U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

BUILDING TECHNOLOGIES OFFICE

Field Assessment of Energy Audit Tools for Retrofit Programs

J. Edwards, D. Bohac, C. Nelson, and I. Smith
NorthernSTAR Building America Partnership

July 2013

Building America
U.S. Department of Energy

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

Printed on paper containing at least 50% wastepaper, including 20% postconsumer waste

Field Assessment of Energy Audit Tools for Retrofit Programs

Prepared for:

The National Renewable Energy Laboratory

On behalf of the U.S. Department of Energy's Building America Program

Office of Energy Efficiency and Renewable Energy

15013 Denver West Parkway

Golden, CO 80401

NREL Contract No. DE-AC36-08GO28308

Prepared by:

J. Edwards, D. Bohac, C. Nelson, I. Smith

Northern*STAR* Building America Partnership

University of Minnesota

2004 Folwell Avenue

St. Paul, MN 55108

NREL Technical Monitor: Stacey Rothgeb

Prepared under Subcontract No. KNDJ-0-40338-02

July 2013

Acknowledgments

This report was prepared for the U.S. Department of Energy Building America program. Funding for this work was provided by the National Renewable Energy Laboratory under contract KNDJ-0-40338-02. Additional funds to support this work were provided by the University of Minnesota and the Initiative for Renewable Energy and the Environment.

The primary authors were Jennifer Edwards, Program Manager at Center for Energy and the Environment (CEE); Dave Bohac, Director of Indoor Air Quality at CEE, Carl Nelson, Program Manager at CEE, and Isaac Smith, Program Assistant at CEE.

The authors would like to acknowledge the contributions and support of our colleagues. Thanks first and foremost to Casey Hilbelink, Steven Schmidt, and Kirk Kohlema for conducting the field work and data collection, and Bob Mello and Heather Hanson for managing the residential field crew. Thanks additionally to our CEE co-workers Lester Shen and Rich Szydlowski for their technical input, and Joshlyn Goepfrich for scheduling home visits. Thank you to Joan Glickman at the Department of Energy for the opportunity to participate in the Home Energy Score pilot, and to Glenn Dickey, Kolleen Kawa and Hannah Wood at SRA International for managing the pilot process. Thanks also to Greg Nahn at WECC for technical assistance and training on REM/Rate. And finally, thank you to Pat Huelman and Tom Schirber at the University of Minnesota for assistance with project management and report editing.

Contents

List of Figures

Unless otherwise noted, all figures were created by NorthernSTAR.

List of Tables

Unless otherwise noted, all tables were created by NorthernSTAR.

Definitions

ACH	Air changes per hour
ALR	Air leakage ratio
CDD	Cooling degree day
CEE	Center for Energy and Environment
cfm50	Cubic feet per minute at 50 Pascals pressure
DOE	U.S. Department of Energy
HDD	Heating degree day
HERS	Home Energy Rating System
EPS	Energy Performance Score
kWh	Kilowatt-hour
RESNET	Residential Energy Services Network

Executive Summary

This project focused on the use of home energy ratings as a tool to promote energy retrofits in existing homes. A home energy rating provides a quantitative appraisal of a home's energy performance, usually compared to a benchmark such as the average energy use of similar homes in the same region. Home rating systems can help motivate homeowners in several ways. Rating systems based on energy performance models, the focus of this report, can establish a home's achievable energy efficiency potential and provide a quantitative assessment of energy savings after retrofits are completed, although their accuracy needs to be verified by actual measurement or billing data. Ratings can also show homeowners where they stand compared to their neighbors, thus creating social pressure to conform to or surpass others.

There are several potential applications for home ratings, and the important characteristics—e.g., speed, accuracy, and clarity—will depend on the type of transaction one is trying to influence. One important consideration for rating tools aimed at the retrofit market is how they will integrate with existing home energy service programs, where technicians perform in-home audits. For residential programs that target energy savings only, home visits should be short and focused on key efficiency measures for that home. In order to gain wide adoption, a rating tool must be easily integrated into the field process, demonstrate consistency and reasonable accuracy to earn the trust of home energy technicians, and have a low monetary cost and time hurdle for homeowners.

This project field-tested three different building performance models of varying complexity, in order to assess their value as rating systems in the context of a residential retrofit program. The major focus was the Home Energy Score, which was under development by the U.S. Department of Energy (DOE) at the time of this project. It is designed to give a complete home performance assessment while simplifying the building measurements to 36–67 data inputs, depending on a home's configuration. The Center for Energy and Environment (CEE) was one of nine national pilot sites in the spring of 2011, and tested the Home Energy Score on 154 Minnesota homes. The goal of these pilots was to provide feedback to the DOE about technician experience, homeowner reaction, and the pattern of scores in a given region of the country, so that DOE could make necessary changes before the national launch of the tool. Numerous changes have been made to the national version, launched in June 2012, including adjustments for several issues raised though this analysis.

This pilot also evaluated the energy modeling performance of SIMPLE and REM/Rate. SIMPLE, developed by Michael Blasnik and Associates, is a spreadsheet-based home energy model that runs on fewer than 50 streamlined inputs. It uses broad classifications for certain home characteristics and allows field technicians to switch between estimations and diagnostic measurements, depending on the scenario. REM/Rate, based on an audit protocol developed by the Residential Energy Services Network (RESNET), requires more detailed building inputs than either SIMPLE or the Home Energy Score. In particular, REM/Rate uses detailed construction characteristics for individual wall, window, foundation, and attic areas to characterize the building shell in detail.

The energy performance of the three models showed similar trends, and importantly, the more detailed characterization of a home did not provide a better estimate of a home's asset energy use:

- Overall, there was systematic overprediction of gas use by the Home Energy Score, SIMPLE, and REM/Rate, although the SIMPLE model results were the closest to utility bill data. The SIMPLE model overpredicted gas usage by 18% on average, compared to 55% and 63% for the Home Energy Score and REM/Rate, respectively.
- The Home Energy Score, SIMPLE, and REM/Rate all overpredicted electricity consumption, and had a larger minimum baseline electricity load than actual usage. The Home Energy Score overpredicted electricity by 23% on average. SIMPLE's average overprediction was 29% and 25% for two alternate versions of the model, and REM/Rate overpredicted by 27% on average. These results include a small number of homes with very high actual electricity use.
- These asset models have similar levels of correlation for predicting natural gas use, from a low R-squared of 0.42 for the Home Energy Score to a high of 0.52 for REM/Rate. The asset models all have a low R-squared values for predicting electricity use, from 0.05 for the Home Energy Score to 0.30 for REM/Rate. The validity of this level of correlation will depend on the application at hand, but it urges caution in interpreting and presenting model results.
- We estimate that, when used as a stand-alone tool, the length of time required for a trained technician to collect and enter data for one home would be approximately 1.5–2 hours for the Home Energy Score, 1 hour for SIMPLE, and 4–6 hours for REM/Rate.

In CEE's experience, the Home Energy Score added 30 minutes of field time to the current 1–1.5 hour residential home visit, as well as 30 minutes for data entry. CEE's home visits are conducted by two field staff. The most time-consuming entry was measuring a home's total window area. The biggest challenge was how to classify certain home features common in the Midwest, in particular, unfinished basements and story-and-a-half style homes. The results showed that homes could not significantly improve their scores if they made the recommended investments, which could prove a challenge for motivating homeowners. Homes scored an average of 5.4 on the Home Energy Score, and implementing recommended upgrades would increase their score by an average of 2 points. No homes had an initial score greater than 8, and none could score a 10 after upgrades. This point distribution is by design, and will be updated before the final version is released.

In addition, the Home Energy Score generates a list of upgrade measures that have a 10-year payback or less for that particular home. CEE compared these automated recommendations to the in-person recommendations made by the field crew, and found that field crews identified cost-effective upgrades 33% more often than the Home Energy Score. Notably, the types of recommendations given by the Home Energy Score did not correlate well with the in-person recommendations. The field crew gave 309 unique recommendations for air sealing, attic and wall insulation, and the Home Energy Score recommended these measures 150 times, or half as frequently. Part of the reason is that field staff can diagnose when small portions of the home require upgrades; e.g., a single side attic, while the Home Energy Score is not as specific.

The results of this pilot favor the streamlined, lower cost approach of SIMPLE and the Home Energy Score for energy retrofit applications. However, each tool must balance the streamlined data collection with a thorough upfront protocol for how to classify homes consistently across a given region. Importantly, it would be insufficient to rely on automated recommendations from a building performance model without the direct input of field staff to diagnose the specific condition of a home. CEE also concluded that the Home Energy Score point system did not provide enough motivation for homeowners to invest in major upgrades, since very few homes could achieve a high score.

1 Overview of Minnesota Pilot

This project focuses on the use of home energy ratings as a tool to promote energy retrofits in existing homes. A home energy rating provides a quantitative appraisal of a home's asset performance, usually against a benchmark such as the average energy use of similar homes in the same region. A rating can motivate homeowners in several ways. First, a rating can clearly communicate a home's achievable energy efficiency potential. Second, it can provide a quantitative assessment of energy savings after retrofits are completed. This is valuable information for a homeowner deciding whether to invest several thousand dollars in their home. And finally, it lets homeowners know how their home rates compared to their neighbors, thus creating pressure to conform to a social standard.

A growing number of residential energy efficiency programs rely on comparative feedback to indicate how a home performs relative to similar homes in the area. Some typical examples are based on actual energy usage from utility bills, such as Opower's Home Energy Reports, which show utility customers how their monthly usage compares to similar homes in their area (Opower 2011). Reviews have shown that these techniques can lower use by approximately 1%–3% as occupants adjust their habits in response to increased awareness (Davis 2011; Carroll et al. 2009). But while billing information is useful for giving residents an overall understanding of their home's performance, it can't be used to diagnose a home's major energy sinks, or to estimate the energy savings from specific retrofit measures.

A home's actual energy use is a result of the fixed building structure and systems (the "asset") as well as occupant habits, such as thermostat set points and the use of lighting or electronics. The goal of many rating systems is to decouple the effect of occupant behavior and rate the asset alone. Numerous quantitative rating tools model a home's asset energy use based on measured building characteristics. Several tools are commercially available, and range in cost and complexity from free Web-based self-assessment tools (e.g., EnergySavvy's online audit tool) to comprehensive tools that require a certified professional rater, such as the Residential Energy Services Network's (RESNET) Home Energy Rating System (HERS), which costs on the order of $400 per home. Each tool can offer distinct value within the home energy efficiency marketplace, such as program recruitment or new home certification for homebuyers.

There are numerous questions about how home rating tools can best drive home energy retrofits. In particular, it is important to consider how rating tools will integrate with existing home energy service programs. For some residential programs that target only energy savings, home visits should be short and focused on key efficiency measures for that home. Adding complexity and detail can reduce program cost-effectiveness and increase the monetary and time hurdle for homeowners to participate. In order to gain wide adoption, a rating tool must also be easily integrated into the field process, and demonstrate consistency and reasonable accuracy to earn the trust of home energy technicians. Some of the consequential features of a rating tool are whether it requires diagnostics versus estimations, the ability of a rater to make general characterizations versus individual measurements, and the accessibility of different data inputs during a typical home visit.

This project evaluated a new option for home rating tools, the Home Energy Score, which is under development by the U.S. Department of Energy (DOE). It is based on the Home Energy

Saver model developed by Lawrence Berkeley National Laboratory and is designed to give a complete home performance assessment while streamlining the building measurements to 36–67 data inputs, depending on a home's configuration. This project piloted the Home Energy Score for 154 Minnesota homes and examined its usability in the field, as well as its role as a motivational tool for homeowners to make retrofit upgrades. This pilot also compared the energy modeling performance of the Home Energy Score with two additional audit assessment tools, the SIMPLE model and REM/Rate.

2 Review of Home Energy Rating Systems

Numerous home energy rating programs exist or are under development. As of 2009, more than 30 countries had policies that mandated time-of sale certification or disclosure requirements for homes (Dunsky et al. 2009). One of the most extensive is the United Kingdom's Energy Performance Certificate, which is required for most residential or commercial buildings constructed, rented, or sold in the United Kingdom (see Figure 1 below).

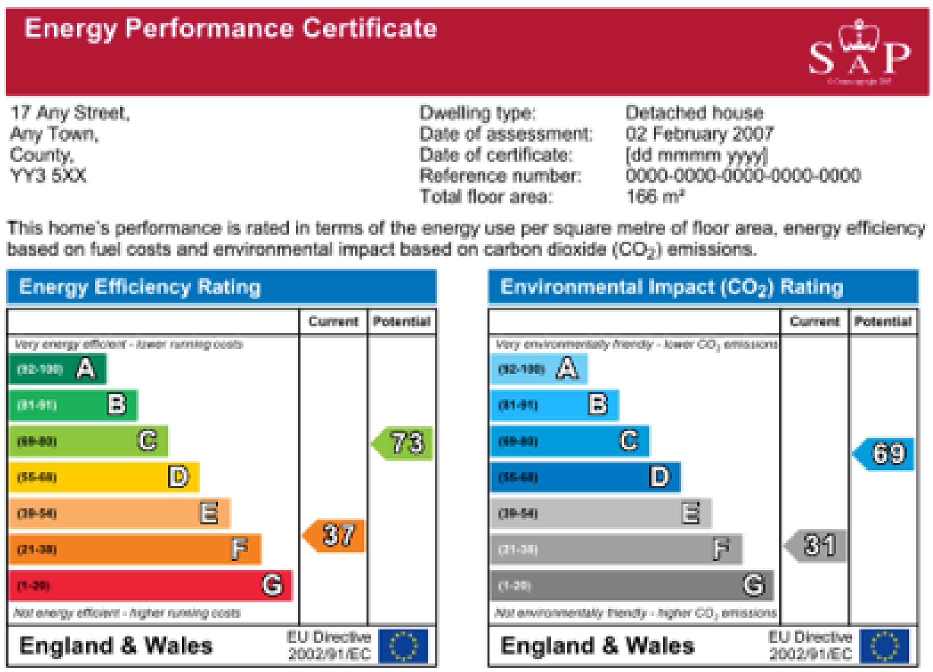


Figure 1. United Kingdom Energy Performance Certificate (Energy Saving Trust)

In the United States, mandatory disclosure of energy performance is less common, although there is a growth of activity at the state and local levels. A label's characteristics and complexity may change depending on the specific point of market intervention (Faesy 2010). Most rating tools have been applied to the new construction market. For example, RESNET's HERS index is used for DOE's EnergySmart Home Scale tool and is a common way for builders to demonstrate compliance with the U.S. Environmental Protection Agency's ENERGY STAR® program for new construction (see Figure 2 below).

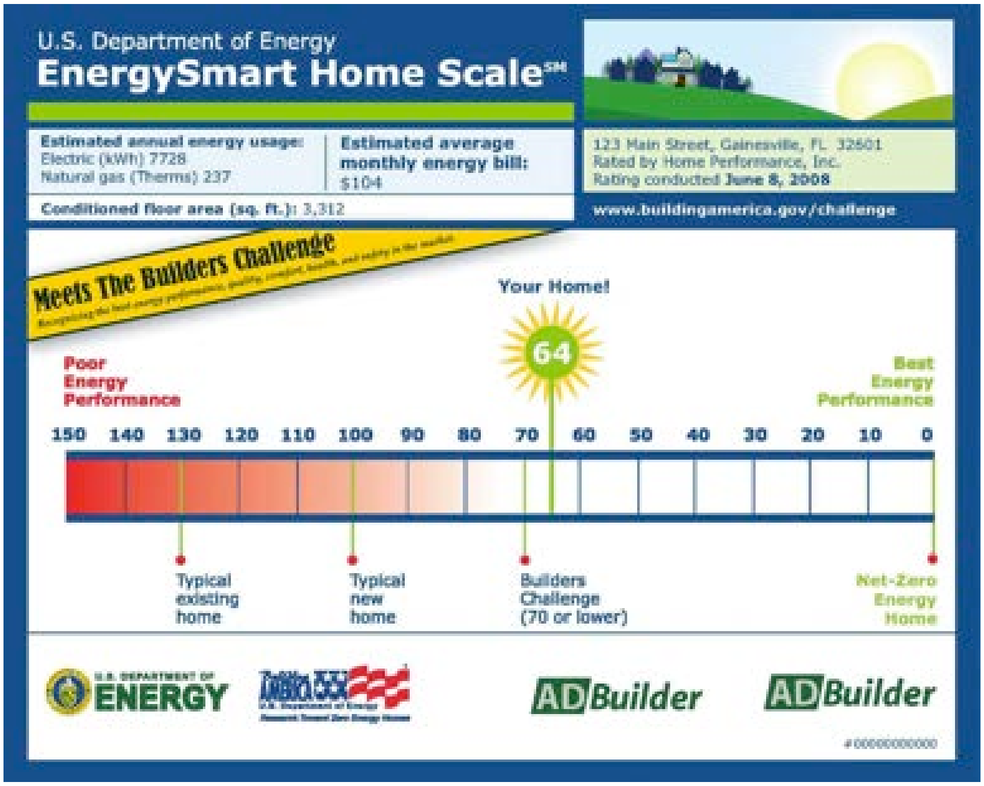


Figure 2. DOE EnergySmart Home Scale (DOE)

Energy rating systems have also been used in the existing home retrofit market, though this is a more challenging application than new home construction. An often-cited barrier to residential energy upgrades is their invisibility, which limits a homeowner's interest, as well as his or her ability to recoup the investment cost through an increased home value at time of sale. In theory, a rating makes unseen upgrades (such as wall insulation) more tangible and therefore creates a market value for the home amenity. One example of a rating system for the existing home market is the Energy Performance Score (EPS), developed by the Earth Advantage Institute and the Energy Trust of Oregon (Figure 3). The EPS was intended as a simplified energy audit tool and has been used or piloted in numerous residential retrofit programs (Earth Advantage 2012). The EPS is based on the SIMPLE model, which was also field-tested as part of this project.

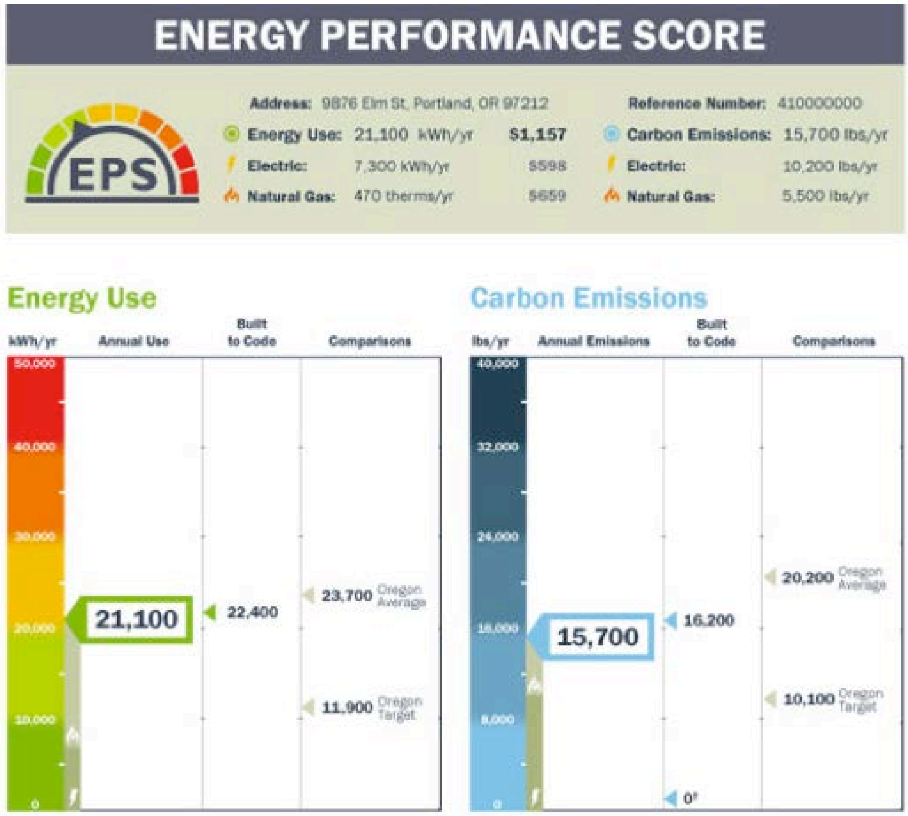


Figure 3. Earth Advantage's Energy Performance Score (Earth Advantage Institute)

Cost is a key concern in the retrofit market, whether assessments are part of a mandatory program or not. Homeowner surveys have shown that the maximum price people are willing to pay for a home asset score is \$100–\$200 (Earth Advantage Institute 2009, Newport Partners LLC 2011). A full-scale home audit is significantly more expensive, costing \$400 or more. Cost is particularly important for utility-sponsored programs, since they often must demonstrate cost-effectiveness as part of the regulatory approval process.

2.1 The Home Energy Score

In 2010, the DOE proposed a national residential rating program to provide a consistent rating for the home retrofit market through standardized training and certification, data collection protocols, a national home database, and a universal residential energy label (DOE 2010). DOE identified several important components of an effective national rating program. First, it should offer a tool that balances sufficient accuracy at a reasonable cost. It should use a consistent metric and method that is comparable across homes, yet allow for some flexibility at the local level. In addition, the rating must strive for a recognized brand identity, in the way ENERGY STAR is a recognized brand for consumer goods. DOE convened several focus groups to determine the most marketable information and design options, which resulted in the Home Energy Score design, described below (Newport Partners LLC 2011).

The Home Energy Score rates a home's asset performance on a scale of 1 to 10, with 10 being the best score (corresponding to the lowest energy use). The score design from the pilot phase is shown in Figure 4. The rating scale depends on a home's climate zone, but it is not adjusted to account for a home's size or age. Therefore, all else being equal, a larger home will receive a lower score than a smaller home. The report also shows a home's potential score if the homeowner completes the recommended retrofit upgrades. This potential and the corresponding annual dollar savings are calculated by estimating the energy savings from the tool's recommended cost-effective upgrades, discussed below. And finally, the score report gives a benchmark to compare the top 20% of similarly sized homes, defined as homes either larger or smaller than 2,200 ft^2 (LBNL 2010).

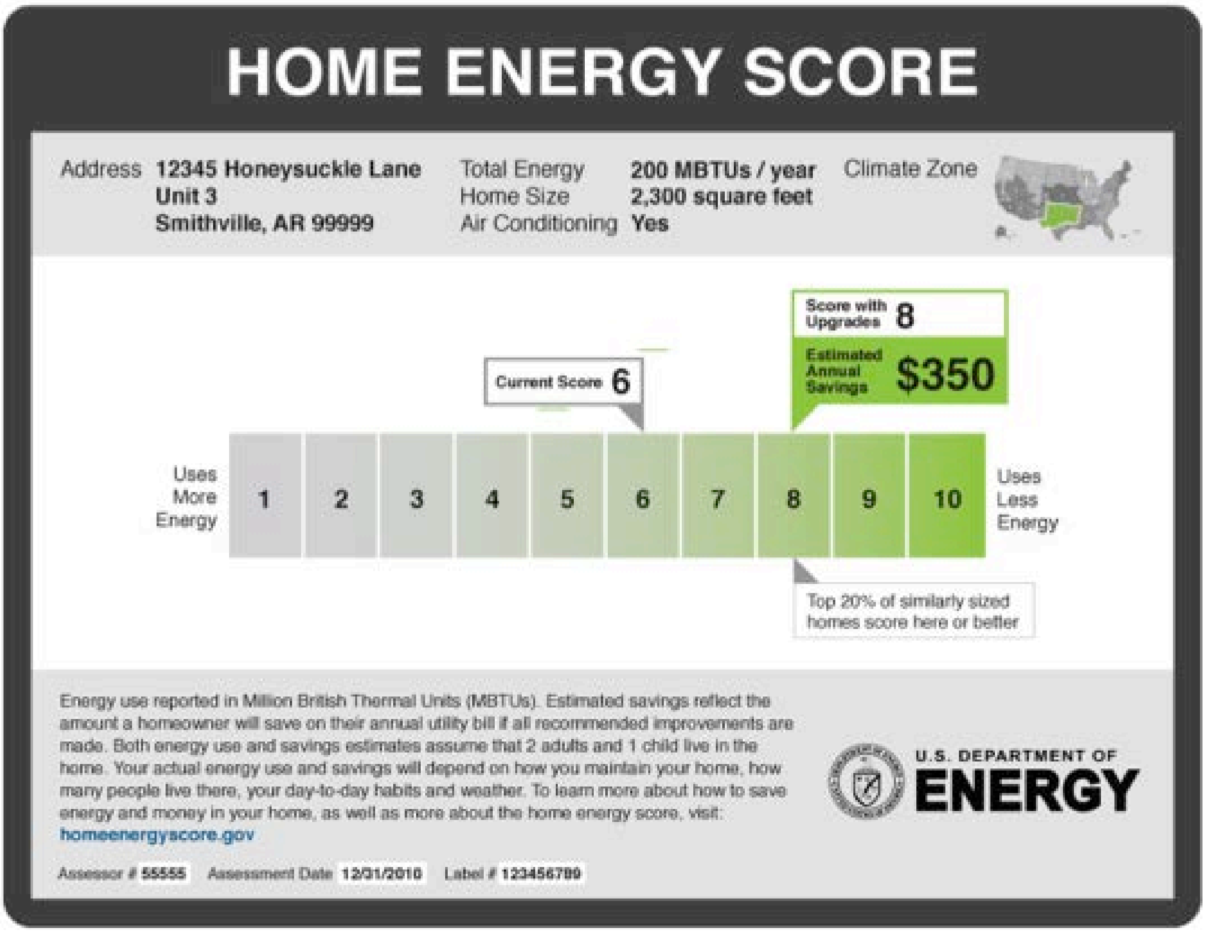
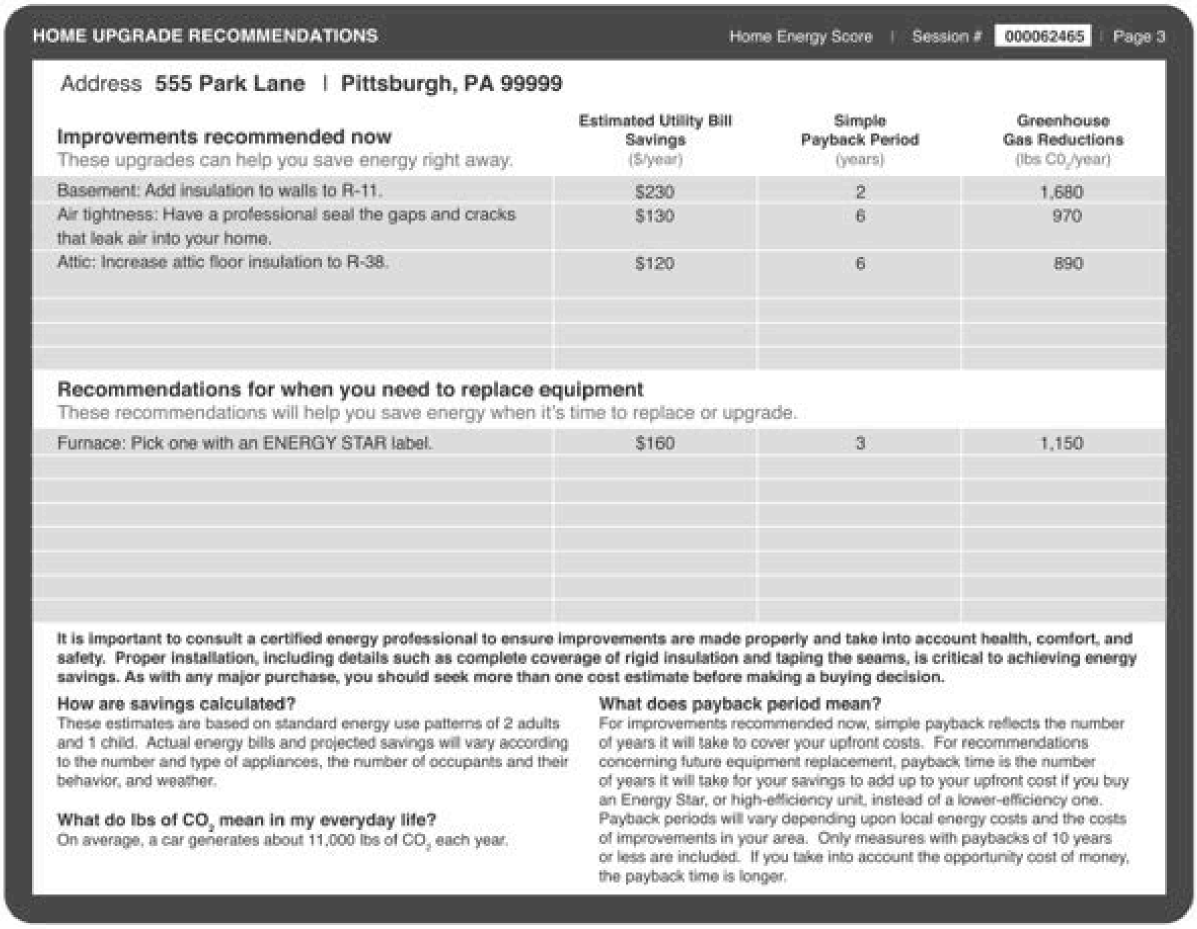


Figure 4. Sample Home Energy Score Report from pilot phase (DOE)

The Home Energy Score provides recommended retrofit measures, along with estimated savings and payback period (Figure 5). Recommended measures tell homeowners about their best investment options, for example, insulating the attic, sealing gaps and cracks, and insulating basement walls. The report provides two types of recommendations: measures to invest in as soon as possible that are specific to the home, as well as general recommendations for when the homeowner needs to replace equipment. All improvements that are recommended immediately have a payback of 10 years or less. Cost information is taken from the National Renewable

Energy Laboratory's National Residential Efficiency Measures Database,[1] and is generalized information that does not reflect local installation costs or available rebates.



HOME UPGRADE RECOMMENDATIONS   Home Energy Score | Session # 000062465 | Page 3

Address 555 Park Lane | Pittsburgh, PA 99999

Improvements recommended now These upgrades can help you save energy right away.	Estimated Utility Bill Savings ($/year)	Simple Payback Period (years)	Greenhouse Gas Reductions (lbs CO_2/year)
Basement: Add insulation to walls to R-11.	$230	2	1,680
Air tightness: Have a professional seal the gaps and cracks that leak air into your home.	$130	6	970
Attic: Increase attic floor insulation to R-38.	$120	6	890
Recommendations for when you need to replace equipment These recommendations will help you save energy when it's time to replace or upgrade.			
Furnace: Pick one with an ENERGY STAR label.	$160	3	1,150

It is important to consult a certified energy professional to ensure improvements are made properly and take into account health, comfort, and safety. Proper installation, including details such as complete coverage of rigid insulation and taping the seams, is critical to achieving energy savings. As with any major purchase, you should seek more than one cost estimate before making a buying decision.

How are savings calculated?
These estimates are based on standard energy use patterns of 2 adults and 1 child. Actual energy bills and projected savings will vary according to the number and type of appliances, the number of occupants and their behavior, and weather.

What do lbs of CO_2 mean in my everyday life?
On average, a car generates about 11,000 lbs of CO_2 each year.

What does payback period mean?
For improvements recommended now, simple payback reflects the number of years it will take to cover your upfront costs. For recommendations concerning future equipment replacement, payback time is the number of years it will take for your savings to add up to your upfront cost if you buy an Energy Star, or high-efficiency unit, instead of a lower-efficiency one. Payback periods will vary depending upon local energy costs and the costs of improvements in your area. Only measures with paybacks of 10 years or less are included. If you take into account the opportunity cost of money, the payback time is longer.

Figure 5. Sample Home Energy Score upgrade recommendations page from pilot phase (DOE)

Several design elements of the Home Energy Score have changed as a result of pilot feedback, although the overall concept and underlying model remain the same. Major changes include:

- The upgrade recommendations page has been redesigned and is now optional.
- Default assumptions about occupant behavior loads have been lowered to better reflect utility data (the non-asset energy uses such as hot water and lighting).
- The score report shows dollar savings over a 10-year period (changed from annual savings).
- The home's source energy use is no longer reported on the first page; it is included on a later page along with a breakdown by fuel type (kWh and therms).

[1] Available at www.nrel.gov/ap/retrofits/index.cfm

3 Project Goals

This project field-tested three different building performance models of varying complexity, in order to assess their value as rating systems in the context of an existing home retrofit program. The major focus was the DOE Home Energy Score, which was piloted at nine locations nationwide during the spring of 2011. The goal of these pilots was to test the tool in the field and provide feedback to DOE about technician experience, homeowner reaction, and the pattern of scores in a given region of the country, so that DOE could make necessary changes before the national launch of the tool. The Minnesota pilot was one of the sites that provided utility bill information to validate the quantitative accuracy of the model and some of the default assumptions.

The two other building performance models tested during the Minnesota pilot were SIMPLE and REM/Rate. SIMPLE is a spreadsheet-based home energy model developed by Michael Blasnik and Associates. It uses a streamlined set of inputs and broad classifications for certain home characteristics, which allows field technicians to switch between estimations or measurements depending on the scenario. Basic house and mechanical information is required, but the model allows users to enter approximations that are calibrated to local data, with the option to override a default when the user requires diagnostic measurements. For example, a technician can estimate a home's airtightness as "leaky," "average," or "tight," but can also enter the measure of pressurized airflow (cfm50) if available. The choice will depend on the requirements of a specific program (this report includes some discussion of our experience with cfm50 estimation). The SIMPLE model includes inputs for occupant behavior, though it can be used as an asset-only model. As mentioned above, SIMPLE is the basis for the Energy Performance Score (Earth Advantage Institute 2009).

REM/Rate, developed by RESNET, requires more detailed building inputs than either SIMPLE or the Home Energy Score. In particular, REM/Rate uses construction libraries to characterize the building shell in detail. Area, insulation R-value, and orientation measurements are required for each wall, ceiling, and rim joist segment. Window area and type can be entered for each individual window, as opposed to a general characterization for each side of the home, as is used for the Home Energy Score. However, REM/Rate also encourages users to aggregate windows of the same construction and orientation.

Table 1 summarizes the three models that were examined in this pilot. Note that while SIMPLE and REM/Rate refer to building performance models, the Home Energy Score is technically a scoring tool built on top of the Home Energy Saver model. The Home Energy Saver (and its companion, Home Energy Saver Pro) have capabilities in addition to those described for the Scoring tool (Bourassa et al. 2012; Parker et al. 2012).

Table 1. Building Performance Models Used in Pilot

	Home Energy Score	SIMPLE	REM/Rate
Number of Inputs	36–67	22–43	Approx. 100
Estimated Length of Home Visit	60–90 minutes	45–60 minutes	3–4 hours
Data Entry and Processing Time	20–30 minutes	15 minutes	1–2 hours
Tool Access	Online (public)	Spreadsheet (proprietary)	Software (proprietary)
Scope	Asset-only	Can include operational information	Can include operational information; asset-only for HERS report

The Home Energy Score, SIMPLE, and REM/Rate are only a subset of the many existing tools that model building energy use (Mills 2002). However, few have been piloted side-by-side to test their integration with residential energy retrofit programs. One example is a pilot study sponsored by the Energy Trust of Oregon that compared four tools: the EPS, based on SIMPLE; REM/Rate; and two versions of the Home Energy Saver (HES-Mid and HES-Full). Their findings for Oregon homes showed that SIMPLE was the easiest to use as well as the best predictor of utility bill data, with a *mean* absolute percent error of 25% for total energy use (Earth Advantage Institute 2009). By comparison the HES-Full mean absolute percent error was 33%, REM/Rate was 44%, and HES-Mid was 97%. When the *median* absolute percent errors were compared, HES-Full had the lowest error at 22% (compared to SIMPLE's 24%).

This project's learning objective was to gain greater understanding of the field performance of these energy asset tools for Minnesota homes. It is not, by contrast, an assessment of model accuracy when vetted and robust data building are available (see, for example, the discussion in Parker et al. 2012). This field performance is a combination of the models' accuracy in predicting energy use as well as their overall usability in the field, including the length of time to collect necessary data and the ability of field staff to consistently represent different types of homes. Specific research questions include:

- What is the distribution of Home Energy Score pilot scores? What is the distribution the tool predicts after upgrades are made?
- How does the electricity and gas energy use predicted by the Home Energy Score compare with utility bill data for pilot homes? How do SIMPLE and REM/Rate outputs compare?
- How well do the Home Energy Score's recommended improvements correlate with the in-person assessments conducted by home auditors?
- What is the overall usability of the Home Energy Score as an add-on to home energy visits and what are major recommended improvements? What type of training will be required for different technicians to use the tool consistently?

- Are field staff able to make consistent estimations of select building characteristics before taking diagnostic measurements, to reliably use SIMPLE's default input fields?
- What do pilot results indicate about the Home Energy Score's effectiveness to motivate homeowners to complete major upgrades?

4 Technical Approach and Field Work

Pilot homes were located in the cities of Minneapolis and Apple Valley, a suburb 20 miles south of the central metro. This sample captured housing types from two distinct eras. Three-quarters of single family homes in Minneapolis were built before 1945,[2] and primarily have 2 × 4 frame construction with wood or stucco siding. Many homes are "story-and-a-half" or "bungalow" style where the second level contains side attics and living space with knee walls and slanted ceilings. Apple Valley homes were built in the 1970s or later with 2 × 6 frame construction with insulated headers, newer windows, and higher levels of insulation in the walls, attic, and basements. All pilot homes have natural gas service.

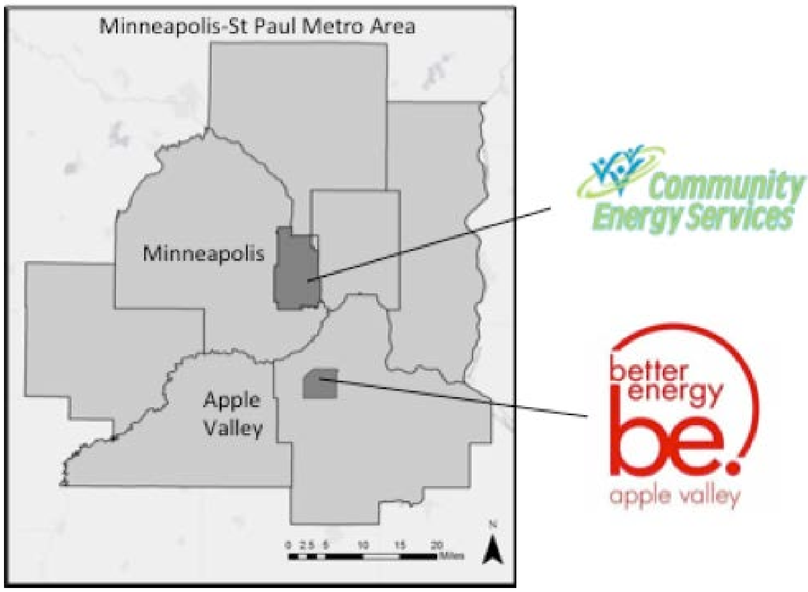


Figure 6. Residential energy programs participating in Minnesota pilot

The field data collection was conducted in two stages. From February through June 2011, 154 homes were scored as part of the Home Energy Score pilot, which was offered to homeowners as an added feature of CEE's residential energy program. Supplemental data were also collected during these visits in order to conduct a SIMPLE level energy analysis on these homes. During the fall and winter of 2011, a subset of 50 homes were recruited to participate in a second, more extensive HERS level audit using REM/Rate. This division allowed field staff and homeowners to evaluate the Home Energy Score independently for the DOE.

4.1 The Center for Energy and Environment's Residential Energy Program

CEE tested audit tools as an add-on to two residential programs: the Community Energy Services program in Minneapolis, and the Better Energy Apple Valley program in Apple Valley, Minnesota. These enhanced programs were offered through 2012 using support funding from the partner cities and the Legislative-Citizen Commission on Minnesota Resources. The programs employed a community-based approach for homeowner recruitment and engagement. Residents

[2] City of Minneapolis Assessor's Office, single-family detached housing data (queried June 2008).

were recruited to attend a two-hour neighborhood workshop where they learned about energy efficiency and ways to improve their homes. Homeowners were then asked to sign up for a home visit and pay the copay on site. In total, 98% of workshop attendees signed up for a home visit, and these two programs combined conducted 6,500 home energy visits. Of the homes that received a visit, approximately 25% of homeowners followed through to complete major retrofit upgrades (Nelson 2011). The program was available only for one- to four-unit owner-occupied homes in designated cities.

CEE's residential programs take a whole-house approach. During the 1½-hour home visit, an energy technician and an energy counselor conduct air leakage and combustion safety diagnostics, inspect the home's mechanical equipment and insulation levels, and install energy-saving devices such as compact fluorescent lamps and faucet aerators.[3] The homeowner is given the results and a list of recommendations for major upgrades. In addition, many pilot homeowners were opt-in participants to CEE's "Ready-Set-Go" program. This program is targeted at homeowners who are likely candidates for wall or attic insulation upgrades.[4] Visits are scheduled for an additional 30 minutes, during which the energy technician conducts a detailed inspection of the home's current insulation and provides a cost quote for any insulation recommendations. This "visit plus bid" gives homeowners a cost estimate honored by participating contractors, and saves them the additional step of scheduling a preliminary visit with a contractor. When willing, "Ready Set Go" participants were prioritized for the Home Energy Score pilot in order to take advantage of the additional time spent on a detailed insulation inspection.

4.2 Pilot Home Selection

The pilot provided the Home Energy Score to 154 homes (approximately 75% in Minneapolis and 25% in Apple Valley). In order to manage logistics for the pilot team, participants were preselected based on schedule and given the option to opt-out of the pilot. As described above, pilot homeowners had a basic knowledge of energy efficiency issues in their home as a result of the recruitment workshop, and learned more by working with the Energy Counselor during the visit. For that reason, this pilot was not an effective measure of the Home Energy Score's utility as a stand-alone educational tool.

Pilot homes do not represent a random sample of homes in the Minneapolis metro region. Since homes are selected from energy program participants, homeowners are not likely to have had recent energy efficiency work done on their homes. Further, since many homeowners were participants in CEE's Ready-Set-Go program, they were more motivated to receive insulation upgrades.

4.3 Data Collection and Analysis

The sections below describe the field data collection process for each of the three modeling tools.

[3] In order to gain access to the Home Energy Scoring Tool, the energy technician is certified as a Home Energy Score Qualified Assessor (QA). He or she must be either BPI Building Analyst or RESNET Provider certified and pass a 20-question test about the Score, administered by DOE.
[4] For example, during winter months, homeowners are encouraged to enroll if their homes have ice dams.

4.3.1 Home Energy Score

The Home Energy Score requires 36–67 data inputs, depending on a home's construction and mechanical systems. The data inputs include basic configuration and insulation levels, window type and sizes, mechanical system information, and optional house air leakage diagnostics. In order to provide a simplified audit tool, many single input values entered into the Home Energy Score tool had to be an amalgam of multiple sections of the home. For example, a home's attic insulation level may not be consistent across the entire space, especially for story-and-a-half home configurations. In this instance, data on the different attic R-values were recorded on paper forms, and an area-weighted average was input into the Home Energy Score model. In addition, the discrete Home Energy Score R-value inputs are often noncontiguous; the options for attic are R-0, R-11, R-13, R-15, R-19, R-21, R-30, R-38, R-49, and R-60.

The Home Energy Score pilot required a significant amount of "learning by doing" as field assessors gained familiarity with the trial version of the tool and received feedback about how to collect data as they encountered new situations. In particular, there were challenges around characterizing basements and the building shell for different home configurations, such as story-and-a-half homes, split-level homes, and underground garages. Assessors adapted their protocol as the pilot progressed in order to fit the model treatment. As a result, the data collection methods were not consistent throughout the pilot. Specific examples are discussed in more detail in Section 5.3, along with the effect on the results.

4.3.2 SIMPLE

The SIMPLE model accommodates diagnostic measurements when available, but also allows users to make estimates and use default values. Where the above Home Energy Score inputs overlapped with SIMPLE inputs they were used directly. However, field staff also provided qualitative estimates of the following home characteristics in order to validate SIMPLE's more streamlined approach:

- Insulation level of attic and above grade walls
- Home airtightness
- Window shading.

Field technicians recorded their estimates of the above characteristics *prior to* making measurements for the Home Energy Score. For example, while the Home Energy Score asks for the total area of windows on each side of the home, the SIMPLE model includes three categories of building window area: low, typical, and high. In this case, the field technicians first inspected the home and one of the three categories for the SIMPLE model, and then measured the actual area for the Home Energy Score. Like the Home Energy Score, SIMPLE requires the user to combine distinct sections of a home if they are built differently.

SIMPLE, like the Home Energy Saver, can be used as an asset-only model, or can accept information about occupant habits. A subset of 53 homeowners was given a one-page survey to collect additional information about their homes, which they completed during the home visits. This survey asked about temperature set points, the style and number of refrigerators, and specific additional high energy uses in the house, such as dehumidifiers, swimming pools, etc.

Occupants also reported on certain household behavior, like shower length and laundry frequency.

This project tested three different approaches to using the SIMPLE model:

- SIMPLE-1: Run model as asset tool using Home Energy Score default assumptions for occupant habits and behavior (150 homes)
- SIMPLE-2: Run model as asset tool using SIMPLE default assumptions for average occupant behavior (150 homes)
- SIMPLE-3: Run model as a combined asset and behavioral tool using occupant survey information about energy habits and plug loads (53 homes).

All analysis used SIMPLE model version 0.9.8, released April 2011.

4.3.3 REM/Rate

A CEE RESNET-certified staff member collected more detailed home characteristics on a subset of 50 homes during a second home visit. Home Energy Score pilot participants were recruited via phone and email to solicit their participation in an extended research project. Initial interest in a return home visit was low, since homeowners had already participated in an energy visit, and they were told that the additional visit would take 4–6 hours. To increase participation, each homeowner was offered a $100 gift card incentive, as well as the promise of additional information about his or her home's performance.

These visits occurred 3–9 months after the Home Energy Score visit. During that time, a total of 28 participants had completed recommended energy upgrades, which included air sealing, attic insulation, and wall insulation upgrades. The REM/Rate analysis "backed out" these energy upgrades, since they would change the predicted energy use and corresponding utility data that were available for each home.

REM/Rate allows both simplified and detailed levels of input. Simplified inputs use general building design characteristics (e.g., house type) and built-in algorithms to determine building shell areas and other characteristics. This project made use of more detailed inputs, which provide the user greater control, primarily over the building shell construction. This includes the area and construction details of individual sections of walls, windows, the foundation, and attic cavities. From a user perspective this was the most significant difference between REM/Rate and the other two models, since it adds time to the data collection and input process. REM/Rate also has detailed input options for HVAC characteristics, duct system characteristics, and air infiltration rates (measured or estimated). This analysis used REM/Rate version 12.96.

Table 2 summarizes the level of detail used to characterize a building within each model for this project. In the case of SIMPLE and REM/Rate, the models offer the option to provide more detailed inputs than what we used here.[5] Our approach was to make use of the minimum level of detail that might be reasonable to collect during a field survey, and not necessarily to optimize a model's energy use prediction. The SIMPLE inputs correspond to version 2, which did not use

---

[5] The same is true for the Home Energy Saver model, which the Home Energy Scoring Tool is built on. Again, the Home Energy Score is not an apples-to-apples comparison with SIMPLE and REM/Rate.

any override options except for blower door test results. Note that many of these SIMPLE inputs use an ordinal scale, such as "high, medium, low." Our data collection for REM/Rate included the standard detailed construction information for foundation, wall, and attic space.

Table 2. Treatment of Key House Characteristics Within Each Building Model

Building Characteristic	Home Energy Score	SIMPLE	REM/Rate
Home Characteristics	Year built, number of stories, number of bedrooms	Number of stories, number of bedrooms	Year built, number of stories, number of bedrooms
House Layout	Total conditioned floor area	Total above-ground floor area	Individual floor areas, individual wall areas, floor plans
Foundation	Type and insulation R-value	Type and location of insulation (ceiling, wall, or none)	Slab area and insulation, wall type and area, rim joist area and insulation for individual sections
Air Leakage	cfm50	cfm50	cfm50
Attic Insulation	Single UA average value	Estimation using 4-point scale	Insulation R-value and rafter size for individual attic cavities
Wall Construction and Insulation	Single construction and UA average value for each of the four sides	Estimation using 5-point scale for whole-home insulation level	Surface area, orientation, insulation R-value, stud size and spacing for individual wall sections
Ducts	Conditioned or unconditioned location; existence of sealing	Percentage in basement or attic; estimated insulation level and leakiness	Duct blaster test where feasible, location, and insulation
Windows and Skylights	Total area for each side of house; number of panes and glass type	Type and estimation of overall area using 3-point scale	Customized window location and construction
Mechanical Systems	Type, fuel and efficiency (or age estimate if not available)	Type and fuel	Type, manufacturer, efficiency, age, capacity, and location

5 Results

Below is a summary of the major results from the pilot:

- Homes scored an average of 5.4 on the Home Energy Score, and implementing recommended upgrades would increase their score by an average of 2 points. No homes had an initial score greater than 8, and none could score a 10 after upgrades.
- Overall, the Home Energy Score, SIMPLE, and REM/Rate overpredicted gas use, although the SIMPLE model was the closest to utility bill data. The SIMPLE model overpredicted gas usage by 16% on average, compared to 55% and 63% for the Home Energy Score and REM/Rate, respectively.
- The Home Energy Score, SIMPLE, and REM/Rate all overpredicted electricity consumption, and had a larger minimum baseline electricity load than actual usage. The average overprediction from the Home Energy Score was 23%, SIMPLE's average overprediction was 25% for the asset version of the model, and REM/Rate overpredicted by 27% on average. The correlation between modeled and measured electricity use was highest for the asset models that scale electricity loads with home size. These results include a small number of homes with very high actual electricity use.
- The asset models have comparable levels of correlation for predicting natural gas use, from a low R-squared of 0.42 for the Home Energy Score to a high of 0.52 for REM/Rate. The asset models all have a low R-squared value for predicting electricity use, from 0.05 for the Home Energy Score to 0.30 for REM/Rate.
- There was negligible variation in the Home Energy Score electricity use for the 14 homes with a furnace and no central air conditioning. This implies that the pilot version of the tool did not properly capture furnace fan electricity use as a function of space heating load.
- Field crews recommended retrofit upgrades for a home more often than the automatic recommendations generated by the Home Energy Score, especially for attic and wall insulation.

5.1 Pilot Home Characteristics

Figure 7 shows the year of construction for pilot homes. The median year of construction is 1929. A majority of homes were built before 1945 (61%), and very few pilot homes were built after 1990 (5%). The distribution closely mirrors the existing housing population in Minneapolis and Apple Valley, as determined from assessor data. The one exception is the higher proportion of pilot homes built after 1970 (23% compared to 11%), which reflects the higher ratio of Apple Valley homes in the pilot compared to the actual ratio of city populations.

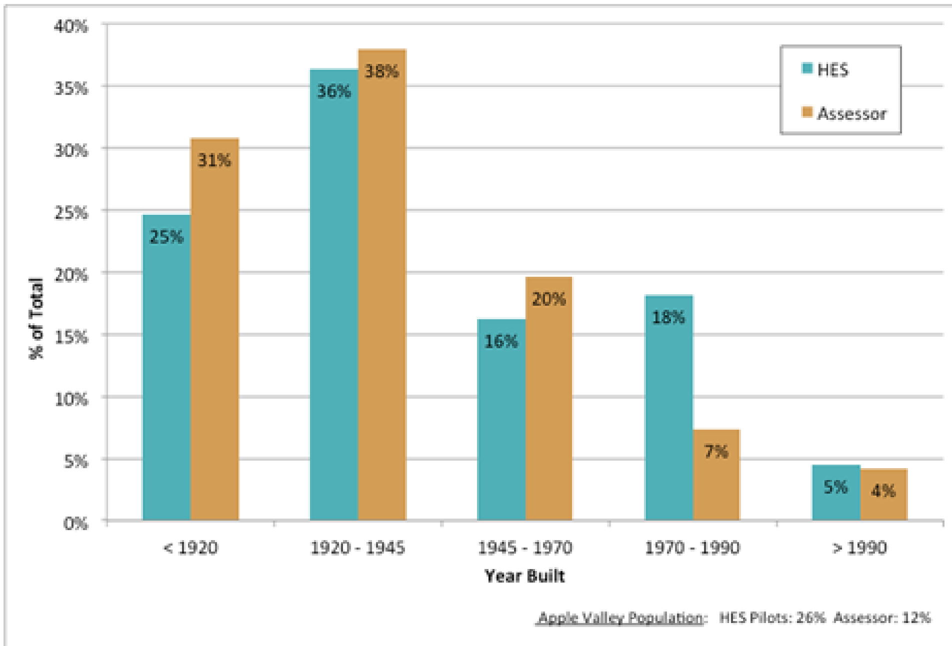


Figure 7. Year of construction of pilot homes (n = 154)

Figure 8 below shows the variation of pilot home floor area. The median home size is 2,203 ft^2 (equal to the national median for metropolitan areas), and there is a fairly consistent distribution of home sizes between 1,750 and 2,750 ft^2. The method for counting basement area changed partway through the pilot, so in 37 cases the floor area includes unfinished basements. This change is explained in more detail below.

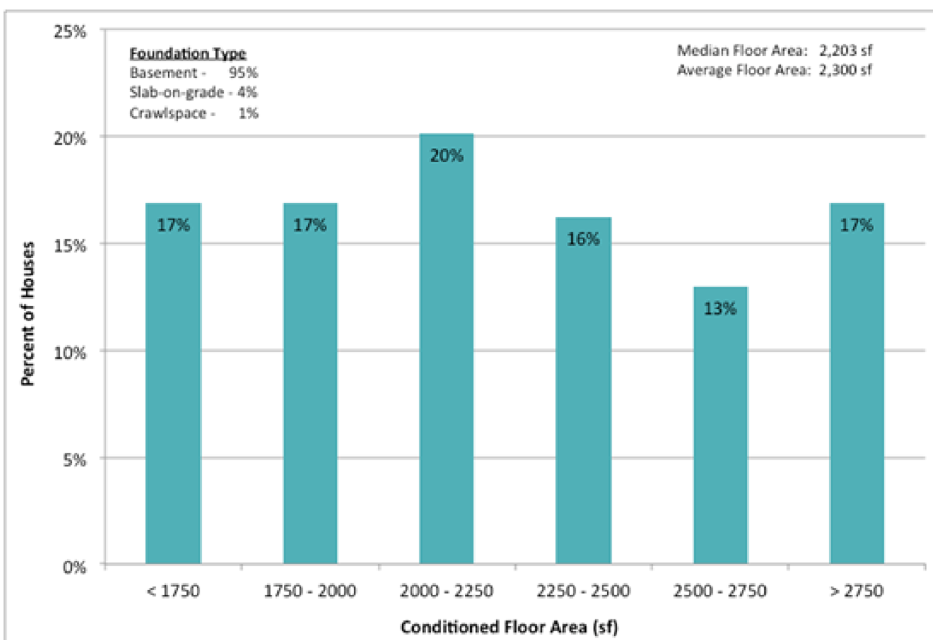


Figure 8. Floor area of pilot homes (n = 154)

Figure 9 shows the distribution of the number of stories for the pilot homes. The largest category is 1.5-story homes (which also includes homes designated as 1.25 and 1.75 stories). This story-and-a-half style is common in prewar Minneapolis neighborhoods. These homes take longer to assess because of their complex and sometimes inaccessible attic configurations. They also have a lower insulation potential than full one- or two-story homes because of their smaller attic cavity areas, which is an important consideration when rating a home's energy potential. The Home Energy Score and SIMPLE do not currently have options for 1.5 or 2.5 stories, so these homes were entered as two- or three-story homes, respectively. All pilot homes were stand-alone single-family homes with the exception of three townhomes, which were counted as two-story homes.

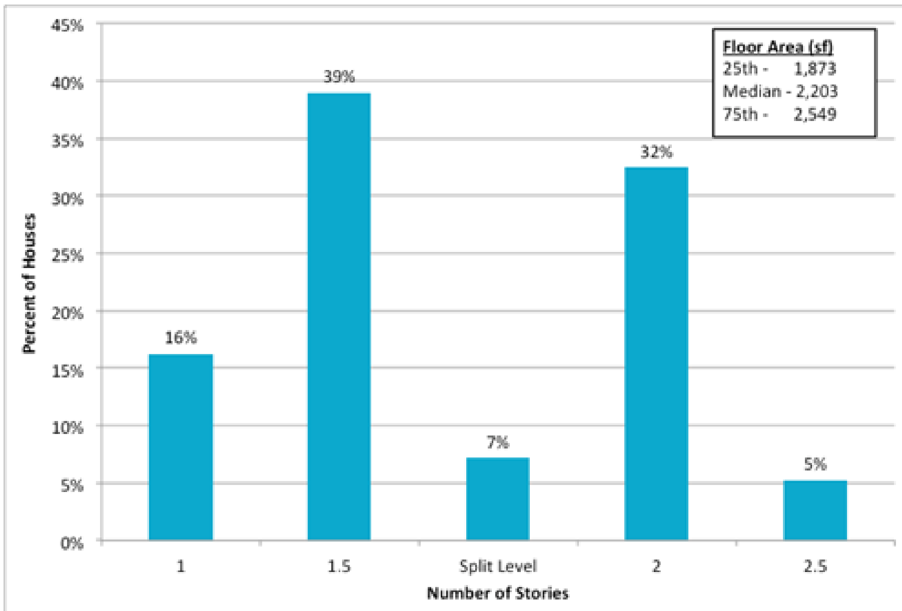


Figure 9. Number of Stories of Pilot Homes (n = 154)

The summary table below lists some additional characteristics of pilot homes. All homes have natural gas heating: 75% have a central gas furnace, and 25% have a gas boiler. Heating systems were approximately 20 years old on average, when age could be estimated. In addition, 71% of homes have central air conditioning.

Table 3. Characteristics of Pilot Homes

Heating System	Furnace	75%
	Boiler	25%
Central Air Conditioning	Yes	71%
	No	29%
Basement	Finished	46%
	Unfinished	50%
	Slab-on-Grade	4%

Figure 10 and Figure 11 below show the distribution of existing wall and attic insulation R-values for the pilot homes. Only 15% of prewar homes have an existing attic insulation level greater than R-30, and 34% have wall insulation of R-11 or higher.[6] Assuming that prewar homes were built with minimal wall insulation, this indicates that more than half of these homes have had insulation upgrades beyond original levels. As expected, newer homes have higher insulation levels that correspond to more stringent building codes.

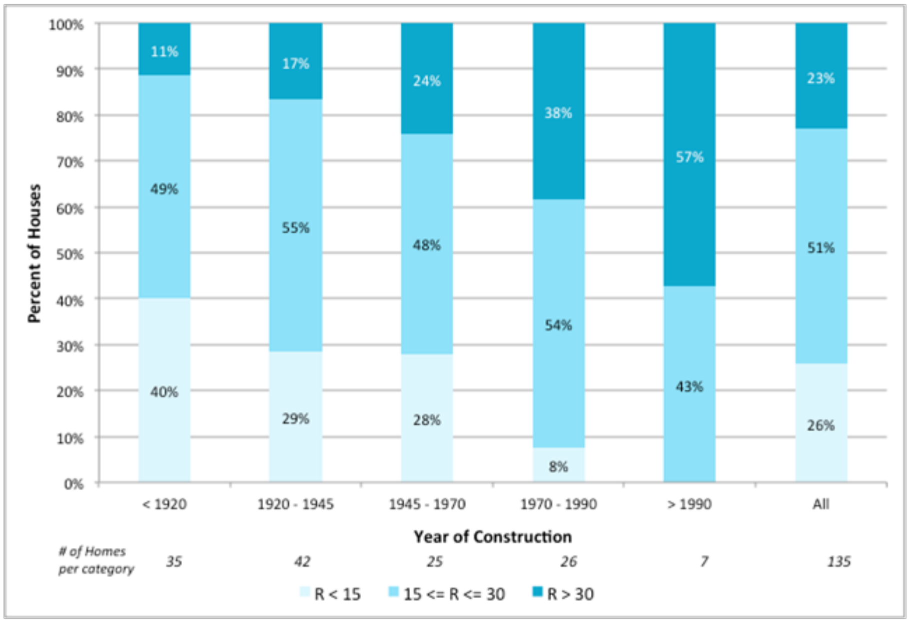


Figure 10. Existing attic insulation by year of construction

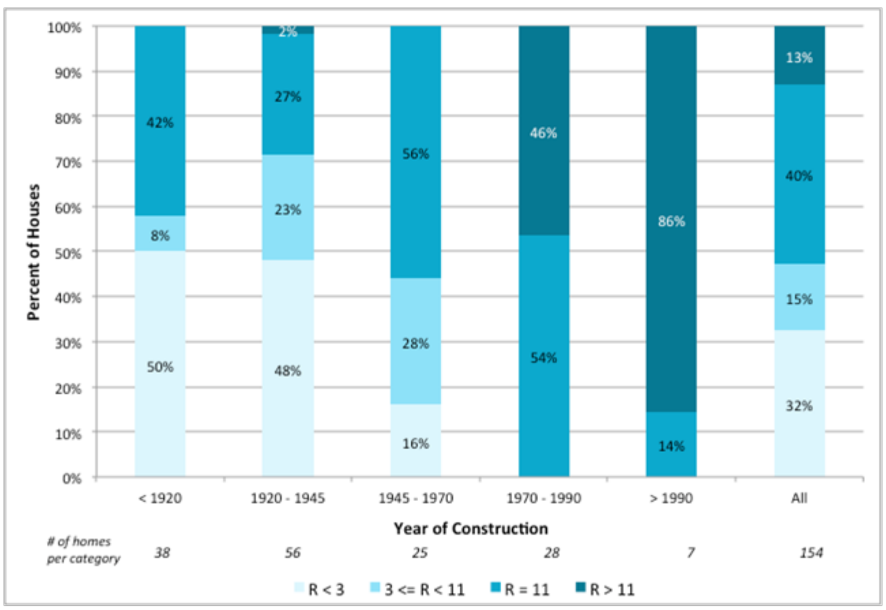


Figure 11. Existing wall insulation by year of construction

[6] Calculated as a weighted average for homes older than 1945.

And lastly, Figure 12 below shows the air leakage ratio (ALR) of pilot homes by year of construction, which follows the same pattern as insulation levels. Fifty percent of homes built before 1920 have an ALR above 1.3. Of the 35 homes built since 1970 only 4 homes, or 11%, have an ALR above 1.3. The ALR is a measure of the home air leakage and the exterior environment divided by the size of the home, derived here from the blower door test (cfm50) and the total floor area of the home.[7]

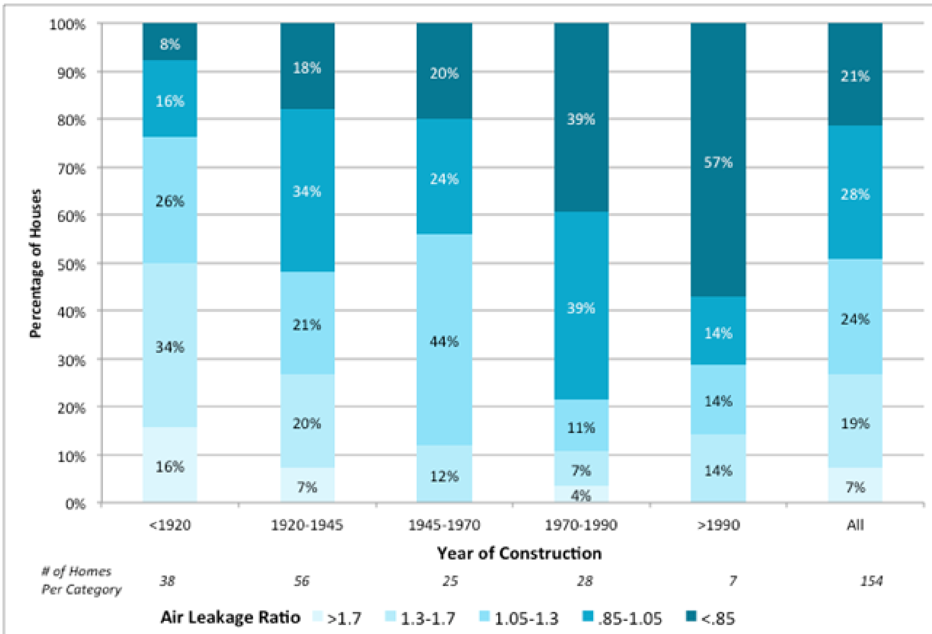


Figure 12. ALR by year of construction

5.2 Home Energy Score Results

Figure 13 shows the distribution of the existing scores for the pilot homes, and how they would score if they completed the recommended upgrade measures. The average score was a 5.4, with 19% of homes scoring a 3 or below, and 41% able to score a 7 or 8. No homes scored above an 8. After upgrades the average score was a 7.4, with 84% of homes able to achieve a score of 7 or above. No homes are able to score a 10 after upgrades, which is by design reserved for very low-energy homes. In two cases, homes could not improve their score above a 1 after upgrades. This implies that the Home Energy Score did not find any cost-effective upgrades.

It is important to note that since the pilot version several changes were made to the scoring process so that homes should now receive higher scores, on average. The retrofit recommendations have also been expanded, which means that homes should show more opportunity to improve their score by investing in upgrades.

[7] For homes with a ceiling height of 8 feet the ACH50, or air changes per hour at a pressure difference of 50 Pa, is simply the ratio multiplied by 7.5. A total of 75% of the pilot homes had a first floor ceiling height of 8 feet and 24% had a height of 9 feet (one home had 10-foot ceilings).

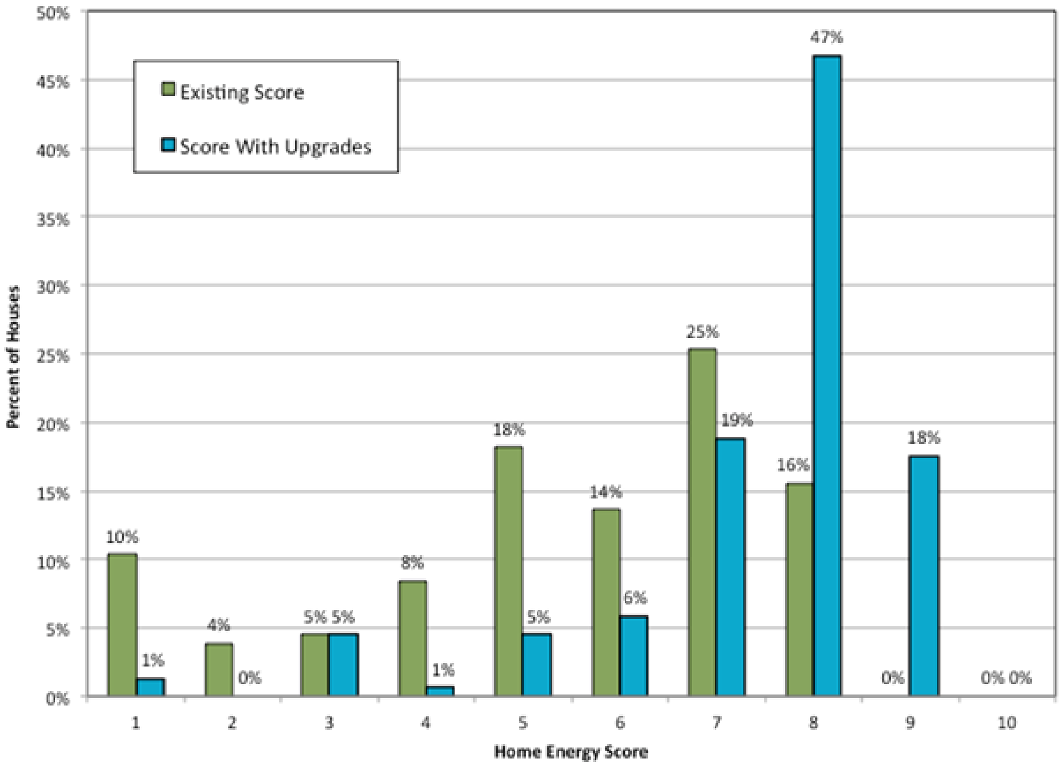


Figure 13. Home Energy Scores for Minnesota pilot homes (n = 154)

Figure 14 below shows a different view of score results. Homes are sorted according to their current scores from lowest to highest (in green), and aligned with their potential scores after upgrades (in blue). In this view, a completely filled-in space would represent a perfect housing sample, where all homes score a 10. This figure indicates that homes currently occupy about half the space (congruent with an average score of 5.4), and that the average increase in score is 2 points. Sixty-eight homes (44%) are able to increase their scores by only 0 or 1 point, while 10 (6.5%) are able to increase their points by 5 or more. This graph also shows how much unachievable potential remains even if all homes completed their recommended upgrades.

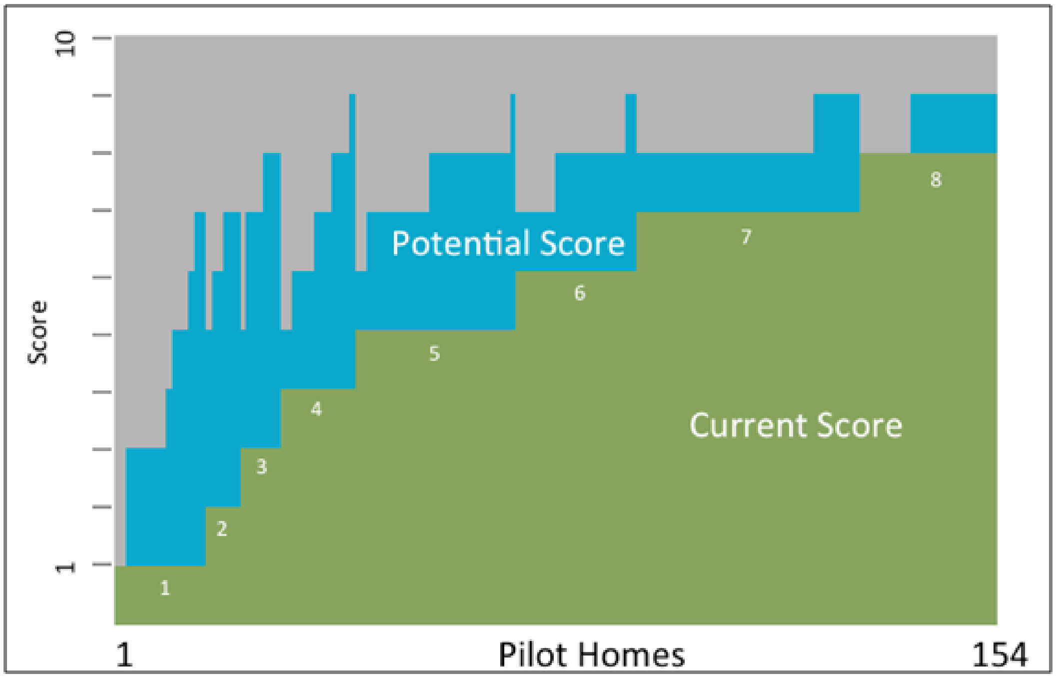


Figure 14. Home Energy Scores for Minnesota pilot homes (stacked column)

Figure 15 shows a map of the scores by location for Minneapolis and Apple Valley, along with the average age of homes in each neighborhood. While there is not a distinct pattern of scores within each city, a disproportionate number of high scores are located in the outer suburb where the median year built is 1983.

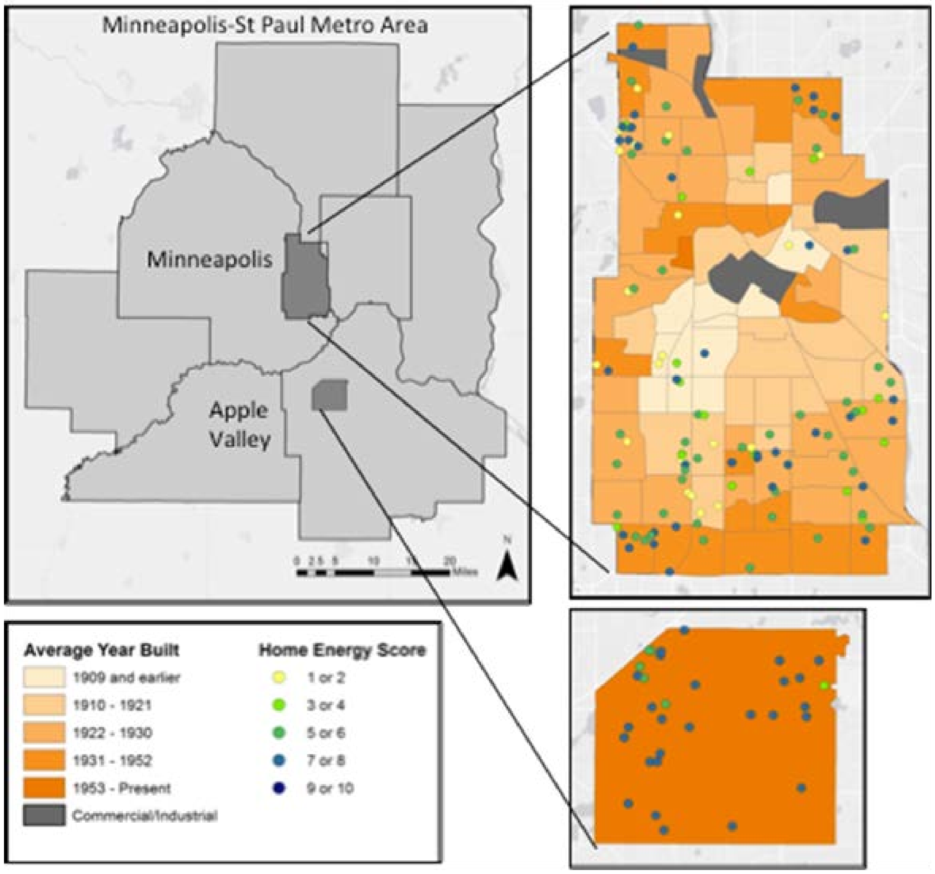


Figure 15. Map of Home Energy Score results

5.3 Energy Modeling Results

The predicted energy use from the three audit tools was compared to weather-normalized customer utility bills for the 12-month period preceding the home visit. Utility bills were weather normalized on an annual basis to match the heating degree day (HDD) and cooling degree day (CDD) assumptions in each energy model. We used annual base 65°F HDD and CDD data from the National Oceanic and Atmospheric Administration for Minneapolis. Gas baseline use was extracted by computing the average of the summer months (June, July, and August) and multiplying by 15. This factor of 15 (versus 12) was used to account for the lower water inlet temperature during winter months. Billing data were considered complete when 12 monthly billing periods were available with no estimated use values.

Complete electricity bill information was available for 138 homes, and natural gas data for 136 homes (approximately 90% of all participants). Seven additional homes were missing gas use data from one summer month. For these cases the missing month was computed as the average use of the remaining summer months (June through September), and these homes are included in results. Data were unavailable for the remaining homes for a number of reasons, including incomplete customer waiver forms or resident occupancy for less than 12 months.

A summary of results is shown in Table 4. For all models the mean annual gas use was greater than the weather-normalized utility billing data. REM/Rate overpredicted a home's gas use by the largest margin, an average of 657 therms, or 63% for each home. On average, the Home Energy Score overpredicted gas use by 603 therms, or 55% more than metered utility bill data. The Home Energy Score results are partially attributable to a misunderstanding over the definition of conditioned basements, which is described further below in this section. The three versions of the SIMPLE model had mean overpredictions of 130, 144, and 106 therms respectively, or 16%, 18%, and 13%.

Table 4. Summary of Energy Output From Three Models Compared to Utility Data

	Home Energy Score	SIMPLE			REM/Rate
		v. 1	v. 2	v. 3	
NATURAL GAS					
Number of Homes	*143*	*143*	*143*	*53*	*51*
Utility Mean (therms/yr)	1,166	1,166	1,166	1,150	1,150
Mean Annual Use (therms)	1,769	1,296	1,310	1,256	1,807
Mean of Differences:					
therms	+603	+130	+144	+106	+657
percent	+55%	+16%	+18%	+13%	+63%
st dev (therms)	673	353	360	351	558
Median of Differences:					
therms	+473	+193	+200	+137	+685
percent	+52%	+20%	+22%	+15%	+73%
Abs. Mean of Differences:					
therms	666	288	302	264	725
R-squared	0.42	0.50	0.48	0.55	0.52
Percent of homes within ± 25%	26%	59%	52%	62%	12%
Percent of homes within ± 50%	46%	90%	87%	96%	33%
ELECTRICITY					
Number of Homes	*138*	*138*	*138*	*53*	*50*
Utility Mean (kWh/yr)	7,796	7,796	7,796	7,038	7,937
Mean Annual Use (kWh)	7,249	7,516	7,522	7,834	8,176
Mean of Differences:					
kWh	-546	-280	-274	+660	+238
percent	+23%	+29%	+25%	+28%	+27%
st dev (kWh)	4,405	4,177	3,925	2,396	3,707
Median of Differences:					
kWh	+764	+628	+410	+809	+735
percent	+12%	+9%	+7%	+12%	+9%
Abs. Mean of Differences:					
kWh	3,302	3,168	2,909	1,991	2,753
R-squared	0.05	0.16	0.24	0.59	0.30
Percent of Homes Within ± 25%	30%	33%	41%	52%	44%
Percent of Homes Within ± 50%	65%	64%	68%	71%	70%

Conversely, the average annual electricity use computed by the Home Energy Score and the asset versions of SIMPLE (SIMPLE-1 and SIMPLE-2) was lower than that from the utility billing data. However, this is affected by a small number of high electricity users. The average and the median percent differences are both positive. SIMPLE-3, which accounted for occupant habits and additional loads, and REM/Rate both overpredicted electricity use. Both models were also used on a smaller random sample of pilot homes; the four outlier electricity users were not part of the SIMPLE-3 population (one was included in the REM/Rate sample). REM/Rate recorded the highest mean annual use, but had the lowest mean absolute difference of the asset models, which reflects that the average electricity use of REM/Rate population is higher. Overall, the mean difference as a percent of the utility bill was similar across all models (between 23% and 29%). The lower median percent variation (7%–12%) underscores the effect of the large users.

Table 4 also reports the R-squared value for the linear regression analysis performed on each model, which indicates the models' ability to predict the variation in actual energy use between homes. The asset models (not including SIMPLE-3) have similar levels of correlation for predicting natural gas use, from a low R-squared of 0.42 for the Home Energy Score to a high of 0.52 for REM/Rate. The asset models all have a low R-squared value for predicting electricity use, from 0.05 for the Home Energy Score to 0.30 for REM/Rate. SIMPLE-3, which takes occupant habits into account, has higher correlation with utility data, with an R-squared of 0.55 for natural gas and 0.59 for electricity. Since SIMPLE-3 is not an asset-only model, these results are not directly analogous.

An interpretation of these results must take into account how basement floor area was treated in the Home Energy Score, which was a point of confusion in terminology for several pilot locations. Two Home Energy Score inputs, conditioned floor area and foundation type, help determine a home's size. The options for foundation type were slab-on-grade, unconditioned basement, conditioned basement, vented crawlspace, and unvented crawlspace. CEE, along with other cold-climate pilot sites, generally categorizes basements as conditioned because the heating season space temperature typically remains between 55°F and 65°F, even if they are unfinished without a direct heating or cooling register.[8] However, the model treats these inputs as finished floor area, and determines a home's structure and heating load accordingly. Therefore, from the model perspective, a "conditioned basement" foundation will be heated, and the appropriate share of the home's floor area will be allocated below grade.

Midway through Minnesota's pilot, DOE feedback modified the basement floor area interpretation so that field staff began to count only finished basement area as "conditioned," but the conditioned basement issue was never clarified. The primary result was that the second set of pilot homes that had unfinished basements were assumed to have a basement floor area that was smaller than the actual value. Neither of these approaches properly models an unfinished basement that is kept at a lower temperature than the upstairs during the heating season. See Table 5 for an overview of the different cases.

[8] There is usually no basement ceiling insulation and a high level of return leaks to the basement that helps circulate basement air to the rest of the house.

Table 5. Treatment of Conditioned Basements in Home Energy Score Pilot Homes

Series	Number of Homes	Collected Data Points	Treatment Within Home Energy Score
Finished Basements	76	Conditioned (finished) floor area; Conditioned basement foundation	Homes represented correctly
Unfinished Set 1	41	Conditioned floor area; Conditioned basement foundation	Floor area correct; basement kept at thermostat temperature set point
Unfinished Set 2	37	Finished floor area; Conditioned basement foundation	Floor area lower by 30 to 50 percent; basement kept at thermostat temperature set point

It should be noted that this issue applies only to the Home Energy Score results. The SIMPLE model specifies above-grade finished floor area and has options for foundations that include both “conditioned” and “conditioned and finished” basements. Home layouts in REM/Rate are customized to the home’s actual structure, including materials and insulation levels, so it was also not affected by this issue.

Figure 16 shows a comparison of the Home Energy Score and utility bill annual natural gas use for different datasets. Equal values would fall along the “1 to 1” line. Results show that most homes fall above this line, meaning the Home Energy Score overpredicted gas use for pilot homes. Overprediction is highest for Set 1 homes with unfinished basements, since the model treated these basements as if they were kept at the thermostat set point. This overprediction is lower for Set 2 homes with unfinished basements, because the smaller floor area has a counterbalancing effect. The slope of the best fit line for homes with finished basements is close to 1, which implies that on average the model more accurately predicts space heating energy use. However, the low R-squared value of 0.36 indicates that the model does a poor job of predicting gas use for individual homes.

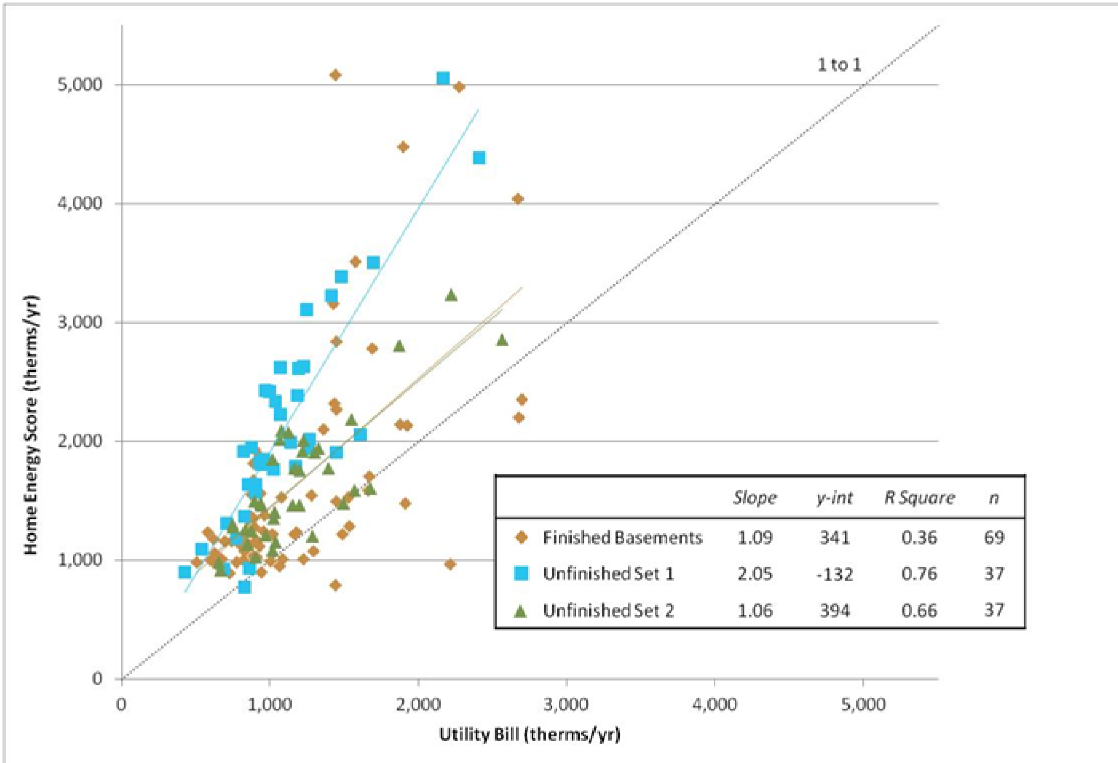


Figure 16. Natural gas comparison for Home Energy Score versus utility data (scatter)

In Figure 17 the Home Energy Score results are sorted from low to high along the x-axis along with the corresponding utility bill use. This pattern shows that much of the overprediction results from sometimes large overestimates for higher use homes. The maximum (weather-normalized) annual utility gas use was 2,697 therms, and the Home Energy Score predicted higher use for 17 homes (12% of the total). These homes were a mix of finished and unfinished basements. Similarly, 21 homes used less gas than the absolute minimum predicted by the Home Energy Score (781 therms).

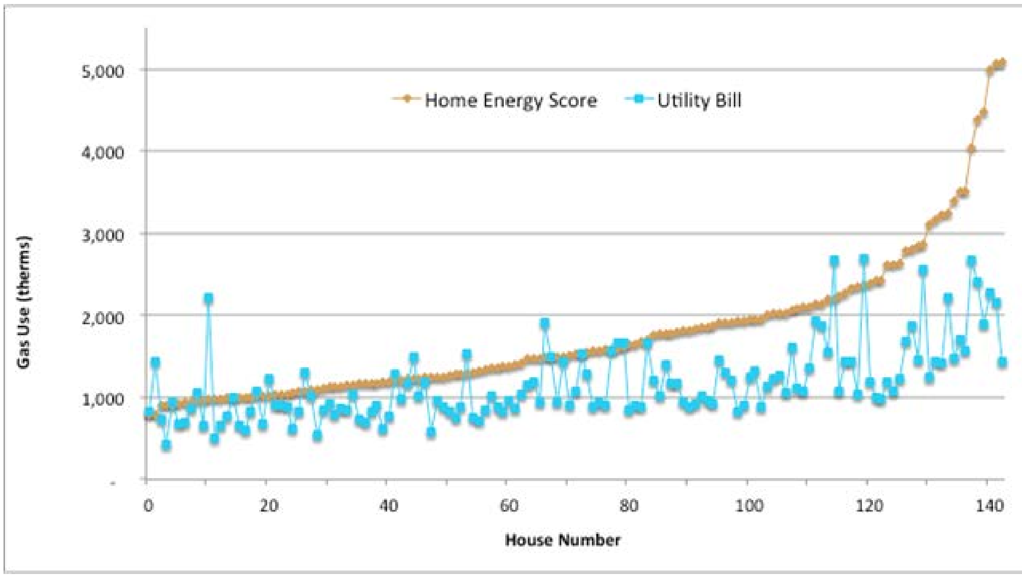


Figure 17. Natural gas comparison for Home Energy Score versus utility data (line graph)

Figure 18 below shows a comparison of the Home Energy Score and utility bill gas results for different categories of existing wall insulation. The categories of insulation reflect the discrete options available for the Home Energy Score inputs. Results show that overprediction is greatest for homes with little or no existing wall cavity insulation (R-3 or less). The regression slope is significantly greater than 1 and the R-squared is only 0.54. The regression slopes decrease for higher wall R-values and are less than 1 for R-values of 11 and 19. In other words, the pilot Home Energy Score overpredicted energy use when modeling low levels of wall insulation, and underpredicted energy use when modeling high wall insulation, a possible indication of model bias. One possibility is that the measurement and treatment of wall heat loss could be improved, especially for the assembly R-value of an empty wall. A larger sample of homes may help researchers better understand the differences between modeled and actual gas use, particularly for homes with wall insulation between R-3 and R-11 and those with R-19 as those samples included only 21 and 19 homes, respectively.

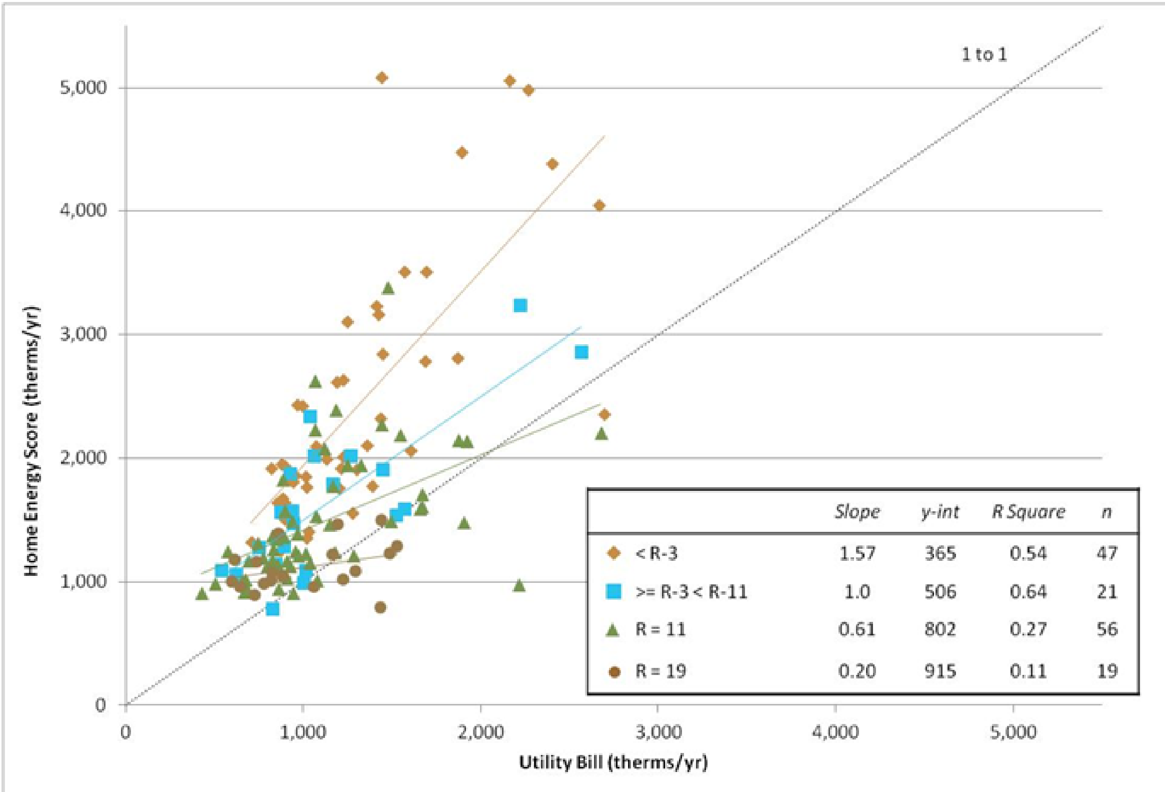


Figure 18. Home Energy Score gas demand versus wall insulation

Figure 19 below shows how the Home Energy Score electricity output compared to customer utility data. In this case the slope of the best-fit line is close to zero, indicating that there is little correlation between the modeled and actual electricity use. There is a much larger spread among utility electricity use than the Home Energy Score predicts. This is expected from an asset model, since a large portion of electricity use is either driven by behavior or lighting and appliance efficiencies that are not included in the model. The Home Energy Score baseline electricity loads are higher than indicated by utility data. The Home Energy Score results show a minimum value of 5,854 kWh, whereas 57 pilot homes, or 41%, used less electricity.

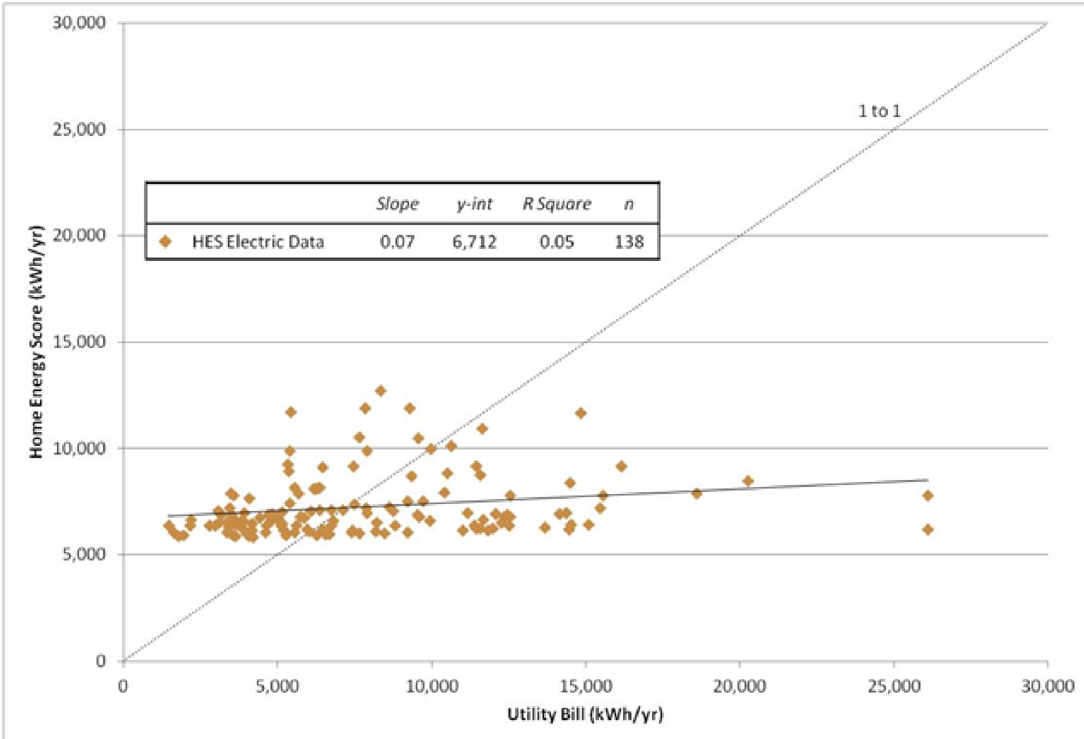


Figure 19. Electricity comparison for Home Energy Score versus utility data (scatter)

The sorted line results in Figure 20 also show the Home Energy Score's high baseline for electricity uses. In addition, this graph displays the variability from home to home, which, unlike the gas use, does not trend upward in step with the building model. This reinforces the challenge of capturing electricity use in an asset-only model, where one does not examine the large drivers of electricity variability: lighting, appliances, or occupant habits. The trending would presumably be more pronounced in regions with higher space cooling loads.

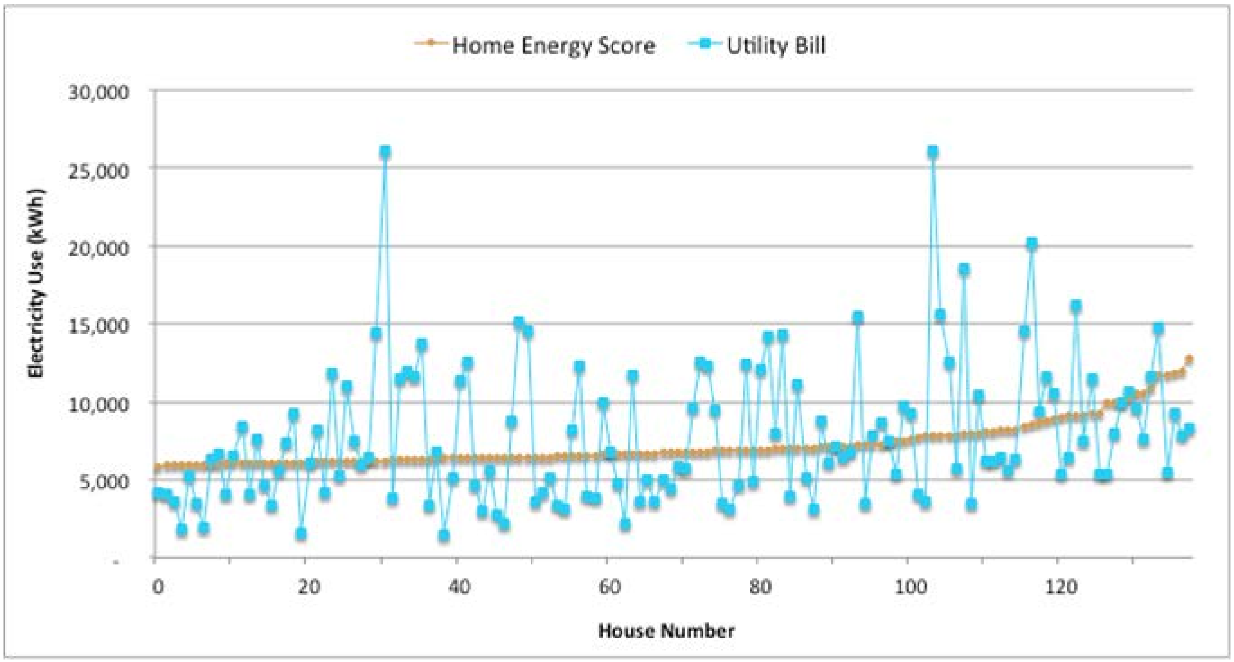


Figure 20. Electricity comparison for Home Energy Score versus utility data (line graph)

One interesting result for homes in the Minnesota climate is how the Home Energy Score's electricity results change with a home's heating system. Figure 21 shows electricity use versus floor area, and categorizes homes by their type of heating system (boiler or furnace) and whether or not they have central air conditioning. Results show a clear pattern for the different categories. For similarly sized homes, those with boilers use more electricity than the ones heated by a furnace. The 14 homes with a furnace and no central air conditioning show a negligible increase in electricity use with floor area. The population average annual electricity use is 5,982 kWh with a standard deviation of 100 kWh (note this is indicative rather than statistically significant for this small sample size). This suggests that the electricity demand from furnace air circulation fans is not being captured by the model and scaled along with the heating load.

Changes to the Home Energy Score methodology since the pilot version have adjusted the model's energy calculations. For results using more recent versions of the Home Energy Score see Roberts et al. 2012. For documentation of the changes made to the Home Energy Score see Bourassa 2012 or the Home Energy Score documentation (LBNL 2010).

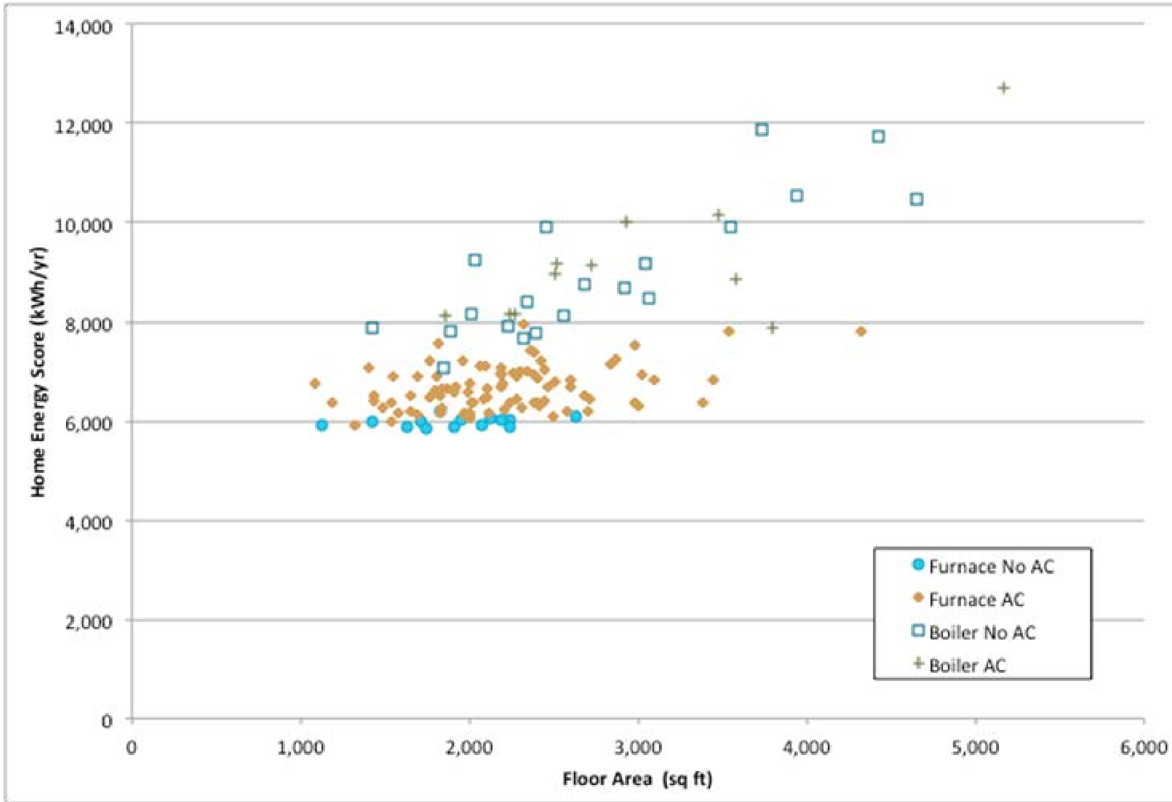


Figure 21. Home Energy Score electricity demand versus type of heating system

Figure 22 shows results from the SIMPLE model. Again, the model versions 1 and 2 refer to the two approaches where SIMPLE was used as an asset model: SIMPLE-1 used the Home Energy Score baseline assumptions, and SIMPLE-2 used SIMPLE's built-in assumptions (see Section 4.3.2). For natural gas use, there is very little difference between the two versions, as expected, since only a small fraction of natural gas use is attributable to the baseline demand. Results show that SIMPLE overpredicted the gas use, but the average difference was only 130 therms per year for version 1 compared to 603 therms for the Home Energy Score (see Table 4). In addition, the average percent difference for version 1 was 16% compared to 55% for the Home Energy Score.

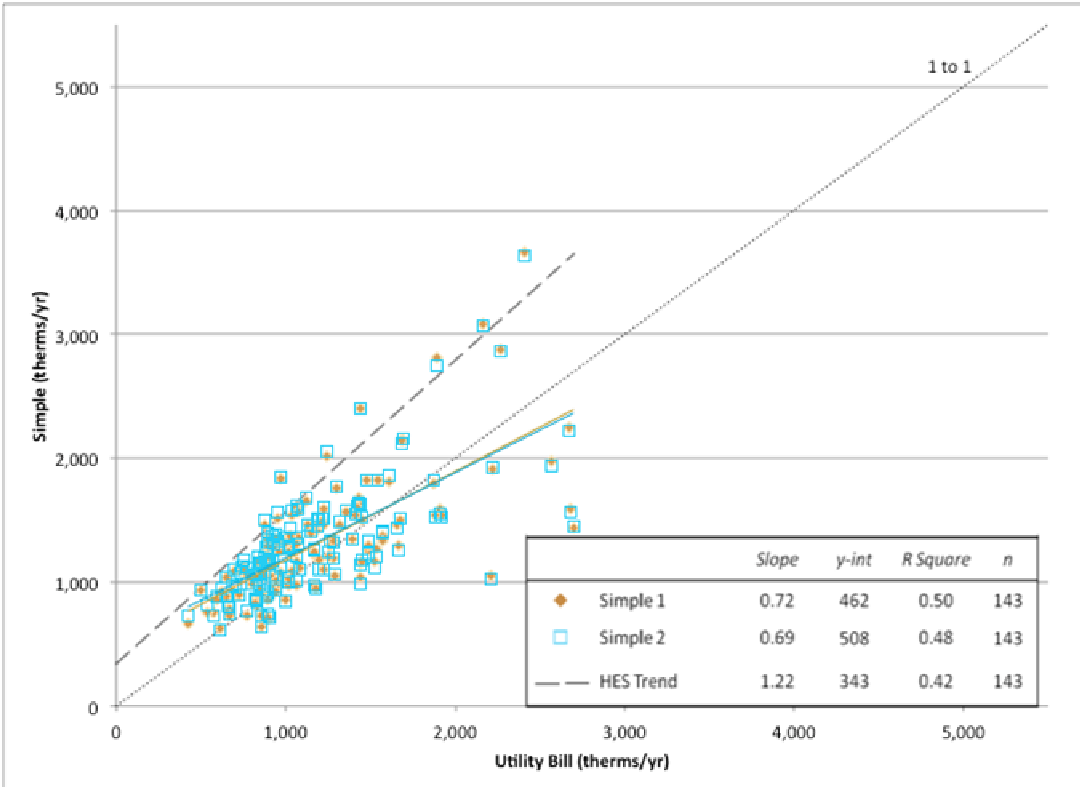


Figure 22. Natural Gas comparison for SIMPLE versus utility data

Figure 23 shows electricity results from SIMPLE. The electricity results also follow a similar pattern to the Home Energy Score. Asset models do not capture the high variation in utility electricity bills because they factor out the major drivers of electricity use, such as occupant-driven use of lights, electronics, and appliances. Figure 23 shows that the SIMPLE-2 model, which used built-in default assumptions, is a slightly better fit with electricity data. Some of these assumptions scale occupant loads with the size of the home.

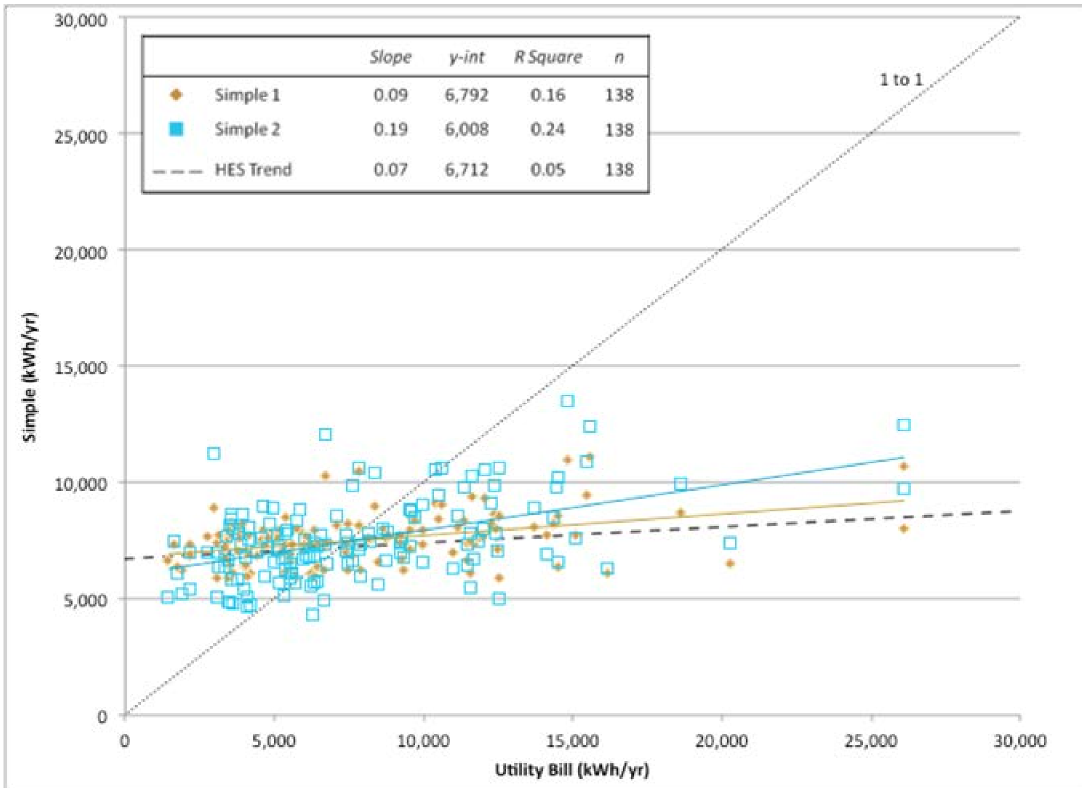


Figure 23. Electricity comparison for SIMPLE asset model versus utility data

The results below compare the third version of the SIMPLE model tested in this pilot, where non-asset factors of a home's energy use are considered, namely, occupant habits and additional large plug loads. A summary of survey results about occupant habits is included in Table 6. The average heating set point is 67.5°F, with 50% of respondents reporting within a range of 2.5° from the average. The average cooling set point is 75°F. There is a spread of 16° between the minimum and maximum, with 50% of respondents reporting within a range of 3°. However, it is important to note that set points are self reported and several surveys were performed during summer months, when winter set points may be misremembered (or optimistic). In addition, 37% of surveyed homes have more than one refrigerator, 67% of homes report one or more large additional electric appliances, and the percent of homes that have electricity as their fuel for cooking and drying clothes are 22% and 20%, respectively.

Table 6. Summary of Responses From SIMPLE Occupant Survey

Survey Question	Summary of Responses	Number of Respondents
Heating Set Point	Average: 67.5°F Minimum: 60°F 25th percentile: 65°F Median: 68°F 75th percentile: 69°F Maximum: 74°F	50
Cooling Set Point ***(if applicable)***	Average: 75°F Minimum: 62°F 25th percentile: 74°F Median: 75°F 75th percentile: 78°F Maximum: 78°F	31
Number of Refrigerators or Freezers	1 unit: 32 homes 2 units: 14 homes 3 or more: 5 homes	51
Primary Refrigerator Style	Freezer on top: 22 Freezer on bottom: 9 Side-by-side: 15	46
Number of Other Large Electric Uses[9]	None: 17 homes 1 additional: 19 homes 2 additional: 10 homes 3 or more: 6 homes	52
Cooking Fuel	Gas: 35 homes Electric: 10 homes	45
Dryer Fuel	Gas: 36 homes Electric: 9 homes	45

Utility data were available for all 53 homes that completed the survey. Results are shown in Figure 24 and Figure 25. The average annual gas use for the subset of 53 houses evaluated using the asset version of SIMPLE (SIMPLE-3) is only 16 therms (1.4%) less than the use for the 140 houses analyzed using the first two versions. The major habitual driver of heating load captured by the SIMPLE-3 survey is thermostat set point, though as shown above, there is not a large range in responses. The electricity results show that the least squares regression line for SIMPLE-3 has a slope closer to 1.0 (0.65) and a better fit (R-squared of 0.59) to the utility data than the first two versions. As expected, this validates the inclusion of additional occupant behavior on improving electricity prediction. Note that the highest electricity users were not captured in this subsample of 53 homes. Therefore, the regression lines for SIMPLE-1 and SIMPLE-2 also have a slope closer to 1.0 when those outliers are removed.

[9] Additional electricity uses include dehumidifiers, pools or electric saunas, window air conditioners, etc.

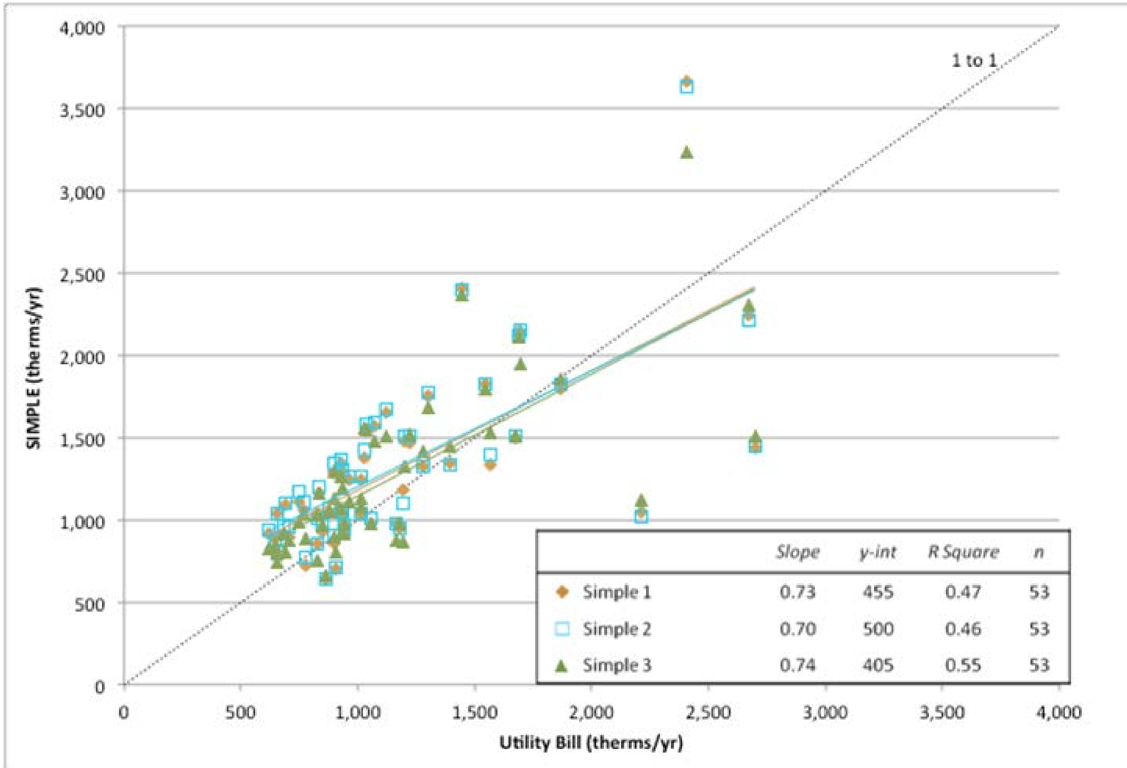


Figure 24. Natural gas output for three versions of SIMPLE model

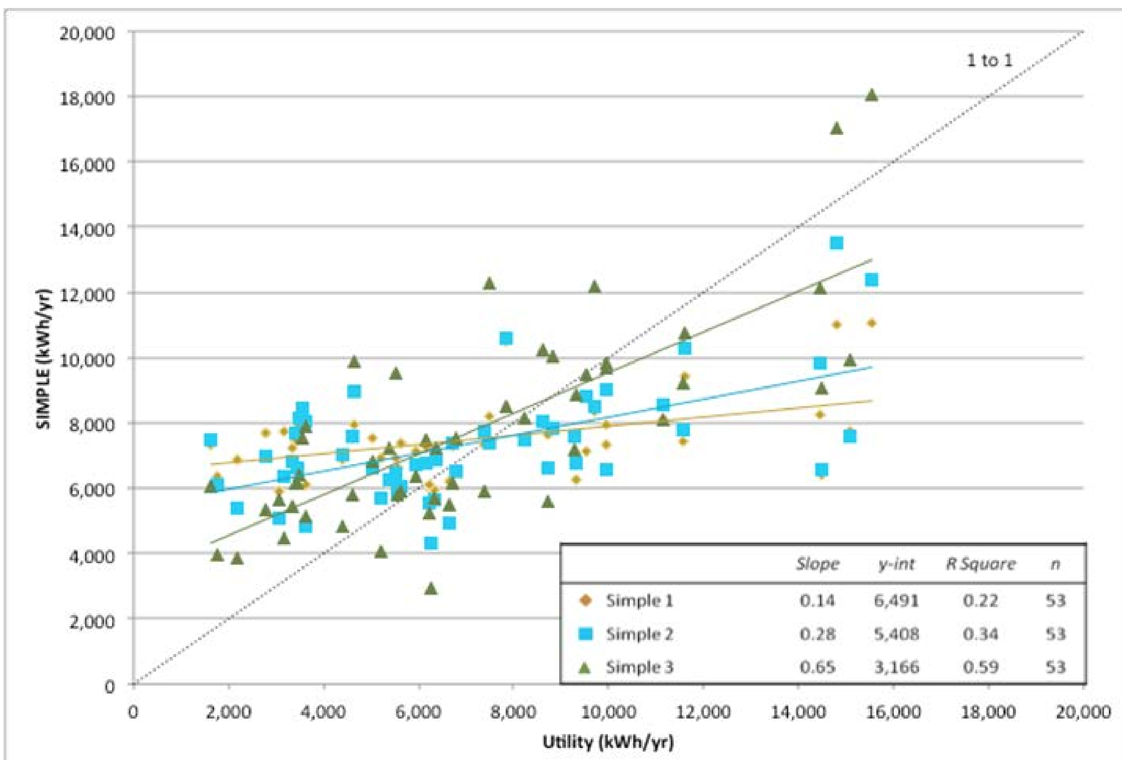


Figure 25. Electricity output for three versions of SIMPLE model

Figure 26 and Figure 27 below compare the REM/Rate annual gas and electricity use to that from utility bills for 51 of the 154 Home Energy Score pilot homes. Figure 26 shows that REM/Rate overpredicts gas use for all except three homes. However, REM/Rate results are scaled closely to the actual usage, with a regression slope of 1.14. The regression slope and standard error of 0.016 fall within the one to one line of agreement. Figure 27 shows the REM/Rate results for electricity use. REM/Rate scaled the electricity results far better than any of the other asset models, producing a slope of 0.37, compared to 0.07, 0.09, and 0.19 for the Home Energy Score, SIMPLE-1, and SIMPLE-2, respectively. REM/Rate collects more detailed information on the home's appliances, including their fuel type, refrigerator age, and percent of fluorescent lighting. REM/Rate also scales some electric loads with conditioned floor area and the number of bedrooms.[10] SIMPLE-2, with a slope of 0.19, also scales some loads with home size and records the energy source for the dryer and oven. Note that the REM/Rate dataset included one of the three large users with an annual electricity demand greater than 20,000 kWh.

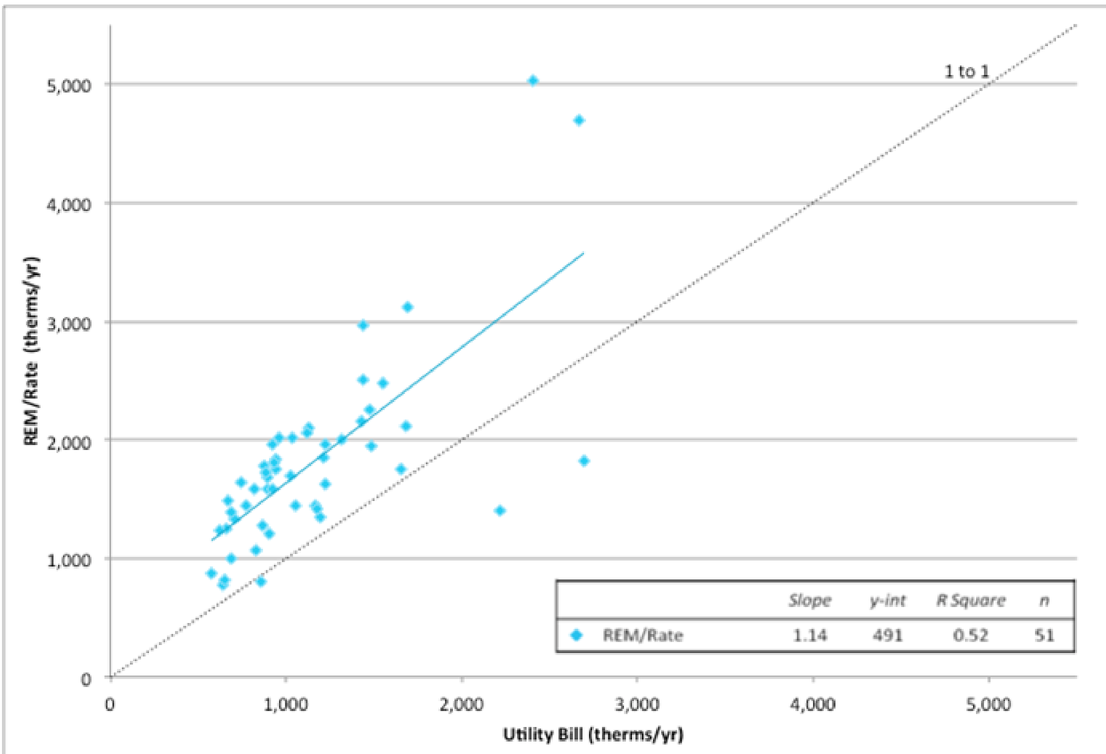


Figure 26. Natural gas bill comparison of REM/Rate results

[10] For more information see Chapter 3 of RESNET standards (2006 referenced for this report).

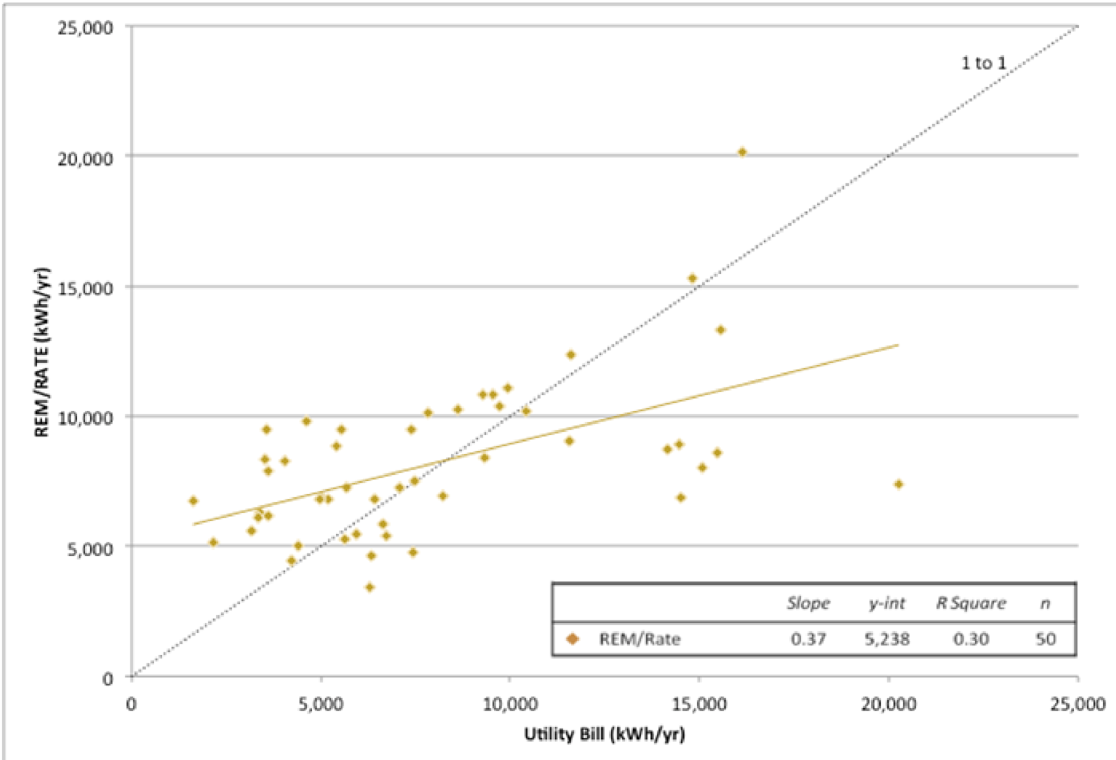


Figure 27. Electricity bill comparison of REM/Rate results

Overall, this comparison of model output to utility data indicates these models perform somewhat well as predictors of actual energy use in Minnesota homes, given the limitations of decoupling asset from operational energy use. However, there is systematic overprediction of natural gas use by all models. In Minnesota's climate, natural gas is the important measure of asset energy efficiency, since air conditioning loads are historically low. This indicates that, for this population of homes, there is an overall bias in the model calculations. This could be partially accounted for by updating the underlying assumptions to account for the consistently lower use by Minnesota homes. [11]

Besides overall bias, the other important factor for model prediction is correlation. The R-squared measure indicates how much of the variation in utility data between homes is accounted for in the model. The three asset models were able to explain 42%–52% of the variation between homes' natural gas use, and 5%–30% of the variation in electricity use. The validity of this level of correlation will depend on the application at hand, but it urges caution in interpreting and presenting model results with overconfidence. This includes, for example, limiting the level of significant figures that are used.

A surprising result for this population of homes is that the more detailed characterization of a home did not provide a better estimate of a home's asset energy use. The REM/Rate model,

[11] Numerous changes have been documented in the Home Energy Saver online documentation (https://sites.google.com/a/lbl.gov/hes-public/home-energy-scoring-tool/release-history) and Bourassa et al. 2012.

which takes (in our estimation) approximately five times longer to execute than the SIMPLE model, overpredicted gas use by three times as much on average (63% compared to 18%, see Table 4). This result indicates that the additional program cost to collect more detailed data is not justified.

5.4 Retrofit Upgrade Recommendations

This study also compared the automated upgrade recommendations generated by the Home Energy Score pilot version to the custom recommendations homeowners received through CEE's residential program. Again, the Home Energy Score recommendations are generated automatically using a national database of measure costs. CEE makes energy efficiency recommendations based on cost-effectiveness, appliance age, and health and safety considerations. Both the Home Energy Score and CEE use a 10-year payback as the threshold for cost-effectiveness. CEE's savings estimates derive from Minnesota utility conservation program plans, approved every three years by the Department of Energy Resources. Results for four weatherization measures are shown in Figure 28.

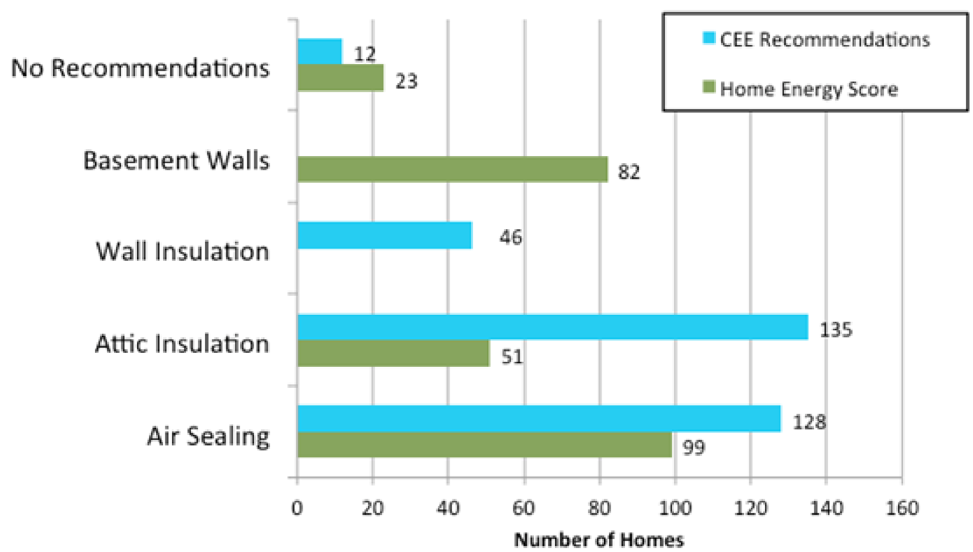


Figure 28. Upgrade recommendations for pilot homes (n = 154)

Overall, the CEE field staff made upgrade recommendations more often than were generated by the Home Energy Score. The Home Energy Score made no recommendations for 23 homes, about twice the number of the field crew. The one exception was the recommendation to insulate basement walls, which the Home Energy Score recommended for more than 50% of homes while the field crew never included that recommendation. This was primarily a result of the confusion over conditioned basements, since the Score treated numerous homes as incorrectly heating the basement. However, it also suggests that the savings from basement wall insulation might be over-predicted or the actual treatment cost might be underpredicted for the pilot region.

The CEE program recommendations also included wall insulation, attic insulation, and air sealing more frequently than the Home Energy Score. It is most surprising that the Home Energy

Score does not recommend wall insulation for homes for which there is no cavity insulation. The lower frequency of recommendations is likely due to a combination of factors. First, the Home Energy Score may overestimate the measure costs in the pilot region. Second, the field crew may recommend insulation for only one portion of a wall or attic space (e.g., a side attic of a story-and-a-half home), which would not be captured by the Score's treatment of attic space as a weighted average R-value. And finally, CEE recommends air sealing for reasons other than the air leakage diagnostic results, most commonly when there is evidence of ice dams or attic moisture. In addition, the CEE field crew recommended furnace upgrades for 37 homes, and water heater upgrades for 25 homes while the Home Energy Score did not recommend these measures.

There are at least two post-pilot updates to the Home Energy Score that will affect which automated recommendations are made. First, the cost information in the retrofit measures database has been updated and will continue to be refined as more data are collected. Second, the Home Energy Score has created microclimate regions to subdivide the 19 climate zones in the pilot version. This allows the Score to use more specific HDD and CDD information for payback calculations. However, the localized installation costs and rebate information that factor into program recommendations are not integrated into the Home Energy Score.

5.5 Energy Audit Tool Usability

The project included a qualitative assessment of the overall field usability of different audit tools, including the estimated data collection time, the availability of certain inputs, and field crew feedback on particular challenges. This information is mainly pertinent to the Home Energy Score and SIMPLE, which were piloted by the same field crew on almost all homes. There was less need to document our REM/Rate experience, since that tool is more commonly used.

Field staff estimated that the additional data collection for the Home Energy Score typically required an extra 30 minutes of time in a home, for a total visit length of 90–120 minutes. This increase was as little as 20 minutes or as much as 60 minutes, depending on the complexity and size of a home. Several of the data points were already collected as part of the standard home visit, in particular the blower door test and some mechanical and insulation values, although more detailed information was required for the audit tools. The Home Energy Score input that took the longest amount of time was measuring total window area. In addition, even if a home characteristic was collected for another purpose, recording that information on a second paper data form added to the logistical time.

In certain instances, the Home Energy Score required information was not readily available. These inputs were therefore also not available for SIMPLE, but since many building characteristics were estimations it was a less obvious issue. The field crew recorded the percent of attic area where the insulation value could be verified (either visually or because records existed of recent work), an occasional limitation of story-and-a-half homes. On average, 68% of attic area could be verified. There were 50 homes where more than 90% of the attic space could be verified, and 23 homes where less than 25% of the area could be verified. Wall insulation was not visually verified on 42% of homes, primarily because the homeowners declined an inspection hole. In addition, the efficiency of mechanical systems was frequently unavailable, in which case the field crew relied on the estimated age of the system. Heating system annual fuel utilization efficiency was available for 64% of homes, water heater efficiency was available for 14% of homes, and air conditioner seasonal energy efficiency ratio was available for 12% of homes.

Several of the Home Energy Score inputs that were a challenge to interpret have already been discussed in the sections above. The most common was the interpretation of conditioned basements, which was an issue for the 50% of homes that had some unfinished basement space. The second most common issue was how to characterize story-and-a-half homes, in particular, which areas to classify as wall insulation versus attic insulation. While less frequent, there were also questions about how to deal with townhomes, split-level homes, and underground garages.

The figures below focus on SIMPLE's streamlined data collection approach and compare the field crew's qualitative estimations of certain house characteristics with the corresponding diagnostic measurement collected later in the home visit. The figures show two cases: airtightness and window area. These diagnostic measurements are two of the more time consuming to collect, and the goal was to gauge the field crew's intuitive reporting of those house characteristics. This comparison should be considered more illustrative than conclusive. CEE had not developed a protocol for how to make estimations, results are from two field staff only, and they made estimations at the beginning of a visit—before fully inspecting a home—to avoid being influenced by the actual measurements.

Figure 29 shows a box-and-whisker plot for field staff estimations of house airtightness compared to the results of the blower door test, represented as ALR.[12] These results show some stratification of the field crew estimations. Homes classified as "average" cover the entire range, but the ranges covered by the 25th to 75th percentiles of "leaky" homes and "fairly tight" homes are distinct, and fall above and below an ALR of 1.0. The two extreme classifications, "very leaky" and "tight" are applied to only three homes each.

[12] The ALR is calculated as the blower door results in cfm50 divided by the building shell area in square feet.

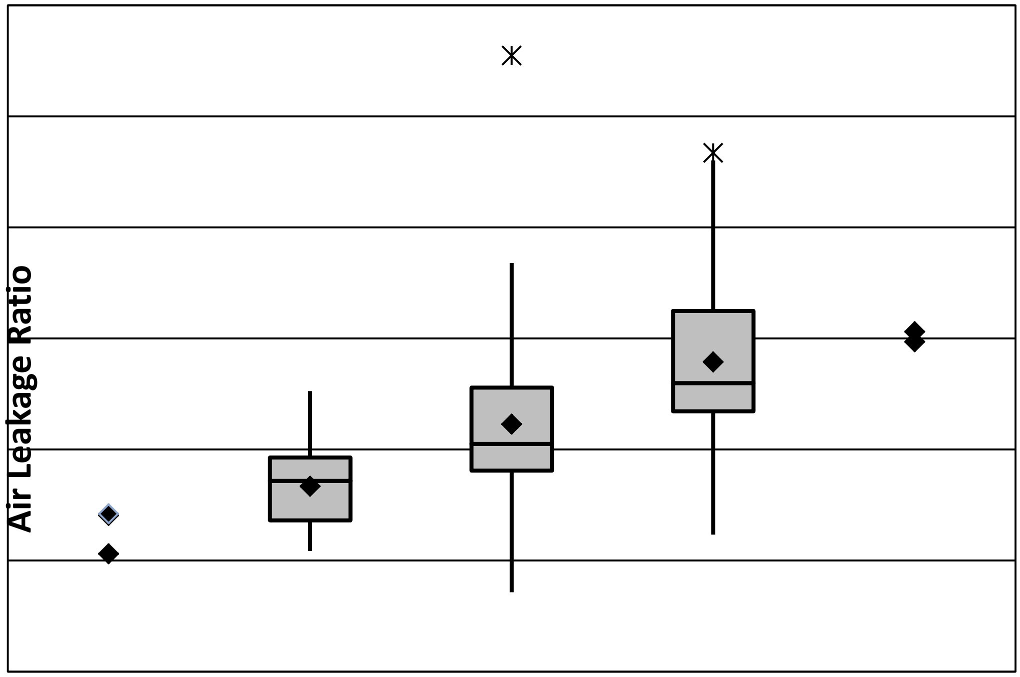


Figure 29. Field crew estimations of airtightness for SIMPLE input

By comparison, the window area estimations shown in Figure 30 have a less distinct pattern. The field crew estimations are shown against the home's ratio of total window area to floor area. This shows that homes classified as having a "high" window area had either a large ratio or could have been large homes, which would lower the ratio of window area to floor area. The characterization of window area as "low" or "typical" was scattered across the range of window and home sizes. A well-defined protocol for estimating window area should improve these outcomes.

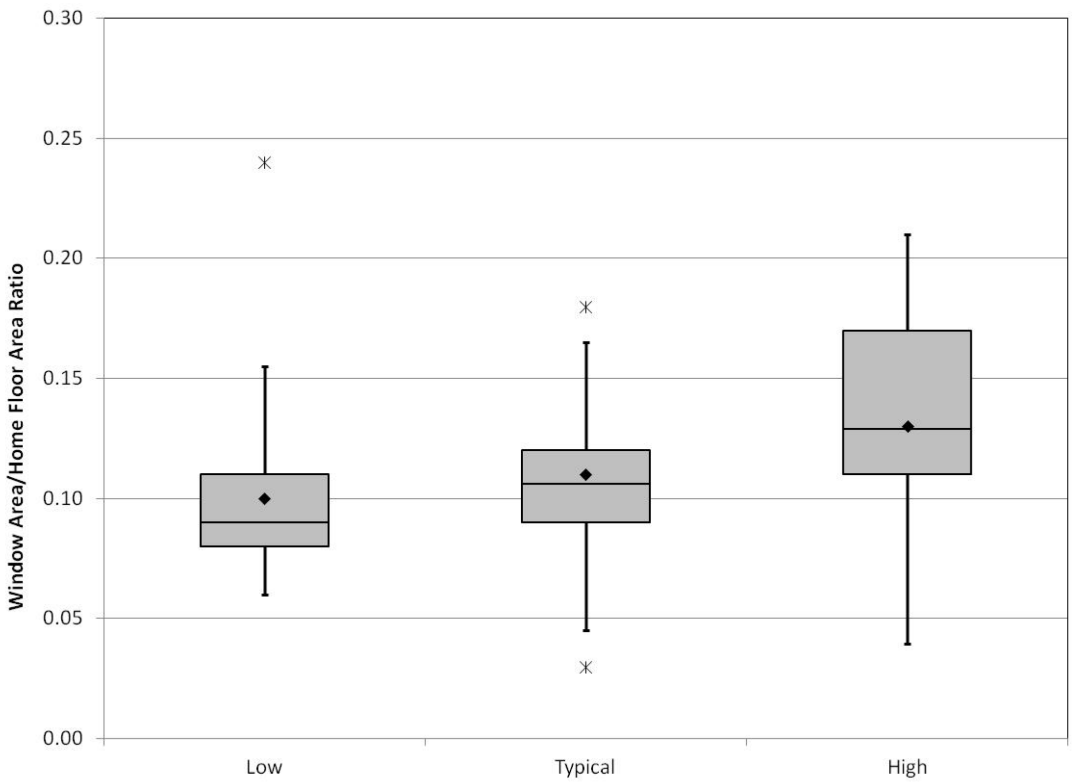


Figure 30. Field crew estimations of window area for SIMPLE inputs

6 Conclusions

This pilot project focused on three audit tools that have a range of complexity in how they characterize building performance: the Home Energy Score, SIMPLE, and REM/Rate. This assessment focused on their ability to deliver useful energy ratings in the existing home retrofit market, which heightens the importance of cost-effective delivery, but also requires a reliable assessment of a home's energy use in order to deliver a trusted result. The estimated length of time required for CEE staff to complete the field data collection and data entry for one home is approximately 1 hour for SIMPLE, 1.5–2 hours for the Home Energy Score, and 4–6 hours for REM/Rate.

The comparison between the models' energy results to utility bills did not show a correlation between the complexity of an audit tool and its ability to predict a home's energy use. In fact, for these pilot homes, the tool with the fewest inputs, SIMPLE, provided the closest prediction of a home's annual natural gas use. The average percent difference between SIMPLE-2 and actual use was 18%, compared to 55% for the Home Energy Score and 63% for REM/Rate. This result indicates that the additional program cost to collect more detailed data is not justified.

The models' ability to predict electricity use was less consistent across homes. This was expected from an asset model that doesn't account for occupant habits and plug loads, the main drivers of variation in home electricity use in Minnesota climates, which have lower air conditioner loads. The average difference between the model and utility bill data was 23% for the Home Energy Score, 25% for SIMPLE-2, and 27% for REM/Rate. However, these relatively low average percent differences hide the fact that many homes use less electricity than the lowest predicted by the models and a handful of homes use considerably more. This raises some doubt about the value of including electricity use in asset ratings in regions without high electric heating or cooling loads.

In general, this comparison of model output to utility data indicates there is an overall bias in the model calculations. This could be partially accounted for by updating the underlying assumptions to account for the consistently lower use by Minnesota homes, as has occurred in the national version of the Home Energy Score. However, there was also low correlation as represented by the R-squared calculations. The three asset models were able to explain 42%–52% of the variation between homes' natural gas use, and 5%–30% of the variation in electricity use. The validity of this level of correlation will depend on the application at hand, but it urges caution in interpreting and presenting model results with overconfidence.

The tradeoff of a simpler audit tool is that field technicians must have a consistent protocol for how to interpret the variety of home features they will encounter. Components of the building structure that were the most challenging to implement in the Home Energy Score were basements, story-and-a-half homes, underground garages, walkouts, and homes with major additions. These were all "outside the box" situations that required approximations, but there were not clear instructions for interpretation that all technicians would have followed in a consistent manner.

Energy upgrade recommendations are challenging to automate within a rating system without direct input from a field technician. There was minimal consistency between the recommendations generated by the Home Energy Score and those that the CEE field crew made in-person. In particular, the field crew recommended insulation and air sealing measures twice as frequently as the Home Energy Score. However, one added value of energy auditing tools can be to assist the auditor to calculate the savings from designated upgrade measures.

The Home Energy Score point system is a potential challenge for motivating home retrofits. Minnesota homes scored an average of 5.4 points out of 10, and could increase their score by an average of 2 points if they invested in recommended upgrades. This small increase might be a disincentive for homeowners who want to feel satisfied by a sense of completion, especially if they are less concerned with comparing their homes across a wide population (compared to new home buyers).

References

Bourassa, N. J.; Rainer, L. I.; Mills, E.; Glickman, J. (2012). "The Home Energy Scoring Tool: A Simplified Asset Rating for Single Family Homes." Proceedings of the American Council for an Energy-Efficient Economy 2012 Summer Study on Energy Efficiency in Buildings. August 13–17, 2012. Asilomar, CA.

Carroll, E.; Hatton, E.; Brown, M. (2009). "Residential Energy Use Behavior Change Pilot." MN Department of Commerce B21383. Prepared by Franklin Energy.

Davis, M. (2011). "Behavior and Energy Savings: Evidence from a Series of Experimental Interventions." New York NY: Environmental Defense Fund.

Dunsky, P.; Lindberg, J.; Piyalé-Sheard, E.; Faesy, R. (2009). "Valuing Building Energy Efficiency through Disclosure and Upgrade Policies: A Roadmap for the Northeast U.S." Dunsky Energy Consulting, November 2009.

Earth Advantage Institute and Conservation Services Group (2009). "Energy Performance Score 2008 Pilot." Prepared for Energy Trust of Oregon. August 2009.

Earth Advantage Institute (2011). "About EPS." www.earthadvantage.org/programs/energy-efficiency/energy-performance-score/about-eps. (Accessed January 10, 2012.)

Faesy, R. (2010). "Building Energy Labeling." Presented at the RESNET Building Performance Conference. Raleigh, NC, February 24, 2010.

LBNl. (2010). "The Home Energy Scoring Tool for Asset Rating - Home Energy Saver: Engineering Documentation." https://sites.google.com/ a/lbl.gov/hes-public/home-energy-scoring-tool. (Accessed November 11, 2010.)

Mills, E. (2002). Review and Comparison of Web- and Disk-Based Tools for Residential Energy Analysis. LBNL-50950. Berkeley CA: Lawrence Berkeley National Laboratory.

Nelson, C. (2011). "Energy Efficient Cities: Using a Community-Based Approach to Achieve Greater Results in Comprehensive, Whole-House Energy Efficiency Programs." Minneapolis MN: Center for Energy and Environment, November 2011.

Newport Partners LLC. (2011). "Motivating Home Energy Improvements: Focus Groups for the U.S. Department of Energy." Review Draft available at http://apps1.eere.energy.gov/buildings/publications/pdfs/homescore/doe_fg_report.pdf

Opower. (2011). "Home Energy Reports." http://opower.com/what-is-opower/reports. Accessed December 22, 2011.

Parker, D.; Mills, E.; Rainer, L.; Bourassa, N.; Homan, G. (2012). "Accuracy of the Home Energy Saver Energy Calculation Methodology." Proceedings of the American Council for an Energy-Efficient Economy 2012 Summer Study on Energy Efficiency in Buildings. August 13–17, 2012. Asilomar, CA.

Roberts, D.; Merket, N.; Polly, B.; Heaney, M.; Casey, S.; Robertson, J. (2012). *Assessment of the U.S. Department of Energy's Home Energy Scoring Tool*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-54074.

U.S. Department of Energy. (2010). "National Building Performance and Rating Program: Residential Effort Stakeholder Webinar." March 9, 2010.

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

DOE/GO-102013-3817 ▪ July 2013

Printed with a renewable-source ink on paper containing at least 50% wastepaper, including 10% post-consumer waste.