agency visits. The full project report provides a detailed review of findings in the following areas.

- Introduction
- Window and Door Work
- Air Sealing
- Attic Insulation
- Wall Insulation
- Basement and Crawl Space Insulation
- Ventilation
- Heating and Cooling
- Hot Water System
- Other Measures – Refrigerator, Dryer, Lighting, Low-Cost Measures
- Professionalism
- Communication and Education

This report provides a sample of the findings in one of the studied areas.

Table 6 displays the rating scale that was used to assess the air sealing work that was done throughout the home. The scale includes aspects of effectiveness, cleanliness, client safety, and worker safety. The first five items in the scale are bolded, indicating that all five of these items must be completed for the rating to be a 4 or a 5.

Table 6. Air Sealing Rating Design

Rating	Quality of Air Sealing in Specific Places
1	Fire rated materials used around chimneys, flue pipes, recessed lighting, etc.
2	Materials integrated with existing finishes
3	Used supply-air respirators when using two-part foams
4	Blower door used to target air sealing
5	Blower door and smoke used to verify air sealing effectiveness
6	Caulk not used to seal gaps larger than 1/8"
7	Large holes and gaps covered with rigid materials
8	Area cleaned before caulk or foam applied
9	All material scraps and packaging removed from home
10	Materials installed consistently and are uniform in appearance
11	Gloves and eye protection used when cutting, using caulk or foam
12	Used hard hats when working in confined spaces

Table 7 displays the ratings for the air sealing work that was completed in various parts of the home. The table shows that mean ratings range from 2.2 for the air sealing between the area of the basement to the conditioned space to a 2.7 for the air sealing of the second floor rim

joist. Most of the air sealing ratings were relatively low. The only area where any contractors scored a four or five on the rating scale was sealing of the attic floor penetrations and the crawl space to the conditioned space. Installers often lost points because they did not use the required personal protective equipment.

Table 7. Air Sealing Ratings

	Number Rated	Percent With Each Rating					Mean Rating
		Needs Improvement	→ Excellent				
		1	2	3	4	5	
Attic floor penetrations	58	10%	41%	34%	9%	5%	2.6
Knee walls	19	5%	58%	37%	0%	0%	2.3
Second floor rim joist	3	0%	33%	67%	0%	0%	2.7
Basement penetrations to outside	24	8%	46%	46%	0%	0%	2.4
Basement to conditioned space	12	17%	50%	33%	0%	0%	2.2
Crawl space to conditioned space	14	0%	64%	29%	7%	0%	2.4
Crawl space to outside	9	0%	56%	44%	0%	0%	2.4
Garage to conditioned space	8	0%	63%	38%	0%	0%	2.4

The measure installation observations and data analysis led to the following conclusions about the strengths of the measure installation process.

- Installer flexibility and dedication often well surpassed the work scope.
- Good problem solving skills were demonstrated.
- High quality work was observed.
- Experienced HVAC contractors were contracted for installations.

The research led to the following conclusions about opportunities for improved installation performance.

- Increased use of blower door when air sealing could improve results.
- Respect for the clients' homes could be improved, such as the use of booties and covering clients' furniture.
- Installer safety practices need improvement.
- Increased assessment of HVAC contractors should be done, rather than routinely accepting their statements and recommendations.
- Explanation of CFLs when installing the bulbs could lead to improved persistence.
- Client education needs improvement.

Final Inspection

There were 91 final inspections observed by the Weatherization Experts. The following aspects of the final inspections are discussed in the full project report.

- Home Walkthrough
- Testing
- Occupant Interaction
- Professionalism

This section provides a sample of the results from the observations of final inspections.

Table 8 displays findings with respect to the combustion safety tests. This table shows that tests on the heating system and water heater were only conducted in about 75 percent of the observed visits, and tests on the gas stove, space heater, and the ambient test were even less frequently conducted. However, these tests were almost always performed correctly when they were conducted.

Table 8. Combustion Safety Tests

	Tests Performed			
	Should Have Been	Should Have Been & Was	% Performed	% Performed Correctly
Heating system	54	39	72%	97%
Water heater	50	38	76%	95%
Gas stove	26	8	31%	100%
Space Heater	7	0	0%	N/A
Ambient	45	18	40%	100%

Table 9 displays the rating scale used to assess the quality of testing. The rating is based upon testing equipment, correctness of test procedures, and safety.

Table 9. Quality of Testing Rating Design

Rating	Quality of Testing
1	Inspector test equipment calibration current
2	Tests done correctly
3	Test results used to grade work quality
4	Inspectors proficient in operating test equipment
5	Combustion appliances disabled during blower door and duct testing

Table 10 shows that, on average, inspectors received mid-level ratings for testing quality.

Table 10. Testing Quality Rating

	Number Rated	Percent With Each Rating					Mean Rating
		Needs Improvement				Excellent	
		1	2	3	4	5	
Testing quality	88	16%	17%	19%	13%	35%	3.3

The final inspection observations and data analysis led to the following conclusions about the strengths of the final inspection process.

- Inspectors did a good job when they closely followed the work orders.
- Installers’ attendance at the final inspection increased efficiency of inspection and follow-up work.
- Inspectors usually provided clients with information on who to call if they encountered any problems with the work.
- Referrals were made to additional assistance programs that could improve client outcomes.

The research led to the following conclusions about opportunities for improved final inspections.

- Increased client education through explanation of installed measures and reinforcement of the client action plan could lead to greater program impact.
- Testing quality showed room for improvement.
- Installation quality, in addition to completeness, should always be assessed.
- Reduced use of auditors to conduct the final inspection could provide additional perspective on the work required in the home. Agencies should use separate inspection staff, or use a different auditor to do the final inspection than the one who did the initial home audit.

Factors

State programs, local agencies, weatherization staff, and contractors differ on numerous dimensions that may impact the quality and effectiveness of services delivered. One of the goals of this study was to determine whether there were characteristics of agency management, staff qualifications, or training that are related to service delivery quality, as assessed in this study. Definitive conclusions cannot be drawn due to the small number of agencies and jobs observed compared to the large number of factors that can influence success, but the study did investigate whether there were any meaningful relationships between potentially determining factors and agency performance.

Audit Factors

The factors studied were based upon agency responses to surveys included as part of the program evaluation. The following factors related to audits were reviewed.

- Use of in-person instruction at the time of the audit.
- Requirements for staff engaged in measure selection.
 - Technical certification
 - Extensive weatherization work experience
 - Extensive weatherization supervision experience
 - Construction experience
- Requirements for staff engaged in diagnostic procedures.
 - Technical certification
 - Extensive weatherization work experience
 - Extensive weatherization supervision experience
 - Construction experience

For the most part, there were no relationships seen between these factors and ratings made by the Weatherization Experts. However, there were some quality factors that were related to higher ratings.

- The audit technical ratings averaged 2.5 for observations at agencies that did require the certification, compared to 2.1 for those that did not.
- Agencies that required extensive weatherization work experience had an average audit education rating of 2.1 compared to 1.6 for those who did not.
- Agencies that required extensive weatherization supervision experience had an average audit education rating of 2.3 compared to 1.9 for those who did not.

Measure Installation Factors

Agencies utilize different models for the installation of measures. While some agencies use their own staff members (weatherization crews) to provide measure installation, other agencies hire contractors to perform the work. Sixty-five percent of the installations that were observed by the Weatherization Experts were done by contractors, 31 percent were done by crews, and four percent were done by both crews and contractors.

We assessed the relationship between whether crews or contractors were used for the installations and the Weatherization Experts' assessments. There were some differences.

- Crews were more likely to have important information at the time of the installation.
 - 91 percent of crews had household demographic data, compared to 22 percent of contractors.
 - 77 percent of crews had the audit report, compared to 34 percent of contractors.
 - 91 percent of crews had a materials list, compared to 49 percent of contractors.

- Crews had higher mean ratings than contractors for efficiency, professionalism and safety practices.
 - Crews had an average mean rating of 4.4 for efficiency, compared to 3.9 for contractors.
 - Crews had an average mean rating of 4.2 for professionalism, compared to 3.7 for contractors.
 - Crews had an average mean rating of 2.9 for safety, compared to 2.3 for contractors.

Final Inspection Factors

The following factors related to final inspections were reviewed.

- Use of innovative program for inspection of weatherized units.¹
- Whether in-person instruction is required at the time of the final inspection.

There were some relationships between these factors and agency final inspection rating scores.

- Agencies that reported they had an innovative program had a mean technical inspection rating of 3.3, compared to a mean rating of 2.3 for those who did not.
- Agencies that reported in-person instruction had higher technical and professional ratings than those who did not.
 - Agencies with in-person instruction had a mean technical inspection rating of 2.6, compared to 1.7 for those who did not.
 - Agencies with in-person instruction had a mean professional rating of 4.2, compared to 3.5 for those who did not.

However, contrary to what would be expected, they did not have higher education ratings. If additional education was done by these agencies, it may have been focused on explaining the program and services rather than discussing energy usage and potential areas for usage reduction.

All Program Area Factors

The following factors related to all aspects of weatherization services were reviewed.

- Use of an innovative education approach.
- Use of an innovative weatherization staff training approach.
- Number of different types of training activities provided by the agencies.

¹ This was based upon the state WAP office’s response to an evaluation survey that asked the state to indicate whether the local agency had an “Innovative and/or particularly effective program for inspecting weatherized units (Quality Assurance).”

- Level of education provided by the agency, determined by the mix of education methods reported by the agencies.
- Level of training provided by the agency, determined by the mix of training activities reported by the agencies.
- Agency quality rating.

The only one of these factors that was related to observation ratings was that agencies that reported they had an innovative training program had higher ratings in all parts of the process than those who did not. Agencies that reported the training program had the following higher scores.

- All audit technical ratings averaged 3.4 compared to 2.3 for those without.
- All installation technical ratings averaged 3.3 compared to 3.0 for those without.
- All inspection technical ratings averaged 3.2 compared to 2.4 for those without.

Impacts

The rationale behind this study's design is that if weatherization providers accurately determine which measures have the greatest potential impact; effectively install selected measures according to best weatherization practices; and carefully inspect the completed job for safety, completeness, and quality, the program will produce high quality work that significantly reduces energy usage.

This study focused on the implementation of the program, rather than the savings that were achieved. However, the program's impact analysis assessed agency-specific results for agencies that had at least eight single family homes with sufficient usage data to assess energy savings. There were 120 agencies that had sufficient data to assess the agency-specific gas savings and eight of the 19 observed agencies were included in this group. There were 103 agencies that had sufficient data to assess the agency-specific electric savings and eight of the studied agencies were included in this group.

Due to the small numbers of observations per agency, and the fact that only eight of the agencies had enough data to compute agency-specific savings results, it is difficult to draw conclusions regarding the relationship between observed performance quality and energy savings impacts. While there is not enough data to draw conclusions, it is noteworthy that three of the top four ranked agencies, over all observation quality ratings, had either gas or electric savings that were in the top half-decile of the agencies that could be ranked by savings values.

Findings and Recommendations

The research conducted showed that there are many strengths of the service delivery process, but also significant opportunity for improvement. Detailed results show the specific areas for focus, but the general findings were that client education, communication through written audit findings, and parts of the technical implementation could be improved. To significantly impact these areas, agency management needs to set the tone for what is expected. The key issues are as follows.

- Prioritize classroom and in-field training for all auditors and installers.
- Conduct assessments to determine where additional training is needed.
- Tailor additional training to quality control findings.
- Emphasize client communication and education as an integral aspect of all stages of service delivery.

Specific training needs that were identified in the Weatherization Experts' observations were as follows.

- Building science fundamentals
- Critical thinking – assessing unique situations
- Combustion safety testing
- Ventilation assessment
- Safe work practices – worker safety and lead safe work
- Client education

Challenges identified with providing adequate training were as follows.

- More clarity is needed on program specifications and procedures.
- Sufficient time must be allocated for training, in addition to the continued emphasis on production.
- Pressure for job production often overwhelms the need for training.
- Agency management does not always prioritize training.

Some strategies for national, state, and local management to improve the availability and quality of training are as follows.

- Increased resources for staff training. This can lead to improved client outcomes and workforce development.
- A combination of classroom and field training can provide all types of learners with the types of experiences needed.
- Greater attendance on attending national conferences can provide information on best practices and increase motivation for improved performance.
- A focus on why tests are conducted can improve understanding of how to properly conduct the tests.
- An emphasis on client education as an important “measure” is needed to improve this aspect of service delivery.
- Reinforcement of training topics with monitoring, assessment, and feedback can lead to improved implementation and service delivery quality.

References

APPRISE. *New Jersey Comfort Partners Comprehensiveness Report*. 2002. <http://www.appriseinc.org/reports/NJCP%20Comprehensiveness.pdf>. Princeton, NJ.

APPRISE. *PPL Electric Utilities Winter Relief Assistance Program Evaluation Report*. 2006. <http://www.appriseinc.org/reports/Final%20PPL%20WRAP%20Evaluation%20Report.pdf>. Princeton, NJ.

Ternes, M., M. Schweitzer, B. Tonn, R. Schmoyer, and J. Eisenberg. 2007. *National Evaluation of the Weatherization Assistance Program: Preliminary Evaluation for Program Year 2006*. ORNL/CON-498. Oak Ridge, TN: Oak Ridge National Laboratory.