Energy Trust of Oregon Residential Pay for Performance Pilot Evaluation Report

Submitted by Apex Analytics LLC
October 11, 2021

Table of Contents

Executive Summary

Energy Trust of Oregon implemented a residential Pay for Performance (P4P) pilot from April 2019 to December 2020. This report presents findings from an evaluation of the pilot that Apex Analytics conducted, under contract to Energy Trust.

Pilot Description

Energy Trust's P4P pilot sought to motivate contractors to take action to increase the energy savings from the projects they installed. Households participating in the pilot installed one or more measures with deemed savings and incentive values (called anchor measures) and received incentives for those measures as they would under the standard residential program.[1] Participating contractors (called aggregators) were then eligible for an additional performance incentive for any measured energy savings their projects achieved above the deemed savings estimates.[2]

The pilot assessed performance incentives on a portfolio of projects, completed over a period of approximately six months.[3] Using monthly billing data, the pilot measured energy savings for each project over the course of a 12-month performance period from the close of the portfolio, at the end of which it paid performance incentives. Energy Trust contracted with Recurve Analytics to provide a dashboard that tracked each project's energy savings, allowing Energy Trust and the aggregators to identify projects with the best performance.

Energy Trust selected three contractors to work with as aggregators for the pilot: an HVAC contractor, a weatherization contractor, and a whole house contractor. The companies invited to participate as aggregators were trade allies that had been highly active in Energy Trust's Residential Program.

The pilot anticipated that the availability of performance incentives, combined with access to project tracking data, would motivate aggregators to take steps to increase the energy savings their projects generated. For example, aggregators might encourage participants to install additional measures not eligible for deemed

[1] Energy Trust relaxed participation requirements for some measures on P4P projects relative to projects in the standard residential program, but otherwise the participation process for households was the same.

[2] This approach contrasts with a "pure" P4P approach, in which all incentives would be based on measured energy savings. By continuing to pay deemed incentives, Energy Trust reduced the risk of participating in the pilot to the aggregators.

[3] While originally intended to follow a six-month schedule, actual portfolio periods varied during pilot implementation.

incentives, or they might identify the customer or project types that result in the greatest savings and focus their outreach efforts to target those types of customers or projects.

Energy Trust chose to end the pilot earlier than originally planned due to challenges around project qualification and data availability, as well as budgetary concerns. The closing of the pilot coincided with the end of the performance period for the first portfolio, which closed in late 2019. As a result, aggregators were eligible for only one performance payment, and did not have an opportunity to adapt their practices in response to that payment.

Research Objectives and Activities

Table 1 summarizes the key research objectives of this evaluation and the corresponding research activities Apex used to address those objectives. Because Energy Trust ended the pilot earlier than anticipated, we were unable to assess some of the objectives Energy Trust defined. Objectives we were unable to assess are listed in grey text in Table 1.

Table 1: Research Objectives and Evaluation Activities

Research Objective	Pilot Staff Interviews	Pilot Data Analysis	Aggregator Interviews	Savings & CE Review	Participant Survey
Does the P4P pilot design enable better targeting of interventions with variable outcomes?	✓		✓		
Does the P4P pilot design improve measure cost effectiveness?	✓			✓	
Does the P4P pilot design create new participation opportunities for lagging markets?	✓		✓		
Is the market ready for a "pure" P4P approach with no guaranteed (deemed) incentives? What is trade ally receptivity to P4P?	✓		✓		
How persistent are the energy savings from P4P?		✓		✓	
What are the benefits and challenges in offering P4P as a track within the program?	✓		✓		
Are participants satisfied with their experience (installation, performance of measures)?					✓
Should the program continue to offer residential P4P?	✓	✓	✓	✓	

As Table 1 indicates, the Apex team conducted five key research activities to address the objectives of this evaluation:

- **Pilot staff interviews:** Apex interviewed key staff involved in pilot delivery at Energy Trust, its program delivery contractor, CLEAResult, and Recurve. We conducted two rounds of staff interviews, in December and January of 2019-

2020 and in March and April of 2021. In both rounds, we completed interviews with 5 staff members.

- **Aggregator interviews:** Apex conducted interviews with the aggregators participating in the pilot at the close of each portfolio period, in January of 2020, June and July of 2020, and March and April of 2021.
- **Pilot Data Analysis:** Apex reviewed and characterized Energy Trust and CLEAResult's tracking data following each portfolio to assess characteristics of P4P projects, including costs, anchor measures installed, and non-anchor measures installed. Following the close of the first portfolio period, we also reviewed and characterized project performance data.
- **Savings and Cost Effectiveness Review:** Apex reviewed Recurve's approach to estimating project-level savings for consistency with CalTRACK guidelines. Apex also reviewed Energy Trust's approach to calculating cost effectiveness for P4P portfolios.
- **Participant Survey:** Apex analyzed data from Energy Trust's Fast Feedback participant surveys to assess participants' experience with the pilot and compare P4P pilot participant satisfaction with that of standard program participants.

Key Findings

Aggregators acknowledged that performance incentives might motivate them to change their approach to the market but did not do so prior to receiving incentives. Aggregators did not report making significant changes to the types of customers and projects they targeted or the measures they offered as a result of P4P performance incentives. Consistent with these reports, the differences between measures installed, homes served, and project cost were limited. While aggregators had access to performance data through the pilot's data portal, the potential to draw meaningful conclusions based on partial-year data was limited and aggregators did not devote effort to monitoring the data. Prior to receiving payments, performance incentives remained somewhat abstract for aggregators. As one said, "Some of that performance will result in me getting paid some money. If I understood that relationship, I would be geeking out on [the performance data] more, but because it is nebulous, I'm not super motivated." Nonetheless, aggregators saw the potential to use performance data and incentives to guide their offerings.

Project disqualifications limited the size of aggregators' eligible portfolios, with data availability challenges a key factor leading to disqualification. The pilot experienced lower-than-expected project volumes overall, leading pilot staff to replace the weatherization aggregator after the first portfolio period and extend portfolio period

timelines in some cases. Data availability and project eligibility challenges compounded this limited volume.

More than half (56%) of the projects submitted for the HVAC aggregator's first portfolio and 39% of the projects submitted for the whole house aggregator's first portfolio were ultimately ineligible to be included in their portfolios assessed for performance incentives. Insufficient baseline data, often associated with a change in utility accounts, was the most common reason projects were ineligible.[4] Using monthly usage data, as the pilot did, a single missing meter reading could call into question two months of data – both the missed month and the subsequent reading if it was unclear whether that reading represented 30 or 60 days of data – which was sufficient to disqualify the project. Aggregators also noted that many of the projects they complete are in homes that have been recently purchased, rental housing that is changing tenants, or part of other, larger changes in the home.

Savings from electric portfolios fell short of deemed estimates, largely due to unrealistic savings estimates for ductless heat pump installations. For both the HVAC and whole house aggregators, projects including ductless heat pump installations were the most likely to achieve measured energy savings below their deemed estimates. Nearly half (47%) of the households installing ductless heat pumps in the HVAC aggregator's portfolio saw their measured energy consumption increase relative to the baseline. These increases in energy use likely reflect households that had not previously had cooling equipment using their ductless heat pumps for cooling or using their heat pump to heat when they might previously have used a supplemental source like wood. Program staff also noted that the milder climate in Southern Oregon, where the HVAC aggregator primarily worked, might reduce heating energy use, and thus savings, relative to the deemed estimates.

Conclusions and Recommendations

Apex draws the following conclusions and recommendations from this research.

- **Conclusion 1: Paying performance incentives 12 months after a portfolio closed created too long of a feedback loop to effectively influence the way aggregators approached the market.** The performance incentive was the pilot's most significant feedback mechanism to inform aggregators about project performance and motivate them to take steps to increase savings. While aggregators had access to project performance data, without a clear sense of the potential financial gain, they had little motivation to devote time and effort to analyze those data and identify opportunities to increase

[4] A change in utility accounts serving a home during either the baseline or performance period was cause for project disqualification.

savings. With performance incentives paid a year after a portfolio closed, aggregators had to wait 18 months or longer after their first P4P installations to receive feedback in the form of performance incentives. Had the pilot continued, aggregators would then have had to wait a further 18 months to experience the effects of any changes they made in their business practices.

- **Recommendation 1a: In future P4P offerings, Energy Trust should find ways to provide more timely performance payments.** For example, use of more frequent, interval billing data may allow for more accurate projection of savings, allowing a program to offer payments based on estimated savings on a quarterly basis. Energy Trust explored options for providing more timely payments in the P4P pilot but determined doing so was not possible given the data available.
- **Recommendation 1b: In future P4P offerings, Energy Trust and implementation contractor staff should be prepared to analyze performance data for aggregators** and present findings that could lead to actions to increase energy savings. Pilot staff presented this type of data to aggregators, and aggregators reported these presentations were valuable. If possible, it would be beneficial to provide these findings and analysis even before a performance period ends. These interim conversations could also provide opportunities to discuss individual projects performing particularly well or poorly.

〉 **Conclusion 2: Use of more frequent interval billing data could mitigate many, but not all, of the challenges the pilot faced.** In particular, use of interval data would help to overcome many of the data limitations that reduced the number of projects eligible for inclusion in P4P portfolios. With monthly data, loss of only a few data points was sufficient to disqualify a project. Daily or more frequent interval data would provide a much larger number of data points, making it easier to compensate for missing data in analysis. A larger number of data points could also allow for more granular savings analysis, and, as noted above, potentially allow the program to estimate savings and make performance payments sooner after a portfolio closes. Even with data issues resolved, however, the pilot's decision to use installation contractors as aggregators limits its potential scale. The participating aggregators were among the contractors submitting the highest volumes of projects to Energy Trust's residential program, and even these high-volume contractors struggled, in some cases, to meet the pilot's minimum requirements for portfolio size.

- **Recommendation 2: Energy Trust should ensure that future P4P offerings will have access to daily or more frequent interval energy usage data.**

〉 **Conclusion 3: Inaccuracies in deemed savings estimates have the potential to obscure the messages performance incentives send to aggregators.** The

pilot's first portfolio – the only one to complete the performance period – fell short of its deemed estimates for electric savings. This underperformance was largely due to ductless heat pump installations and appeared to stem from two factors: shortcomings of the deemed estimates,[5] and incompatibility between deemed estimates based on a current practice baseline and P4P savings estimates based on an existing conditions baseline.[6] With these factors, which are outside of the aggregators' control, limiting their potential to earn performance incentives, aggregators would have little motivation to increase the energy savings their ductless heat pump projects generate.

- **Recommendation 3: Energy Trust should be prepared to address challenges related to deemed values in future P4P efforts that offer incentives for savings above deemed estimates.** For example, it may be necessary to remove projects based on certain measures from P4P portfolios if Energy Trust determines that deemed values are not a reasonable benchmark for their performance.

[5] Pilot staff noted that southern Oregon, where many of the pilot's heat pump installations took place, has a milder climate than other parts of the state and may require less heating energy consumption, and thus generate less heating energy savings, than the deemed values assume.

[6] A notable share of ductless heat pump (DHP) installations saw energy consumption increase in the performance period relative to the baseline period, suggesting that households that had not previously had space cooling equipment were using their DHPs to cool and/or households were using their DHPs to offset heating with wood or some other fuel. While Energy Trust's latest deemed estimate for DHPs assumes that a portion of participants would have installed an air conditioner had they not purchased a DHP, P4P savings estimates are based only on past usage, adjusted for weather and change in comparison group energy use. They do not consider alternative equipment options.

Memo

To: Board of Directors

From: Mark Wyman, Sr. Program Manager – Residential
Scott Leonard, Program Manager – Residential
Sarah Castor, Program Manager – Evaluation & Engineering

cc:

Date: January 7, 2022

Re: Staff Response to the Residential Pay for Performance Pilot Evaluation

Energy Trust launched the Residential Pay for Performance pilot in 2019 with the goal of investigating the suitability of site-specific metered savings within the Residential energy efficiency program. The pilot was intended to run through 2021. Evaluation activities were designed to be carried out and reported throughout the pilot so the Residential program could get an early read on progress and viability. The program's experience and early evaluation results indicated that aggregators (participating contractors) submitting projects to the pilot were finding it difficult to submit enough qualifying projects to meet the data sufficiency requirements and that it was unlikely to get easier during the remainder of the pilot. The savings analysis from the first portfolio period was delivered in Q4 2020, and except for gas projects for one aggregator, there were no additional savings over deemed. The program made the decision to discontinue the pilot at the end of 2020, after the third portfolio period closed.

The requirements for projects significantly limited the viability of the pilot. Homes that changed heating or water heating fuel were automatically disqualified from the pilot because the metered savings did not accurately reflect the changes in the home's energy use. In addition, homes that changed ownership during the 12 months before or after the Pay for Performance project were also disqualified. These disqualifications were the result of legal requirements around confidentiality of the utility usage data of previous owners. In addition, the need to use an existing conditions baseline for all projects, rather than the baseline called for in the measure approval documents for prescriptive measures, ended up reducing the measured savings and cost-effectiveness of projects. However, the participating aggregators were satisfied with their participation experience and interested in opportunities to receive feedback on the performance of their projects.

Energy Trust gained valuable experience and insight on running a pay-for-performance program by implementing the pilot. There are no further plans to offer a pay-for-performance design for residential customers at this time.

1. Introduction

This report presents findings from an evaluation of Energy Trust of Oregon's residential Pay for Performance (P4P) pilot. Energy Trust contracted with Apex Analytics to conduct this evaluation.

1.1 Pilot Description

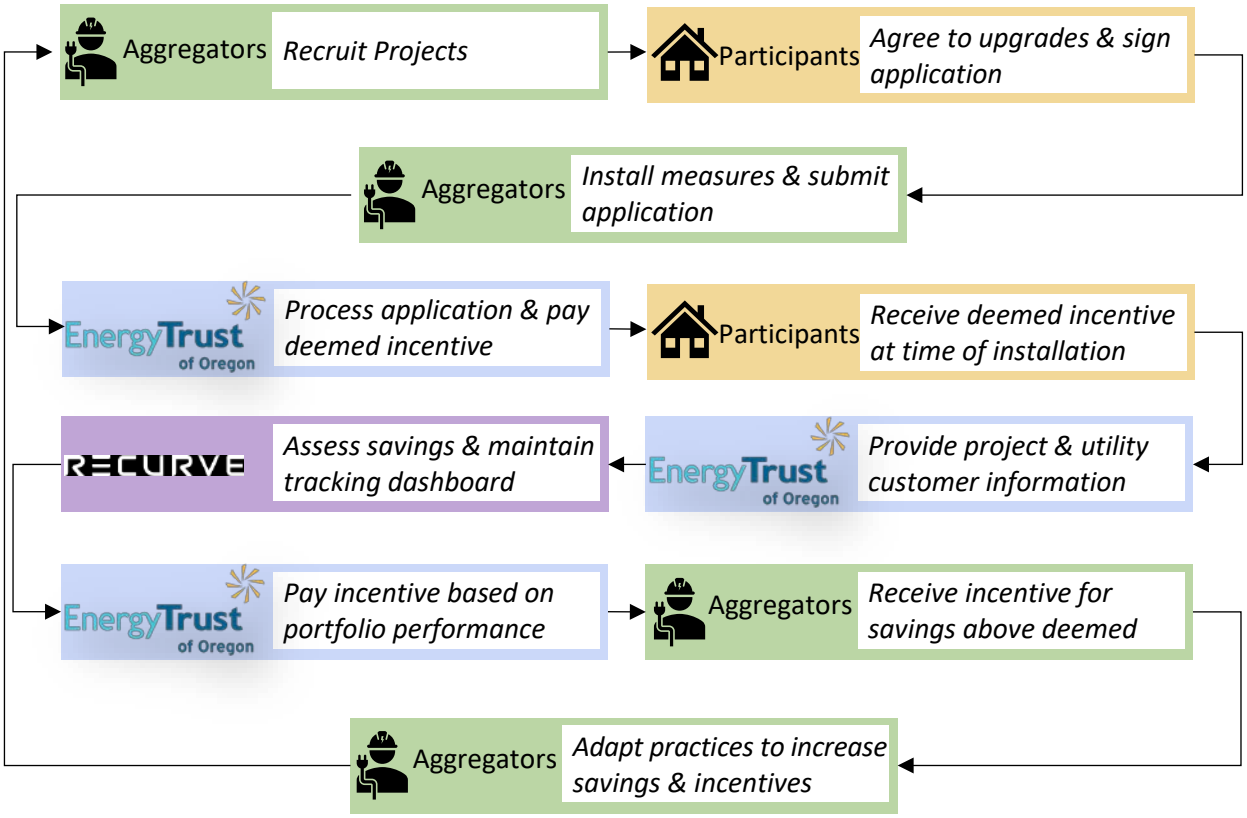
Energy Trust's P4P pilot sought to motivate contractors to take action to increase the energy savings from the projects they install. Households participating in the pilot installed one or more measures with deemed savings and incentive values and received incentives for those measures as they would under the standard residential program. Each project was required to include at least one deemed measure as an "anchor measure." Energy Trust relaxed participation requirements for some measures on P4P projects relative to standard Residential Program projects, but otherwise the process for participants was the same.[7]

The pilot offered participating contractors, called aggregators, an additional performance incentive for measured energy savings their projects achieved above the deemed savings estimates. The pilot anticipated that contractors might install additional energy-saving measures, target projects likely to provide the greatest energy savings, or take other steps to increase savings and maximize their performance incentives. Energy Trust selected three contractors to work with as aggregators for the pilot: an HVAC contractor, a weatherization contractor, and a whole house contractor. The companies invited to participate as aggregators were trade allies that had been highly active in Energy Trust's Residential Program.

Energy Trust contracted with Recurve to provide a dashboard that tracked each project's energy savings, allowing Energy Trust and the aggregators to identify projects with the best performance. The pilot anticipated that aggregators would be able to use this information to determine which measures and project types have the greatest savings potential and pursue more projects of those types. Energy Trust provided utility customer information (UCI) and project tracking data to Recurve to support the analysis underlying the dashboard. Figure 1 summarizes the P4P pilot process.

[7] The P4P pilot eliminated the beginning R-value requirements included in the standard program and did not require that floor and wall insulation in gas heated homes be installed in conjunction with attic insulation.

Figure 1: Key Program Activities



Energy Trust assessed aggregators' performance across a portfolio of projects to determine the savings eligible for performance incentives. To allow for a sufficient comparison group to develop a statistically valid savings estimate, portfolios were required to contain a minimum of 65 projects. The pilot initially anticipated that portfolios would remain open for six months (beginning April 1 and October 1 of each year).[8] In some cases, the pilot extended portfolio periods to allow aggregators to complete a sufficient number of projects. Energy Trust assessed performance incentives for a given portfolio of projects based on the projects' savings performance for one year after portfolio closure.[9]

Energy Trust selected different types of contractors as aggregators, anticipating that each aggregator's portfolios would focus on distinct project types: weatherization, HVAC, or whole home (projects combining insulation and heating and cooling equipment). Energy Trust developed distinct assumptions around

[8] As initially designed, the pilot would close an aggregator's portfolio if it reached 130 projects before the end of the portfolio period, but this did not occur during the pilot implementation.

[9] Energy Trust does not penalize aggregators if monitored portfolio savings are less than deemed estimates.

measure life and other key factors for each project type, which informed the performance incentives Energy Trust offered.

Energy Trust launched the P4P pilot in April 2019. Energy Trust initially planned to run the pilot through the end of 2021. Energy Trust decided to end the pilot early, at the end of 2020, due to challenges around project qualification and data availability, as well as budgetary concerns. As a result, the pilot ran for three enrollment periods. The end of the pilot coincided with the end of the performance period for the first portfolio, which closed in late 2019.

1.2 Research Objectives

Energy Trust specified a variety of research objectives for this evaluation. A key focus of the research was understanding the potential for the P4P pilot structure to provide feedback and motivation that allows aggregators to more strategically target customers and measures to increase energy savings. The evaluation also reviewed Energy Trust's approach to assessing savings and cost effectiveness. Table 2 lists the evaluation research objectives and maps them to the key evaluation activities. Because the pilot ended early, this evaluation was unable to fully address some research questions. For example, as the performance payment for the first portfolio coincided with the end of the pilot, we are unable to assess the extent to which aggregators might have shifted their practices to increase future performance payments. Research objectives we were unable to address are listed in grey text in Table 2.

Table 2: Research Objectives and Evaluation Activities

Research Objective	Pilot Staff Interviews	Pilot Data Analysis	Aggregator Interviews	Savings & CE Review	Participant Survey
Does the P4P pilot design enable better targeting of interventions with variable outcomes?	✓		✓		
If yes, what are the targeting criteria employed?			✓		
Does the P4P pilot design improve measure cost effectiveness?	✓			✓	
What are the most frequent anchor measures? Which are infrequent?		✓	✓		
What additional measures are most commonly associated with each anchor measure?		✓	✓		
Do savings improve as a result of better installation and targeting practices? If yes, what are the practices driving improved impact?		✓	✓	✓	
Do trade allies install additional measures beyond those that receive standard incentives? If yes, what do they do?		✓	✓		
What are the costs associated with P4P projects and how do they compare with standard projects?		✓	✓		
Does the P4P pilot design create new participation opportunities for lagging markets?	✓		✓		
Do the trade ally aggregators serve prior non-participants, and what are the characteristics of those customer groups?		✓			✓
Do data collected through the pilot support the use of a different baseline (such as existing conditions) for P4P projects than for standard measures, in terms of different customer or home characteristics?		✓		✓	
Is the market ready for a "pure" P4P approach with no guaranteed (deemed) incentives? What is trade ally receptivity to P4P?	✓		✓		
Does the Automated Meter Data Analytics (AMDA) platform add value to the users' business practices?	✓		✓		
How persistent are the energy savings from P4P?		✓		✓	
What are the benefits and challenges in offering P4P as a track within the program?	✓		✓		
Are participants satisfied with their experience (installation, performance of measures)?					✓
How does their satisfaction compare to participants through the standard program offering?					✓
Should the program continue to offer residential P4P?	✓	✓	✓	✓	
Are there any changes that would improve P4P?	✓		✓		✓

2. Evaluation Approach

Apex conducted evaluation activities over the course of the pilot period, with research efforts following the close of each portfolio period.

2.1 Pilot Staff Interviews & Document Review

Apex reviewed key pilot documents to understand the pilot design and key pilot changes and decisions documented throughout the pilot process. We also conducted two rounds of interviews with key staff members involved in pilot implementation. In both rounds, the interviews included Energy Trust staff, its Program Management Contractor, CLEAResult, and Recurve, the organization tracking savings for the pilot. Table 3 provides details of these interviews.

Table 3: Pilot Staff Interview Details

Interview Round	Timeframe	Total Number of Staff Interviewed
1	December 2019 – January 2020	5
2	March – April 2021	5

In addition to these formal interviews, the Apex team met with pilot staff periodically throughout the pilot period to receive updates on pilot progress and discuss issues pilot staff were considering.

2.2 Aggregator Interviews

Apex interviewed aggregators by phone following each portfolio period, for a total of three rounds of interviews with the HVAC and whole house aggregators, and two rounds of interviews with the weatherization aggregator, who joined the program following the first portfolio period (Table 4). Aggregator interviews focused on aggregators' experience with the pilot and any changes they had made in their approach to the market in response to the availability of performance incentives.

Table 4: Aggregator Interview Details

Interview Round	Timeframe	Total Number of Aggregators Interviewed
1	January 2020	2
2	June – July 2020	3
3	March – April 2021	3

2.3 Pilot Data Analysis

Apex drew on three sources of pilot data to characterize pilot progress and savings achievements:

- Energy Trust Project Tracking data provided measure-level data on projects installed through the P4P pilot, as well as projects installing similar measures through the standard residential program. These data were limited to measures eligible for deemed incentives (anchor measures).
- CLEAResult tracking sheets included data on the non-anchor measures installed as part of P4P projects, recorded by aggregators.
- Savings performance data for the first HVAC and whole home portfolios, which completed the full 12-month performance period.

3. Market Influence

The P4P pilot's logic anticipates that a combination of incentives and feedback will motivate contractors, acting as aggregators, to take steps to increase the energy savings performance of their projects. Aggregators could take a variety of actions to increase their projects' energy savings, including:

- Recommending additional measures, beyond the "anchor measures" that qualify for prescriptive incentives from Energy Trust.
- Identifying the types of households or projects likely to achieve the greatest energy savings and targeting outreach to those household or project types.[10]

This section presents findings related to the energy savings feedback contractors received and any resulting changes in their approach to the market.

3.1 Feedback on Project Performance

Providing feedback on project performance is central to P4P program logic, which anticipates that aggregators will use that information to target projects or recommend measures that will generate the greatest savings and increase their performance incentives. The extent to which performance incentives meet an aggregator's expectations or goals constitutes one form of feedback. In addition,

[10] Following-up with participants whose projects are delivering less energy savings than anticipated to identify and address any potential installation issues is another action aggregators could take to increase the energy savings of their portfolios. However, Energy Trust staff noted that, during the pilot period, they discouraged aggregators from reaching out to customers in this way to avoid raising customer concerns about privacy.

Energy Trust contracted with Recurve to maintain an online portal providing pilot staff and aggregators project-level performance data, updated monthly.

3.1.1 Aggregator Use of Performance Data

Aggregators participating in the P4P pilot were interested in knowing how their projects performed in terms of energy savings. Aggregators cited the potential to gain a greater understanding of project performance as a motivation for participating in the pilot and as one of the pilot's strengths. According to one aggregator, "We think we are doing a whole bunch of good. Having the real data can help us identify projects that do not go well or upgrades we should steer away from because they do not result in energy savings."

While aggregators had access to the pilot's data portal, which allowed them to monitor the energy savings performance of individual projects, they did not report using it regularly. Aggregators noted that it was difficult to draw meaningful conclusions without a full year of performance data. Aggregators also reported they perceived drawing conclusions from the portal data would require significant effort and analysis. According to one aggregator, "My suspicion is that it might take more analysis than anyone around here has time to do to figure [reasons for project overperformance or underperformance] out, unless the trends are obvious." One aggregator suggested that they might devote more effort to understanding the performance data if they had more experience with performance incentives and a clearer sense of how project performance would impact those incentives.

Pilot staff acknowledged the complexity of the Recurve portal and reported plans to meet with aggregators to discuss project performance in an ongoing way, had the pilot continued. One pilot staff member stated that "A complex website that does not give you all the information you need wouldn't be something you would probably log into a lot to try and figure out," referring to the difficulty of drawing meaningful conclusions about project performance before the full performance period closed.

Another pilot staff member noted that the aggregators involved in P4P programs in other jurisdictions were more often larger engineering and efficiency program implementation firms rather than installation contractors. These larger firms are more likely to have staff who are trained in data analysis and are familiar with energy savings metrics.

Rather than relying on the data portal, aggregators reported looking to pilot staff to identify anomalies or trends in their project performance data and bring them to the aggregators' attention. One aggregator said, "We are busy, and I'm probably not going to make time in my day to look at that Recurve dashboard. But if I have a meeting with [pilot staff] and part of the agenda is to review the dashboard, then it happens." Expressing a similar view, another aggregator expressed a desire for

the pilot to provide an account manager, who could walk them through the performance data and identify projects for closer analysis that could provide insight into savings outcomes. Pilot staff reported plans to have these types of conversations with aggregators, had the pilot continued.

3.1.2 Timeliness of Feedback Provided

Providing timely feedback on project performance was a key challenge for the P4P pilot. The most important factor contributing to this challenge was the need to use monthly energy usage data, rather than daily or more frequent interval data. Pilot staff noted that there was a delay of approximately two months for billing data to be processed and available in the data portal, and that a full year of data was typically necessary to draw meaningful conclusions about a project's performance.

Pilot staff reported that other pay-for-performance programs with access to daily or more frequent usage data are able to develop savings estimates with greater confidence based on shorter measurement periods. For example, some programs provide quarterly estimated performance payments to aggregators.

Providing more frequent performance payments could play an important role in engaging aggregators, as interview findings suggest the performance payments themselves may be a meaningful feedback mechanism. In a summer 2020 interview, one aggregator speculated that they would be motivated to track project performance more closely once they had a better sense for the potential financial outcomes. According to this aggregator, "Some of that performance will result in me getting paid some money. If I understood that relationship, I would be geeking out on that more, but because it is nebulous, I'm not super motivated."

Program staff described similar expectations that aggregators would monitor performance more closely once they had begun to receive performance incentives. One staff member said, "In theory, when you start to see those market changes is after you have been through enough cycles of the aggregators getting paid what they thought they would get paid or more or less than that. That is the trigger, when the price signals and market forces would improve the program."

3.2 Customer, Project, and Measure Targeting

Aggregators did not report making significant changes to the types of customers and projects they targeted or the measures they offered as a result of P4P performance incentives. Throughout the pilot, all the aggregators reported submitting all of their potentially eligible projects for P4P incentives, allowing the program to determine which ones did not qualify for P4P and would instead be part of the standard residential program. Aggregators also reported they did not typically consider performance incentives in deciding what measures to recommend or which customers to target.

While they did not report changing their practices during the pilot period, aggregators reported seeing potential to use performance data and incentives to guide their offerings. According to one aggregator, "I can't imagine it would make us quit carrying or quit doing a certain thing, but I can see it as causing me to put my thumb on the scale with [sales] consultants as far as what they are recommending more with higher priority."

One aggregator also noted that, while they had not changed their approach due to performance incentives, the P4P pilot's changed eligibility requirements for some measures influenced the measures they installed for certain customers. For example, one aggregator noted that they would prioritize floor insulation for a customer whose existing attic insulation made them ineligible for incentives from the standard program, while they would prioritize increasing insulation levels in the attic under the P4P pilot, which allowed for higher baseline insulation levels.

3.2.1 Measures Installed

All P4P projects were required to include at least one measure eligible for Energy Trust deemed incentives, which the pilot called anchor measures. Table 5 lists the anchor measures installed as part of projects included in the P4P pilot.

Table 5: Count of Anchor Measures Installed in P4P Projects

Aggregator Focus	Measures Installed	Count of P4P Installations			
		Portfolio #1	Portfolio #2	Portfolio #3	Total
HVAC	Heat Pump Advanced Controls	72	52	43	167
	Ductless Heat Pump	51	60	28	139
	Heat Pump Replacement	41	42	13	96
	Gas furnace	17	19	14	50
	Floor insulation	0	0	0	0
	Ceiling insulation	0	0	0	0
	Thermostat	1	0	0	1
	Central AC	0	1	1	2
	Other Measure	6	10	0	16
	Total Projects*	165	159	93	417
Whole House	Ceiling insulation	43	52	30	125
	Knee wall insulation	11	1	0	12
	Windows	12	21	7	40
	Floor insulation	10	20	9	39
	Wall insulation	7	38	12	57
	Ductless heat pump	7	8	8	23
	Heat pump replacement	2	3	4	9

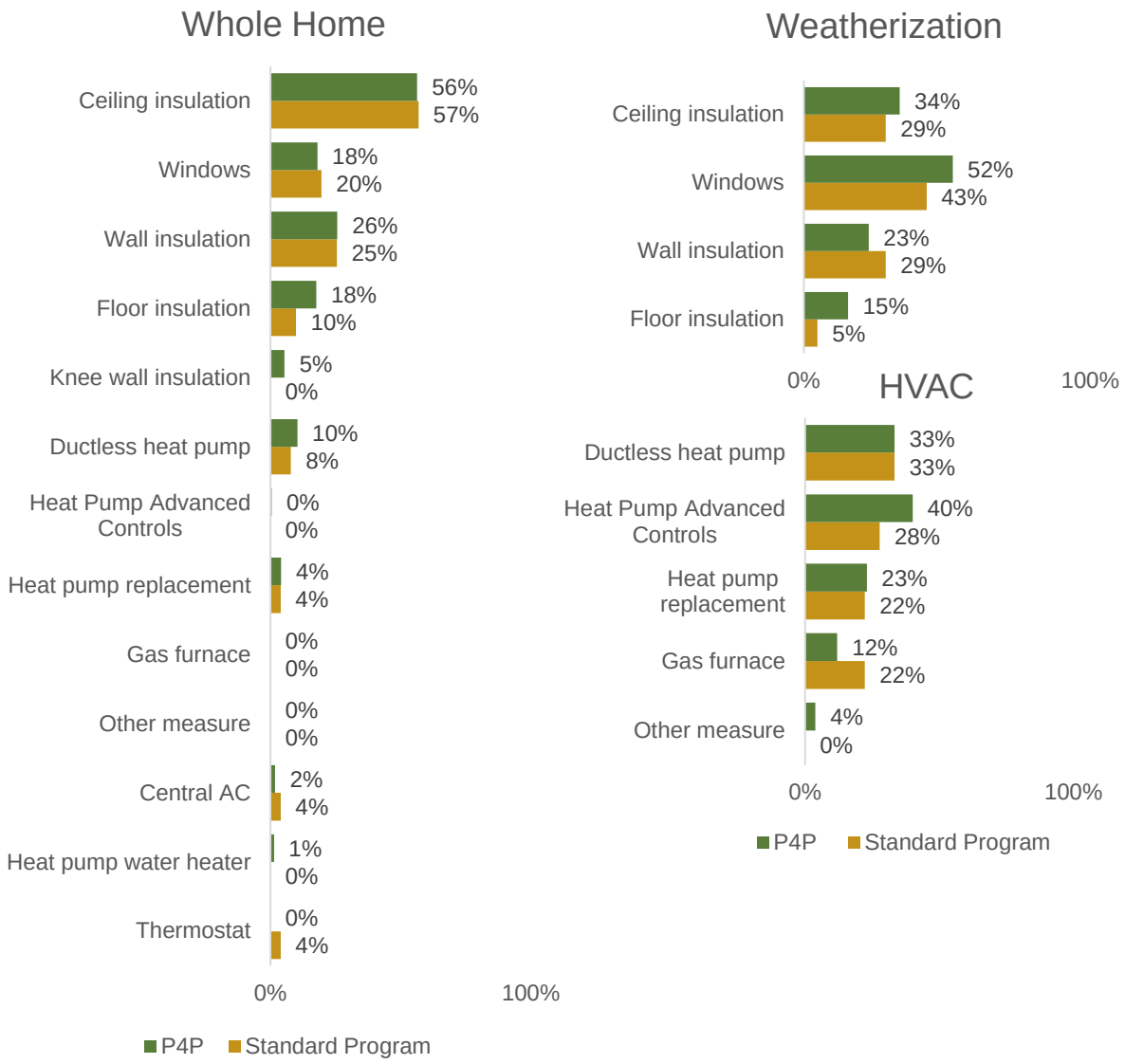
Aggregator Focus	Measures Installed	Count of P4P Installations			
		Portfolio #1	Portfolio #2	Portfolio #3	Total
	Central AC	0	1	3	4
	Heat Pump Advanced Controls	0	1	0	1
	Heat pump water heater	0	2	1	3
	Total Projects*	65	101	56	222
Weatherization	Windows		43	58	101
	Ceiling insulation		33	32	65
	Wall insulation		29	15	44
	Floor insulation		19	11	30
	Total Projects*		93	101	194

* Projects may include multiple measures.

Table 5 includes only projects processed through the P4P pilot. While the P4P aggregators reported submitting all of their potentially eligible projects through the P4P pilot, each had some projects that were processed through the standard program. These projects were likely subject to firm disqualifiers, like fuel switching, the presence of solar, or participation in other Energy Trust programs that made them ineligible for P4P.

There were few notable differences in the anchor measures aggregators installed as part of P4P projects and projects processed through the standard program (Figure 2). Both the whole home and weatherization aggregators were more likely to install floor insulation in P4P projects than projects processed through the standard program, while the HVAC aggregator was more likely to include gas furnaces in standard program projects.

Figure 2: Share of Projects Including Anchor Measure, P4P Pilot & Standard Program

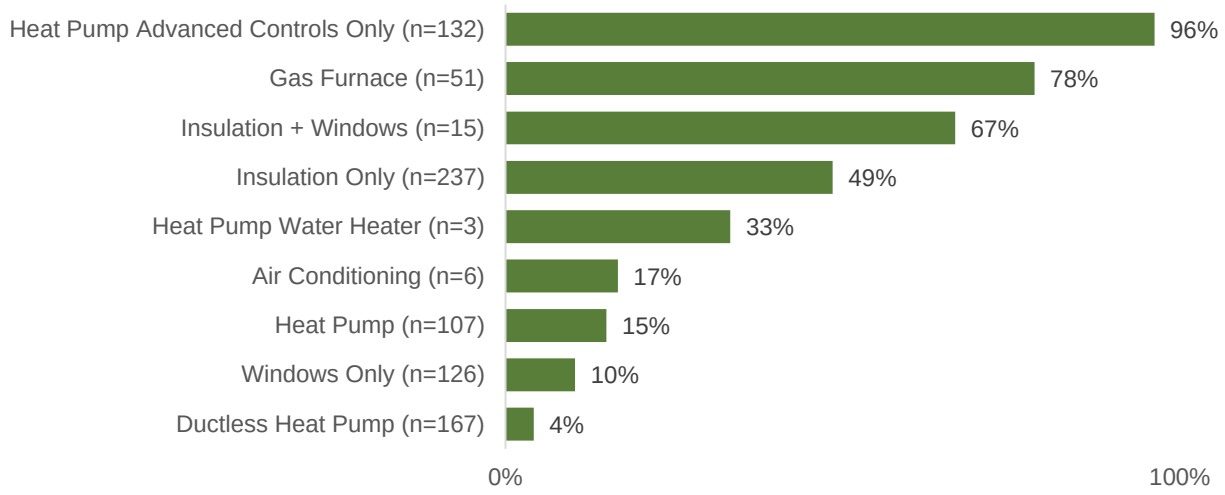


The P4P pilot tracked measures aggregators installed that did not qualify for deemed incentives in addition to the deemed, anchor measures. Energy Trust anticipated that performance incentives might motivate aggregators to install more of these non-anchor measures in order to increase their projects' savings above the estimates based on deemed measures alone. Almost all projects with heat pump advanced controls as the sole anchor measure also included non-anchor measures (Figure 3).[11] The HVAC aggregator frequently listed heat pump advanced controls

[11] Energy Trust's heat pump advanced controls measure is a contractor-installed thermostat with auxiliary heat lockout.

as an anchor measure when completing heat pump installations that did not qualify for Energy Trust incentives.

Figure 3: Share of Projects Including Non-Anchor Measures by Anchor Measure Type



The non-anchor measures aggregators installed in P4P projects that included HVAC anchor measures were generally concentrated among a small number of measure types (Table 6). As noted above, almost all heat pump advanced control installations included a heat pump as a non-anchor measure. The non-anchor measures installed in projects with building shell measures were less concentrated among certain measure types. While air sealing was the most common non-anchor measure installed with insulation projects, more than a third of insulation projects with non-anchor measures did not list air sealing. In contrast, only 5% of gas furnace installations did not list an air conditioner installation as a non-anchor measure.

Table 6: Most Common Non-Anchor Measure Associated with Anchor Measure Types

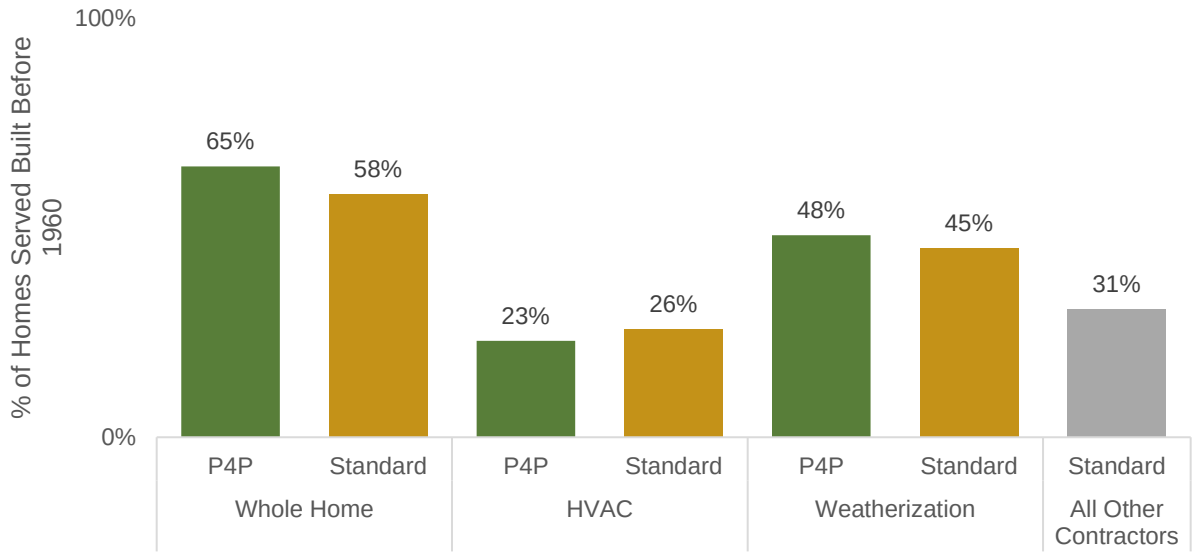
Project Type	Installations w/ Non-Anchor Measures	Most Common Non-Anchor Measure	Share Non-Anchor Meas. Installations w/Most Common Meas.
Insulation Only	115	Air Sealing	63%
Windows Only	13	Air Sealing	31%
Insulation + Windows	10	Air Sealing	80%
Heat Pump Water Heater	1	Heat Pump	100%
Ductless Heat Pump	7	Air Sealing	43%
Heat Pump Adv. Ctrls. Only	127	Heat Pump	99%
Heat Pump	16	Duct Sealing	75%
Gas Furnace	40	Air Conditioner	95%
Air Conditioning	1	Duct Sealing	100%

3.2.1 Home & Project Characteristics

There were some differences in the types of homes aggregators served through P4P and those that were processed through the standard program. These differences likely reflect underlying differences between households eligible for P4P and those that did not qualify for the pilot.

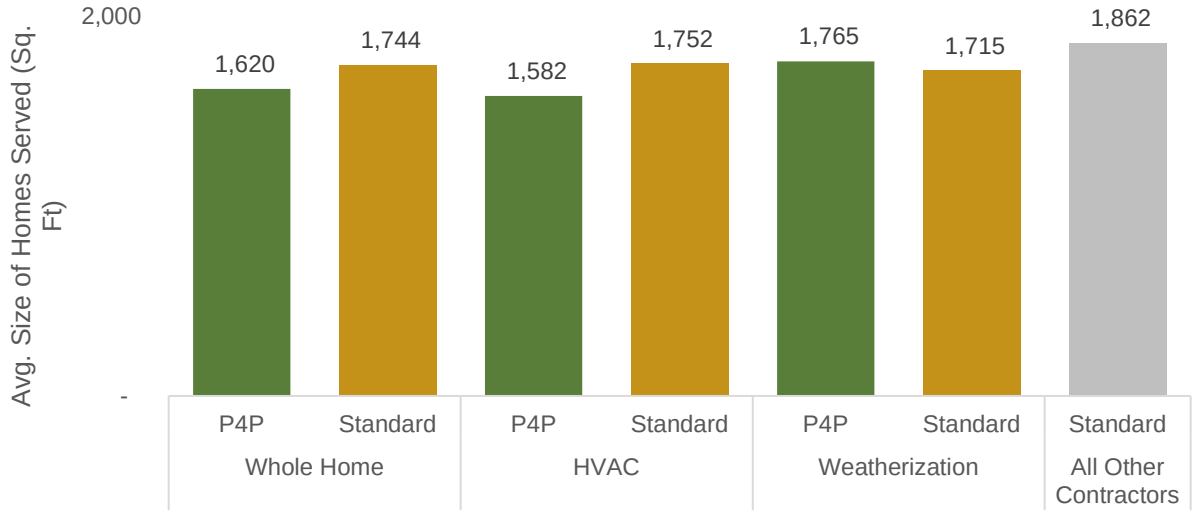
The whole home aggregator served the largest share of older (pre-1960) homes, and the homes they served through the P4P pilot were slightly more likely to be older than the homes they served through the standard program. The HVAC aggregator, in contrast, served the smallest share of older homes (Figure 4).

Figure 4: Pre-1960 Homes as Share of Projects



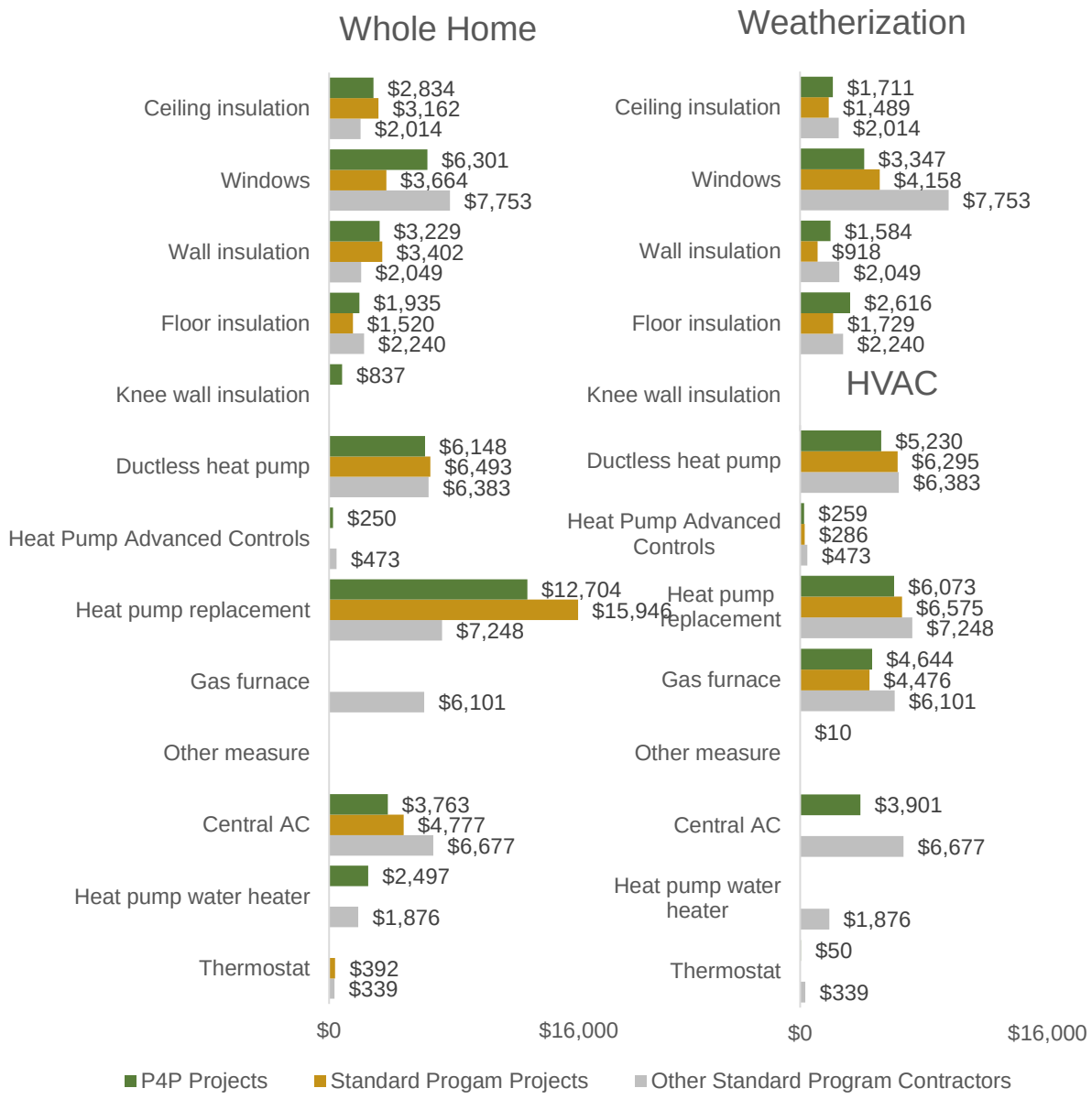
The homes the whole home and HVAC aggregators served through the P4P pilot were slightly larger, on average, than the homes they served through the standard program. All the P4P aggregators served homes that were, on average, smaller than the homes served by other standard program contractors.

Figure 5: Average Size of Homes Served by Program Path



The measure costs aggregators reported were largely consistent with these differences in home size. The whole home aggregator's insulation projects that were processed through the P4P pilot were slightly less expensive than their projects processed through the standard program (Figure 6), likely reflecting the slightly smaller, on average, homes they served through the pilot. The weatherization aggregator, in contrast, reported slightly higher costs for P4P projects, consistent with the slightly larger average home size they served through the pilot.

Figure 6: Average Anchor Measure Installed Cost



3.3 Savings Outcomes

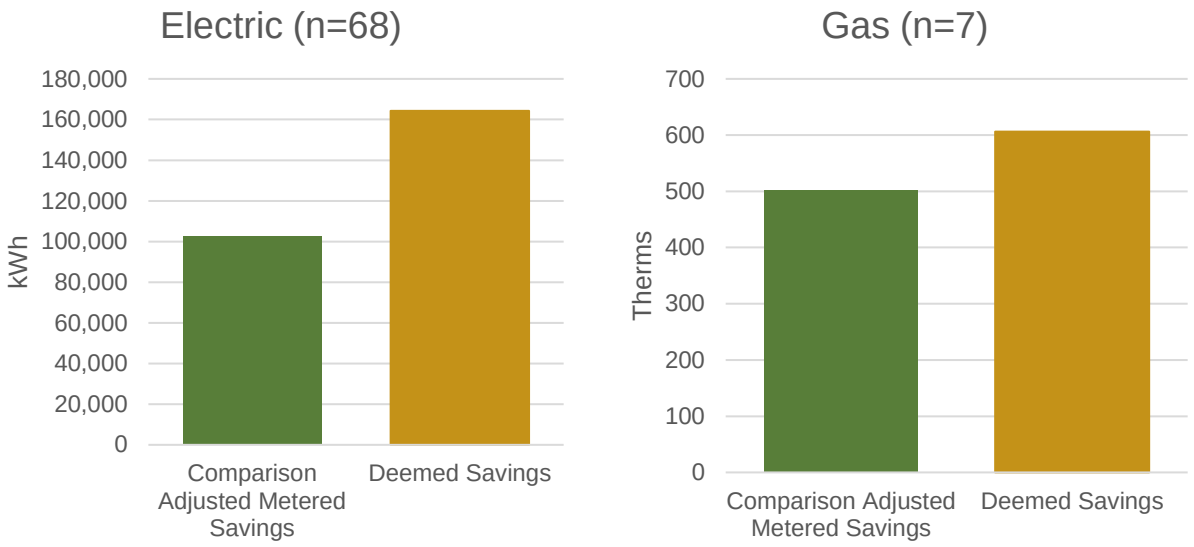
At the end of 2020, when Energy Trust ended the pilot, only the first portfolio, which closed in late 2019, had undergone a full, 12-month performance period. The weatherization aggregator participating in the first portfolio period did not submit any projects to the P4P pilot, and Energy Trust selected a different contractor to serve as the weatherization aggregator beginning in the second portfolio. As a result, complete savings performance data are available only for the first HVAC and

whole home portfolios. This section describes the savings performance of those portfolios.

3.3.1 HVAC Portfolio

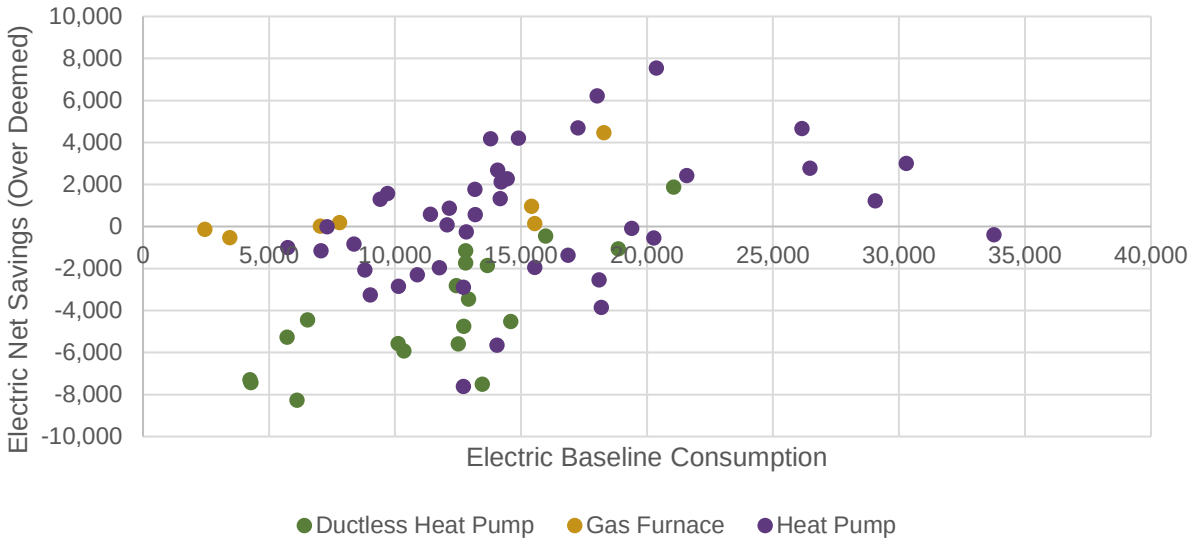
The HVAC aggregator's portfolio primarily consisted of projects achieving electric savings through heat pump installations. In total, there were 68 qualified projects providing electric savings and seven projects providing gas savings in the HVAC portfolio. For both fuels, overall metered energy savings fell short of the deemed estimates. Figure 7 summarizes the aggregate performance of the HVAC portfolio. Given the relatively small number of gas savings projects, the remainder of this section focuses on electric savings outcomes.

Figure 7: HVAC Portfolio Overall Performance



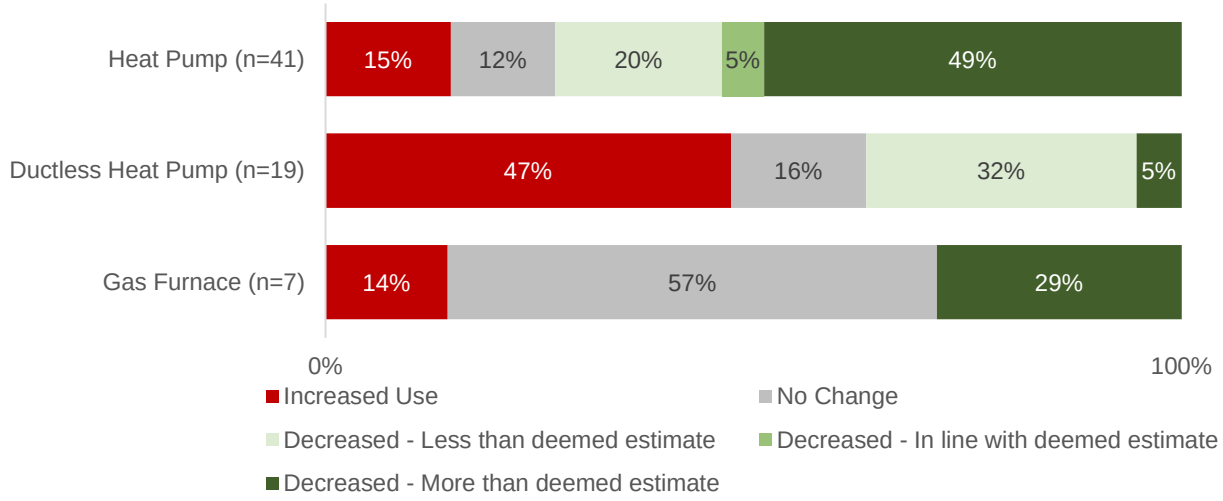
Program staff reported that the performance of the HVAC portfolio had fallen short of the deemed estimates because the milder climate in Southern Oregon, where the HVAC aggregator operates, leads to lower baseline energy consumption and thus lower energy savings. Further analysis of the HVAC performance data provides some support for this hypothesis. There is a relationship between a home's baseline consumption and its savings performance relative to the deemed estimates (Figure 8), although there remains considerable variation in savings performance between homes with similar baseline consumption levels.

Figure 8: Relationship Between Electric Savings Performance and Baseline Consumption (HVAC Portfolio)



Ductless heat pump installations were particularly likely to fall below the deemed estimates. Nearly half of the households that installed ductless heat pumps had their energy consumption increase in the reporting period, relative to the baseline period (Figure 9). That increase in energy consumption likely reflects households that had not previously had cooling equipment using their ductless heat pumps for cooling, or households offsetting wood heating or other fuels with their ductless heat pumps. Households installing ducted heat pumps were more likely to save energy, with nearly half exceeding the deemed savings estimates.

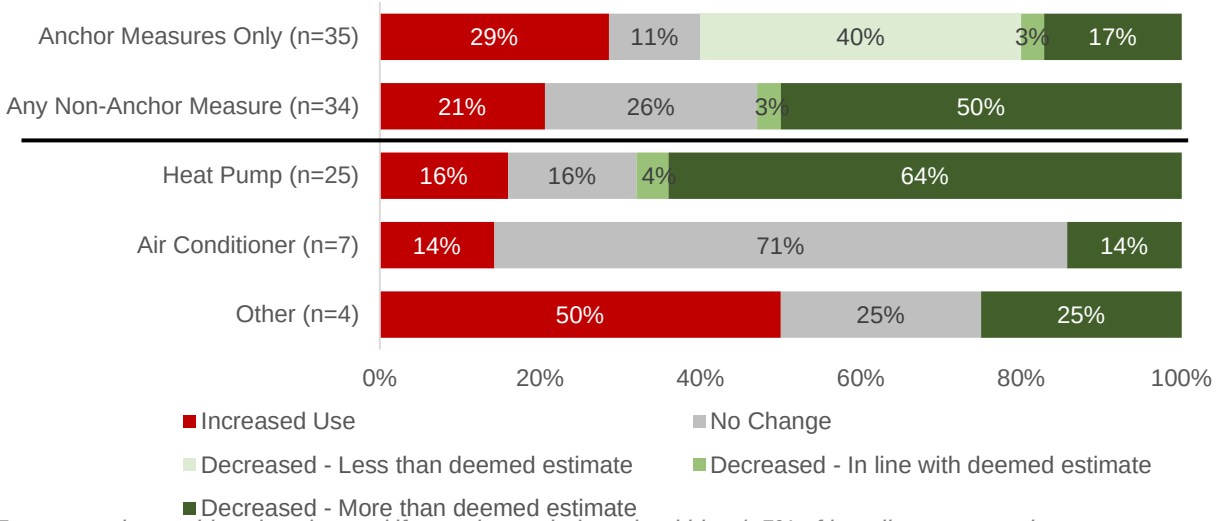
Figure 9: Change in Electric Energy Consumption by Heating System Type Installed (HVAC Portfolio)



Energy use is considered unchanged if reporting period use is within +/- 5% of baseline consumption. Likewise, savings are considered in line with deemed estimates if they are within +/- 5% of those estimates.

Households installing non-anchor measures were notably more likely to exceed the deemed savings estimates for their projects than households installing anchor measures only (Figure 10). Households installing non-anchor measures were also less likely to see their energy consumption increase.

Figure 10: Change in Electric Energy Consumption by Non-Anchor Measure Installed (HVAC Portfolio)



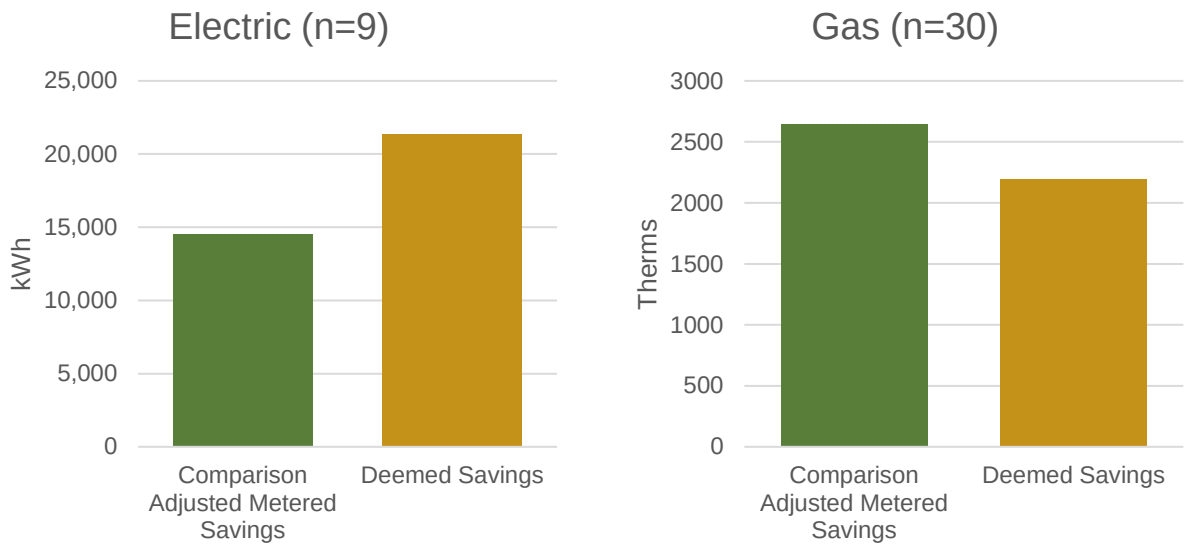
Energy use is considered unchanged if reporting period use is within +/- 5% of baseline consumption. Likewise, savings are considered in line with deemed estimates if they are within +/- 5% of those estimates.

The HVAC aggregator installed heat pumps as non-anchor measures in some cases in which the heat pump installation did not qualify for Energy Trust incentives. A majority of these projects exceeded the deemed savings estimates, which were based on relatively small savings associated with heat pump advanced controls.

3.3.2 Whole Home Portfolio

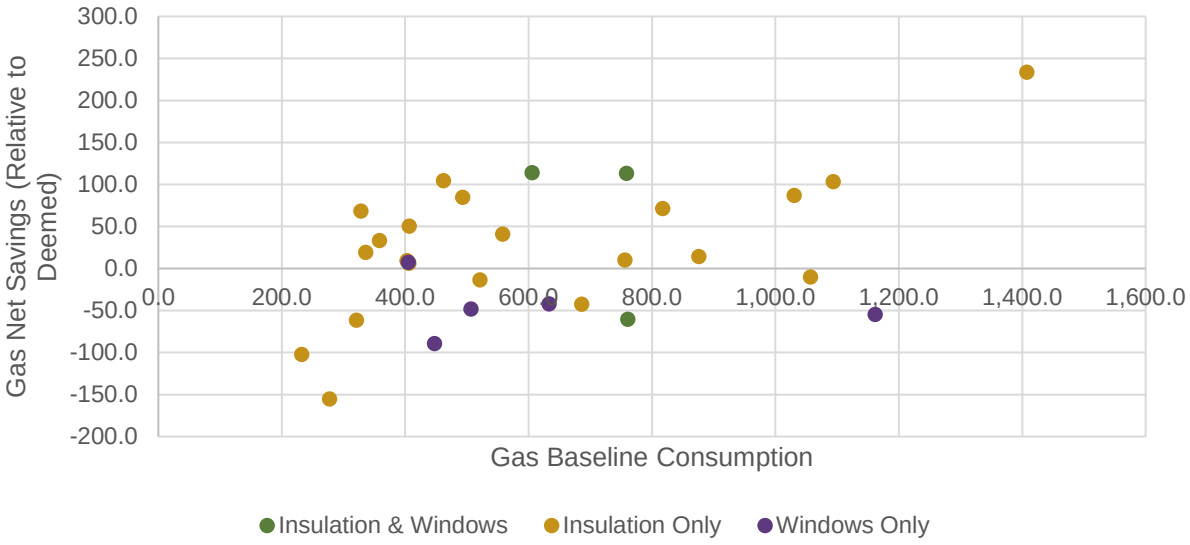
The whole house aggregator's portfolio primarily included projects that achieved natural gas savings through insulation and windows upgrades. A smaller number of projects also provided electric savings from installation of heat pumps (ductless and ducted), in some cases in addition to shell measures. In total, there were 30 qualified projects providing natural gas savings and 9 providing electric savings. As with the HVAC aggregator, electric savings fell short of the deemed estimates. The whole home portfolio's natural gas savings exceeded the deemed estimates. Figure 11 summarizes the performance of the whole home portfolio as a whole.

Figure 11: Whole Home Portfolio Overall Performance



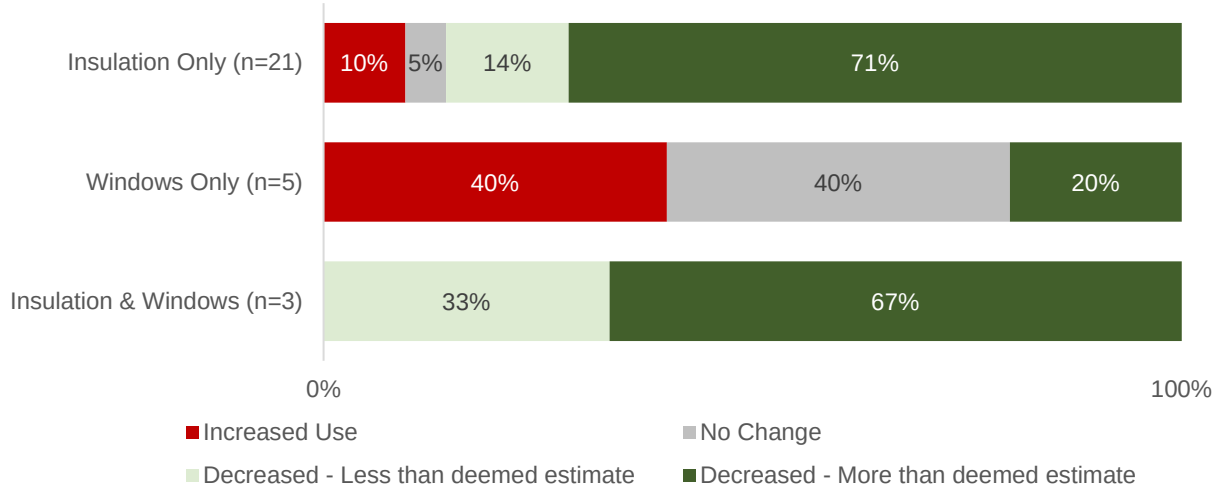
The relationship between natural gas energy savings relative to deemed estimates and baseline energy consumption was less clear for whole home projects than for HVAC projects. While the lowest-performing projects were in homes with the lowest baseline consumption and the highest performing project was in the home with the highest baseline consumption, these patterns did not hold throughout the portfolio (Figure 12).

Figure 12: Relationship Between Gas Savings Performance and Baseline Consumption (Whole Home Portfolio)



More than two-thirds of insulation projects, including insulation-only projects and projects in which both insulation and windows were installed, exceeded their deemed savings estimates (Figure 13). The five projects that included windows only did not perform as well, with only one project showing energy savings.

Figure 13: Change in Natural Gas Energy Consumption by Building Shell Measures Installed (Whole Home Portfolio)

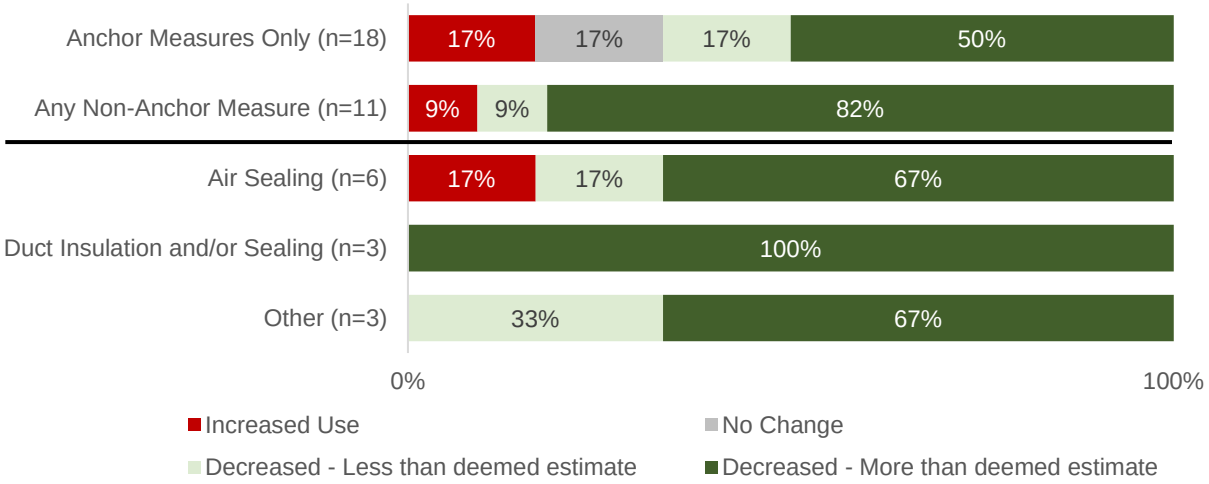


Energy use is considered unchanged if reporting period use is within +/- 5% of baseline consumption. Likewise, savings are considered in line with deemed estimates if they are within +/- 5% of those estimates.

Households installing additional measures as part of their whole house upgrades were notably more likely than those installing only anchor measures to exceed the

deemed savings estimates for their projects (Figure 14). Majorities of households installing each type of non-anchor measure exceeded the deemed savings estimates.

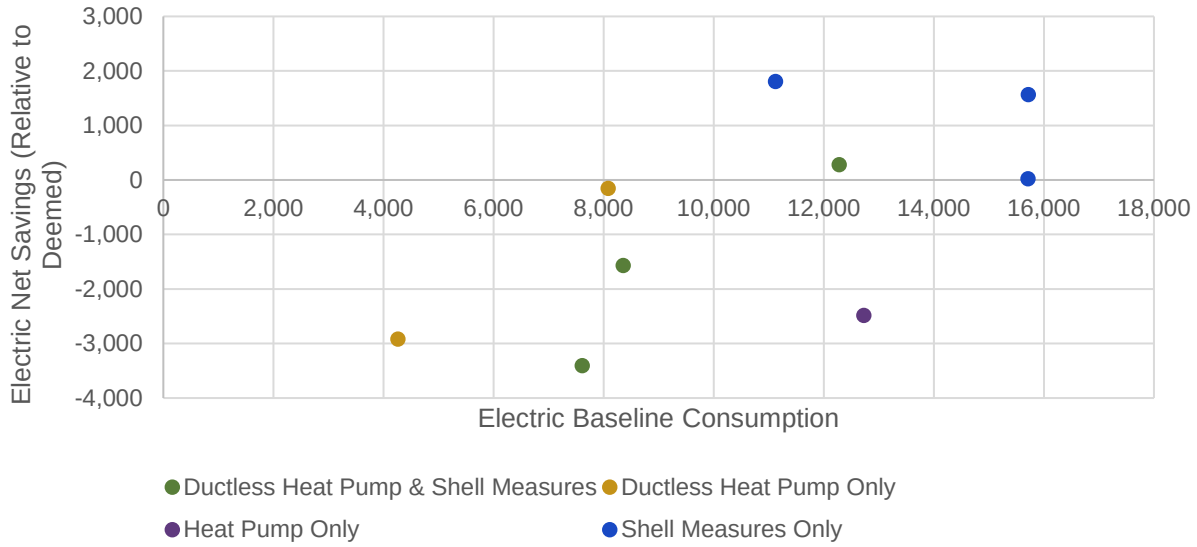
Figure 14: Change in Natural Gas Energy Consumption by Non-Anchor Measures Installed (Whole Home Portfolio)



Energy use is considered unchanged if reporting period use is within +/- 5% of baseline consumption. Likewise, savings are considered in line with deemed estimates if they are within +/- 5% of those estimates.

Relatively few projects in the whole home portfolio included electric efficiency measures, with the electric efficiency measures installed primarily consisting of heat pumps. Consistent with the HVAC portfolio, most heat pump installations achieved savings below the deemed estimates (Figure 15). All of the stand-alone shell measure installations (windows and insulation) that provided electric savings achieved savings in line with or above the deemed estimates.

Figure 15: Relationship Between Electric Savings Performance and Baseline Consumption (Whole Home Portfolio)



4. Pilot Experience

This section assesses three key aspects of pilot design: the role of installation contractors as aggregators, the participation experience for aggregators and end-use customers, and the availability and sufficiency of energy consumption data.

4.1 Role of Installation Contractors as Aggregators

4.1.1 Aggregator Motivations

The installation contractors participating in the P4P pilot reported that the opportunity to better understand the energy savings performance of their installations and the potential to earn additional incentives motivated them to participate. In a representative comment, one aggregator reported they were motivated by "The idea of being able to track our projects through an online platform and actually see how many kWh we are saving…and, honestly, at the end of the period, after they have reviewed everything, the opportunity to have a little more money for us."

One aggregator also described P4P as an opportunity to overcome some of the constraints around Energy Trust's prescriptive incentive offerings, saying they were drawn to "Being able to do work, and potentially subsidize work, that we know is going to result in net global energy savings, without having to shoehorn those into the strictures of the current Energy Trust program." This aggregator also stated that a more market based P4P structure might be more stable than prescriptive

incentives, which are subject to changing incentive amounts and eligibility requirements.

4.1.2 Role of Installer-Aggregators

Energy Trust's decision to use individual installation contractors as aggregators set its residential P4P pilot apart from pilots in other jurisdictions. Pilot staff reported that other P4P programs have used large engineering firms or implementation contractors, who work with multiple installation contractors, as aggregators. In a "pure" P4P approach, in which the program administrator did not offer deemed incentives, these third-party aggregators would bear the upfront cost of offering marketing support and potentially incentives, favorable financing terms, or other inducements to recruit participants. The aggregator would then receive compensation through the performance incentive.

Working with installation contractors as aggregators, as Energy Trust did, offers three key benefits, particularly during a pilot period:

- **It provides performance incentives to market actors directly involved in selling projects and installing measures.** The contractors acting as aggregators reported that the opportunity to earn performance incentives had the potential to motivate them to change the types of measures they recommend or take other steps to increase their energy savings. This type of motivation may be less direct under a third-party aggregator design.
- **It leverages actors already active in the market.** A third-party aggregator operating under a "pure" P4P approach represents a new business model in the market. That business model would likely require a relatively large upfront capital outlay to support any marketing, incentives, or other inducements the aggregator would offer to recruit participants. It may be difficult to find actors willing to take on this unproven business model.
- **It allows for integration with ongoing, standard program offerings.** The contractors acting as aggregators in the P4P pilot continued to work with customers and offer incentives very similar to those they offered under the standard program. A third-party aggregator approach might lead to greater differences between P4P offerings and standard program offerings, resulting in greater friction, and potentially competition, between P4P and the standard program.

The pilot's experience suggests two challenges with the approach of using installation contractors as aggregators: limited project volumes and competing business priorities, as discussed in the following sections.

4.1.2.1 Project Volumes

The number of projects individual installation contractors complete over a given time period limits the number of projects they can include in a P4P portfolio. The contractors Energy Trust invited to participate in the P4P pilot as aggregators were among those submitting the highest volume of standard program projects. As Table 7 demonstrates, all of the P4P aggregators were among the five highest-volume contractors of their respective types in the period from June to December of 2020. The whole home aggregator in particular far exceeded any other whole home contractor in project volume, with the next highest-volume contractor submitting 40 projects during that period.

Table 7: P4P Aggregator Project Volume June – December 2020

Aggregator Type	June – December 2020	
	Total Projects Submitted	Project Volume Rank (1 = highest)
HVAC	134	5
Weatherization	120	2
Whole Home	104	1

In total, in the final six months of 2020, 33 contractors met the threshold of 65 projects the P4P pilot set as a portfolio minimum in order to allow for a statistically valid comparison group analysis to control for exogenous effects. These contractors represented 53% of all projects submitted during that period. This presents a potential challenge in scaling up a P4P program using individual installation contractors as aggregators, since nearly half of projects submitted to Energy Trust's standard residential program were completed by contractors who would be unable to meet a portfolio minimum threshold.

Further, these estimates assume that all of the projects submitted would be eligible for P4P. The pilot experienced data challenges (discussed in section 4.3, below) that disqualified a relatively large share of projects. Even if Energy Trust could overcome those challenges, some projects will inevitably be ineligible for P4P.[12] As a result, contractors will need to submit a larger total number of projects to achieve a portfolio of 65 eligible projects, further limiting the number of contractors able to

[12] Projects in homes with solar photovoltaic (PV) systems, homes that participate in other Energy Trust programs during the baseline or performance period, or projects flagged due to data validity checks would remain ineligible even if other data issues were resolved.

act as aggregators and the share of projects eligible to be included in P4P portfolios.[13]

4.1.2.2 Business Priorities

The P4P pilot's aggregators' primary business is selling and installing HVAC, weatherization, or whole home projects. While they were motivated to maximize their performance incentives, other business priorities could take precedence. For example, after the second portfolio period, one aggregator reported they had been doing a high volume of projects in multifamily buildings. According to this aggregator, this multifamily work was "great for business for us, we probably make more money on that anyway, but it took half the company's install capacity away from doing anything that would qualify for the pilot."

The pilot also changed insulation aggregators following the first portfolio period. The aggregator originally selected to focus on insulation projects underwent a significant reorganization of their business during the first portfolio period, during which they largely stopped completing existing home retrofit projects. The pilot selected a different aggregator to participate beginning in the second portfolio period that could provide a greater volume of projects.

4.2 Participation Process

Energy Trust designed the P4P pilot participation process to largely parallel the standard program participation process for both customers and aggregators. As detailed in the following sections, both groups indicated minimal difficulty navigating the pilot's participation process.

4.2.1 Customer Participation

As noted above, with the exception of expanded eligibility requirements for certain measures, the P4P pilot's design sought to maintain a participation process for end-use customers that was as similar to the standard program process as possible. Energy Trust assessed customer experience with P4P through its monthly Fast Feedback surveys, which ask participants a brief set of questions about program satisfaction and influence shortly after they participate in an Energy Trust program.[14] Apex reviewed these survey findings to assess whether they reflected

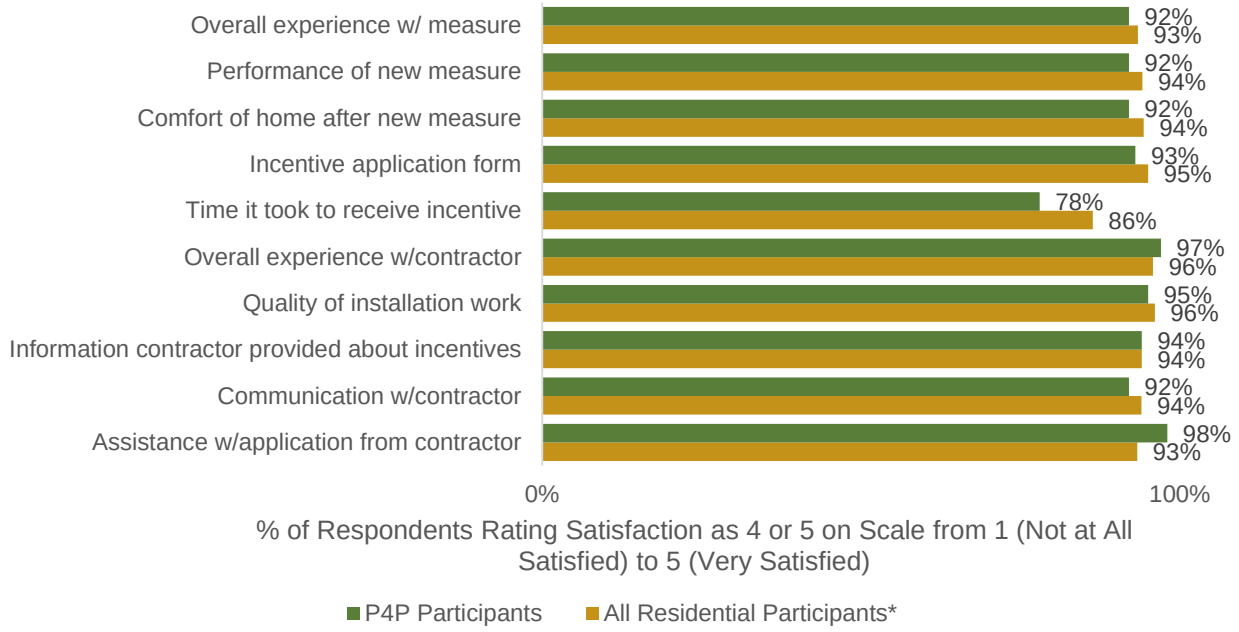
[13] For example, if 20% of projects submitted are ineligible for P4P, a contractor would need to submit 82 total projects to meet the portfolio minimum of 65 eligible projects. Twenty-six contractors, representing 46% of all projects submitted, achieved that project volume in the second half of 2020.

[14] Findings reported here draw on the *2020 Fast Feedback Survey End of Year Report*, dated March 22, 2021, prepared by ADM Associates.

differences in program experience between P4P participants and residential program participants overall.

Consistent with the pilot's design and aggregators' reports that they did not adjust program delivery for P4P participants, there were few notable differences in satisfaction between P4P participants and standard program participants installing similar measures (Figure 16).[15] Somewhat fewer P4P participants were satisfied with the time it took to receive their incentive, but P4P participants were somewhat more satisfied than other participants with the assistance they received from their contractors in completing the application.

Figure 16: P4P Participant Satisfaction Ratings



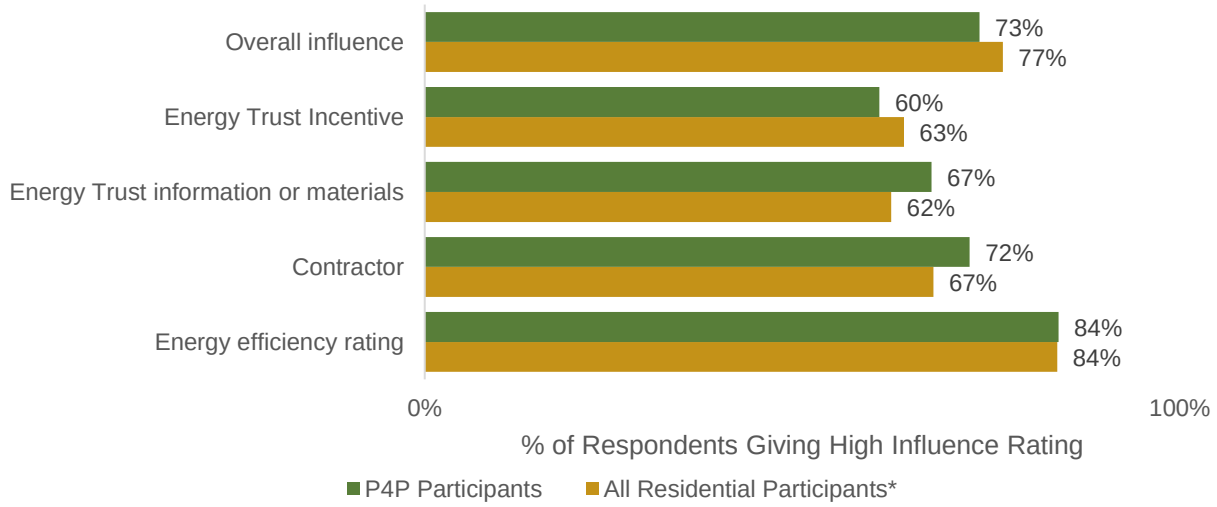
* Satisfaction ratings for all participants weighted to reflect distribution of measures installed among P4P sample.

There were also minimal differences in respondents' ratings of the influence of various program elements in their decision to install the measure between P4P participants and standard program participants. P4P participants were slightly more likely to rate their contractors and Energy Trust information and materials as influential in their decision. Given the limited range of contractors participating in

[15] Ratings differed between respondents installing different types of measures. To facilitate comparison with the P4P respondent sample, the Apex team weighted responses of the full participant sample to match the distribution of measures installed by surveyed P4P participants. P4P survey respondents installed ceiling insulation (n=29), other insulation (n=22), ducted heat pumps (n=16), windows (n=15), ductless heat pumps (n=12), gas furnaces (n=3), and smart thermostats (n=4).

P4P as aggregators, these differences likely reflect the contractors' approach to the market to a greater extent than any elements of the P4P pilot design.

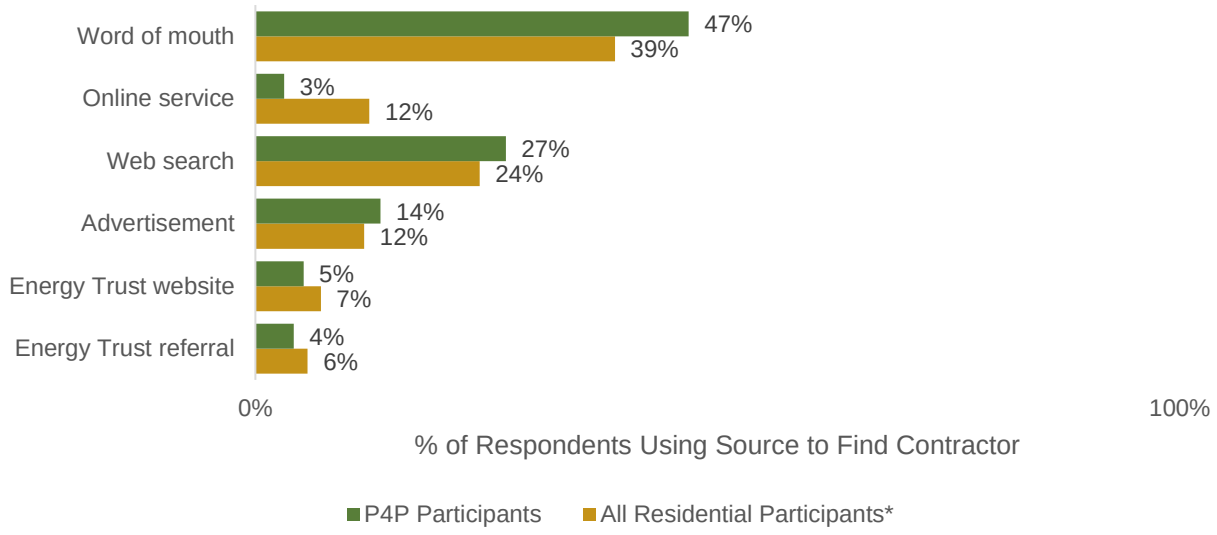
Figure 17: P4P Participant Ratings of Influence of Program Elements in Installation Decision



* Influence ratings for all participants weighted to reflect distribution of measures installed among P4P sample.

Word of mouth was the most common way participants found their contractors, both among P4P participants and other participants, although P4P participants were more likely than residential participants overall to report finding their contractor through word of mouth (Figure 18). P4P participants were less likely than residential participants overall to report using an online service like Angie's List to find their contractor.

Figure 18: P4P Participant Source Used to Find Contractor



These findings could have implications for future P4P efforts. Staff noted that some P4P efforts in other jurisdictions use performance data to both identify eligible households and target those with the greatest potential to save. These strategies would require aggregators to conduct active, targeted outreach to eligible, customers with high savings potential. The Fast Feedback survey did not include response options that would reflect this type of active, targeted contractor outreach to participants. However, the prevalence of word of mouth and web searches as a source of contractors suggests that active, targeted outreach could represent a shift in contractors' approaches.

4.2.2 Aggregator Experience

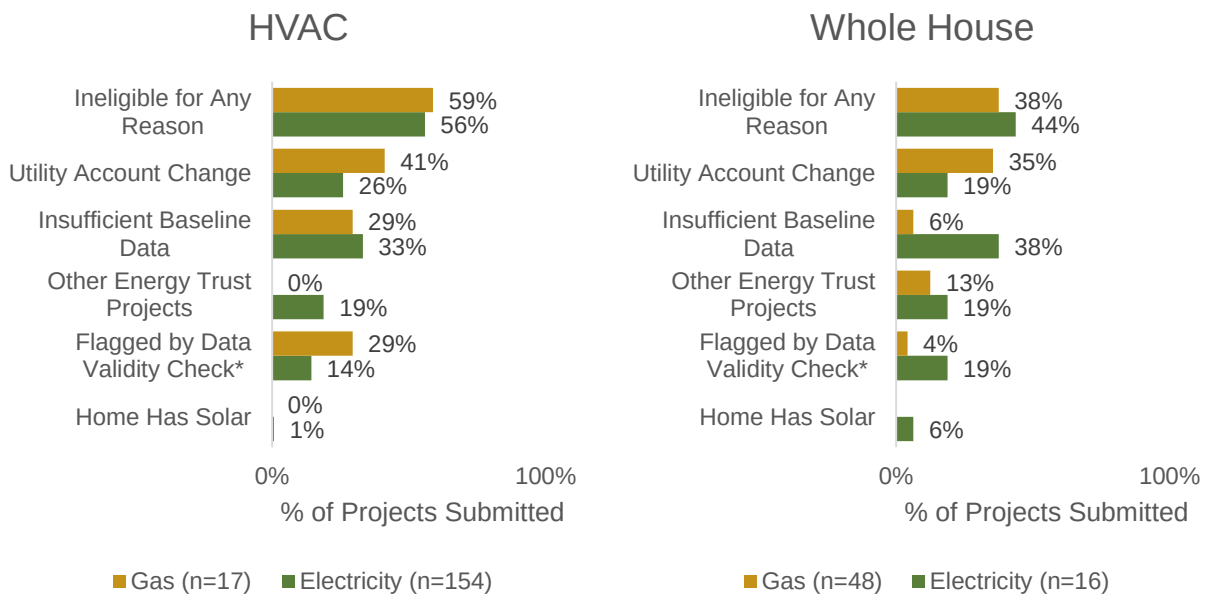
Differences in the participation process for aggregators between the P4P pilot and the standard residential program were relatively minor, and aggregators reported the difficulty of participation was not notably different between P4P and the standard program. Aggregators reported that they had to scan and email forms to Energy Trust as part of the P4P pilot rather than using a web portal as they do for the standard program, which the aggregators found slightly less convenient. Aggregators noted, however, that convenience was offset by the benefit of having a dedicated program staff member receiving and reviewing their applications.

Pilot staff held individual presentations to orient each aggregator to the pilot and met with aggregators following the close of each portfolio period to review the pilot data portal and encourage aggregators to track the performance of their projects. Aggregators reported these interactions were helpful and addressed the questions they had about P4P. Aggregators did not identify any aspects of the participation process that caused them difficulty or confusion. Both aggregators and program staff reported that the pilot worked smoothly within the standard residential program.

4.3 Data Availability & Sufficiency

A relatively large share of the projects P4P aggregators submitted to the pilot were ultimately ineligible to be included in their portfolios assessed for performance incentives. As Figure 19 shows, the most common reasons projects were ineligible had to do with utility account changes at the property in either the baseline or performance period or the availability of baseline data.

Figure 19: Reasons for Project Disqualification from Portfolio 1[a]



[a] Figures listed include all projects disqualified for each reason. An individual project with multiple reasons for disqualification would be included in percentages for each reason.

*Data validity check items included a change in comparison adjusted metered savings greater than 50%, projects in the first or 99th percentile for comparison adjusted metered savings and a reporting period model CVRMSE value greater than or equal to one.

Aggregators reported that many of the projects they complete are in homes that have been recently purchased, rental housing that is changing tenants, or part of other, larger changes in the home. According to one aggregator, "One thing I had never really thought of before…is how often there is more going on in the house. Rentals changing tenants, sales, people remodeling, just other stuff going on, other than…whatever measure upgrade you have done."

Pilot staff identified the prevalence of projects with insufficient baseline data in the spring of 2020 and investigated its causes. At the time, nearly a third (29%) of P4P projects had gaps in their energy consumption data, particularly during the baseline period.[16] Pilot staff classified the data gaps they encountered into three categories, listed in Table 8. As Table 8 notes, the customer changes described above were often also associated with data gaps.

[16] This analysis, conducted on all projects completed to date, as of summer 2020, draws on a slightly different dataset than the findings in Figure 19, which are based on final Portfolio 1 data. The projects identified as containing data gaps largely correspond to those listed in Figure 19 as containing insufficient baseline data and/or a utility account change.

Table 8: Frequency of Data Gaps Identified by Pilot Staff in July 2020

Data Gap Type	Definition	Pilot Staff Assessment	Count of Projects (n=559)	% of Projects
Off Cycle Read	Meter readings 35-40 days after previous, just beyond the CalTRACK 35-day threshold.	Likely represents actual usage: daily consumption typically comparable to same period in previous year.	25	4%
Missing Read	Meter readings more than 40 days after previous, median length 63 days.	Likely represent missing data: daily consumption typically roughly half of same period in previous year.	39	7%
Customer Change	Change in customer identifier between meter readings, often with a gap in time between readings.	Likely represent change in occupancy, potentially including a period of vacancy between occupants.	121	22%
		Any Data Gap	**160**	**29%**

Source: CLEAResult analysis presented July 2020.

Staff noted that the potential to resolve each type of data gap varies. Off cycle reads are likely the easiest to resolve because, while they fall outside of CalTRACK guidelines, they largely appear to reflect actual household usage.[17] Staff reported that there may also be potential to restore data from a missing read, although doing so would be more difficult, potentially requiring considerable manual data cleaning. Missing data due to periods of vacancy during a customer change is not possible to resolve.

Pilot staff noted that data privacy limitations also factor into projects in which a customer changes has occurred. While P4P participants agree to share energy consumption data with Energy Trust and the aggregator, a previous resident would not have had occasion to provide consent for this type of data sharing. Pilot staff stated that finding a way to overcome these limitations, whether by anonymizing data in some way or reaching out to a previous resident for consent, would have increased the number of projects eligible for performance incentives.

Pilot staff stated that access to daily or more frequent interval energy consumption data, rather than the monthly energy usage data the pilot relied upon, likely would have helped to resolve many of the data issues the pilot encountered. With a greater number of datapoints, the pilot could better fit energy consumption models to the data and more accurately estimate savings, potentially over shorter time periods. In addition, the pilot would be better able to manage individual, missing data points. Pilot staff pointed out that a single missing read could call into question two months of data – both the missed month and the subsequent reading if it was

[17] CalTRACK presents a standardized methodology for measuring energy savings from normalized metered energy consumption. Additional information is available at caltrack.org.

unclear whether that reading represented 30 or 60 days of data – which was sufficient to disqualify the project.

Pilot staff and aggregators described Energy Trust's restrictions against fuel switching as another notable factor limiting the number of projects eligible for P4P, although Apex was unable to assess its prevalence based on the data we received.[18] One aggregator in particular reported that they advocate for customers to replace gas furnaces and central air conditioners with a high-efficiency heat pump, describing fuel switching projects as "our best energy saving projects." This aggregator noted that, without projects involving fuel switching, the projects that qualified for their P4P portfolio were limited.

5. Savings and Cost Effectiveness

This section presents Apex's review of Energy Trust's savings calculation and cost effectiveness approaches.

5.1 Savings Calculations

Recurve calculated the energy change for each individual meter using the CalTRACK 2.0 methodology. These methods specify steps for data cleaning, weather station selection, regression, and weather normalization, and are documented online at www.caltrack.org. Energy Trust's P4P pilot also included a matched comparison group to control for exogenous factors that might affect energy use between the pre- and post-period. A comparison group is not explicitly part of the site level methodology specified by CalTRACK but can be incorporated if data from non-participants is available.

Using Energy Trust data from non-participants, Recurve selected 6 comparison meters for each treated meter using least distance matching of pre-period consumption data and some quality control criteria, then assessed the quality of these matches using the Kolmogorov-Smirnov test. Energy use change from the matched comparison meters was applied on a percentage basis for a normal (TMY3) year to the treatment group baseline before calculating adjusted energy savings. This type of adjustment is an appropriate method for incorporating a comparison group into site level normalized annual consumption calculations to account for exogenous energy use changes.

[18] Any project that included fuel switching, whether for space heating or water heating, was disqualified, including any associated measures, even if they were not directly related to the fuel switching measure. For example, a project that included a gas storage to heat pump water heater conversion as well as insulation upgrades could not participate in P4P on the basis of the insulation measures alone.

5.2 Cost Effectiveness

Energy Trust sought to understand whether the P4P pilot design improved measure cost effectiveness.[19] Energy Trust staff developed a cost effectiveness screening test to assess P4P measures and projects. Apex reviewed the cost effectiveness results Energy Trust staff developed to validate the input assumptions, review the various costs and benefits, and offer feedback about the general approach.

A summary of the various cost and benefit components feeding into the cost effectiveness tests are displayed in Table 9 below. Energy Trust uses the Total Resource Cost (TRC) test as the primary test of measure-level cost effectiveness; measures with values of 1.0 or greater are cost-effective. Measures must also have a Utility Cost Test (UCT) value of 1.0 or more, which can be achieved by setting a maximum incentive level.

Table 9. Cost Effectiveness Components

CE Component	Used in Cost Test	Deemed Measure (anchor)	P4P
Measure Upgrade Costs	TRC	Incremental Cost	Full cost
Measure Incentive Costs	UCT	Prescriptive incentives	Prescriptive customer incentives + P4P contractor incentives
Admin/overhead costs	Both	Excluded	Excluded
Benefits (Electric + Natural Gas Savings)	Both	Prescriptive	Prescriptive + P4P

There were a few notable differences between the cost effectiveness approaches for traditional, prescriptive measures and the approach for P4P projects. For costs, the prescriptive program received the incremental costs for measure upgrades. In contrast, because incremental costs are unknown for the P4P non-anchor measures, Energy Trust used the full measure cost of the upgrade for P4P projects. This approach is consistent with Energy Trust's practice for characterizing measure costs for existing condition baseline measure types. This approach therefore aligns the P4P costs with the assumed savings, since claimed P4P savings are relative to an existing condition baseline.

Another related element of Energy Trust's cost effectiveness approach for P4P projects was to include only projects that achieved savings beyond the deemed values, and thus generated performance incentives, in the cost effectiveness

[19] We should note that while the P4P objective was to improve measure cost effectiveness, the P4P tests were only developed at the portfolio-level since savings are reported at a project and not measure-level.

calculation. The intent was to screen the cost effectiveness of the savings claimed, which were solely resultant from projects exceeding the deemed savings values.

Energy Trust developed the P4P cost effectiveness tests at the aggregator portfolio level. To compare the cost effectiveness of installations completed through the P4P pilot relative to the standard residential program, Apex compared the prescriptive deemed TRC values relative to the P4P values. To aggregate the measure level prescriptive TRCs for each aggregator, we linked the prescriptive measure level TRCs with the tracking database and then generated fuel-neutral (MMBtu) lifetime savings-weighted results. The results of the cost effectiveness comparison show that – at least directionally – the P4P pilot did not have a positive influence on the cost effectiveness of the portfolio of measure offerings.

Table 10. Prescriptive versus P4P TRC Results

Company	Prescriptive TRC	P4P TRC
Whole Home aggregator	1.21	0.51
HVAC aggregator	1.34	0.86

These cost effectiveness results are based on the aggregators' first portfolio of projects, completed in 2019. Aggregators reported they did not closely monitor the Recurve data portal, and only a few months of performance data were available at the time the portfolio closed. Thus, there was limited opportunity for aggregators to adapt to feedback on the performance of their projects.[20] Due to the pilot ending early, we are unable to determine whether additional performance feedback, including incentive payments, would have led to improved cost effectiveness in subsequent portfolios.

In addition, as described above, performance data indicate that deemed estimates may overestimate savings for ductless heat pumps. To the extent that this is the case, the prescriptive TRCs based on those savings values would also overestimate true values.

6. Conclusions and Recommendations

Apex draws the following conclusions and recommendations from this research.

- **Conclusion 1: Paying performance incentives 12 months after a portfolio closed created too long of a feedback loop to effectively influence the way aggregators approached the market.** The performance incentive was the pilot's most significant feedback mechanism to inform aggregators about

[20] Aggregators claimed that the highest savings projects were excluded due to fuel switching, or other disqualifying factors. While aggregator portfolios would be more cost effective if they could select for the highest-yield projects, doing so would conflict with the P4P criteria.

project performance and motivate them to take steps to increase savings. While aggregators had access to project performance data, without a clear sense of the potential financial gain, they had little motivation to devote time and effort to analyze those data and identify opportunities to increase savings. With performance incentives paid a year after a portfolio closed, aggregators had to wait 18 months or longer after their first P4P installations to receive feedback in the form of performance incentives. Had the pilot continued, aggregators would then have had to wait a further 18 months to experience the effects of any changes they made in their business practices.

- **Recommendation 1a: In future P4P offerings, Energy Trust should find ways to provide more timely performance payments.** For example, use of more frequent, interval billing data may allow for more accurate projection of savings, allowing a program to offer payments based on estimated savings on a quarterly basis. Energy Trust explored options for providing more timely payments in the P4P pilot but determined doing so was not possible given the data available.
- **Recommendation 1b: In future P4P offerings, Energy Trust and implementation contractor staff should be prepared to analyze performance data for aggregators** and present findings that could lead to actions to increase energy savings. Pilot staff presented this type of data to aggregators, and aggregators reported these presentations were valuable. If possible, it would be beneficial to provide these findings and analysis even before a performance period ends. These interim conversations could also provide opportunities to discuss individual projects performing particularly well or poorly.

〉 **Conclusion 2: Use of more frequent interval billing data could mitigate many, but not all, of the challenges the pilot faced.** In particular, use of interval data would help to overcome many of the data limitations that reduced the number of projects eligible for inclusion in P4P portfolios. With monthly data, loss of only a few data points was sufficient to disqualify a project. Daily or more frequent interval data would provide a much larger number of data points, making it easier to compensate for missing data in analysis. A larger number of data points could also allow for more granular savings analysis, and, as noted above, potentially allow the program to estimate savings and make performance payments sooner after a portfolio closes. Even with data issues resolved, however, the pilot's decision to use installation contractors as aggregators limits its potential scale. The participating aggregators were among the contractors submitting the highest volumes of projects to Energy Trust's residential program, and even these high-volume contractors struggled, in some cases, to meet the pilot's minimum requirements for portfolio size.

- **Recommendation 2: Energy Trust should ensure that future P4P offerings will have access to daily or more frequent interval energy usage data.**

〉 **Conclusion 3: Inaccuracies in deemed savings estimates have the potential to obscure the messages performance incentives send to aggregators.** The pilot's first portfolio – the only one to complete the performance period – fell short of its deemed estimates for electric savings. This underperformance was largely due to ductless heat pump installations and appeared to stem from two factors: shortcomings of the deemed estimates,[21] and incompatibility between deemed estimates based on a current practice baseline and P4P savings estimates based on an existing conditions baseline.[22] With these factors, which are outside of the aggregators' control, limiting their potential to earn performance incentives, aggregators would have little motivation to increase the energy savings their ductless heat pump projects generate.

- **Recommendation 3: Energy Trust should be prepared to address challenges related to deemed values in future P4P efforts that offer incentives for savings above deemed estimates.** For example, it may be necessary to remove projects based on certain measures from P4P portfolios if Energy Trust determines that deemed values are not a reasonable benchmark for their performance.

[21] Pilot staff noted that southern Oregon, where many of the pilot's heat pump installations took place, has a milder climate than other parts of the state and may require less heating energy consumption, and thus generate less heating energy savings, than the deemed values assume.

[22] A notable share of ductless heat pump (DHP) installations saw energy consumption increase in the performance period relative to the baseline period, suggesting that households that had not previously had space cooling equipment were using their DHPs to cool and/or households were using their DHPs to offset heating with wood or some other fuel. While Energy Trust's latest deemed estimate for DHPs assumes that a portion of participants would have installed an air conditioner had they not purchased a DHP, P4P savings estimates are based only on past usage, adjusted for weather and change in comparison group energy use. They do not consider alternative equipment options.