Realization Rates of the National Energy Audit

Linda G. Berry **Michael B. Gettings**

ABSTRACT

Engineering estimates of savings resulting from installation of energy conservation measures in homes are often greater than the savings actually realized. A brief review of prior studies of realization rates prefaces this study of rates from an engineering audit tool, NEAT, developed for the Department of Energy's Low-Income Weatherization Assistance Program and used in a New York state utility's low-income program. Estimates of metered and predicted savings are compared for 49 homes taken from a database of homes that participated in the first year of the utility's program. Average realization rates ranging from 57% to 69% result, depending on the data quality. Detailed examinations of two houses using an alternative engineering method, the DOE-2 computer program (considered an industry standard), seem to indicate that the low realization rates mainly result from factors other than inaccuracies in the audit's internal algorithms. Causes of the low realization rates are examined, showing that the strongest single factor linked to the low rates in this study is the use of secondary heating fuels that supplement the primary heating fuel. This study, like other similar studies, concludes that engineering estimates are valuable tools in determining ranked lists of cost-effective weatherization measures, but they may not be accurate substitutes for measured results in evaluating program performance.

THE NATIONAL ENERGY AUDIT (NEAT)

The National Energy Audit (NEAT) is a computer audit tool developed at Oak Ridge National Laboratory for the U.S. Department of Energy's Weatherization Assistance Program (the Program). The Program, which provides energy-efficiency services to about 200,000 low-income households per year, is implemented in all 50 states. Currently, more than half of the states use NEAT. NEAT's engineering algorithms are based on those developed for Lawrence Berkeley National Laboratory's Computerized Instrumented Residential Audit (CIRA).

The National Energy Audit is designed to select the most cost-effective combination of energy-efficiency measures for installation in each weatherized home. According to the *National Energy Audit Manual* (Krigger et al. 1997), there are several steps in the measure selection process, which include (1) entering data that describe the dwelling; (2) computing heat loss, heat gain, and energy required to keep the house at a specific thermostatic set point; (3) checking the applicability of 33 possible measures to the house as described by the auditor; (4) calculating savings for each applicable conservation measure and computing its discounted savings-to-investment ratio (SIR); (5) ranking the energy-efficiency measures in order of their SIR; (6) determining interacted SIRs; (7) again ranking the conservation measures by their interacted SIRs and preparing a list of essential materials; and (8) using, if desired, an adjustment feature that requires the entry of actual fuel consumption data. If fuel consumption data are entered, NEAT will adjust predicted measure savings to actual (as opposed to modeled) fuel consumption. To use this feature, the auditor must obtain data on actual fuel consumption from the gas or electric utility company that supplies heating/cooling fuel to the dwelling.

REALIZATION RATES: DEFINITION AND SOURCES OF ERROR

A realization rate is the ratio of measured savings to audit-predicted savings. It is usually expressed as a percentage. If the predicted and measured savings match exactly, the realization rate would be equal to 100%. When measured savings

Linda G. Berry is on the research staff and **Michael B. Gettings** is a staff engineer at Oak Ridge National Laboratory, Oak Ridge, Tenn.

exceed predicted savings, the realization rate is greater than 100%. When measured savings are less than predicted savings, the realization rate is less than 100%. If an audit's realization rates were consistently close to 100%, its predictions of savings could be used to estimate program performance. If not, evaluations based on measured fuel consumption are necessary to correctly characterize program performance. In this paper, we will discuss the results of efforts to determine realization rates for NEAT and to identify possible reasons for realization rates that differ from 100%. We will also assess the advisability of using NEAT predictions of savings as a substitute for evaluation results based on measured fuel consumption data.

Previous Research on Audit Realization Rates

In general, audit-predicted savings tend to be at least somewhat higher than measured savings. Engineering algorithms and models used to predict energy savings usually produce estimates of energy savings that are higher than those produced by the analysis of metered fuel consumption data (Nadel and Keating 1991). In their review of 42 impact evaluations conducted by electric utilities on a variety of types of conservation programs, Nadel and Keating found that engineering estimates of savings were typically higher than impact evaluation estimates (based on metered data). Among the 11 residential retrofit program evaluations they examined, for example, engineering estimates were significantly higher then measured savings estimates in all but two cases. In 8 of the 11 cases, the realization rates were less than 50%. Nadel and Keating (1991) report rates as high as 117% and as low as 15%. The mean realization rate in the 11 residential retrofit validation studies they reviewed was 44%, and the median was 40%. A more recent example of the tendency of engineering audits to estimate higher savings than impact evaluations can be found in a study of a utility's Low-Income Energy-Efficiency Program (ULIEEP) in which Barakat and Chamberlin (1995) found realization rates of 55% to 63%.

Another study of realization rates in the demand-side management programs of the four largest California utilities summarized the results of more than 50 evaluation studies of programs operating between 1990 and 1992 (Brown and Mihlmester 1994). The programs reviewed included efforts in several end-use sectors (including commercial, industrial, agricultural, and residential). Some of the programs offered direct assistance (measures are installed at no cost to the customer) and some shared savings incentives (where both the utility and the customer invest in measures and share the benefits of reduced bills). Brown and Mihlmester (1994) found that the lowest realization rates were in residential direct assistance programs (with a mean of 68% and a median of 53%). The realization rates for commercial, industrial, and agricultural programs were higher than residential realization rates, and the shared incentive programs had higher realization rates than the direct assistance programs.

As Nadel and Keating (1991) point out, many factors affect the relationship between engineering estimates of savings and measurements of savings in impact evaluations. In several studies that they reviewed, the discrepancies between predicted and realized savings were attributed to post-retrofit decreases in the use of secondary fuels. For example, houses using electricity as a primary heating fuel and wood as a secondary fuel began using much less wood after weatherization. Because significantly less wood was used after retrofit, the measured decreases in electricity consumption were small, much smaller than had been predicted. In other validation studies, discrepancies were attributed to inaccurate assumptions used in the engineering calculations and to performing calculations on prototype buildings without considering how much energy the occupants of the actual buildings were using. Quality control problems in measure installations were another identified cause of low realization rates.

Previous research shows that the National Energy Audit (NEAT), like many other residential engineering energy audits, has typical realization rates well below 100% (Sharp 1994; Dalhoff 1996). A recently completed report provides an additional measurement of NEAT's realization rate by using a data set from the New York gas utility's low-income weatherization program. This study estimates NEAT's realization rate by comparing NEAT predictions for 49 houses in the New York utility's program to the results of a billing analysis based on the Princeton Scorekeeping Method (PRISM) for the same houses. Using this data set, the reasons for any observed discrepancies between the NEAT-predicted and measured savings also are explored (Gettings et al. 1998).

Many Potential Sources of Error in the Determination of Realization Rates

It is difficult to determine a realization rate with precision because the results are influenced by so many possible sources of error. Some potential external and internal sources of error are listed below.

External error sources may include the following:

- *Errors in audit input data*, such as differences between actual weather at the building site and the weather input to the model or errors in obtaining and recording building characteristics and thermal properties.
- *Errors in application*, such as applying the audit to prototype buildings without considering how much energy the actual building is using.
- *Errors in implementation*, such as installing measures that were not recommended by the audit, not installing measures that were recommended by the audit, or poor quality of installations. The quality of installations can be affected both by contractor errors and by details of construction that make it impossible to achieve the level of impact assumed by the audit. For example, when wall insulation is installed, NEAT assumes that complete

cavity coverage is achieved. In reality, achieving an even distribution over the entire area may not be possible.

- *Errors due to the influence of occupant behavior,* such as changes in household composition, appliance stock, usage habits, fuel switching, or other differences between actual occupant behavior and the behavioral patterns assumed by the model. The reduced post-weatherization usage of secondary heating fuels has been identified as a reason for lower than predicted savings in several studies.
- *Errors in billing data,* such as missing or incomplete data, poorly fitted models, or extreme outliers among the data points.

Internal sources of error may include the following:

- *Errors in assumptions* about physical mechanisms, such as thermal transfer mechanisms, or about thermostat set-points, R-values, internal gains, etc. One incorrect assumption made by most engineering models is the empty house assumption. Like most engineering audits, NEAT essentially models an empty house and assumes that walls and floors are bare. In reality, homeowners often have a portion of their walls or floors covered by furniture, shelving, cupboards, drapes, decorations, storage (particularly in attics or basements), etc. Some of these items may add resistance to heat flow through these surfaces that is not accounted for in the engineering calculations. The higher the pre-retrofit R-value of a building component, the less actual savings are obtained from adding more insulation to the component. This is particularly true when the initial value is low. Thus, actual savings from insulating building components can be less than those predicted based on the assumption of empty houses.
- *Errors in internal algorithms,* such as errors in numerical solutions or in coding.

There are many possible levels of model validation, depending upon how much control over the potential sources of error one implements. The simplest form of validation compares actual building energy use to model predictions with no attempt to control sources of error. Even if good agreement is obtained in such a comparison, however, it is difficult to interpret the results because all of the possible sources of error are operating simultaneously and may be offsetting in their effects. In general, realization rates tend to be less than 100% because models assume ideal or perfect conditions, while in reality many sources of error are present.

In our study, we paid particular attention to evaluating data quality and potential error sources. We were able to exercise some control over external sources of error. For example, we corrected input errors in both the audit and billing data whenever possible and concluded that, for our sample, this was a relatively minor source of error. We also checked for implementation errors and found them to be minimal with respect to the match between recommended and installed measures. However, we had no information on installation quality, so this remains an uncontrolled and potentially important source of error. We had limited information with which to assess behavioral effects. For example, the utility database recorded the average thermostat set points indicated by the occupants during the initial interview but had no information regarding possible changes in these settings after the retrofit. The use of secondary fuels seemed to be an important behavioral influence on realization rates, but no data on the extent of secondary fuel use were available. We were unable to assess other behavioral effects.

To assess the probable importance of internal sources of error, we compared NEAT's predictions of energy consumption and savings with the predictions of the DOE-2.1E model. DOE-2.1E is an hourly energy use simulation model that is a more sophisticated and detail-intensive tool than NEAT. DOE-2.1E, which is an industry standard, is designed for the skilled energy modeler/researcher rather than the residential energy auditor. To the extent that the two models produce similar predictions, one can conclude that NEAT's internal assumptions and algorithms are sound.

FINDINGS

NEAT-Predicted Savings

NEAT predictions of annual savings based on the corrected input files, without billing adjustment, averaged 104.7 MBtu per dwelling. Adjustment of NEAT predictions of savings to the amount of consumption shown in the billing data produced lower predicted savings of 88.1 MBtu per dwelling, as is shown in Table 1. The predicted savings with billing data adjustment are used to calculate the realization rates shown in Table 2. Most NEAT users do not use the optional billing data adjustment feature because of the extra effort involved in obtaining and entering billing data. Our study showed, however, that the use of this feature will move predicted savings closer to actual savings. In addition, having access to billing data is useful for targeting more resources to dwellings with the most potential for savings and as a baseline for later evaluations. In the two previous studies of NEAT realization rates, the one in North Carolina used the billing data adjustment feature, while the one in Iowa did not (Sharp 1994; Dalhoff 1996). There was little difference in the results of these two studies. One reported a realization rate of 57% (Sharp 1994) and the other a rate of 54% (Dalhoff 1996).

NEAT's Realization Rates

To determine NEAT's realization rate, we compared predicted savings to the estimated savings obtained by applying the Princeton Scorekeeping Method (PRISM) to the billing data. PRISM is a computer program developed by the Center for Energy and Environmental Studies of Princeton

TABLE 1
NEAT Predictions of Savings with and without Billing Data Adjustment

	With Billing Data Adjustment	Without Billing Data Adjustment
Average Predicted Savings (in MBtu)	88.1	104.7

TABLE 2
Realization Rates Calculated with Data Sets with Varying Levels of Data Quality

Data Sets	Average Realization Rate
High Data Quality (n-12)	69%
High and Medium Data Quality	57%
All House—High, Medium, and Poor Data Quality (n=49)	66%

University. It analyzes monthly residential utility bills, normalizing energy use to a standard weather, to allow comparison of pre- and post-retrofit consumption to estimate savings resulting from weatherization. The process of checking for errors in NEAT input files and billing data files made it apparent that some houses had more consistent, reasonable, and understandable information than others. Efforts were made to improve the correctness and consistency of both data sets where possible. Nevertheless, a good deal of variation in data quality remained. Because data were available on only 49 houses, we did not wish to drop any of them from the sample. On the other hand, realization rates for individual houses varied from less than zero to more than 200%, and it seemed important to consider how realization rates would change as data quality improved or declined. Accordingly, a scheme was developed for rating the overall quality of the data upon which a dwelling's realization rate would be calculated.

Separate criteria were established to rank the confidence placed in the quality of the building input files for the NEAT runs and to rank the confidence placed in the quality of the metered consumption data that were the input to PRISM. The criteria used for the NEAT input files included whether or not the house layout could be determined from the input data and the number and significance of the inconsistencies or problems. For example, in some homes, the totals of ceiling and floor areas did not match or the recorded furnace efficiencies were unusually high or low. For the billing data, the criteria were the R^2 from the PRISM run, the magnitude of difference between the estimated pre- and post- base loads, the degree of scatter, and the amount of adjustment, if any, made to outlying data points. An example of an adjustment in the billing data occurred when monthly use was totally inconsistent with the surrounding months' consumptions and the weather data. This may have resulted from an error in reading the meter or an actual situation, such as the family temporarily leaving the home for vacation.

After the separate ratings for building data and billing data quality were completed, they were combined into a single index. Using this index, three groups of houses were defined: dwellings with high, medium, and low data quality. It was expected that the highest data quality group would have the highest realization rate. However, as Table 2 shows, the realization rates (with billing data adjustment) turned out to be about the same for the high data quality group (n = 12) and for the total sample of 49 houses (69% and 66%, respectively). The average realization rate for the combined set of high and medium data quality houses is somewhat lower at 57%.

Because it was surprising to see little relationship between data quality and realization rates, a closer examination of the patterns of variation seemed warranted. When the standard deviations associated with the realization rates for the three groups of houses (dwellings with high, medium, and low data quality) presented in Table 2 were examined, they showed that there was no statistically significant relationship between data quality and realization rates. In other words, the standard deviations (which were as large as the mean values) for the three groups were so great that the differences among the group means were not significant.

The patterns of variation in the realization rates, when examined for smaller subsets of dwellings, showed that the lack of statistical association between realization rates and data quality occurred primarily because realization rates above 100% were concentrated in the houses with the poorest data quality. Among the ten houses with the lowest data quality, more than half had realization rates above 100%. Among the houses with high or medium data quality, in contrast, only about 15% of the houses had realization rates above 100%. Therefore, when the houses with poor data quality are dropped and only the medium and high data quality houses remain, the average realization rate decreases from 66% to 57% while the standard deviations remain unchanged.

When both the poor and medium data quality houses are dropped and only the 12 houses with the highest data quality are considered, the realization rate increases to 69%. Although the standard deviation for the 12 houses with high data quality was about the same as for the other two groups shown in Table 2, most of the variation occurs in the homes with lower data quality rank within this group. In particular, among the five houses in ranks 8 through 12, realization rates vary from 2% to 216%. Among the seven houses with the best data quality (ranks 1 through 7), in contrast, the range is from 21% to 138%. Only one house among the top seven has a realization rate above 100%, and this group's standard deviation is comparatively low (less than 0.36).

Realization rates for several other subsets of dwellings were examined, too. For example, if all of the 11 dwellings, out of the total sample of 49, that had realization rates above 100% are excluded, the average realization rate falls to 42%. If the 11 dwellings with the lowest realization rates are excluded (this subset included 8 houses with realization rates of less than 25%), the group average rises to 83%. Finally, if the 11 houses from both the highest and the lowest part of the distribution are excluded, then the average realization rate becomes 54%.

Clearly, realization rates vary a great deal depending upon which houses the analyst decides to exclude from the sample. It is unclear which set of houses provides the best estimate of NEAT's realization rate. When only three groups were examined (dwellings with high, medium, and low data quality), there was no association between data quality and realization rates. A closer examination of smaller subsets of houses suggested that higher data quality may be associated with lower standard deviations.

The overall distribution of PRISM estimates of savings vs. NEAT predictions of savings for our New York sample is shown in Figure 1. Clearly, NEAT tends to predict higher savings than are actually realized. This is the case even if the extreme values are eliminated from the plot shown in the figure. Plots analogous to Figure 1 using subsets of houses corresponding to the various data quality groupings revealed no additional information of value.

ANALYSIS OF REASONS FOR VARIATIONS IN REALIZATION RATES

Previous studies of NEAT realization rates found that the audit-predicted savings were more closely matched to actual savings in some types of houses than in others. Sharp (1994) concluded that for houses with pre-weatherization consumption of less than 80 MBtu and audit-predicted savings of less than 40 MBtu, the NEAT predictions were, in most cases, very close to the measured values. On the other hand, in houses with pre-weatherization consumption of more than 80 MBtu and audit-predicted savings of more than 40 MBtu, NEAT-predicted consumption and savings were usually higher than measured values. Dalhoff (1996) reported that much of the difference in audit-predicted savings and measured savings in Iowa occurred because of NEAT's tendency to overpredict pre-weatherization consumption and to underpredict post-weatherization consumption. He suggests that NEAT overpredicts consumption for houses with lower overall thermodynamic efficiency and underpredicts it for houses with better efficiency. Our New York data are not consistent with the idea that NEAT overpredicts consumption in houses with lower efficiency. The average measured pre-weatherization consumption, based on a PRISM analysis, for the New York gas utility program houses was 266 MBtu. NEAT predictions of average pre-weatherization consumption were very close to the PRISM values at 254 MBtu. Thus, we saw a slight tendency to underpredict pre-weatherization consumption. This probably occurred because the gas utility program targeted very high usage customers. On the other hand, as is consistent with Dalhoff's results, we, too, saw a tendency to underpredict post-weatherization consumption. The PRISM-based value for post-weatherization consumption was 213 MBtu, while the average NEAT prediction for post-weatherization consumption was 149 MBtu. The discrepancy between measured and NEAT-predicted savings occurred because of

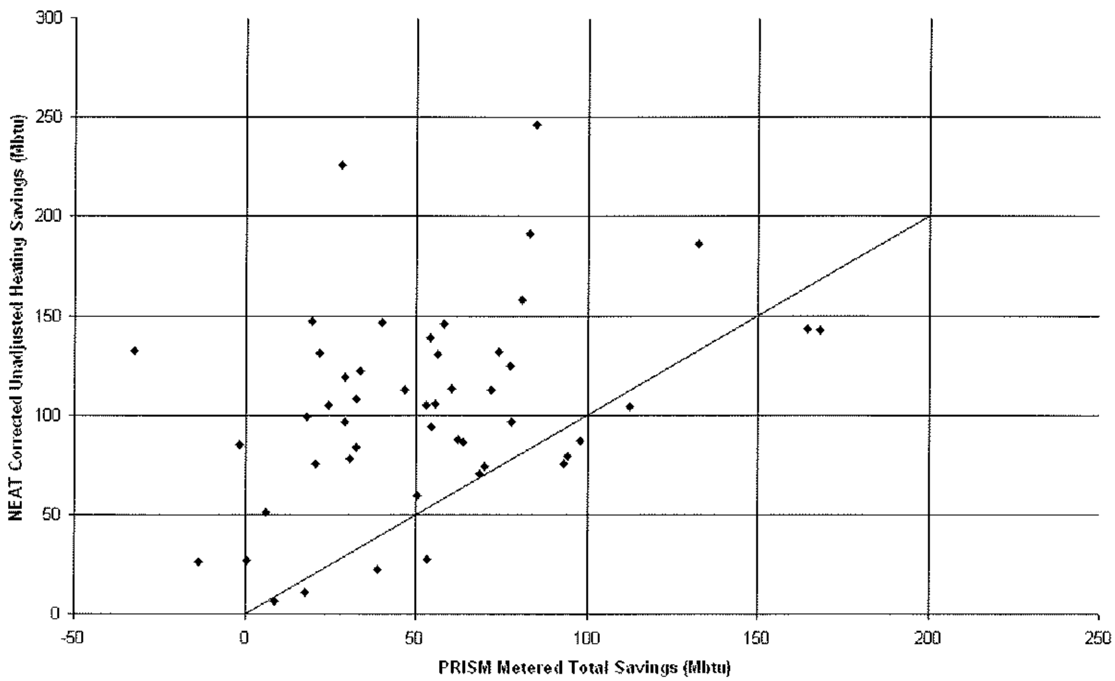

Figure 1 Comparison of savings estimated in PRISM analysis to savings predicted by NEAT.

the significantly higher than predicted post-weatherization consumption.

In this study, several subgroups of dwellings were examined to see if certain subgroups had higher or lower realization rates than others. We found few significant relationships. Realization rates did not appear to be associated with the amount or type of insulation installed, nor were they correlated with pre-weatherization consumption levels or the pre-weatherization energy use index of Btu per square foot per heating degree-day. Other factors we examined, which showed little or no relationship to realization rates, included the heated floor area, the number of occupants, and the thermostat set point. The one factor that appeared to have a strong relationship to the realization rates was the use of secondary heating fuels. Among the 41 houses with no use of secondary heating fuels, realization rates averaged 70%. Among the eight houses that used a secondary heating fuel, the average realization rate was 48%. Thus, our data, like previous studies of realization rates, suggested that post-weatherization decreases in the use of secondary fuels may be an important source of error in the determination of realization rates. Unfortunately, the data indicated only that secondary heating sources existed, but not to what degree they were utilized. Thus, no adjustment could be made in the consumption data to reveal the extent of the effect this factor had on the realization rates of individual houses.

COMPARISONS OF NEAT VS. DOE-2.1E MODELING RESULTS

To assess the accuracy of NEAT''s internal algorithms, we compared its results to those obtained with another model, DOE-2.1E, which is an industry standard. Using identical building descriptions, the same two test houses were analyzed with both models. The primary goal was to assess the accuracy of NEAT's engineering algorithms by comparing their performance to the results obtained with DOE-2.1E. A secondary goal was to assess how well each model's predictions matched the actual savings measured with billing data. The methods of model setup and analysis are discussed in detail in Gettings et al. (1998).

The pre-and post-retrofit consumption and the estimated savings obtained with the two engineering models and with a billing analysis are compared below for two test houses. Comparisons of the NEAT and DOE-2 results make it possible to assess NEAT's ability to accurately estimate the savings associated with individual measures as well as the savings associated with the complete package of retrofit measures.

Pre- and Post-Retrofit Consumption and Estimated Savings Comparisons for the Three Methods

In both houses, the two engineering models and the billing analyses all produced very similar results for pre-retrofit consumption. This result is significant because it demonstrates that the basic algorithms and default data used by NEAT are reasonable and that the collected data are sufficient to define a building's energy use characteristics, at least for these two case study buildings. This also means that savings estimates calculated later should not be biased by patterned high- or low-use baseline models. The fact that the engineering model simulations of pre-retrofit consumption match so well, both with each other and with the billing data, suggest that the basic engineering algorithms in NEAT are sound. It must be noted that the utility database indicates the presence of an electric space heater as a secondary heat source in the second house. Again, however, no estimate of the extent of use is available.

Although the pre-retrofit consumption estimates agreed well across methods, there was less agreement among the savings estimates (Table 3). In particular, for the first house, NEAT predicted savings of 966 ccf, while DOE-2 predicted savings of 1,024 ccf, which is a good match. The billing analysis estimate of savings, however, was lower than either model's prediction at 795 ccf. In the second house, the NEAT estimate of savings was 260 ccf and the DOE-2 estimate was 253 ccf, again showing excellent agreement. The savings estimate from the billing analysis, however, at 163 ccf was much lower than either of the model predictions.

In the first house, the two engineering models produced similar estimates of post-retrofit consumption (938 ccf and 1102 ccf), while the billing analysis estimated higher post-retrofit consumption (1252 ccf) than either of the models. Here, the billing analysis estimate of savings was lower than the savings estimated by the models mainly because of the higher post-retrofit consumption shown by the billing analysis. For the pre-retrofit consumption estimates, the billing analysis results (2047 ccf) fell between the model estimates of 1904 ccf and 2126 ccf.

In the second house, the two engineering models produced estimates of post-retrofit consumption (1456 ccf and 1625 ccf) that fell on either side of the billing analysis estimate

TABLE 3
Comparisons of Pre-Retrofit Consumption, Post-Retrofit Consumption, and Savings Predictions Obtained from NEAT, DOE 2.1 and Billing Data

	NEAT	DOE 2.1	Billing Data
House 1 (ccf/year)			
Pre-Retrofit Consumption	1904	2126	2047
Post-Retrofit Consumption	938	1102	1252
Savings	996	1024	795
House 2 (ccf/year)			
Pre-Retrofit Consumption	1716	1878	1699
Post-Retrofit Consumption	1456	1625	1536
Savings	260	253	163

of 1536 ccf. The pre-retrofit consumption estimate from the billing analysis (1699 ccf) was lower than the model estimates of 1716 ccf and 1878 ccf. This lower pre-retrofit consumption for the billing analysis, without a correspondingly lower post-retrofit consumption, causes the savings estimate from the billing analysis to be less than those from the models.

In summary, the similarity of NEAT and DOE-2 results suggests that typical NEAT realization rates of less than 100% are probably not due to inaccuracies in its internal engineering algorithms. The lower savings estimated by the billing analysis, as compared to both of the engineering models, are likely due to implementation and behavioral factors.

Measure-by-Measure Comparisons

Billing analysis cannot separate the effects of individual weatherization measures. However, the separation of effects by measure can be explicitly performed with NEAT and DOE-2. Comparisons of the measure-by-measure savings predicted by the models in each of the case study houses showed excellent agreement. Generally, the measure-by-measure examination revealed comparable savings, with a 28% standard deviation of the absolute difference between the models. Variation was lower for straightforward measures (infiltration reduction, envelope insulation) and higher for measures with complex interactions among building components (duct insulation, heating system replacement, rim joist insulation). This was most likely due to the fact that the two modeling systems deal with the complex interactions in somewhat different ways (Gettings et al. 1998).

NEAT IS A VALUABLE TOOL FOR SELECTING MEASURES

NEAT's value for selecting cost-effective measures and improving program performance has been demonstrated in an experimental study in North Carolina (Sharp 1994) and in a meta-evaluation of state-level program evaluations (Berry 1997). In the North Carolina study, the introduction of NEAT increased heating energy savings from 18% to 23% (Sharp 1994). In her meta-evaluation, Berry (1997) attributes much of the recent improvement in program performance to the post-1989 introduction and now widespread use of advanced audits. In 1989, the program was not yet using advanced audits. Today more than half of the states use them. States that use NEAT and other advanced audits achieve higher savings than those that do not (Berry 1997).

In all the different NEAT runs we conducted, with and without various modifications to the setup files, and various corrections to the input data and assumptions, the same measures were almost always recommended in the same order. The fact that the set and sequence of recommended measures are insensitive to reasonable variations in input parameters and setup configurations suggests that NEAT's ability to rank the cost-effectiveness of measures is not usually compromised by many of the sources of error that may affect realization rates. Since performing this measure selection process is NEAT's primary purpose, its value as a tool in the field is not seriously affected by the finding that its typical realization rates are well below 100%.

Although the set and order of recommended measures show little sensitivity to realization rates, the intensity or frequency of recommendations may be sensitive to realization rates. Dalhoff (1996) reported that, when realization rates are low, NEAT may recommend more attic insulation and wall insulation than is actually cost-effective, at least by that measure of cost-effectiveness used in the program. It also may indicate greater cost-effectiveness for infiltration reduction work than is actually obtained. He also noted that the cost-effectiveness of attic insulation and infiltration reduction measures was more sensitive to overestimates of savings than was the cost-effectiveness of wall insulation.

Our data show that the cost-effectiveness of foundation and wall insulation measures was most sensitive to variations in realization rates, mostly because these measures tended to have SIRs closer to 1.0. Because the ranking of measures is not sensitive to variations in realization rates, the potential effects of overpredicted savings can be reduced by installing only the higher ranked measures. Alternatively, setting a higher cutoff value for the SIR in the NEAT setup files would eliminate questionable measures. With an assumed realization rate of 50%, the measure cutoff SIR should be set to 2.0. With an assumed realization rate of 70%, the SIR should be set to 1.4. This will eliminate recommendations to install those measures of questionable cost-effectiveness.

AUDIT-PREDICTED SAVINGS NO SUBSTITUTE FOR EVALUATIONS BASED ON MEASURED DATA

In general, realization rates tend to be less than one because engineering models assume ideal or perfect conditions that do not fully reflect a more complex reality. For example, audits assume that installation quality meets ideal standards, while real installations are imperfect. Similarly, audits assume no changes in occupant behavior, while real occupant behavior often changes significantly. Audits assume empty houses or houses whose envelope surfaces are unencumbered with a wide variety of furnishings, decorations, and stored items. Because of the many potential sources of error in the determination of realization rates, audit predictions of savings are not a reliable substitute for evaluations based on measured fuel consumption data. Even the application of a correction factor to audit predictions is problematic because the magnitude of the gap between predicted and actual savings is not constant from one situation or application to another. In other words, because realization rates vary widely with different samples of houses taken from different locations and time periods, it is not possible to specify a universally applicable correction factor.

The fact that the estimates of NEAT's realization rate reviewed in this paper varied from 42% to 83% means that unvalidated NEAT predictions of savings should not be used to characterize program performance. Engineering audit predictions of savings are not accurate substitutes for evaluations based on measured fuel consumption data.

REFERENCES

Barakat and Chamberlin. 1995. Impact and process evaluation of the Consolidated Edison Utility Low-Income Energy Efficiency Program (ULIEEP): Second year report, final. Prepared for Consolidated Edison, New York.

Berry, L. 1997. State-level evaluations of the weatherization assistance program in 1990-1996: A metaevaluation that estimates national savings. ORNL/CON-435, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Brown, M.A., and P.E. Mihlmester. 1994. Summary of California DSM impact evaluation studies. ORNL/CON-403, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Dalhoff, G. 1996. An evaluation of the performance of the NEAT audit as implemented in the Iowa weatherization program. Prepared by Greg Dalhoff and Associates for the Wisconsin Energy Conservation Corporation.

Gettings, M.B. 1990. The North Carolina field test audit: Description and operation, Draft version. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Gettings, M.B., L.G. Berry, J.B. Maxwell, and M.A. Beyer. 1998. Validation of the National Energy Audit (NEAT) with data from a gas utility residential weatherization program for low-income customers in New York State. ORNL/CON-457, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Krigger, J., N. Adams, and M. Gettings. 1997. *National Energy AudiT (NEAT) users manual.* Oak Ridge National Laboratory, ORNL/Sub/91-SK078/2.

Nadel, S., and K. Keating. 1991. Engineering estimates vs. impact evaluation results: How do they compare and why? pp. 24-34 in Energy program evaluation: Uses, methods, and results. 1991 International Energy Program Evaluation Conference, Proceedings, CONF-910807, Fifth International Conference in Chicago.

Sharp, T. 1994. The North Carolina field test: Field performance of a preliminary version of an advanced weatherization audit for the Department of Energy's Weatherization Assistance Program. ORNL/CON-362, Oak Ridge National Laboratory, Oak Ridge, Tenn.