What do they know? Strengths and limits in survey data collection about residential structures and equipment

Ingo Bensch, Energy Center of Wisconsin, Madison, WI
Karen Koski, Energy Center of Wisconsin, Madison, WI
Patrick Michalkiewicz, Peoples Gas & North Shore Gas, Chicago, IL

ABSTRACT

Data collected from household surveys form the basis of many energy efficiency goals, program designs, and evaluations. Potential studies use surveys to understand the penetration of efficient equipment and options for structural upgrades and set goals that efficiency programs are expected to meet. Market research uses surveys to understand homes, households, appliances, practices, and customer tendencies with the goal of creating program designs that will be effective. Program evaluations draw on survey responses to establish pre-program baselines and post-program saturations and to understand why people made the choices they did. In other words, survey data are fundamental to the energy efficiency field, but are the data accurate? What data can be reliably collected by customer as self-reports and what requires a field visit?

The authors of this paper conducted a large study of energy efficiency potential for two natural gas utilities in the Midwest that included a natural opportunity to investigate the ability of telephone surveys to accurately identify:

- building characteristics
- the presence of different types of natural gas-using equipment
- the type, age, and efficiency level of the heating system
- the type and efficiency level of the water heater
- the type of thermostat
- temperature settings that households maintain.

This investigation is based on 1,772 telephone survey completions with residential customers and site visits to the homes of 87 of these respondents. The site visits were intended to collect more detailed data than can be gathered by telephone, but they provided a convenient opportunity to check the accuracy of data that is commonly collected by energy program researchers and evaluators through telephone surveys.

The paper will discuss what survey data proved to be accurate and what did not. The authors will also analyze the source of errors in self-reports, highlight potential improvements, and discuss inherent limits for data collection of building and equipment data through telephone surveys. The ultimate take-aways for readers are empirically based insights on how far customer surveys can go in collecting housing and household information, the degree of accuracy of these data, and what metrics require field visits.

Introduction

Household surveys have become an essential data source for many studies and program evaluations in the energy efficiency field. Researchers, evaluators, and utilities use surveys routinely for such diverse purposes as characterizing markets and buildings, understanding load profiles, estimating the available efficiency potential, determining program impacts on participants' energy usage, and tracking customer and participation satisfaction. We contact randomly chosen households to ask questions on such topics as the respondents' awareness of particular efficiency programs, the equipment

they have in their homes, what energy-using appliances they have bought recently, how they chose from among the available models, how much a particular program, brand, or feature influenced their choices, how consistently they do (or don't) turn off equipment that is not in use, their perceptions of the relative merits of potential energy-saving measures, their interest in saving energy, and how they feel about their utility.

Surveys are a convenient and relatively inexpensive tool to obtain answers about a wide range of opinions and facts concerning utility customers, program participants, and people whom energy efficiency advocates would like to influence. They cost less and allow the collection of a much greater number of data points than either interviews or onsite audits.

However, the convenience of surveys as a data collection tool comes at a different sort of cost ... accuracy. As we have seen in past studies, there are some questions people just cannot answer accurately. For example, Pigg and Nevius found that, while homeowners were willing to tell us how much insulation they had in their attics, they were often unable to do so accurately. The majority of homeowners for whom an energy audit found inadequate insulation thought their home was adequately or well-insulated (Pigg 2000). Similarly, in a comparison of two types of self-reports—visual questions showing a graphic and questions relying only on words—Peters found that verbal questions alone may not be enough to produce accurate self-reports (Peters 2002). This was particularly true for the Energy Star logo, which would be difficult to describe verbally, and devices that respondents may not recognize consistently by the term we would use in a survey question.

So, what, we wondered, are the limits of telephone survey-based data collection? What energy-related questions can people answer accurately in response to simple survey questions? When are more sophisticated survey methods or other forms of data collection called for?

A Potential Study Offers a Convenient Natural Experiment

During the fall of 2012, the Energy Center of Wisconsin conducted fieldwork for a potential study for Peoples Gas and North Shore Gas, the natural gas utilities that serve the city of Chicago and some of its northern suburbs. Primary data collection consisted largely of a combination of a telephone survey of a stratified random sample of 2,096 residential customers and field visits to 111 of the surveyed households. The study also comprised surveys and visits to non-residential customers, but we focused on the residential sector for this paper.

The residential telephone survey comprised 51 separate questions, ranging from housing and equipment characteristics to thermostat practices and past remodeling efforts. Our survey questions resembled those asked by many other surveys in the energy efficiency field. The respondent was the natural gas account holder or another adult who claimed to be familiar with natural gas usage in the home.

The onsite visits were intended to collection additional data about a smaller sample of Peoples Gas and North Shore Gas customers that that would require technical measurements, such as blower door tests, or that we didn't think customers would be able to answer consistently. However, in the process, we collected field data on a few of the questions we had already asked during the telephone survey, thereby providing an opportunity to determine the accuracy of the survey data.

Our Methodology

Our methodology for the analysis presented in this paper was quite simple. First, we took an inventory of all questions from the telephone survey for which we also had data from the field visits that could corroborate or refute the household's self-report. Then, we compared the responses. Given that our data from the site visits was based on observations taken by residential technical experts and/or photographs taken during the site visits, we took the observed data as correct whenever observed data and survey responses disagreed.

We limited our analysis to a comparison of survey responses and observations from the residential sector because (1) we had more overlap between the telephone survey and onsite data collection within the residential component of the study and (2) households are more uniform and thus can be analyzed as a single unit. Within the residential sector, we analyzed data for single- and multi-family housing, but excluded housing units in master-metered buildings.

Results

We were able to compare survey respondents' self-reports to observed values from our onsite visits for 14 different metrics. We had asked about each of these metrics in the telephone survey using one of three different question types: yes/no questions, multiple choices questions, and short-answer questions. We have detailed the results from several of these comparisons below and summarized them all in Table 6 at the end of this section of the paper.

Building characteristics

Building characteristics and sizes were important data elements for our study. We established these with the three following questions and also recorded these characteristics during our onsite visits:

Is the building at this address...
a one-family house detached from any other house
a one-family house attached to one or more houses
an apartment or condo building with two or more units
something else

How many stories does (your home at address) have? Please exclude basements or other spaces below ground level.

What is the approximate square footage of the living space of (address)?

Respondent self-reports agreed with on-site observations for 95 percent of the 73 cases in which we had data on the building type to compare, but only in 84 percent of the 87 cases in which we had data on the number of stories in the building. Thirteen percent of survey respondents reported a larger number of stories (three instead of two or two instead of one) than we observed during our visit, possibly because they ignored our request to exclude basements and other spaces below ground.

Conversely, however, survey respondents underreported the sizes of their homes when we asked for the amount of living space. Only 30 percent of the 64 respondents[1] for whom we have self-reports and onsite measurements provided a home size within 500 square feet of our measurement. Figure 1 shows a comparison of self-reported to measured home sizes. Most of the self-reported values are below the unity line, indicating that people underreported their home sizes.

[1] This number excludes 26 respondents who indicated that they did not know the size of their home and five respondents for whom we were missing either the self-report or a field measurement from the onsite visit.

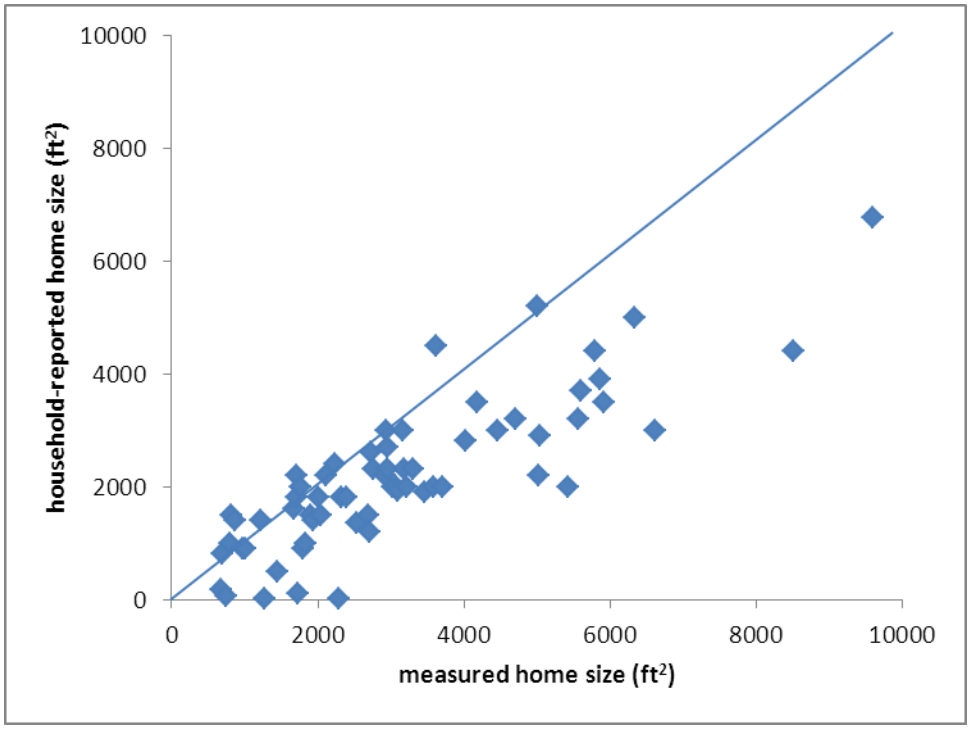

Figure 1. Comparison of self-reported to measured home sizes

Natural gas-using equipment

A series of questions early in the survey instrument established the end-uses for which the respondent's household uses natural gas. The inquiry was structured as a root question with a list of end-uses to which respondents were asked to provide yes/no answers:

> *Next, I will list some applications for which some customers use natural gas. For each one, please tell me whether you use natural gas for this purpose in your home.*[2]
>
> *heating the home*
> *heating water for activities like bathing, clothes washing and dishwashing*
> *cooking (range top or stove top with gas burners)*
> *baking (oven)*
> *drying clothes*
> *fueling a natural gas fireplace*
> *heating water for hot tubs or spas*

Survey responses matched our onsite observations in the vast majority of cases for space heating, water heating, and cooking with accuracy rates of 97 percent or better. However, data agreement was lower—generally in the 80 to 90 percent range—for the remaining end-uses, as noted in Table 1 below. Survey responses tended to report natural gas use for fewer ovens and fireplaces than we observed, but more natural gas clothes dryers and hot tubs/spas than we saw. Overreports of natural gas-fueled hot tubs and spas were substantial, amounting to 14 percent of our observed cases. (In the case of hot tubs, the misreporting was more often about the presence of a hot tub or spa rather than the fuel used to heat its water.)

[2] Respondents in single-family-attached homes and those in multi-family buildings were also told that we wanted to know only about uses of natural gas that are part of their natural gas account.

We hypothesized that some of the apparent underreporting of appliances using natural gas could be the result of respondents incorporating their knowledge of whether they actually use their ovens and fireplaces into their response, while our onsite team could only observe the presence of those appliances. That is, respondents might have been telling us about the ovens and fireplaces they use rather than the ones they have. The apparent overreports of spas and hot tubs could be the result of ambiguity of what features constitute a hot tub spa or socially desirable response bias.

Table 1. Fuel source for various appliances: comparison of self-reported and observed conditions[3]

Appliance: Oven (n=91)			
		observed condition	
		oven heated with natural gas	oven heated with something else
self-reported condition (telephone survey)	oven heated with natural gas	61 (67%)	3 **(3% false positives)**
	oven heated with something else	7 **(8% false negatives)**	20 (22%)
Appliance: Clothes Dryer (n=86)			
		observed condition	
		clothes dryer heated with natural gas	clothes dryer heated with something else
self-reported condition (telephone survey)	clothes dryer heated with natural gas	60 (70%)	10 **(12% false positives)**
	clothes dryer heated with something else	4 **(5% false negatives)**	12 (14%)
Appliance: Fireplace (n=86)			
		observed condition	
		fireplace heated with natural gas	fireplace heated with something else
self-reported condition (telephone survey)	fireplace heated with natural gas	24 (28%)	5 **(6% false positives)**
	fireplace heated with something else	9 **(10% false negatives)**	48 (58%)
Appliance: Spa (n=88)			
		observed condition	
		has spa/hot tub	does not have spa/hot tub
self-reported condition (telephone survey)	has spa/hot tub	1 (1%)	12 **(14% false positives)**
	does not have spa/hot tub	1 **(1% false negatives)**	74 (84%)

Heating system types

Once we established the presence of natural gas space heating, we asked a multiple choice question about the home's heating system type. The question was formulated as:

[3] Number of cases analyzed varies by metric because we excluded cases in which survey respondents answered "don't know" or did not provide a response for the applicable question. We also excluded a small number of cases for which we did not have usable on-site data.

What type of heating system is used to heat your home?[4]

forced air or warm air furnace
boiler
not sure

As shown in Table 2, 92 percent of respondents answered this question correctly with equal shares of respondents falsely identifying a furnace as a boiler or a boiler as a furnace. (Because furnaces are more common, this amounts to some bias in favor of boilers.)

Table 2. Heating system type: comparison of self-reported and observed conditions

n=81[5]		observed condition	
		forced air	boiler
self-reported condition (telephone survey)	forced air	61 (75%)	3 **(4% false positives)**
	boiler	3 **(4% false negatives)**	14 (17%)

Furthermore, as a proxy for whether people's heating systems are condensing units, and thus more efficient than average, we asked respondents whether a white plastic pipe extends from their furnace to an outside wall. (Past experience had shown that people are unable to tell us whether their system is condensing or give us its efficiency level in AFUE[6] terms.) Our survey question was:

Is there a white plastic pipe that goes from the furnace (or boiler) to the outside of your house or building, either through the wall or into a chimney?

We analyzed this question only for respondents in single-family homes because they are more likely to be able to answer detailed questions like this about their heating system. Of the 56 respondents for whom we had data, only 66 percent were able to give us an accurate answer.

Interestingly, though, respondents' ability to give us accurate answers was greatly dependent on whether the respondent actually had such a white plastic pipe. Nearly all of the respondents with condensing heating systems (the kind that would have such a pipe) indicated that they have a white plastic pipe, and the remaining three respondents said that they did not know. However, among those with non-condensing heating systems, answers split roughly evenly between yes (22), no (19), and don't know (16).

[4] Survey implementers were given clarifying statements to use as needed. In those statements, we explained that a central furnace would blow warm air into individual rooms, usually through ducts and that a boiler is a central appliance that heats water or makes steam and distributes it through radiators or pipes into individual rooms.
[5] Excludes responses from a small number of households that answered "don't know" or did not answer the question.
[6] annual fuel utilization efficiency

Table 3. Condensing heating system (white plastic pipe): comparison of self-reported and observed conditions

n=56 *(Limited to single-family homes; totals and percentages exclude "don't know" responses.)*		observed condition: has a condensing heating system	observed condition: does not have a condensing heating system
self-reported condition (telephone survey)	has a white plastic pipe	15 (27%)	19 **(34% false positives)**
	does not have a white plastic pipe	0 **(0% false negatives)**	22 (39%)
	don't know	3	16

Also, because local code that governs the clearance needed around a heating system combustion air exhaust point appeared to pose a barrier to the condensing heating systems, we asked survey respondents whether there are at least ten feet of space between their building and the neighboring building on either side. The survey question was formulated as:

> *On at least one side of (address) is there more than ten feet between (address) and the neighboring building?*

Seventy-two percent of respondents gave answers we could corroborate with measurements during our onsite visits. Of the remainder, nearly all (25%) overreported the space to the neighboring building.

Table 4. Ten feet of space on at least one side of house/building: comparison of self-reported and observed conditions

n=36		observed condition: has 10 feet or more	observed condition: does not have 10 feet
self-reported condition (telephone survey)	has 10 feet or more	17 (47%)	9 **(25% false positives)**
	does not have 10 feet	1 **(3% false negatives)**	9 (25%)

Thermostat types

After a series of questions about the home's heating system, we confirmed that the respondent had a thermostat and inquired about the type: programmable or manual. The question, which had been used in multiple prior surveys, was stated as a simple yes/no question in which we defined programmable thermostats and asked whether the respondent had one. The question was stated as:

> *A programmable thermostat is one that automatically changes the temperature during different times of the day. Regardless of whether you use this feature, do you have a programmable thermostat? (Some people call them a clock thermostat.)*

For thermostat types and programming, our field verification consisted of a second set of questions of the household and the photographing of the thermostat. Of the 80 telephone respondents for whom we have valid data for comparison, only 76% gave an answer to the telephone survey that was

consistent with the field team's observation during their onsite visit. As shown in Table 5, most of the inconsistent responses were false positives.

Table 5. Thermostat types: comparison of self-reported and observed conditions

n=80[7]		observed condition	
		has programmable thermostat	does not have programmable thermostat
self-reported condition (telephone survey)	has programmable thermostat	37 (46%)	16 **(20% false positives)**
	does not have programmable thermostat	3 **(4% false negatives)**	24 (30%)

Interestingly, the spot checks we did of households' responses *during the onsite visit* all matched what we could determine from the photographs of the thermostats. Only the telephone survey responses seemed to overreport the presence of programmable thermostats.

Summary

Table 6 summarizes these results and a few others not presented here in detail. Note that the table is organized by type of question and sorted within question types by degree of accuracy.

Among yes/no questions in which a respondent simply reacts to a statement with a positive or negative answer, we experienced a range of accuracies. When there are inaccuracies, we tended to see more false positives than false negatives, and this tendency toward false positives was very strong for a handful of questions.

With one exception in our small sample of questions analyzed, we experienced somewhat greater accuracy among multiple choice and short-answer questions. More importantly, there seemed to be less systematic bias toward positive or negative overstatement among inaccurate responses for these two types of questions.

[7] Excludes a small number of households that responded "don't know" or did not answer the question.

Table 6. Accuracy of questions by type and topic

Question type	Topic	Accuracy Rate	Tendency among inaccurate reports				
			strongly false negative	false negative / understatement	accurate reporting (a) or balance in direction of error (b)	false positive / overstatement	strongly false positive
yes/no	gas-heated oven	89% (n=91)		x			
yes/no	gas-fueled fireplace	86% (n=86)		x			
yes/no	gas-heated spa or hot tub	85% (n=88)					x
yes/no	gas-heated clothes dryer	84% (n=86)				x	
yes/no	thermostat type	76% (n=80)					x
yes/no	distance to neighboring building	72% (n=36)				x	
yes/no	PVC venting system for furnace/boiler	66% (n=56)					x
yes/no	PVC venting system for water heater	66% (n=56)					x
yes/no	thermostat programming	64% (n=36)				x	
multiple choice	building type	95% (n=73)			a		
multiple choice	heating system type	92% (n=81)			a		
open-ended	temperature setting	85% within 3°F (n=82)			b		
open-ended	number of stories	84% (n=87)				x	
open-ended	home size	30% within 500 ft^2 (n=64)		x			

potential confirmation bias

Implications

Our results point out that not all data are created equal. Even within telephone surveys, not all questions and question types are created equal. This is nothing new, but our analysis points to the degree of accuracies and biases that may exist in telephone survey responses to different types of questions, not just in our study, but in energy efficiency evaluation and research generally. We see several implications from this analysis.

Beware of confirmation bias

We interpret the high rate of false positives for yes/no questions as likely confirmation bias, or the tendency of people to agree with statements presented to them unless they are clearly false. Daniel Kahneman's book *Thinking Fast and Slow*, in particular, makes the case that the human brain is wired to accept statements presented to it unless the person hearing the statement is paying particular attention or the statement is clearly false (Kahneman 2011). Survey research also points to a tendency to agree with statements presented in a survey (Krosnick 2010).

Hence, questions like our inquiry whether people have a white plastic pipe running from the furnace to the outside or whether a respondent's home has 10 feet of space to adjacent buildings are likely to result in overreports. In both of these cases, most respondents are unlikely to be entirely confident, and a greater share of people who are uncertain will provide a positive response than a negative one. For example, people may think of their basement when asked about a white plastic pipe and visualize a hodgepodge of pipes, ducts, and other clutter that makes it easy to imagine that there is a white plastic pipe there.

On the other hand, we would expect questions that generate clear-cut answers in respondents' minds to have less confirmation bias. For example, whether or not a household has natural gas-fueled space heating should (and did) provide more accurate responses.

Knowing that confirmation bias exists is helpful, but what can we do about it? We have three suggestions. First, results from survey questions with potential confirmation bias should be interpreted with that possibility in mind. In our question about the white plastic pipe, for example, we found not only that positive responses were overstated, but those who answered "don't know" almost never had a plastic pipe. Hence, positive responses were really a "maybe," while uncertain responses (like "don't know") were almost always a "no" and negative responses were a clear "no." Understanding in what direction responses are biased allows one to apply appropriate filters when interpreting data.

Second, when higher accuracy is needed, survey questions could be asked in more rigorous ways to avoid confirmation bias. This could be done by asking respondents for additional detail, framing the inquiry as a "short answer" question, or even asking the respondent to look at the equipment in order to answer questions about it. For the white pipe question, we could have simply asked whether there is a plastic pipe leading from the furnace to an outside wall and then inquired what its color is. Or, we could have asked the respondent to describe what kinds of pipes lead away from the furnace. Finally, we could have asked the respondent to go to the basement during the relevant part of the survey and answer the question from there. All of these options would lengthen the survey time, but the additional effort is likely to provide greater accuracy. Given the accuracy rates we experienced for some of our questions and the prevalence of these same types of questions in many surveys in our field, we think it would be worthwhile to increase the sophistication and rigor of the questions we collectively ask in household surveys.

Third, one can collect the data through a different mechanism than surveys. We present one such option next.

Consider onsite verification

Surveys can only be as accurate as the ability of the respondents to answer the questions asked. Some questions are simply beyond the capability of most households to answer, and others are a stretch or might be answerable by only some households or only specific people within certain households. In those cases, onsite data collection provides a more accurate alternative and sometimes the only reasonable way to collect the needed data.

For the kinds of data we gathered, we found that households were generally able to answer questions, albeit with some degree of error. We seemed to encounter the greatest inaccuracies for the questions that are furthest removed from households' everyday lives and thoughts.

While collecting these kinds of data in-person for a large sample of homes would be prohibitively expensive, we found that onsite visits to a sample of homes provided a good complement to the telephone survey. It gave us the means to verify self-reported data while collecting information that households would have been unable to provide. Having verified data for a sample of telephone respondents opens the door to better interpretation of the survey results and even creates the option to quantify respondent-based biases in the data or even to make adjustments accordingly.

Recognize uncertainty beyond sampling error

In the energy evaluation and research fields, we are accustomed to present the sampling error for any sample-based data we analyze and report. That is, we commonly say that our results have a margin of error of +/- x% at either a 90 or 95 percent degree of confidence. This practice is essential to give users of our data a realistic sense of how close our estimates based on a sample are likely to be to the actual underlying value for whatever metric we are presenting for a larger population.

However, given the results of our analysis above—in particular, the existence of systematic biases in a single direction for some types of questions—we think it would be worthwhile to also acknowledge (and estimate, where possible) the degree of error that might have been introduced by self-reports. Depending on the self-reporting bias, the real uncertainty may be much greater than just the sampling error.

Improving telephone survey accuracy: a hypothetical example

The accuracy needed for energy efficiency projects varies. However, to illustrate the need for additional rigor in the way some standard survey questions are asked in our field, we invite the reader to consider impact evaluations or attribution interviews for net-to-gross adjustments. In that context, consider a standard survey question that asks households whether they have a programmable or manual thermostat using the question we described above. In our survey of more than a thousand households, we found a self-reported penetration of programmable thermostats of 70 percent with a margin of error of about 2.5 percent. However, there was a tendency for false positive self-reports. In fact, the false positives we experienced among the homes we visited were high enough that, if we adjust the population-wide results accordingly, our estimate of the penetration of programmable thermostats drops from 70 percent to about 53 percent, again with a margin of error of about 2.5 percent. The sampling error alone only tells a part of the story and, in this particular case, misses the main plot altogether.

References

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York, NY. Farrar, Straus and Giroux.

Kihm, S. 2013. *Peoples Gas Light and Coke Company Energy Efficiency Potential Study: Program Years 4, 5, and 6*. Madison, Wis. Energy Center of Wisconsin.

Kihm, S. 2013. *North Shore Gas Company Energy Efficiency Potential Study: Program Years 4, 5, and 6*. Madison, Wis. Energy Center of Wisconsin.

Krosnick, J. and Presser, S. 2010. "Question and Questionnaire Design." Handbook of Survey Research (Second Edition). Ed. Peter Marsden and James Wright. Bingley, UK. Emerald Group Publishing Limited.

Peters, J. 2002. "Effects of Using WebTV Graphics on the Accuracy of Self-Reports." *In Proceedings of ACEEE 2002 Summer Study on Energy Efficiency in Buildings*, Asilomar, Calif. American Council for an Energy-Efficiency Economy.

Pigg, S. and Nevius, M. 2000. *Energy and Housing in Wisconsin: A Study of Single-Family Owner-Occupied Homes*. Madison, Wis. Energy Center of Wisconsin.