Self-Reported Virtual or Third-Party Onsite Impact Verification: A Case Study Test of Feasibility, Logistics, and Accuracy of Findings

Cherish Smith, Guidehouse, Ann Arbor, MI
Shannon Dorato, Guidehouse, San Francisco, CA
Andrew Stolberg, Guidehouse, Hinesburg, VT
Timea Zentai, Guidehouse, Boulder, CO
Denise Allard, DTE Energy, Detroit, MI

ABSTRACT

Customer onsite savings verification is one of the costliest energy efficiency program evaluation activities. To mitigate against health and safety concerns, impacts related to the COVID-19 pandemic may have increased the momentum of virtual verification (VV) practices. As utilities grow their energy efficiency portfolios, they must identify opportunities to streamline the customer experience, reduce costs, and optimize resources to focus on higher value research—all while maintaining confidence in the accuracy of energy savings. As such, DTE Energy and Guidehouse tested the feasibility and accuracy of using customer self-reported VV in tandem with traditional onsite third-party impact verification. This paper discusses the methods used and considerations needed for an effective customer self-reported VV, using an online, survey-based approach with the inclusion of self-reported pictures of installed equipment. This research covers a residential direct install home assessment program where we compared the differences in results between VV and onsite verification and worked to address the program challenges. The study targeted 90% confidence and 10% precision at the residential direct install program-level separately for surveys and onsite visits (or 80 completes for each method employed). This research should help utilities and evaluators understand if virtual self-reported surveys versus third-party onsite visits for impact evaluation are consistently reliable, how best to engage customers in the VV process, and any key challenges that may persist and how best to overcome them.

Introduction

Traditional onsite impact verification or standard field verification (SFV) is often one of the costliest energy efficiency program evaluation activities.[1] Even with the high cost, SFV is commonly employed because it yields highly reliable and accurate third-party verification data. The COVID-19 pandemic drove the need to explore alternatives to SFV to mitigate against health and safety concerns. Additionally, as utilities continue to grow their energy efficiency portfolios, they must identify opportunities to streamline the customer experience, reduce costs, and optimize resources to focus on higher value research—all while maintaining confidence in the accuracy of energy savings.

DTE Energy (DTE) along with Guidehouse, a third-party independent evaluator, took the pandemic as an opportunity to test the feasibility and accuracy of using customer self-reported virtual verification (VV) in tandem with traditional onsite third-party impact verification. The two organizations identified the Home Energy Consultation (HEC) program as an ideal program to pilot VV. The HEC program offers a no-cost in-home assessment (the Consultation) and direct installation of free energy efficiency products in single-family residences to help reduce their energy use and provide energy efficiency education. The products include programmable thermostats, LED bulbs, pipe wrap, low flow showerheads, LED nightlights, bathroom faucet aerators, kitchen faucet aerators, and advanced power strips. HEC receives

[1] Contributing to roughly 30% of the DTE Energy direct install program evaluation budget

SFV as part of its annual evaluation of impact savings. In 2020, the HEC program also received VV in parallel to SFV to compare the two approaches and identify opportunities to reduce the cost and customer burden of verification while obtaining similar verification results.

Scope

DTE and Guidehouse implemented VV, an online survey-based approach with the inclusion of self-reported pictures of installed equipment. The study targeted 90% confidence and 10% absolute precision for the HEC program-level Installation Rate Adjustment Factor (IRAF) to calibrate the survey results. This research enabled both the utility and evaluator to understand the reliability of virtual self-reported surveys versus third-party onsite visits for impact evaluation, how best to engage the customer in the VV process, and any key challenges that may persist and how best to overcome them.

Installation Rate Adjustment Factor Methodology

Guidehouse developed an online survey to collect installation data to answer the following key impact questions:

- What are the fuel-neutral and product-level installation rate adjustment factors (IRAFs; see Equation 1)?
- What are the main factors driving these IRAF results (leave behinds, customer uninstalling, misreporting, etc.)?
- What are the product-specific part-use factors (PUFs; see Equation 1)?

Equation 1

$$\begin{aligned} &Installation\ Rate\ Adjustment\ Factor\ (IRAF) \\ &\quad = Installation\ Rate\ (IR) \times Size\ Correction\ Factor\ (SCF) \\ &\quad \times Part\ Use\ Factor\ (PUF) \end{aligned}$$

In this equation, the installation rate is either one or zero depending on if the product was installed or not, respectively. The SCF is the adjustment to the reported size of the product (e.g., the linear feet of pipe wrap or the square footage controlled by a thermostat). The PUF is the usage correction, specifically for thermostats, to capture if the product is used in a way that saves energy.

For specific products, energy savings are achieved when the end user, in this case the program participant, have the product installed correctly and are also using it correctly. For the HEC program, programable thermostats are one of these specific products, and require the participant to have the thermostat properly programed to setback or setup the temperature during times of reduced activity in the home, for example, at night or during the day when the occupants of the home are typically away. Guidehouse collects data from each participant to calculate a PUF of 1 or 0, which is then multiplied with the savings to either give full credit or negate the reported savings. To determine the PUF for thermostat products, Guidehouse verified if the thermostat was in program mode, and if so, determined that the program settings changed the temperature by at least 1°F throughout the week. If the thermostat was not in program mode, customers were asked a series of questions to determine how they typically used their new thermostat to better understand if they qualify to receive a PUF of 1.

Sample Design

To develop the verification sample, quotas were assigned for each product type offered through the program to achieve the target of 90% confidence and 30% absolute precision at the product-level

(Table 1). The same sample target was used for VV and SFV. Guidehouse completed 82 VV surveys with a random sample of program participants, verifying 337 total products. There were 40 site visits completed for SFV, verifying a total of 190 products. Participants were targeted for verification of a single product but had the opportunity to verify any other products that they received through the program. This approach allowed Guidehouse to collect additional data for every stratum. As Table 1 shows, some customer data was removed during quality control; however, all product-level stratum targets were achieved for VV and SFV.

Table 1. Sample design

Strata	Target[2,3]	Achieved (VV)	Achieved (SFV)
LED Bulbs	13	106	62
Programable Thermostat	11	23	11
Pipe Wrap	11	34	18
Shower Products[4]	10	47	25
Other Products[5]	10	89	54
Smart Strips	10	38	20
Total	65	337	190

Source: Guidehouse analysis, 2020.

Verification Approach

Guidehouse used the survey software Qualtrics to program and field the VV survey. After sending the initial email invitations to participate in the survey, Qualtrics used two reminder emails to follow up with customers to complete the survey. Each customer received a personalized survey link that contained embedded data on the utility reported number and types of products that were installed during the consultation.

Like SFV, the survey's first questions verified whether the survey reached the correct and most qualified respondent. These questions asked if the respondent was familiar with and present during the consultation and if they were familiar with the products installed. The customer was not required to be present during the consultation but needed to be familiar with the consultation and the products installed. Unlike SFV, respondents who were not familiar with the consultation or products installed were asked for the contact information of a person who is most familiar and then they were screened out of the survey. Respondents were asked to complete the survey on a mobile phone or tablet to ensure photos can be easily taken and uploaded with the same device the customer used to complete the survey.

The survey included a total of 72 questions covering nine products. The average customer had four to six products installed during their consultation and were provided the opportunity to respond to questions about each of these products. Over 90% of customers responded to questions about all the products they received, although they were only required to respond to questions regarding a single product based on the product-stratum to which they were randomly assigned.

[2] The units in this table are products, not customer sites.

[3] Guidehouse targeted customers based on a specific single measure that they received; however, customers had the opportunity to verify any additional products they received through the program. In total, 65 participants were targeted to verify a minimum of 65 products.

[4] Shower products include low flow showerheads.

[5] The "Other" products stratum consists of LED nightlights (achieving 51 VV, 29 SFV), kitchen faucet aerators (achieving 28 VV, 12 SFV), and bathroom faucet aerators (achieving 10 VV, 13 SFV). Each of these three products met or exceeded the target of 10 observations. However, our targets are based on these measures being grouped together under a stratum, and not at the individual product-level.

After verifying that the products included in the database matched the actual installation in the customer's home, the survey asked customers four questions regarding most products:

1. Are products still installed and functioning?
2. How many of each product is installed?
3. Why were products uninstalled (if applicable)?
4. Request for an uploaded photo of each type of installed product.

Guidehouse asked customers to verify information about the products installed through their consultation in the survey and showed customers pictures of each measure type for easier verification. To ensure precise counts, customers were asked to walk around their homes, complete a visual count, and take pictures of installed products. Customers were not required to take pictures of all measures (i.e., all nine bulbs installed), but only one of the installed measures. This process is similar to SFV; however, a trained field technician is more likely to verify that every measure is installed and working correctly. Additionally, field technicians will ask probing questions to better understand how a measure is typically used or why it is not installed.

For health and safety reasons, participants were advised not to use a ladder or climb up on things to verify product quantities. For thermostat products, additional questions were asked about settings and setpoints to help determine PUF. Customers were also asked to upload photos of the thermostat setpoints. Guidehouse reviewed survey responses daily as they came in to ensure high data quality.

Survey and Onsite Fielding

Guidehouse offered a $25 incentive for customers to complete the 15-minute online survey. The average customer completed the survey in 16 minutes. The survey was fielded in 3 weeks from January 25, 2021, to February 17, 2021. Around 5% of the survey sample contained email addresses with various versions of the word "none" or "no," implying these are not real emails. Guidehouse removed those customers from the sample. Even after the email cleaning, around 9% of emails were undeliverable.

After the soft launch, the survey was sent out in three waves, one wave per week, to around 300 customers per wave and a total of 959 program participants overall. There were 82 surveys completed, achieving a 9% completion rate overall. Roughly 19% (185 customers), opened and started the survey but did not complete it. This may be due to the aggressive 3 weeks schedule set for data collection. Partially completed surveys were not included in the analysis. The average partial respondent stopped at the first product-level question. Around 20% of partial survey respondents completed 50% or more of the survey. About half of those in the sample did not open their emails, while 20% only read their emails and did not click on the survey link.

For comparison, Guidehouse offered a $50 incentive to complete SFV at a site. A typical SFV visit took between 30 and 60 minutes. Field technicians used Fulcrum, a data collection software, and tablets to collect data while in the field. Data was reviewed daily during SFV for reasonableness, completeness, and accuracy. SFV took place in two waves. Wave one was scheduled from October 1, 2020 to December 31, 2020. However, field work for wave one was suspended after the Thanksgiving weekend due to the holidays, a spike in COVID-19 cases throughout the State of Michigan, and a lack of willingness from customers to participate in SFV, likely due to COVID-19. Wave two occurred from February 1, 2021 to March 31, 2021. Again, likely due to COVID-19, customers were less likely to participate in the SFV than in previous years for wave two. Over the 14 weeks of field work, a total of 316 participants were contacted to see if they were willing to participate in the SFV process. Of these participants, 40 site visits were completed, achieving a 13% completion rate overall.

Data Processing

Upon customer completion of the VV Survey, Guidehouse collected and processed the data. The data came in three forms: a Microsoft Excel file with aggregated results, photos that were provided via upload to the survey link, and a PDF digital printout for each individual participant.

The PDF included the questions and individual responses from each participant, a preview of each photo that the individual participant uploaded, and a link to the full-sized version of each photo. This PDF was the primary means of reviewing each individual response. At the product-level within each completed survey, Guidehouse reviewed the verified product quantity as well as proper use and installation. If the preview version of a photo in the PDF did not show proper installation, the full-sized, higher resolution photo was downloaded and reviewed for further analysis.

During the review of the PDF digital print out, any adjustments made to survey responses were documented and updated in the corresponding survey Excel file. An example of an adjustment made could consist of a survey participant verifying four A-line LED bulbs but uploading a photo of four program candelabra LED bulbs. In this case, if the participant verified A-line bulbs but not candelabras, an adjustment would be made to record the verification of candelabras and not A-line bulbs. Of the 82 surveys collected, 22 needed some sort of adjustment.

Any observed ambiguity in the verified quantity or installation of a specific product was removed from the analysis. This was possible due to the oversampling built into the sample design and had little effect on Guidehouse's ability to meet the target sample for each stratum. Some products, and even entire surveys, were deemed unusable. Around 5% of verified product observations were removed due to:

- Ambiguous survey responses
- Product verification deemed inaccurate
- Participants missing a component of verification

Specific cases where Guidehouse deemed the product verification inaccurate or needing to be removed from the analysis include the following:

- Uploading a random picture (e.g., of their feet) or a handwritten note instead of a photo of the product installed (two instances of each)
- Uploading a picture taken off the internet (e.g., not an original photo taken by the participant) (two instances)
- Answering "Don't Know" for the verified quantity of all products (one instance)
- Reusing the same photo multiple times (e.g., posting a photo of an A-line bulb for all verified bulb types) (one instance)
- Uploading the wrong file (e.g., uploading a PDF pamphlet of an HMO healthcare plan) (one instance)
- Altering a product in a way that prevents its verification (e.g., painting the product) (one instance)

Although removal of collected data is not considered a best practice because it can introduce bias into a set of results (whether intentional or unintentional), Guidehouse determined that keeping the ambiguous or erroneous data would adversely bias the results as well. It was decided that removal of this data would be the most appropriate way to minimize bias overall.

Upon completion of the PDF review and after making all product-level verification adjustments, Guidehouse aggregated verified results to calculate product-level and program-level IRAFs.

Results

When comparing the results from the VV to the SFV, Guidehouse found similar results for most products, meaning the confidence intervals overlap. Table 2 compares IRAFs at the product-level and overall program-level.

Table 2. Product-level and program-level installation rates

Product	VV IRAF	SFV IRAF
LED Bulb	0.96 ± 0.02	0.86 ± 0.07
Programable Thermostat	0.17 ± 0.13	0.75 ± 0.21
Pipe Wrap	0.87 ± 0.09	0.94 ± 0.06
Shower Products	0.81 ± 0.09	0.88 ± 0.10
Other[6]	0.85 ± 0.09	0.84 ± 0.13
Smart Strips	0.77 ± 0.12	0.53 ± 0.21
Program-Level Total	0.73 ± 0.04	0.84 ± 0.05

Source: Guidehouse analysis, 2020.

Factors Influencing Installation Rates

There were 337 instances of a product type being reported as being installed in the VV sample. Of these 337 instances, 58 (or 17%) were found to have a discrepancy that contributed to an installation rate being other than 1.0. Of these discrepancies, customer removal due to personal preference was the most common reason why a product could not be verified as installed. These customer removals made up 57% of all discrepancies found in VV, or 9% of all VV observations. This was also the most common reason for a product to not be verified during the SFV, making up 46% of discrepancies found or 15% of all SFV observations.

The second most common reason why a product received an installation rate other than 1.0 was due to misreporting. Misreporting refers to any time a product type or quantity is verified other than what was reported. This can include under-reporting of a quantity or over-reporting of a quantity installed. Misreporting is often due to human error and is commonly found in all energy efficiency programs. Although this was the second most common reason for both evaluation methods, SFV saw almost 4 times more discrepancies due to misreporting compared to VV (45% and 26% of total discrepancies, and 14% and 4% of total observations, respectfully).

The gap in misreporting between SFV and VV occurred most with LED bulbs, were SFV saw 14 instances of misreporting but VV saw only four. This could be a result of VV participants not fully understanding that the reported quantity embedded in the survey could be incorrect, a result of there being a recording error on the part of the implementing contractor during the time of the installation of the product. VV participants may have been more likely to report only on a change in reported quantity that occurred because of customer removal or product malfunction. Figure 1 shows the breakdown of driving factors behind installation rates being other than 1.0 between the VV and SFV evaluation methods.

[6] The "Other" stratum consists of those products whose total savings represented less than 10% of the total fuel neutral savings. The products in "Other" include LED nightlights and kitchen and bathroom faucet aerators.

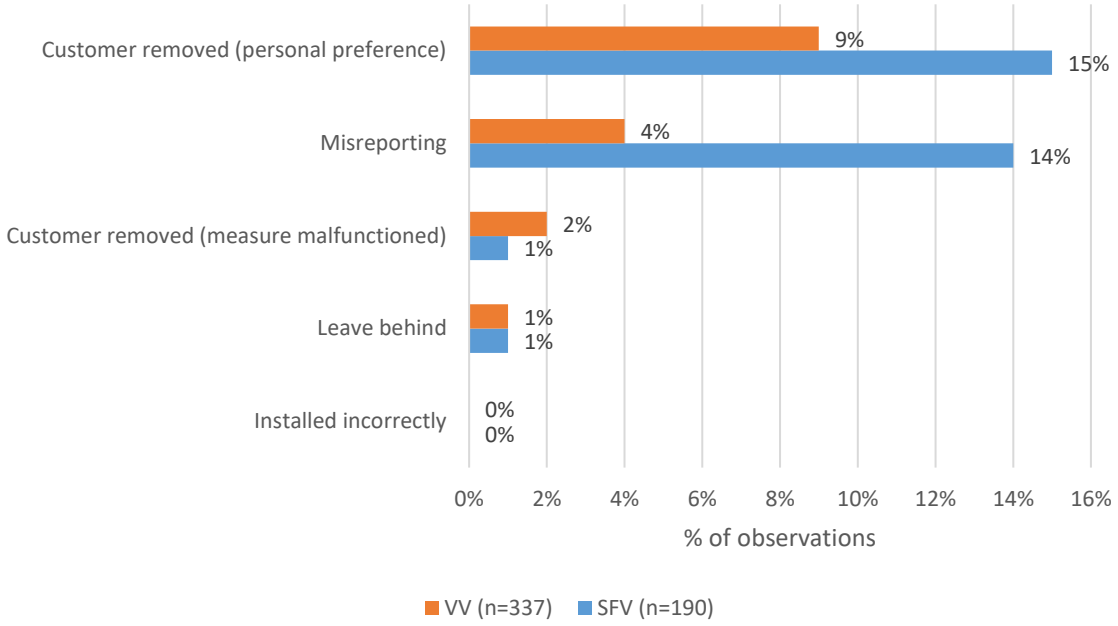


Figure 1. Driving factors behind installation rates being other than 1.0. *Source:* Guidehouse analysis, 2020.

Thermostat Discrepancies

In addition to PUF, an IR and SCF also go into calculating verified savings for programable thermostats. The IR was 100% for all thermostats verified through the VV and SFV and the SCF verifying the reported size of conditioned space was 98%. PUF was, and typically is, the driving force behind the low IRAF for programable thermostats. Figure 2 shows what qualifies for a PUF of 1, and the breakdown of PUF between VV and SFV.

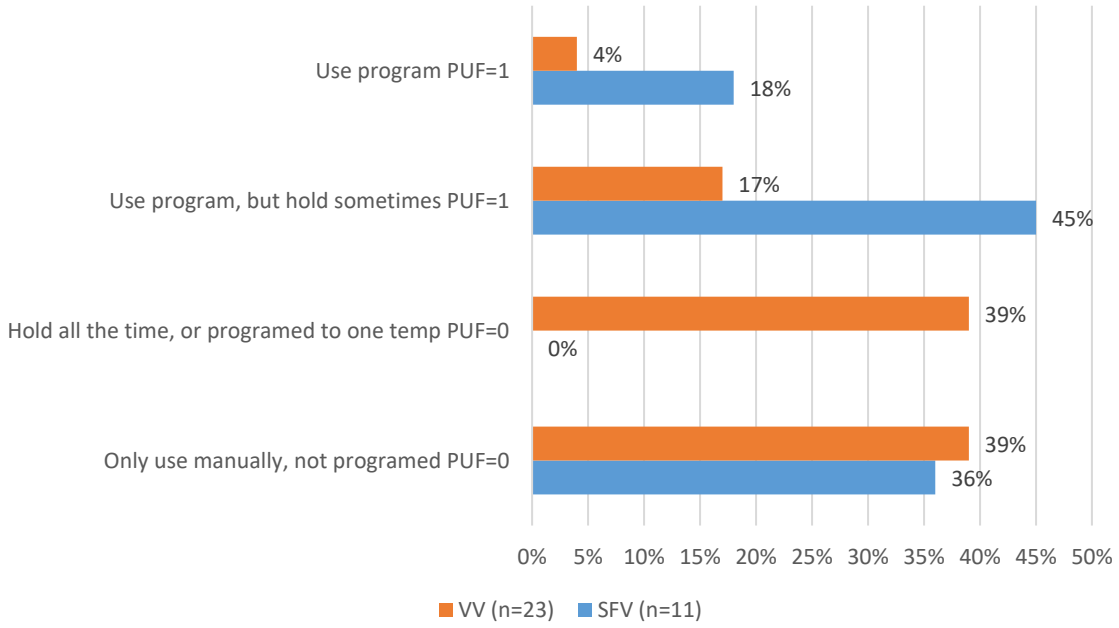
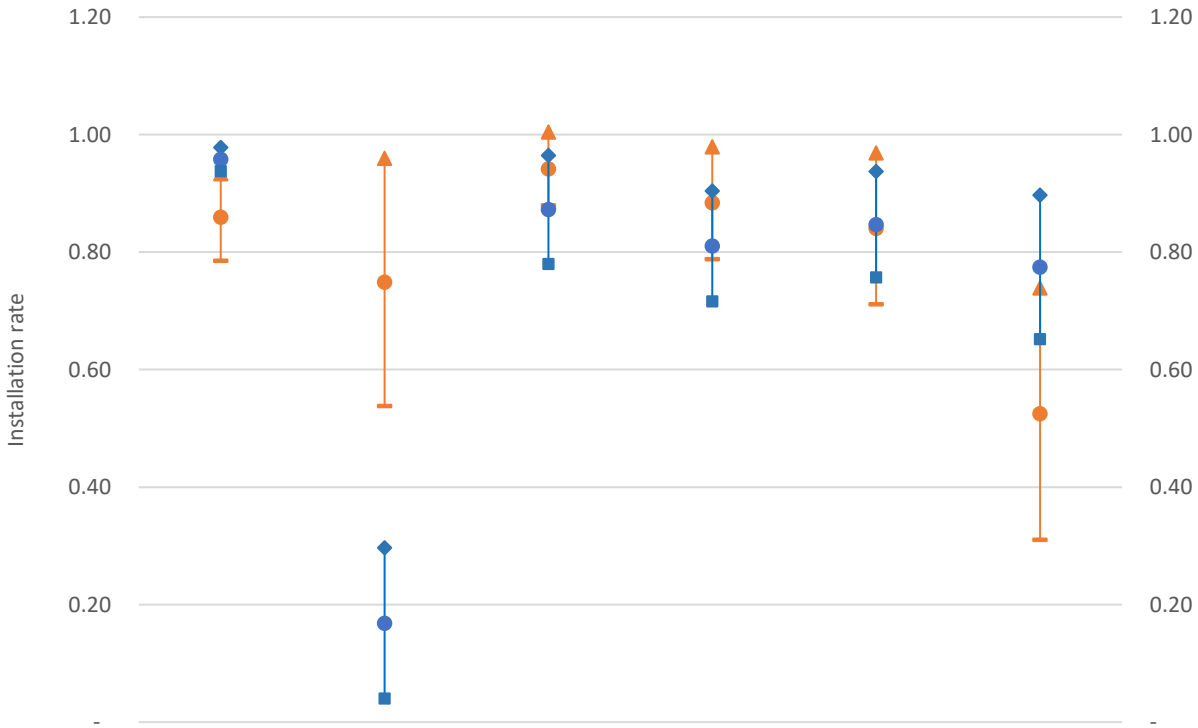


Figure 2. The distribution of PUF for programmable thermostats and breakout between VV and SFV. *Source:* Guidehouse analysis, 2020.

PUF was the driving force behind the low product-level IRAFs (17% for VV and 75% for SFV) for programable thermostats in the 2020 program year. Of the 23 VV responses for programable thermostats, only 21% were self-reported to be using the thermostat in a manner that received a PUF of 1.0. Of the observations where a customer received a PUF of zero, 39% had their thermostat programed to a single temperature with no setback temperature and 39% were using the thermostat as a manual thermostat only. The biggest difference in PUF between VV and SFV was that VV participants reported using the thermostat in "hold mode all the time or programed to one temperature," whereas none of the SFV participants were found to be using their thermostat in this way.

Comparison of Verification Methods

Between the two verification methods, most product-level IRAFs fell within each other's error bounds, in other words, they were similar. Figure 3 visualizes these error bounds.



	LED Bulb	Programable Thermostat	Pipe Wrap	Shower Measure	Other	Smart Strip
(+) IR SFV	0.93	0.96	1.00	0.98	0.97	0.74
IR SVF	0.86	0.75	0.94	0.88	0.84	0.53
(-) IR SFV	0.79	0.54	0.88	0.79	0.71	0.31
(+) IR VV	0.98	0.30	0.96	0.90	0.94	0.90
IR VV	0.96	0.17	0.87	0.81	0.85	0.77
(-) IR VV	0.94	0.04	0.78	0.72	0.76	0.65

Figure 3. Product-level error bounds.[7] *Source:* Guidehouse analysis, 2020.

Two products, LED bulbs and programable thermostats, had no overlap in confidence intervals. Guidehouse suspects these differences are a result of how participants were questioned regarding these two products in the VV survey, compared to how data is collected through the SFV method.

For LED bulbs, participants were primed with the reported quantity of bulbs installed within their home for VV. It appeared that participants were not inclined to physically verify this reported quantity, but rather subtracted any LED bulbs that they removed or that malfunctioned from the reported total presented to them. In contrast, SFV often discovers additional discrepancies in what was installed versus what was recorded as installed, which is common within all energy efficiency programs, because LED bulbs are the most common product in terms of quantity installed per site.

For programable thermostats, participants are asked a series of questions to determine if the thermostat is installed, if the size of conditioned space is the same as reported, and if they are using the thermostat with programed settings that change the temperature by at least 1°F throughout the week. VV survey questions asked the participant to describe how they typically use the thermostat and then asked them to take a series of pictures showing at least two different setpoints. The participants' responses were often contradicted by the photos that they uploaded. There were more opportunities for a thermostat to be categorized as not being used properly, whereas the field technician would be able to decipher the nuances of how the thermostat was being used during an SFV.

Historical Results

The 2020 program year had several obstacles, most notably the COVID-19-related impacts on participant behavior and time spent at home. For many people, the pandemic resulted in them having to work from home, not participating in leisure activities, or refraining from travel. This affected the amount of time they spent in their homes compared to a typical year, and most notably how they interacted with their thermostat. Guidehouse hypothesizes that participants who had programed their thermostats to account for them being away during the day were now overriding and putting them in hold mode or reprogramming them to one temperature. It is believed that this decreased the IRAF for thermostats compared to other years. Table 3 compares the VV results compared to the historical SFV results.

Table 3. Comparison of 2020 VV IRAFs to historical SFV IRAFs

	SFV IRAF				VV IRAF
Product	2017	2018	2019	2020	2020
LED Bulb	0.90	0.92	0.89	0.86 ± 0.07	0.96 ± 0.02
Programable Thermostat	0.46	0.38	0.34	0.75 ± 0.21	0.17 ± 0.13
Pipe Wrap	0.85	0.90	0.84	0.94 ± 0.06	0.87 ± 0.09
Shower Products	0.89	0.80	0.85	0.88 ± 0.10	0.81 ± 0.09
Other[8]	0.72	0.75	0.79	0.84 ± 0.13	0.85 ± 0.09
Smart Strips	N/A	N/A	N/A	0.53 ± 0.21	0.77 ± 0.12
Program-Level Total	0.75	0.74	0.75	0.84 ± 0.05	0.73 ± 0.04

[7] The "Other" stratum consists of those products whose total savings represented less than 10% of the total fuel neutral savings. The products in "Other" include LED nightlights and kitchen and bathroom faucet aerators.

[8] The "Other" stratum consists of those products whose total savings represented less than 10% of the total fuel neutral savings. The products in "Other" include LED nightlights and kitchen and bathroom faucet aerators.

Source: Guidehouse analysis, 2020.

Guidehouse speculates that SFV was also affected by the COVID-19 pandemic, which artificially inflated IRAFs for most products and yielded a program-level IRAF that was 10 points higher than previous years. The theory behind why IRAFs were higher for SFV revolve around the difficulty in recruiting participants to allow an outsider into their homes during a global health crisis. As a result, the SFV effort might have inadvertently recruited a greater number of people who value EE more than the average participant. These energy conscious participants may have been more likely to open their door to an outsider if it meant continued participation in an EE program, and thus skewing the results.

Additional Findings

In addition to the VV IRAF results compared to the SFV results, Guidehouse identified the following additional findings when comparing the two verification approaches:

- VV provides an easier approach for achieving overall and stratum-level targets than SFV from a sample design perspective. Guidehouse virtually verified more customers and products (337 products and 82 customers) than field work (190 products and 40 customers).
- Although VV is customer reported data as opposed to third-party verified (which is more reliable), the ability to oversample and screen out questionable or erroneous data based on a third-party review allows for more accurate product verification without compromising the statistical significance for most products.
- VV provides a more cost-effective evaluation approach over SFV. VV eliminates costs associated with scheduling and field technicians, and cuts incentive costs in half compared to SFV. In total, VV was less than 10% the cost of SFV to administer.
- VV addressed a potential bias that field work introduces into the data it collects. SFV disproportionately collects data during weekday-daytime hours, whereas VV collects data during evening and weekend hours.
- VV was completed in one-fourth of the time needed to conduct SFV, providing a more streamlined approach. The average customer completed the online survey in 16 minutes, while a typical field visit is 30-60 minutes. VV was completed in 3 weeks, while SFV took place over 14 weeks. Overall, it is easier to implement VV more frequently throughout the year than it would be to conduct SFV in more than two waves, and better enables evaluators to provide timely feedback to utilities.
- Guidehouse achieved a 9% completion rate for the VV survey compared to the 13% completion rate for field work. Around 20% of partial survey respondents completed 50% or more of the survey. Although respondents stopped the survey at various points throughout, the two main places where respondents got stuck and stopped the survey were at the first photo upload and the photo upload of the thermostat setpoints.
- Guidehouse lost 14% of the VV survey sample due to erroneous email addresses. Guidehouse recommends email addresses be reviewed for accuracy to improve future virtual data collection efforts.

Conclusions

When comparing VV to SFV, the IRAFs are similar for all products, meaning they fall within each other's error bounds, except for LED bulbs and programmable thermostats. The VV LED bulb IRAF captured less misreporting and customer removal compared to the SFV. The VV programmable thermostat IRAFs were significantly less than SFV primarily due to lower PUFs captured during VV. With some recommended improvements for VV, Guidehouse would like to determine if the IRAF for LED

bulbs and programmable thermostats could be improved to be as reliable and statistically the same as SFV. As a result, Guidehouse recommended pursuing VV in parallel to SFV again for the HEC program in 2021, to continue collecting VV data considering it provides multiple process improvements over SFV—including that it is a safer verification approach because it does not require physical entry into a home, it is less costly than SFV, and it reduces the time commitment from the customer.

Future Research and Recommendations

To improve the survey completion rate in the future, Guidehouse recommends increasing the incentive amount, especially for any customers that verify thermostats since there are more questions for these products. More time should be allowed (at least 2 weeks) for each VV wave so customers can finish their responses. Furthermore, there should be customized reminder emails (as opposed to the general email used in 2020) sent to customers who completed more than half of the survey, encouraging them to finish. Additionally, the no email option should be standardized to assist with data scrubbing and sampling efforts. To avoid customer survey fatigue and the drop off at the 50% completion rate, customers should be able to easily submit the survey if they verify the required measure they were sampled for; any other measure verification should be optional. Finally, survey instructions should be updated for all questions where a photo upload is needed to help encourage customers to finish their surveys. Instructions might include a video showing specific steps on how to check programmed settings on thermostats.

Misreporting made up only 4% of reported discrepancies for VV compared to 14% for SFV. It is presumed that participants assume the reported quantity embedded in the survey is accurate and only report if they personally removed a product. Guidehouse recommends updating the VV survey questions to capture better quality data going forward. Guidehouse and DTE should consider revising the VV survey to include a two-part installation question rather than only asking: “are the products still installed and functioning,” ask “did you receive the <reported quantity> of products,” and “are they still installed and functioning.” Participants are advised not to use a ladder or climb up on things to verify product quantities. This made verifying precise pipe wrap length difficult. The VV survey can be further revised to include pipe wrap quantities in increments of three linear feet to improve accuracy of customer reporting.

Guidehouse suggests DTE considering multiple approaches to improve VV for thermostat products going forward. The survey should be updated to ask thermostat setpoint questions first, then following-up with questions about how the participant uses their thermostat to encourage more alignment between the thermostat behavior questions. Questions should be added about home occupancy during the day, evening, and weekend to understand if there is a correlation between PUF and occupancy.

An instructional video also could be developed as an example of how to verify thermostat setpoints in addition to requesting video verification of the thermostat setpoints from the participant instead of verification pictures. A video could be more comprehensive and less burdensome for the customer since it can capture all the program setpoints in a single upload.

Finally, an additional layer of verification could be used for thermostat products by conducting follow-up phone calls for all participants that verified a thermostat to confirm and increase the reliability of PUF assignments. Additional incentives could be offered to account for any extra time required by the customers. Future studies should also consider a greater thermostat sample targeting a high confidence and precision, since there continues to be high variability for this product year-over-year and between verification methodologies.

Acknowledgements

The authors of this paper gratefully acknowledge the generous and timely assistance of the following staff at DTE and Guidehouse for their contributions throughout this project: Denise Allard, Jason Kupser, Kevin Bilyeu, Dennis Mullan and Jeff LeBrun (DTE), Craig McDonald, Debbie Brannan, Josh Rego (Guidehouse).

Guidehouse also acknowledges the timely assistance and openness of the Walker Miller staff, including Cameron Seeley.