April 20, 2022

REPORT #E22-329

Heat Pump and Air Conditioner Efficiency Ratings: Why Metrics Matter

Prepared For NEEA:
Christopher Dymond,
Sr. Product Manager

Prepared by:
Bruce Harley, Principal

Bruce Harley Energy Consulting, LLC
665 Boulger Road
Stamford, VT 05352

Northwest Energy Efficiency Alliance
PHONE
503-688-5400
EMAIL
info@neea.org

Contents

Executive Summary ... i
Introduction—Heat Pump and Air Conditioner Metrics ... 1
Conventional Ratings and Variable-Speed Heat Pumps ... 2
Performance Variability in the Literature ... 6
Reasons for Variability in Performance ... 8
Impact on Programs ... 15
Variability and Savings Realization ... 15
Energy Savings Potential ... 20
Demand Savings Potential ... 21
EXP07 as a Preferred Metric ... 24
Sources ... 26
Presentation ... 29

Executive Summary

This white paper examines performance metrics (ratings) for residential and small commercial heat pumps and explores the potential benefits and impacts of improving those metrics. It presents evidence that existing metrics are not reliable, particularly for modern, variable-speed heat pumps and air conditioners that are expected to be more efficient but don't always live up to their ratings. Moreover, these systems' built-in firmware—the internally programmed set of operating instructions—can have a significant impact on their real-world performance, yet the firmware operation is explicitly excluded from current rating procedures. This paper presents the case that a much better rating metric would utilize a load-based testing procedure that fully characterizes heat pump performance under realistic operating conditions, including the systems' built-in firmware. The Canadian Standards Association (CSA) has developed such a test procedure (EXP07[1]) that clearly reveals significant differences in performance that are missed by the current heating seasonal performance factor (HSPF) and seasonal energy efficiency ratio (SEER) ratings. An improved metric could significantly increase savings, substantially improve heat pump utility program realization rates, provide better quality information to a wide range of stakeholders, and reduce wasted spending of program resources on under-performing models.

Performance metrics (ratings) that do not accurately represent in field performance not unique to heat pumps. Much of the observed performance of new computer controlled equipment such as clothes washers, clothes dryers, televisions and thermostats are not captured by older test procedures develop for previous generations of products. The lessons learned and evidence presented in this whitepaper should be considered broadly for other product categories.

[1] *CSA EXP07-19 Load-based and climate-specific testing and rating procedures for heat pumps and air conditioners.* 2019, CSA Group. Toronto, Ontario, Canada. www.csagroup.org/store/product/CSA EXP07:19/ An updated edition is expected to be published in early 2022.

Introduction—Heat Pump and Air Conditioner Metrics

A metric is “a standard of measurement” (as defined by Merriam-Webster). Metrics for appliances and mechanical equipment may be defined by federal statute, voluntary agreements, or other standards and most commonly include efficiency *ratings.* For air-source heat pumps and air conditioners, the most common standard ratings provide consumers, salespeople, contractors, engineers, utility programs, industry groups, and government simplified metrics for comparing one product’s performance to others and demonstrating minimum product efficiency as required by law. In the case of heat pumps and air conditioners, efficiency is critically important to energy efficiency and carbon-reduction strategies, and many programs promote higher-than-minimum products to support these strategies. Heat pumps and air conditioners are currently rated using the heating seasonal performance factor (HSPF) for heating and the seasonal energy efficiency ratio (SEER) for cooling.

This paper examines the relevance of these currently available metrics and explores an alternate metric defined by an alternate test procedure called EXP07, published by the Canadian Standards Association (CSA). EXP07 tests heat pumps using their built-in firmware and produces ratings specific to heating and cooling efficiency in eight North American climates. EXP07 is a load-based test that is likely to be more representative of real-world performance, as discussed below. The use of improved performance metrics based can provide higher utility program realization rates,[2] increase electricity, fuel, and demand savings, and decreased carbon emissions as compared with programs that utilize static load based ratings such as HSPF and SEER to qualify products and estimate savings.

---

[2] A realization rate is the ratio of measured savings to savings predicted by engineering estimates. When measured savings exceed predicted savings, the realization rate is >100%. When measured savings are less than predicted savings, the realization rate is <100%.

Conventional Ratings and Variable-Speed Heat Pumps

HSPF and SEER are, effectively, measurements of the equipment hardware; the test procedures that generate these ratings use fixed operating conditions to measure the heating or cooling energy output (capacity) and input (electricity) needed under those conditions. This defines the efficiency: at a given capacity, less input power equals higher efficiency. The test results are then projected across a range of operating conditions to estimate seasonal efficiency. This approach was originally developed for single-stage air conditioners and was later adapted to heat pumps. Single-stage air conditioners and heat pumps turn on and off as the thermostat responds to indoor temperature. However, modern, high efficiency heat pumps and air conditioners are far more complicated. They use variable-speed motors for the compressor and indoor and outdoor fans and other technologies to deliver higher heating and cooling efficiencies, especially at a lower output, and in some cases to provide extended performance for heating in very cold climates.

These significant, tangible benefits of variable-speed technologies can be realized, but they rely on complex algorithms, the built-in internal operating instructions, or “firmware” that controls all their moving parts in a coordinated fashion, much like the computer that controls a modern automobile. Field and laboratory testing have demonstrated that the operating firmware can significantly impact operating efficiency.

Although the test procedures used for HSPF and SEER have been adapted to accommodate variable-speed systems, the basic process is decades old and has not evolved to keep up with the current technology. In his analysis of air conditioners with high SEER ratings, Kavanaugh notes that “prediction of the actual energy or demand savings of modern cooling and heating equipment with a single indicator can be ‘seriously inaccurate’” (Kavanaugh 2002). In comparing an 18 SEER unit to a 10 SEER unit, Kavanaugh showed that the actual SEER improvement was much smaller than the advertised 80%—ranging between 45% and 76%—even while dehumidification capacity decreased, particularly during “low-speed, high efficiency operation.”

Most telling, during standard rating tests, the critical control firmware is intentionally disabled to simplify the process, and the units are tested in fixed test modes. These special test modes facilitate the test process, but they prevent variable-speed systems from being rated on the same scale as their conventional counterparts by allowing the compressor, indoor fan, and outdoor fan speeds to be selected by the manufacturer for each test condition. Bypassing the built-in control firmware ensures that large variations in their real-life operating efficiency can be systematically ignored during the rating process, which can have significant consequences. Proctor notes that "the SEER for a variable speed machine and the SEER for a single speed machine are arrived at through different tests and calculation methods. The result is that…the SEER and HSPF ratings of [variable-speed] machines appear overly enthusiastic compared to ratings of single speed machines. This makes direct comparison based on these metrics impossible" (Proctor 2016). "The SEER and HSPF test methods…lock the system at certain discrete speeds instead of operating the systems under the control programming that would be used in a field installed system. This produces VCHP [variable-capacity heat pump] system operation and performance [during the rating procedure] that is fundamentally different from the operation and performance that occurs when the system is installed in the field and controlled by the onboard programming" (Wilcox 2019).

Variable-speed systems are not being rated on the same scale as their conventional counterparts. Large variations in their real-life efficiency are systematically missed during the rating process, which can have significant consequences.

Further, although HSPF and SEER are both intended to represent seasonal efficiencies, they are applied to only one prototypical climate in the United States and Canada despite the wide range of actual climates. This simplification may be reasonable for single-speed units, though still highly misleading in more extreme climates such as Houston, Phoenix, Minneapolis, or Winnipeg (Fairey 2004). "If the equipment is located in a climate that differs significantly from the climate selected for publication of HSPF values, the seasonal efficiency of the heat pump may be very different from the stated rating. The differences between a specific climate and the rating conditions can result in either an improvement or a reduction in seasonal performance" (Francisco 2004). For single-speed systems, at least the *relative* efficiency of one unit compared with another is likely to hold across a range of climates. However, variable-speed products have much more diverse response profiles programmed into the built-in firmware that affect performance for varying temperatures and load conditions. As a result, a metric that is not climate-specific can introduce significant errors in climates that are colder or hotter than the ones on which HSPF and SEER are based.

Ratings are fundamentally useful to a wide range of market stakeholders when they reasonably represent systematic comparisons across similar types of products. If a consumer believes that one product is likely to use half the energy of another, based on their respective ratings, they may choose to pay more for that product; utility companies and government agencies may provide greater incentives; manufacturers may market specific products; and engineers, designers or contractors may specify them based on these superior metrics. When ratings reliably predict real-world performance, they provide foundational market direction for those that choose the equipment. A reliable heat pump rating should, at a minimum, be *directionally* correct, meaning that higher-rated products should consistently perform better than lower-rated products when used in similar applications. Ideally, a heat pump rating should allow reasonable calculation of annual energy use in a range of climates. Unfortunately, for the reasons noted above, in most cases of variable-speed heat pumps and air conditioners, this does not happen.

Ratings do have limits in terms of what can be expected from them. Ratings are *proxies* for in-field performance and are never perfect representations of field-installed efficiency because they are tested in laboratories under standardized conditions. On the one hand, ratings cannot capture the wide range of variation in user behavior, equipment application and design, installation practices, and weather that affect real-world performance. On the other hand, when ratings are produced in a manner that systematically misrepresents performance benefits, they misguide people (in all corners of the marketplace) into making decisions that result in the misapplication of selected systems, discomfort, and missed energy savings. This consequence can significantly dilute efforts to reduce energy use, cut carbon emissions and other pollution, manage peak energy demands, and reduce operating costs. Over time, under-performance at the portfolio level, driven by even larger shortfalls of individual products, may lead to an unfounded loss of confidence in the entire product class among programs, consumers, and other stakeholders.

Figure 1. HSPF and SEER Ranges—Federal Minimum to Maximum Available

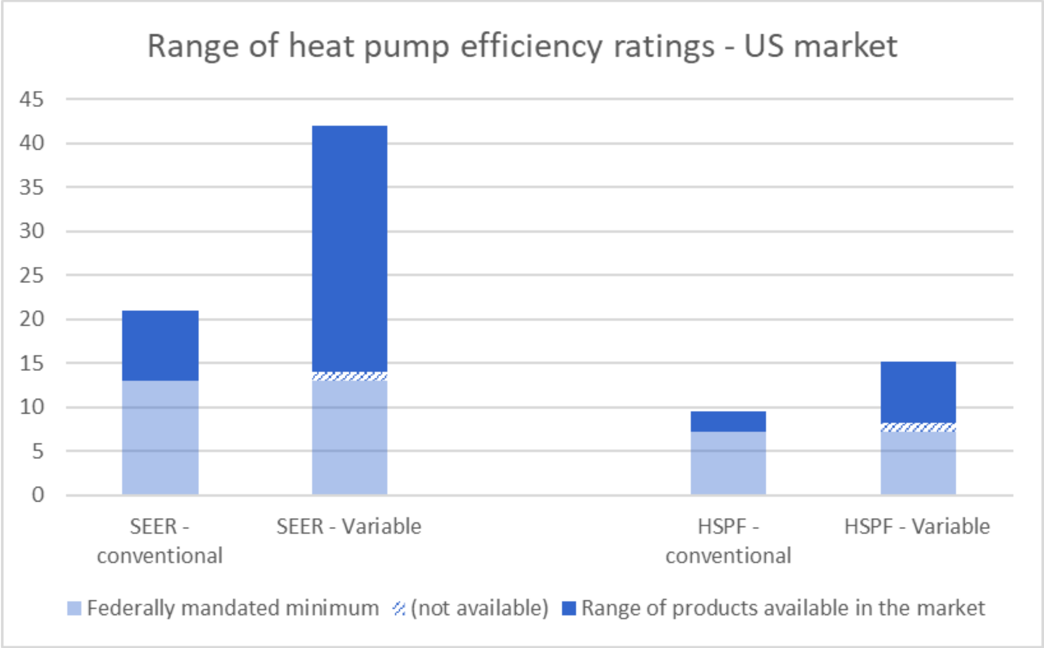


Figure 1[3] shows the range of SEER and HSPF ratings currently available from conventional (1- or 2-speed) heat pumps and variable-speed heat pumps. The current ratings indicate that the lowest-efficiency variable-speed units are slightly better than the lowest-efficiency conventional systems (which must meet federally-mandated minimums). The ratings suggest that for cooling, the best variable-speed units are three times as efficient as the minimum mandated efficiency and more than twice as efficient as the most efficient conventional units available. For heating, the ratings suggest that the best variable-speed units are over twice as efficient as the minimum required by regulation and 60% more efficient than the best conventional systems. However, field and laboratory testing reveal that systems installed in actual homes (or in simulated real-use environments) have not been nearly as efficient compared with code-minimum equipment in any climate as current ratings suggest they should.

[3] Source: AHRI certification directory, www.ahrinet.org/certification/directory

Performance Variability in the Literature

The literature provides numerous examples of disappointing results from program evaluations of heat pump savings and a very wide range of actual performance of variable-speed heat pumps in the field. For example, a 2012 study of 13 homes showed the range of annual heating savings among ductless heat pumps replacing electric heat in the main living area averaged 4,442 kWh per year but varied from 768 to 17,007 kWh—a factor of 22 (Baylon 2012). Among seven heat pumps monitored in the Northeast, measured heating efficiencies (coefficients of performance, or COPs)[4] ranged from 1.1 to 2.3, even though the HSPF ratings varied by only 16% (Williamson 2015).

Evaluation results of utility incentive programs for high-efficiency heat pumps have been relatively disappointing overall. Savings are typically estimated with simple equations using the installed SEER and/or HSPF compared with some assumed baseline efficiency and an assumed amount of usage. The Bonneville Power Administration's recent ductless heat pump program evaluation showed a 50% realization rate for heating savings of units replacing electric furnaces and 84% for units replacing electric baseboard units (Dorato 2018). A 2016 Massachusetts and Rhode Island impact evaluation of ductless variable-speed heat pumps found virtually no correlation between either HSPF or SEER and measured efficiency; although expected program savings were not reported for comparison, the full-load hours were found to be between 37% and 61% of those used in the statewide Technical Reference Manual (TRM[5]), suggesting savings realization rates in that range (Korn 2016). A 2017 impact evaluation in Vermont found a realization rate of 48% for heating (Walczyk 2017). A 2019 study for Efficiency Maine showed a realization rate for heating savings of 67% (West Hill 2019). Other impact

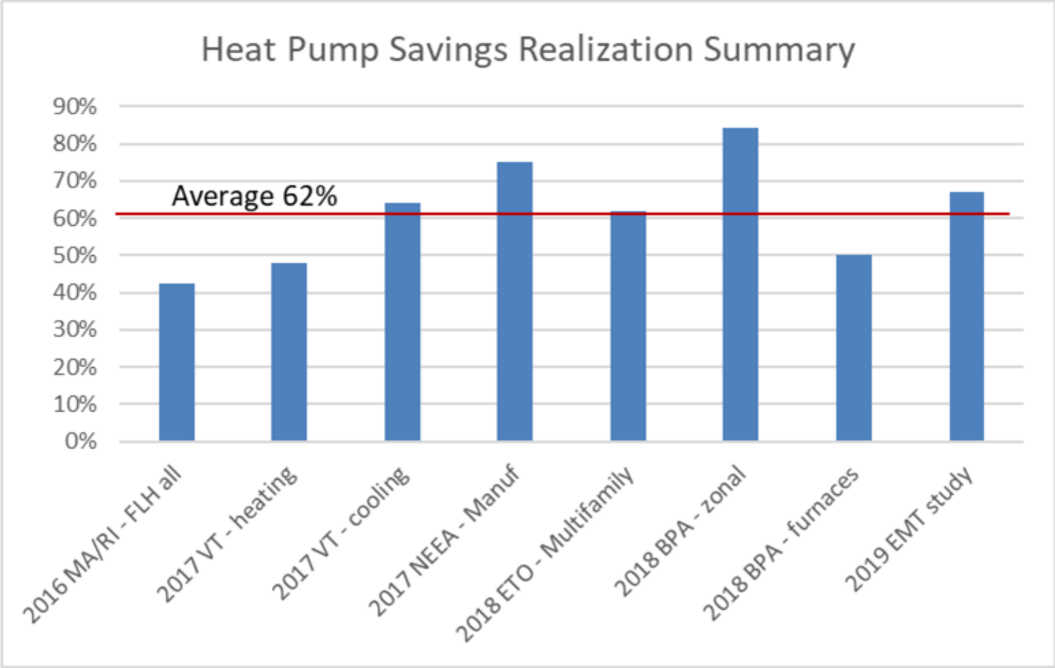
---

[4] The coefficient of performance (COP) of a heat pump (or air conditioner) is the ratio of the heating (or cooling) provided, divided by the electrical energy required. Higher COPs indicate higher efficiency, lower energy consumption and thus lower operating costs. Electric resistance heat has a COP of 1.0.

[5] "A Technical Reference Manual or Technical Resource Manual ("TRM") is a document containing agreed-upon methodologies for calculating electric and gas energy savings resulting from installations of energy efficiency measures and technologies." Source: https://mosaves.com/statewide-energy-efficiency-resources/missouri-technical-resource-manual/

studies have shown realization rates of 75% for NEEA (Hardman 2017) and 62% for Energy Trust of Oregon (Rubado 2018). Figure 2 shows a summary of these realization rates.

Figure 2. Realization Rates of Recent Heat Pump Program Evaluations



Reasons for Variability in Performance

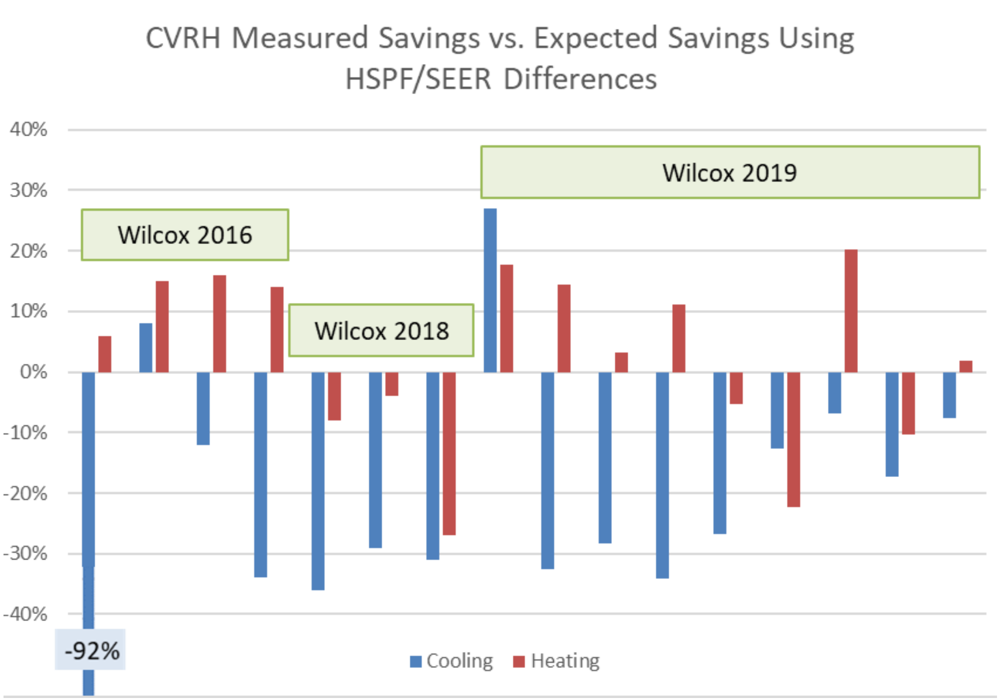
In addition to deviations in real efficiency from the ratings, many reasons can contribute to lower realized performance. Reasons include installation failures (e.g., duct leaks and improper refrigerant charge) and consumer behavior, which in some evaluations included far less use of the heat pumps than anticipated; such issues can negatively affect both savings and efficiency. While ratings cannot account for the variability of performance caused by in-field design and application choices, they should not cause additional performance uncertainty.

Current HP ratings can be as significant a contributor to product under performance and performance variability as the design and installation practices of the installing contractor.

The Central Value Research Home (CVRH) studies by Pacific Gas and Electric (PG&E) (Wilcox 2018, 2019) was able to avoid most of the uncertainties of installation and occupancy Over several years, this research has measured heat pump performance in three unoccupied homes, using a range of ducted and ductless systems, alternating weekly between variable-speed and conventional systems in each house. Different types of control settings, distribution systems, and other operating parameters were tested in various periods. The high-level results include a comparison between expected savings, calculated based on differences in SEER and HSPF, and the measured savings, for each pair of systems.

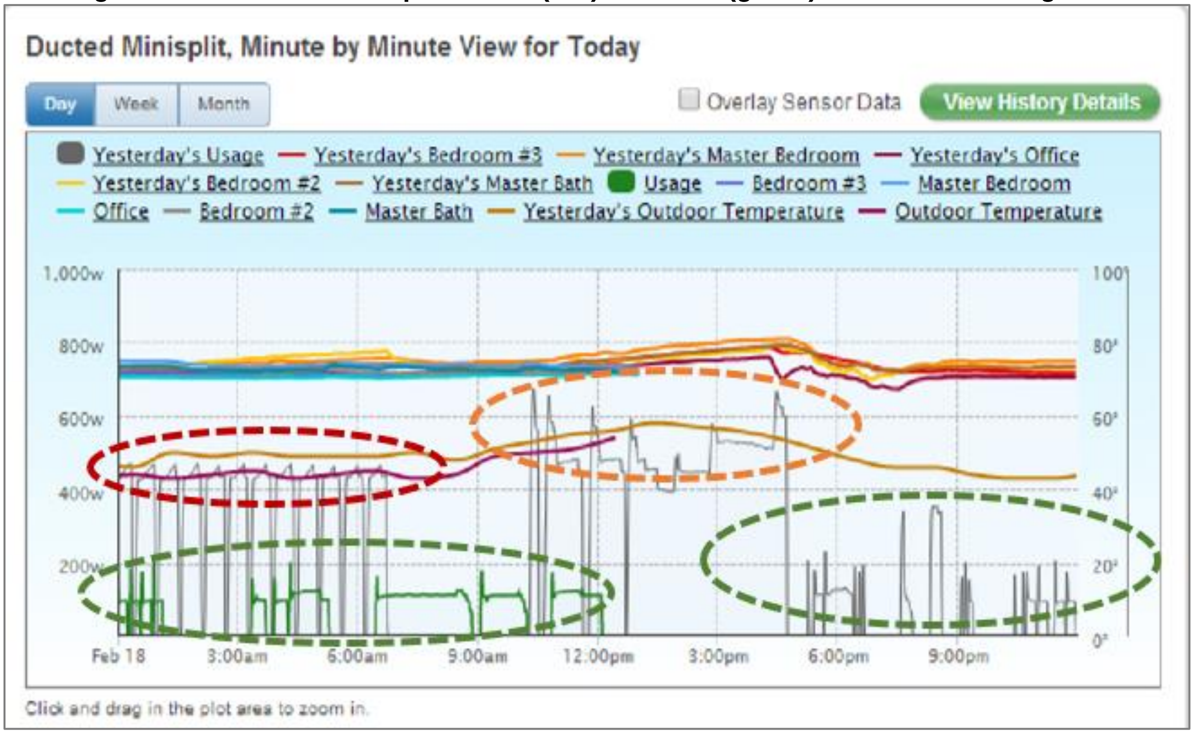
The range of savings of the individual heat pump tests is shown in Figure 3. The results demonstrate a substantial degree of in-field performance variation compared with the expectations calculated using conventional metrics. In many cases, the heating savings exceeded expectations, but the cooling savings fell short in all but two. The degree of variability is especially notable considering that the controlled installation and testing process largely eliminated many confounding uncertainties such as occupant behavior, improper installation of equipment, thermostat setting variability, and duct losses outside conditioned space often present in field studies.

Figure 3. Differential Between Measured Savings and Savings Calculated Using HSPF and SEER Differences, CVRH Projects



One explanation for the wide range of performance in variable-speed heat pumps generally appears to be the built-in operating firmware. As early as 2014, it was observed that a the controlling logic “firmware” could make a dramatic difference in performance in an actual installation. Firmware used by the heat pump affects how the system responds in the field to calls for heat, defrost, part load, preheat and other events that are not present in current full load test procedures. Figure 4 provides a specific example from Oregon, showing that a firmware update resulted in roughly a 50% decrease in energy consumption under low load conditions.

Figure 4. Electric Power Input before (red) and after (green) a Firmware Change



The example in Figure 4 shows the electricity measurements made on a 9,000 Btu/h ductless system on February 17–18, 2014. The outdoor temperature (on the 17th in orange, on the 18th in magenta) is fairly consistent both days, at about 45–50°F overnight and near 60°F at midday. The grey line shows power consumption from February 17, the day the firmware was changed. The green line shows electricity use on February 18 ("today" in the graphic). The compressor power during the morning of the 17th cycled between about 475 Watts and off (gray trace highlighted with a red dotted line). The unit was powered down at 7:30 AM, and the firmware was updated between about 7:30 AM and 10:00 AM. When it is turned back on, the power input increases and the unit runs quite a bit for the first few hours (10:00 AM to 5:00 PM, orange dotted line). Presumably, it was catching up from the heat being turned off for several hours and adapting the brand-new firmware to its operating parameters.

Once the system restabilizes and continues to operate under the new firmware (circled by the green dotted line, right side, from 6:00 PM to 11:59 PM on the 17th, and green dotted line, left side, from 12:00 AM to 12:00 PM on the 18th), the unit modulates at around 110 Watts, with longer run times and some periods of shutting off completely. Even though the outdoor temperature on the 18th is slightly *lower,* indicating a slightly higher heating load, the energy consumption of the same heat pump hardware is about half what it was with the old firmware. Subsequent long-term monitoring of the unit showed that the weather-adjusted *annual* heating consumption dropped by about 35% after the firmware change (Stephens 2016).

The CVRH project reports have touched on the same issues: “VCHP [variable-capacity heat pump] system controls are complex, with many settings adjustments that are not well documented. Settings adjustments may produce unexpected results. …The default control settings are not necessarily optimized for energy performance. Control settings adjustments reduced energy use for both of the VCHP systems tested… However, it is unlikely that installers would be able to identify the optimal settings…many of the control settings on the VCHP units tested lacked adequate documentation to inform installers how they affect system operation” (Wilcox 2019). Some of these control settings adjustments reduced heating energy use by 15–23% and cooling energy use by 6–7%.

A predominant cause for the wide range of performance appears to be the operating firmware… the poor performance was entirely missed by the standard metrics.

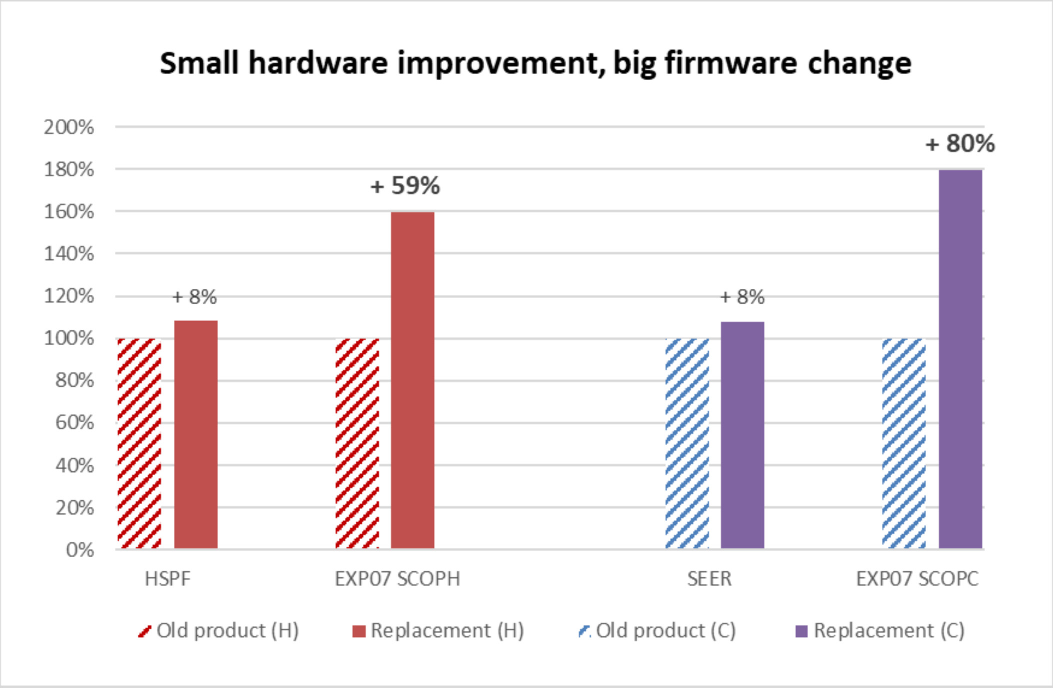
Load-based laboratory testing using variable-speed heat pumps (CSA’s EXP07 test procedure) has also revealed a link between operating firmware and performance variability. Underwriters Laboratory (UL) conducted lab testing of 13 variable speed heat pumps using dynamic laboratory testing that uses the system’s built-in controls, concluding that “the apparent explanation for the wide variability in response and performance is the embedded control algorithms” (Harley 2020).

Subsequent testing after the report was written revealed an even more dramatic example of the impact of the firmware. Figure 5 illustrating the differences between two generations of the same product line. The testing was done on two different model-year heat pumps. These units were built by the same company, from the same product line, with the same capacity and the same refrigerant. The HSPF and SEER data were generated using the static full load test method of AHRI 210/240, and the EXP07 data were generated using the load-based test procedure of CSA EXP07:2019.

The “old product” columns (cross-hatched) refer to a unit built in 2018, and the “replacement” columns (solid) refer to a unit built in 2020. The replacement product is, for all practical purposes, the “new” version of the old product, with some mechanical changes and new built-in firmware.

The two machines in Figure 5 were mechanically very similar as indicated by only an 8% change in the HSPF and SEER ratings. Testing these products with the load based test procedure (EXP07) revealed the impact of the firmware changes, as the resulting seasonal efficiency of the new unit improved by 59% in heating and 80% in cooling. That equates to 37% less seasonal energy use for heating and 44% less for cooling – far more than can be attributed to mechanical changes reflected by the SEER and HSPF values. This is a clear example how the method of test can have a profound impact on efficiency metrics.

Figure 5. Improvement in Heating and Cooling Efficiency of Updated Product, as Measured by Standard Metric (HSPF/SEER) and Load-Based Metric (EXP07 $SCOP_H$ and $SCOP_C$) (normalized)



While the 8% increases in HSPF and SEER in the product update indicate that the hardware components *were* incrementally improved, the embedded firmware explains the much larger change in performance shown in the load-based test. To illustrate the dramatic impact of the built-in operating instructions, Figure 6 shows one test of the old product (NEEA5) using EXP07. The unit cycles on and off repeatedly to maintain the target indoor comfort condition, and in each cycle, the power swings rapidly up and down before settling. The cycling occurs because it begins with too high a cooling output and then has to shut off to avoid over-cooling the room; each cycle wastes some energy. Figure 7 shows the updated product (NEEA10) under the same test condition, working in the way variable-speed systems are intended to work: it ramps up a bit, then reduces the compressor power until it finds a steady (and much lower) cooling output, that maintains the room temperature at the desired setting. These dramatic differences in behavior between the two units occurred at almost every test condition in both heating and cooling modes.

Figure 6. Original system behavior shows the significant power consumption and high-power cycling. Efficiency for this test period is a COP of 2.73.

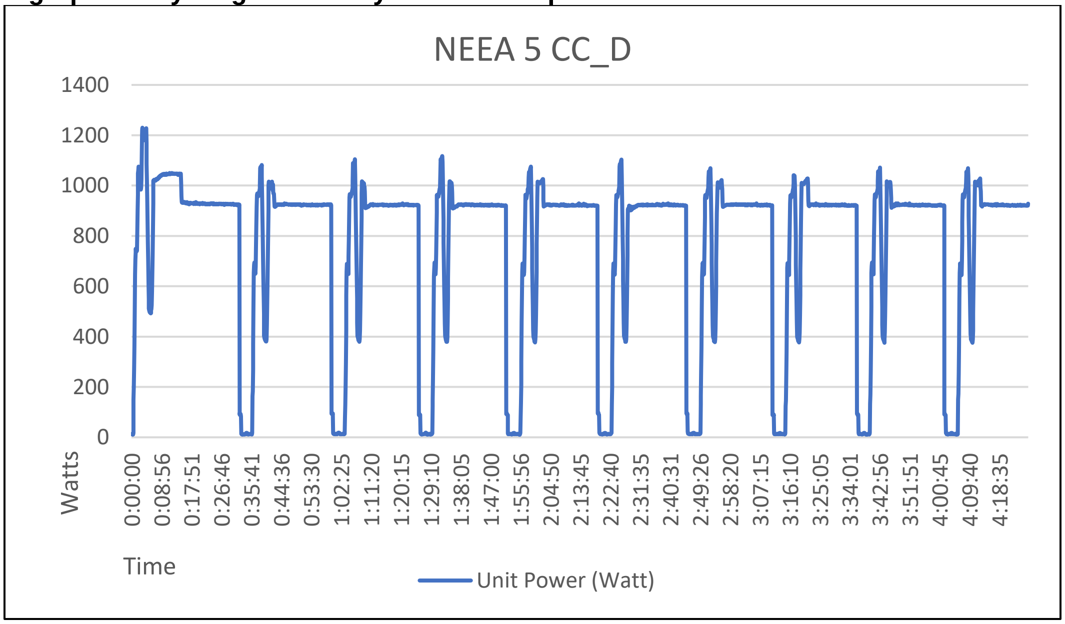
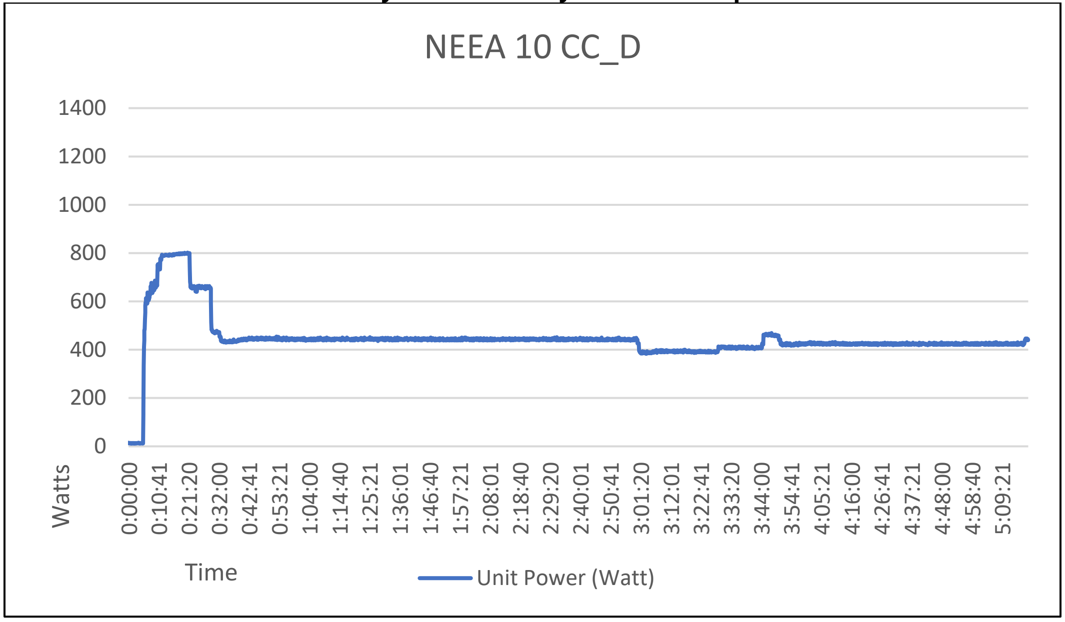


Figure 7. Updated system behavior shows much better modulation. The unit settles at about 450W and does not cycle. Efficiency for this test period is a COP of 4.36.



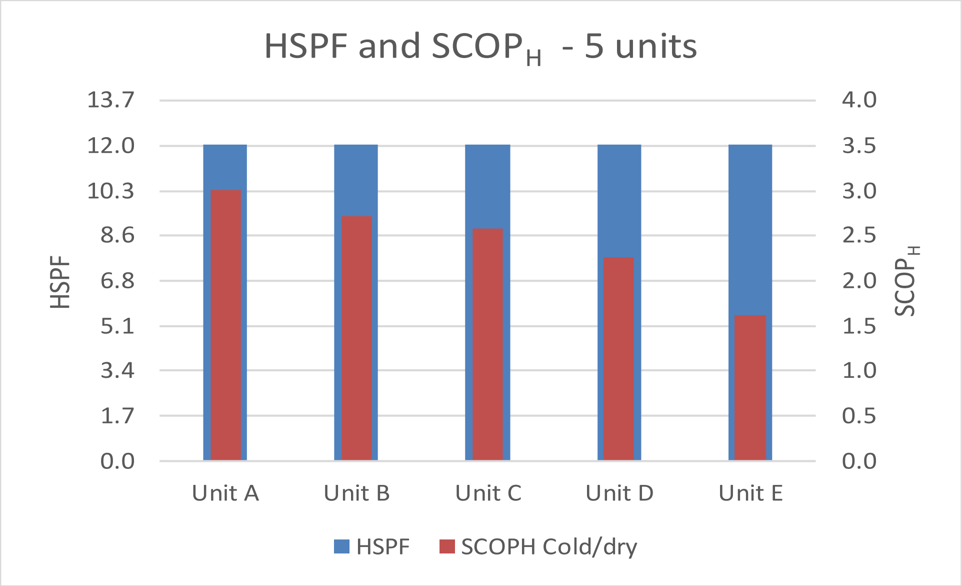
Notably, the manufacturer was made aware of the very poor test results of the first system, caused by high-power short cycling during most operating conditions, as revealed in the first round of testing using EXP07. The manufacturer endeavored to correct the problem in the new product, apparently with a solid degree of success. Therefore, the vast majority of the improvement was clearly due to the updated firmware, and the lab testing process confirmed that the short-cycling was largely curtailed in the update. However, the standard metrics of HPSF and SEER entirely missed the poor performance of the original design and the degree of improvement.

Impact on Programs

Variability and Savings Realization

The results of load-based lab testing can be framed in terms of the potential impact on programs, using two examples. Figure 8 shows five units, each with an HSPF rating of 12. The blue columns show HSPF, and the red columns show the results of the load-based testing in the Cold/Dry climate, which is close to the DOE Region IV climate used for HSPF[6].

Figure 8. Comparison $SCOP_H$ Values of Five Units Tested Using EXP07 to Their Rated HSPF Values



If 12 HSPF were used as the criterion for rebate eligibility in a cold climate, all five would qualify; however, but the resulting utility program realization rate would be 73% due to the overstatement of efficiency embedded in the HSPF ratings for all five units. Units D and E would receive incentives but assuming the load based testing was accurate, they would lower performance than current federal minimum standards require. This miss-alignment between

[6] Both the seasonal COP value for heating ($SCOP_H$), and HSPF are climate-specific bin-hour weighted ratings. The HSPF and $SCOP_H$ are aligned to account for the conversion between the units of HSPF (Btu/kWh) and SCOP (kWh/kWh).

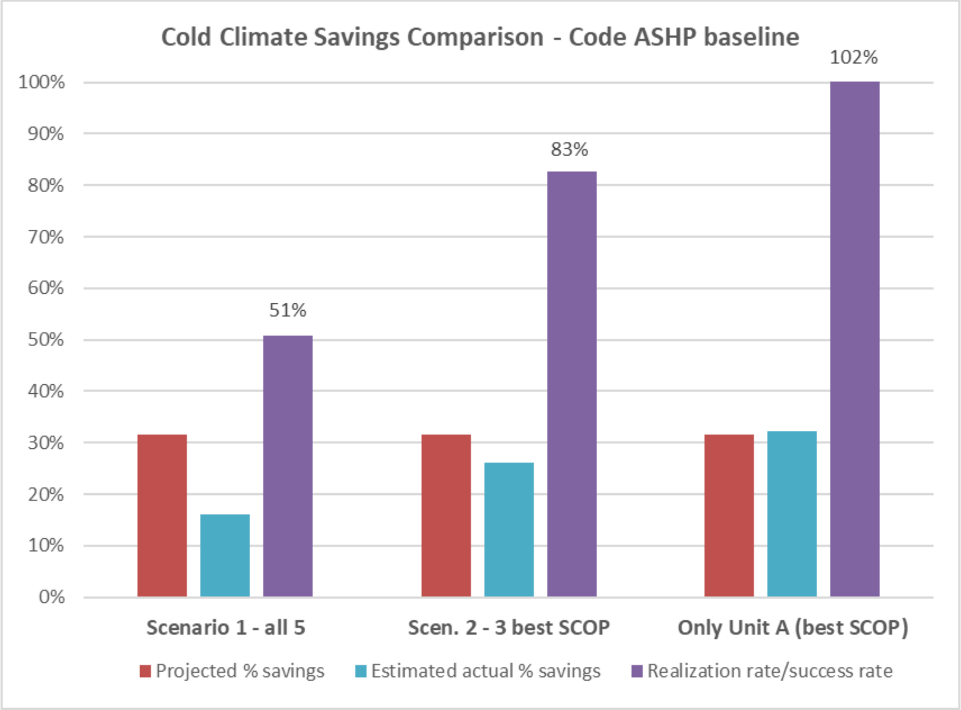
how a system tests in the laboratory and how it performs under its own control logic would burden a utility incentive program with systems do not generate actual savings. The result is the program may fail to achieve its intended cost effectiveness, placing the program's cost recovery at risk.

It is important to note that the comparison in Figure 8 is meant to illustrate the magnitude of program scenarios using load-based rather than fixed-test performance metrics. Illustration is based on a the assumption made is that a load based test procedure (such as EXP07) would accurately represent in-field performance. Figure 8 is based on a limited sample of convenience rather than a product mix from actual program(s). Current research is underway to validate if in fact the load based test procedure approach does represent in-field performance[7].

Figure 9 shows how variability in performance relative to the HSPF rating can lead to unanticipated utility program consequences. An HSPF of 12.0 suggests that each unit will save 32% relative to the federal minimum standard 8.2 HSPF. Figure 9 shows both the calculated savings (based on EXP07 test results) and realization rates of three scenarios: all five units (averaged), as well as two alternate groupings of units. The red columns represent the assumed savings using HSPF, compared with the code baseline (all scenarios show 32% because the HSPF is identical). The blue columns show the theoretical "actual" savings for each scenario using the EXP07 results. The purple columns represent the program-wide realization rate for each scenario. For example, if the engineering savings estimate per heat pump were 1,000 kWh per year and 10,000 units were incentivized, the shortfall in savings from installing all five heat pumps in equal numbers (first scenario) would be almost half of the estimated 10M kWh: ((1.00 – 0.51) x 1,000 x 10,000) = 4.9M kWh/yr, leaving only 5.1M kWh/yr in realized savings. However, if units D and E were excluded (second scenario), the realized savings would be 8.3M kWh/yr, yielding an additional 3.2M kWh, though still 17% short of the program's estimated savings.

[7] The "Rating Rating Representativeness Project" is a 2-year multi-organization funded project designed to generate both field and lab data that can evaluate how well test procedures represent in-field performance. The project is being managed by Northeast Energy Efficiency Partnerships (NEEP) and is expected to be completed in late 2023.

Figure 9. Heating Savings Projections, Estimated Actual Savings, & Resulting Realization Rates for Five HSPF12 Units in Three Scenarios—Chicago Climate



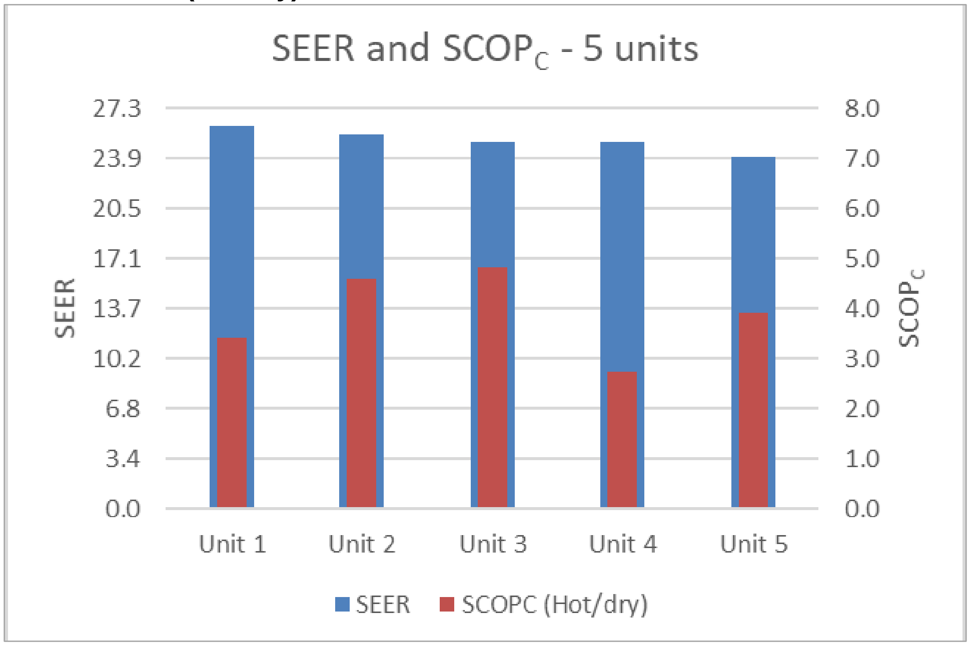
The “Only Unit A (best SCOP)” scenario shows what would happen if the utility program excluded all but the best performer in this group; savings would actually exceed the program assumptions by 2%. Unfortunately, the current metric of HSPF offers no way to filter out the poor performers, because they all share the same HSPF. If the utility program had access to a more representative metric such as $SCOP_H$, it could set the program eligibility threshold at 2.5 or 3 to boost the savings to the levels shown by the other scenarios, respectively.

These realization rates are not out of line with actual program evaluation results, as shown in Figure 2. This is not to suggest that all evaluation shortfalls result solely from an unsuitable rating metric, but that the nature of the HSPF rating inherently misses some very significant contributors to overall field performance. The Cadeo study (Baker 2019) proposes that when a test procedure is updated to characterize current equipment better, significant new energy savings can result from “better characterizing the energy use of the product.” The savings stem from eliminating the less-reliable ratings that mischaracterize units as highly efficient, and

Cadeo's recommendation to NEEA sets heat pump ratings (specifically EXP07) as a "high" priority characterized by a "high" magnitude of new energy savings.

Similar examples for cooling in a hot/dry climate (Sacramento) are shown in Figure 10 and Figure 11. Figure 10 shows SEER and the hot/dry climate $SCOP_C$ (similar to Sacramento CA) for five tested units. While the SEER values of the units tested are not identical, they fall within a narrow range from 24 to 26.1. The $SCOP_C$ values range more widely and exhibit a different rank order from the SEER ratings.

Figure 10. $SCOP_C$ (Hot/Dry) and SEER for Five Tested Units



As with the heating example illustrated in Figure 8, projected savings and realization rates compared with code minimum SEER are shown in Figure 11 (first scenario group) for the average of all five units. Conventionally "projected % savings" (red columns) for these different scenario groupings would vary only between 61% and 62% because the SEER ratings are fairly close. The "estimated actual % savings" (blue columns), if based on SCOP, would average 26% for the five and range between 16% and 40% for the other groupings.

Figure 2. Cooling Savings Projections, Estimated Actual Savings, Resulting Realization Rates for Five High- SEER Units in Four Scenarios—Sacramento

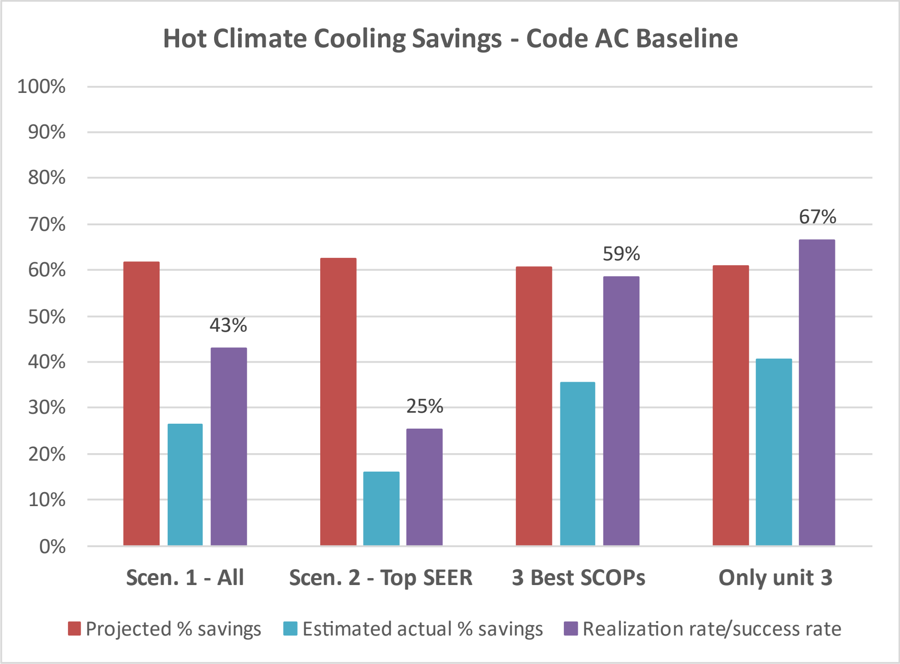
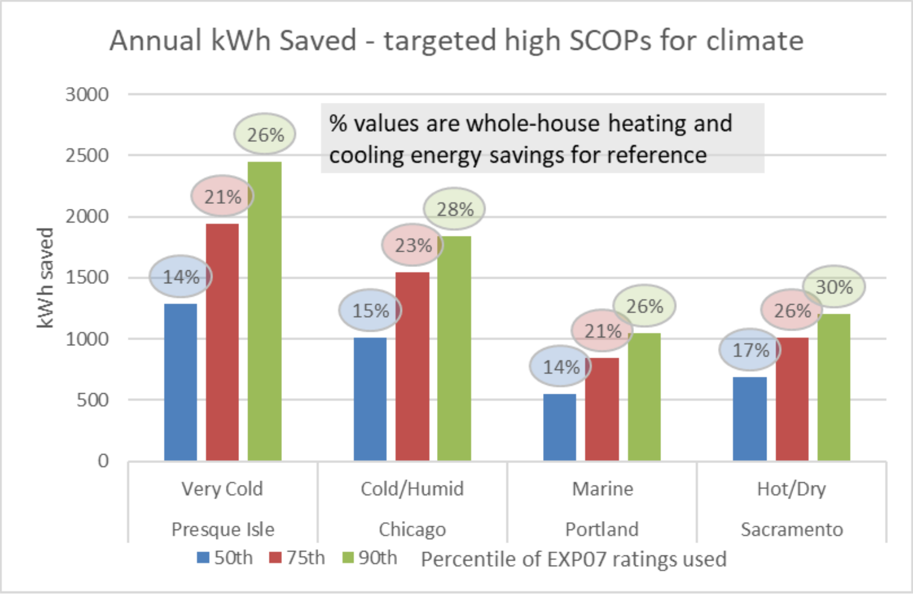


Figure 3. Estimated Savings Potential, kWh/year per House, for Whole-House Heat Pumps Selected for Program Qualification Using Heating and Cooling SCOP (EXP07) vs. Standard HSPF/SEER (4 climates)



The purple columns in Figure 11 show the realization rates of several program scenarios. The first scenario shows a realization rate of 43%, assuming an equal distribution in the program of all five units. The next scenario shows the possible consequence if the program had targeted only the best-performing unit according to SEER; the realization rate would drop to only 25%, exactly the opposite of the intent. In the third scenario, focusing incentives only on the three units with the highest values of $SCOP_C$ would improve the realization rate to 59%.

The key point of this analysis is to show that targeting the best-performing units based on EXP07 (the three best, or Unit 3 alone in the right-hand scenario) can deliver the most savings with a realization rate of up to 67%. As with heating, the signal provided by SEER appears to be counterproductive, and the lack of access to a better metric makes it nearly impossible for programs to optimize savings.

These examples are not meant to represent entire programs or available product offerings; rather, they use a sample of convenience, chosen from lab tested units to EXP07 between Nov 2018 and Nov 2019 (Harley 2020), plus six more subsequently tested by the same lab. They are all variable-speed systems with high HSPF and SEER ratings typical of leading incentive program offerings. Most are small (1- to 1.5-ton) ductless, and a few are larger ducted units. The two groups in these examples were chosen for having the same or very similar HSPF and SEER ratings, respectively, to facilitate the analysis using cohorts of tested units that would plausibly qualify for participation in a fictional incentive program.

A significant magnitude of additional savings and carbon reduction could be achieved by obtaining and utilizing a better performance metric.

Energy Savings Potential

As the previous examples show, energy savings estimates are impacted by the accuracy of the rating used. A rating metric that is more representative of in-field performance can significantly improve energy savings achieved by incentive programs or other promotions of high-efficiency heat pumps and air conditioners. Figure 12 shows more generally, with a larger selection of heat pump models, how much a load-based and climate-specific rating could impact savings estimates and influence where program performance requirements are set.

Figure 12 shows the analysis results based on 17 units tested by UL using EXP07. The estimates show potential savings from implementing program requirements based on EXP07 thresholds rather than HSPF and SEER thresholds and how this can vary depending on climate. The heating minimum HSPF of the group is 9, and the average is 11.9. The minimum SEER is 17, and the average is 22.9. The baseline is calculated as if all 17 units were installed in equal numbers, using the expected energy consumption based on the EXP07 results for the respective climates.

The three columns in each climate represent a selection of only those units that show higher performance based on EXP07, using the 50th, 75th, and 90th percentiles of the EXP07 ratings from the group, respectively.[8] The analysis included estimated heating and cooling consumption for the full sample and the selected improved groupings to arrive at an estimated kWh savings in each climate relative to using the full sample. The largest absolute kWh savings occur in the coldest climate (Very Cold) and the least (about half of the kWh) in the mildest climate (Marine). Notably, the percent of whole-house heating and cooling savings is similar—between 14% and 30%, depending on the degree of selectivity used, across all four climates. By using the three percentile values shown, with ranges of estimated annual savings from 600–1,100 kWh/yr in the mildest climate up to 1,300–2,400 kWh/yr in the coldest climate, these estimates demonstrate a significant level of additional savings and carbon reductions that could be achieved through the use of a better performance metric. As with the other examples above, these results are based on the premise that EXP07 results are representative and, using a sample of convenience, are illustrative of the range of potential benefits of establishing program participation on a better metric.

Demand Savings Potential

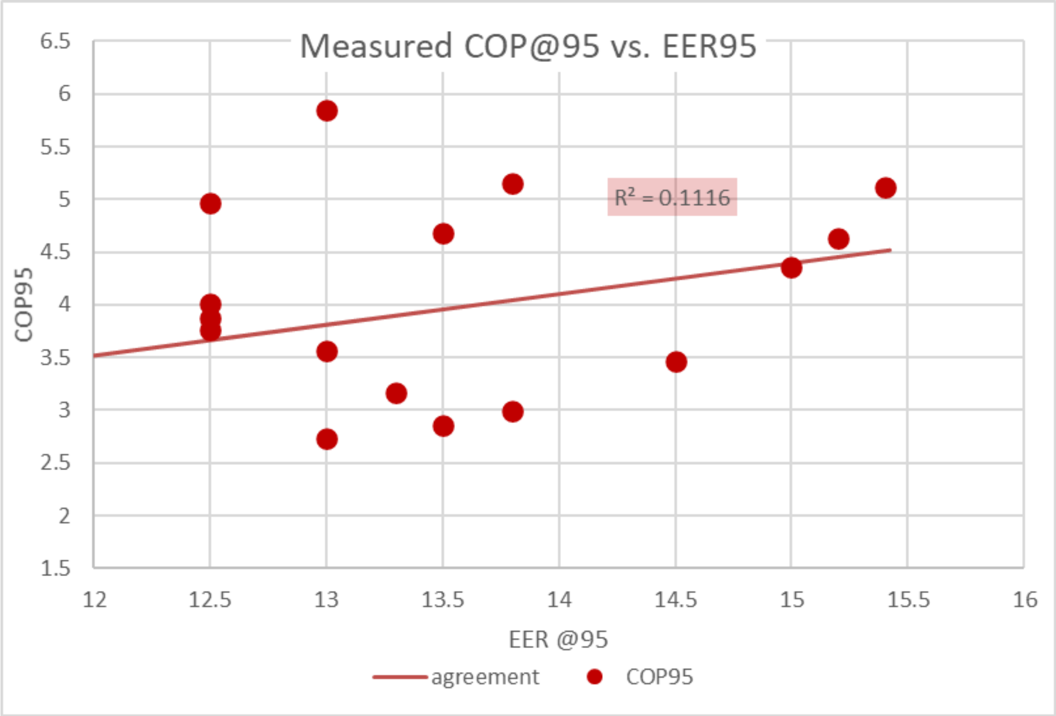
The Energy Efficiency Ratio (EER) rating at 95°F[9] is often used to estimate demand savings impacts of installed equipment compared with a baseline, such as code minimum equipment. Compared to EXP07 load-based tests, demand savings estimates based on static test

[8] Note: The selection by percentile of the highest-performing units varies by climate, because they are based on the EXP07 SCOPs individually appropriate for each respective climate.

[9] Like SEER and HPSF, EER at 95°F is based on AHRI 210/240. It is a steady-state cooling efficiency measured at a single point, rather than a seasonal efficiency.

conditions such as EER at 95°F also appear to be inaccurate. The EXP07 procedure includes a test at 95°F under realistic loading conditions, so the test results should be comparable; however, the rated EER has little bearing on the performance under the load-based test. Figure 13 shows the range of COP at 95°F measured using EXP07, compared with the rated EER at 95°F.

Figure 13. COP Measured at 95°F (dry) vs. EER at 95°F



For cooling, the group of tested heat pumps using the conventional EER at 95°F can be compared to the demand impact of selecting units that showed a higher measured COP at 95°F using the load-based test. Using analysis similar to the previous kWh savings examples, choosing the top 50% of these units based on COP at 95°F from EXP07 would result in additional demand savings of about 0.15 kW per ton, or about 0.45 kW with a typical 3-ton system (as compared to the entire group of 17 units tested with EXP07:2019[10]). By comparison, selecting only the top 50% of the units based on EER at 95°F would result in virtually no peak demand savings compared with the entire group—only about 0.01 kW (10 W) per ton. Based on the EXP07 lab tests, EER does not appear to accurately predict the efficiency of variable-speed units operating under their own controls. For climates with a higher temperature at peak cooling demand, the same analysis selecting half the units with load-based COP at 104°F instead, would yield a slightly more modest demand savings of 0.13 kW/ton.

For heating (winter peak demand), a parallel analysis would be more complicated due to the absence of rated low-temperature performance metrics in the current DOE rating process. The revised DOE test procedures[11] include an *optional* test at 5°F, but such data are not currently available. Utilities in cold climates concerned about winter peak could still perform a similar analysis on the heating efficiencies at 5°F. Peak winter demand however is more typically dependent on the lockout temperature above which auxillary heating from electric resistance heating or fossil fuel is prevented. Peak demand can be significantly reduced if accurate capacity data is provided from a better metric.

The benefits testing high-performance products with a load based test procedure increases product confidence, and can lead to better choices and improved results for all stakeholders.

---

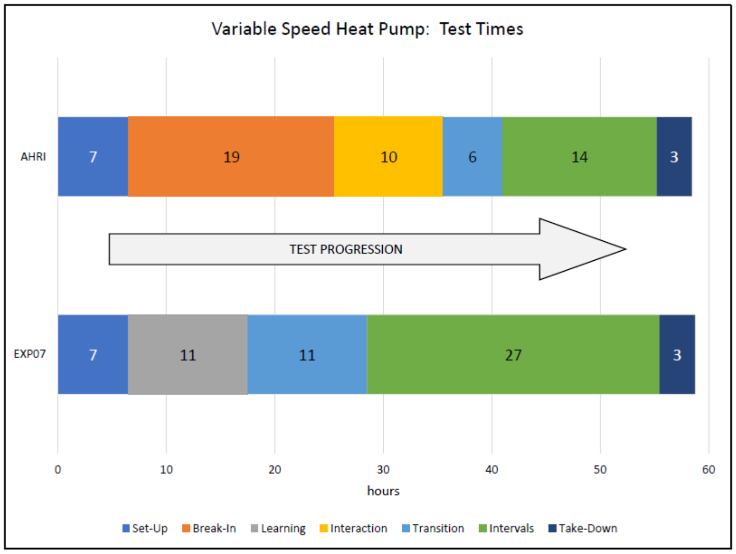
[10] A mixture of 17 ductless and ducted split systems were tested for NEEA and NRCan at the UL's Plano TX facility during 2019-2021.

[11] CFR Title 10, Chapter II, Subchapter D, Part 430, Subpart B, Appendix M1 http://www.ecfr.gov/current/title-10/chapter-II/subchapter-D/part-430/subpart-B/appendix-Appendix%20M1%20to%20Subpart%20B%20of%20Part%20430 Commonly referred to as "Appendix M1," this new procedure becomes effective in January 2023.

EXP07 as a Preferred Metric

Currently, EXP07 is a voluntary test intended for manufacturers of premium products with high efficiencies to qualify their products for incentive programs, achieve energy savings, decarbonization goals, or compliance with performance-based building codes. Some have suggested that load-based testing creates a much higher testing burden than the standard DOE-mandated tests. Analysis by the Underwriters Laboratory (UL) test facility that has done extensive testing using EXP07 indicates that the overall test time is comparable to the existing test regime. Figure 14 shows the timeline comparison of EXP07 testing compared with AHRI 210/240, the standard test method.

Figure 4. Comparison of Test Times, AHRI 210/240-2017 and EXP07-2019



- **Set-Up**: Similar times between AHRI and EXP07. Both procedures follow ASHRAE 37 requirements.
- **Break-In**: Manufacturer-specified for AHRI. Most require 16–20-hour break-in periods. Not required for EXP07
- **Learning**: Required cycles (cooling Dry and Heating Continental) for EXP07. Not required for AHRI.
- **Interaction**: Required technician interaction with the UUT to set compressor and fan speeds. No interaction between the UUT and the technician is required for EXP07.
- **Transition**: Getting the psychrometric chambers on conditions prior to the conducting of learning cycles and test intervals. Transition times from one interval to the next are equal between AHRI and EXP07 testing. The total transition times are based on the number of intervals.
- **Intervals**: 5 cooling and 6 heating intervals are conducted for AHRI and EXP07. The total time associated with each method is based on actual averages compiled from test data
- **Take-Down**: Similar times between AHRI and EXP07

The EXP07 test is currently being revised based on a wide range of technical feedback. The revision, to be published in 2022, will likely result in somewhat shorter test times than the 2019 edition.

The benefits to manufacturers of voluntarily testing their high-performance products to EXP07 are many, including:

- Increased potential for participation in utility and other incentive programs,
- Far better differentiation of their best performers from their competitors' products,
- Increased sales of the best-performing systems in appropriate climates,
- Improvements to their products based on feedback about the influence of controls on the equipment's performance,
- Improved metrics that can be effectively used by consumers, designers, program managers and other stakeholders, and
- Higher savings, reduced demand, and higher realization rates for programs.

The wider range of realistic test conditions used in EXP07 also provides a far more complete performance "map" of the heat pump in heating and cooling modes than the current DOE test procedure. In addition to the eight standard climate ratings, the performance map provides a much more useful tool for modeling, building code compliance, and other analysis (program planning). It provides much greater confidence in results than is currently available using simple single-value metrics such as HSPF and SEER, which cannot translate into more detailed models for the wide range of equipment in the marketplace. That increased confidence can lead to better choices and improved results for all stakeholders.

As of January 2022, a group of utilities and other interested parties has begun a field study to facilitate direct comparisons of equipment performance in actual homes compared with EXP07 and other rating methods. This (along with other studies currently underway) will help determine the extent to which EXP07 is indeed superior to existing metrics in representing in-field performance and will increase confidence in load-based testing as a high-quality metric for residential air conditioners and heat pumps.

Sources

Baker, Nate, Jonah Hessels and Sarah Widder (Cadeo), 2019. *Evaluation of Energy Savings from Test Procedures & Specifications.* NEEA, Development of Standards and Ratings for Commercial and Industrial Equipment (Contract No. 51033) – Task 5. June 2019.

Baylon, David, Ben Larson, Poppy Storm, and Kevin Geraghty, 2012. *Ductless Heat Pump Impact & Process Evaluation: Field Metering Report.* NEEA, E12-237. http://neea.org/resources/ductless-heat-pump-impact-process-evaluation-field-metering-report

Dorato, Shannon, Pace Goodman, Marilla Yaggie, and Ariel Esposito (Navigant Consulting, Inc), 2018. *Impact Evaluation of Ductless Heat Pumps and Prescriptive Duct Sealing.* BPA. September 2018. www.bpa.gov/EE/NewsEvents/presentations/Documents/180911_Impact_Evaluation_of_Ductless_Heat_Pumps_and_Prescriptive_Duct_Sealing.pdf

Francisco, Paul, David Baylon, Bob Davis and Larry Palmiter, 2004. *Heat Pump System Performance in Northern Climates*. American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc. Publication AN-04-5-3. 2004. http://ecotope-publications-database.ecotope.com/2004_005_HeatPumpSystemPerformance.pdf

Hardman, Trent, Alex Chamberlain, Matei Perussi, Cynthia Kan and Karen Horkitz, 2017. *Existing Manufactured Homes Heat Pump Pilot Evaluation Final Report*. Energy Trust of Oregon. November 7, 2017. www.energytrust.org/wp-content/uploads/2017/12/XMH-Heat-Pump-Pilot-Eval-Final-Report-wSR.pdf

Kavanaugh, Steve P, 2002. *Limitations of SEER for Measuring Efficiency.* ASHRAE Journal, July 2002. www.techstreet.com/standards/limitations-of-seer-for-measuring-efficiency?product_id=1719622

Korn, Dave, John Walczyk, Ari Jackson, Andrew Machado, John Kongoletos, and Eric Pfann (Cadmus), 2016. *Ductless Mini-Split Heat Pump Impact Evaluation*. Electric and Gas Program Administrators of Massachusetts and Rhode Island. December 30, 2016. www.ripuc.ri.gov/eventsactions/docket/4755-TRM-DMSHP Evaluation Report 12-30-2016.pdf

Fairey, Philip, Danny Parker, Bruce Wilcox and Mathew Lombardi, 2004. *Climate Impacts on Heating Seasonal Performance Factor (HSPF) and Seasonal Energy Efficiency Ratio (SEER) for Air Source Heat Pumps. ASHRAE Transactions*, American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc., Atlanta, GA, June 2004. www.fsec.ucf.edu/en/publications/html/FSEC-PF-413-04/

Harley, Bruce, 2020. *EXP07:19 Load-based and Climate-Specific Testing and Rating Procedures for Heat Pumps and Air Conditioners: Interim Lab Testing and Rating Results.* NEEA Report #E20-314. July 7, 2020. http://neea.org/resources/exp0719-load-based-and-climate-specific-testing-and-rating-procedures-for-heat-pumps-and-air-conditioners

Proctor, John, 2016. *The Next New Thing -- Is it really that good?* 2015 International Energy Program Evaluation Conference, Long Beach. www.iepec.org/wp-content/uploads/2015/papers/060.pdf

Rubado, Dan, 2018. *Utility Billing Analysis of 2013-2014 Multifamily Ductless Heat Pump Retrofits*. Energy Trust of Oregon. July 11, 2018. www.energytrust.org/wp-content/uploads/2018/08/MF-DHP-Billing-Analysis-Final-Report3.pdf

Stephens, Charles (Retired, NEEA). 2016. Personal communication.

Walczyk, John, Jim Grevatt, Brian McCowan, and Katie Champagne (Cadmus), 2017. *Evaluation of Cold Climate Heat Pumps in Vermont.* VT Department of Public Service. November 3, 2017. https://publicservice.vermont.gov/sites/dps/files/documents/Energy_Efficiency/Reports/Evaluation%20of%20Cold%20Climate%20Heat%20Pumps%20in%20Vermont.pdf

West Hill Energy and Computing, 2019. *Efficiency Maine Trust Home Energy Savings Program Impact Evaluation 2014-2016.* Efficiency Maine Trust. August 23, 2019.

www.efficiencymaine.com/docs/HESP-Evaluation-8-23-19.pdf

Wilcox, Bruce A., Abram Conant and Rick Chitwood, 2016. *Central Valley Research Homes – Variable Compressor Speed Heat Pumps*. ET Project Number ET14PGE8761. Pacific Gas & Electric. September 13, 2016. www.etcc-ca.com/reports/variable-compressor-speed-heat-pumps

Wilcox, Bruce A., Abram Conant and Rick Chitwood, 2018. *Central Valley Research Homes – Variable Capacity Heat Pumps, Evaluation of Ducted and Ductless Configurations 2016-2017*. ET Project Number PG&E 2018_3. Pacific Gas & Electric. May 9, 2018. www.etcc-ca.com/reports/central-valley-research-homes-evaluation-ducted-and-ductless-configurations-variable

Wilcox, Bruce A., Abram Conant and Rick Chitwood, 2019. *Variable Capacity Heat Pumps – Evaluation of Sizing and Controls Settings 2017-2018*. ET Project Number PGE 2019_1. Pacific Gas & Electric. May 3, 2019. www.etcc-ca.com/reports/central-valley-research-homes-evaluation-sizing-and-controls-settings-2017-2018

Williamson, James, and Robb Aldrich, 2015. *Field Performance of Inverter-Driven Heat Pumps in Cold Climates*. Building America/NREL Report #KNDJ-0-40342-05. August 2015.

www.energy.gov/sites/prod/files/2015/09/f26/ba-case-study-inverter-driven-heat-pumps-cold.pdf

Heat Pump Performance Ratings.

Why are better ratings so important?

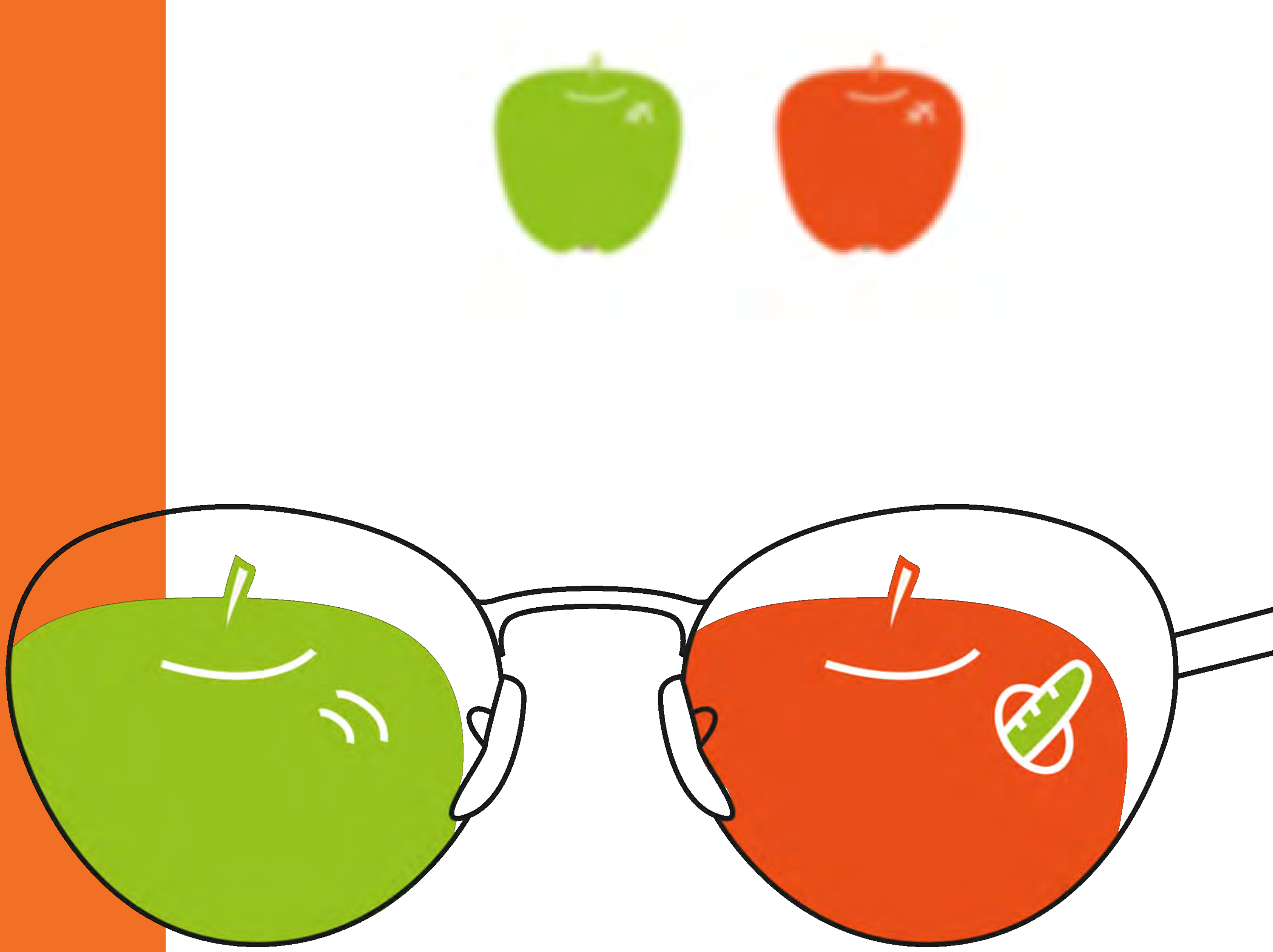
PUBLIC

Good quality ratings are like glasses

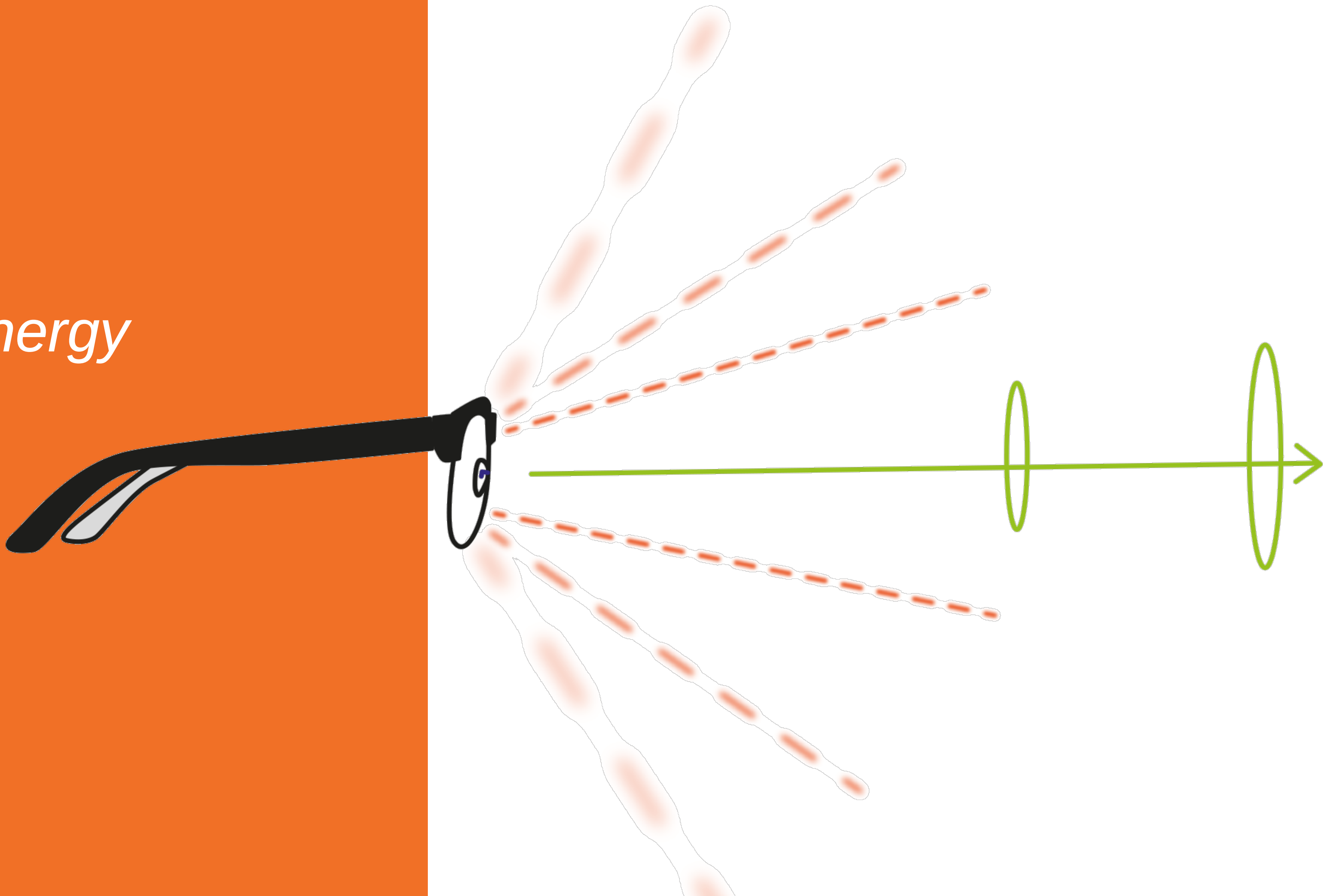
They help you see differences and make informed choices.

Accurate ratings …

… let us see key details we otherwise might have missed.



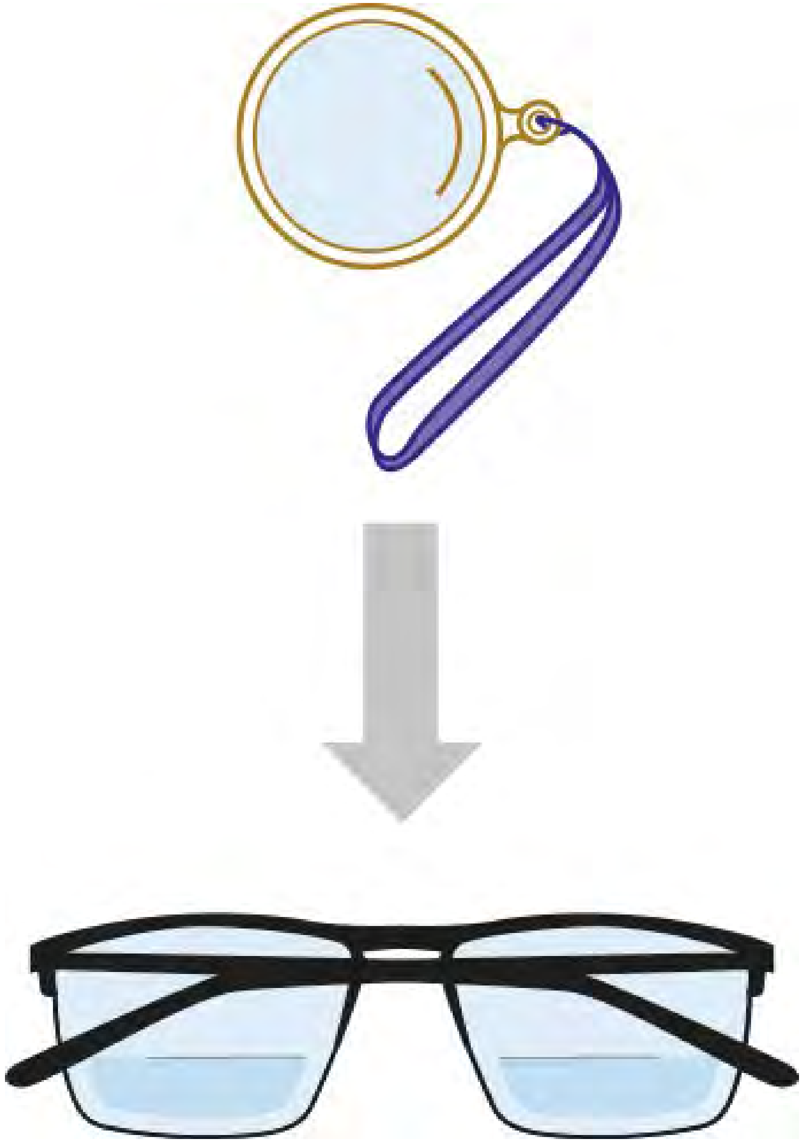
… forecast energy savings,



…help consumers choose the best product for their home or business,

 |

… and encourage manufacturers to improve product performance.

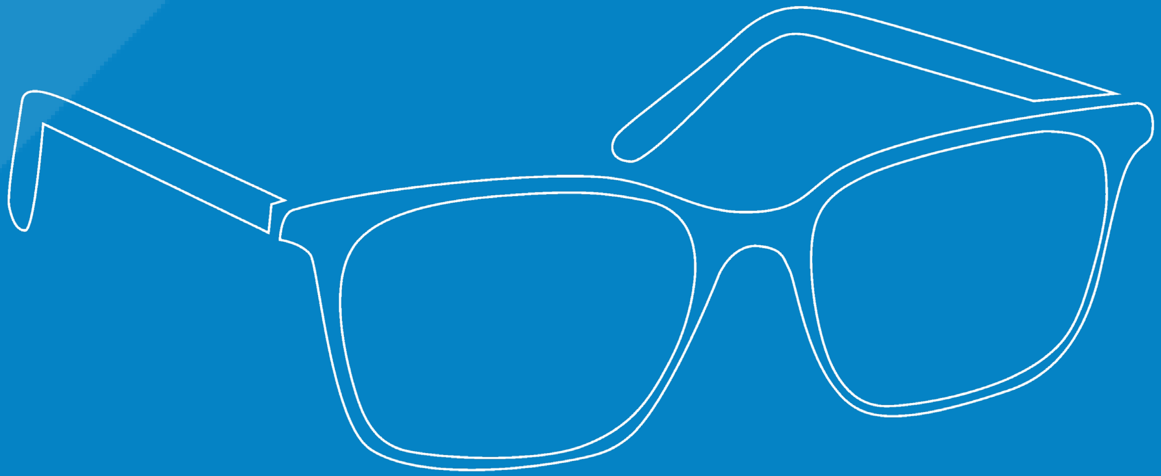


 |

So, what is a good rating?

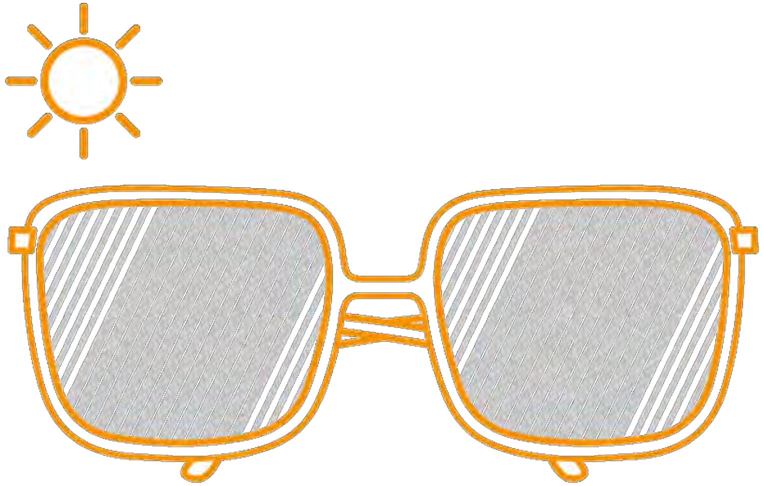
It depends who is looking!

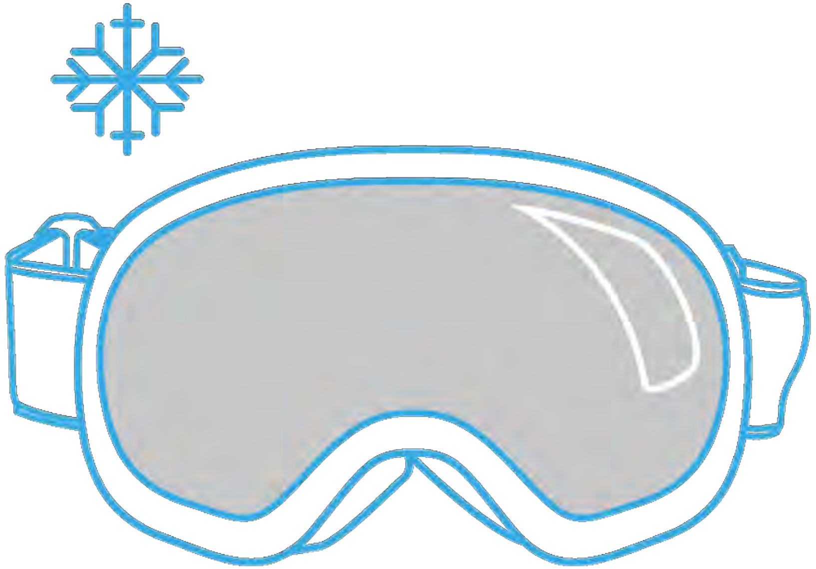
Customers, Utilities or Manufacturers



A GOOD RATING IS:

Easy to understand.



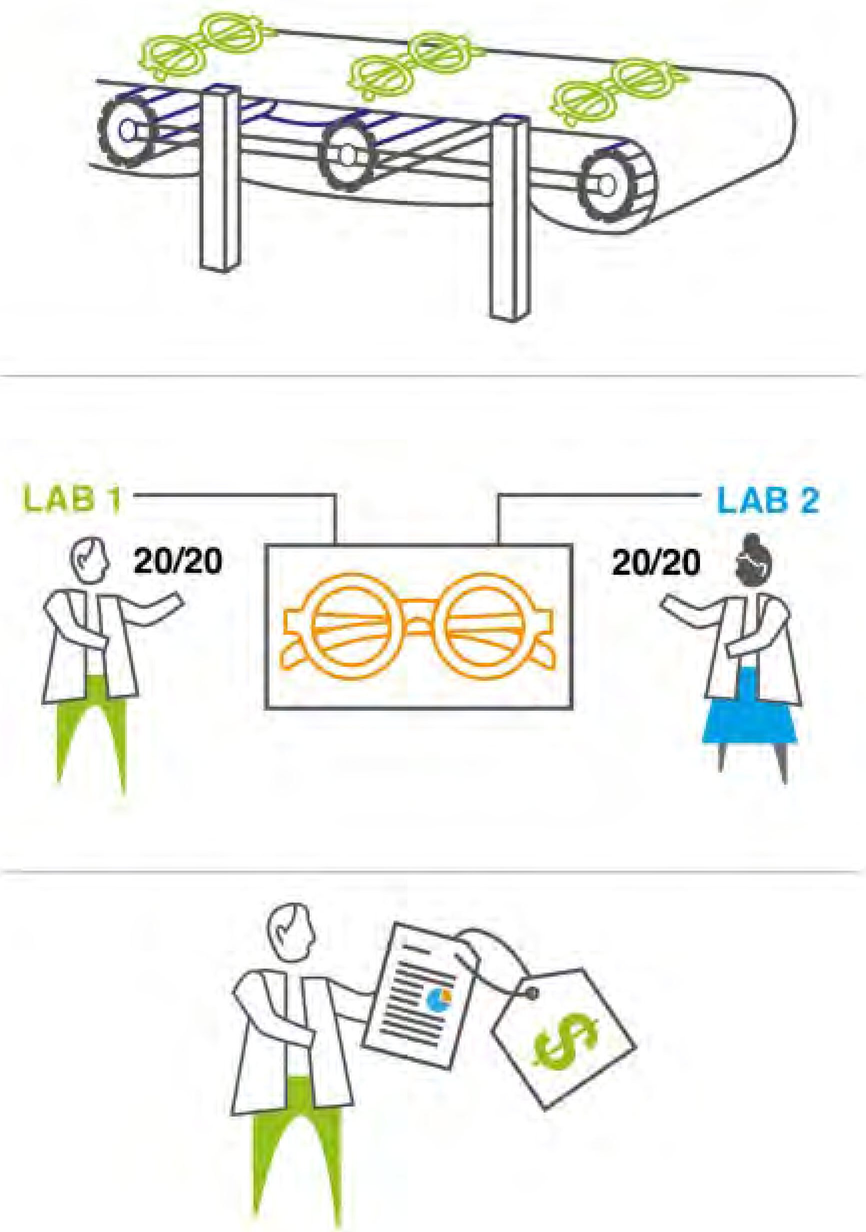


A GOOD RATING IS:

– Repeatable (test to test)

– Reproducible (lab to lab)

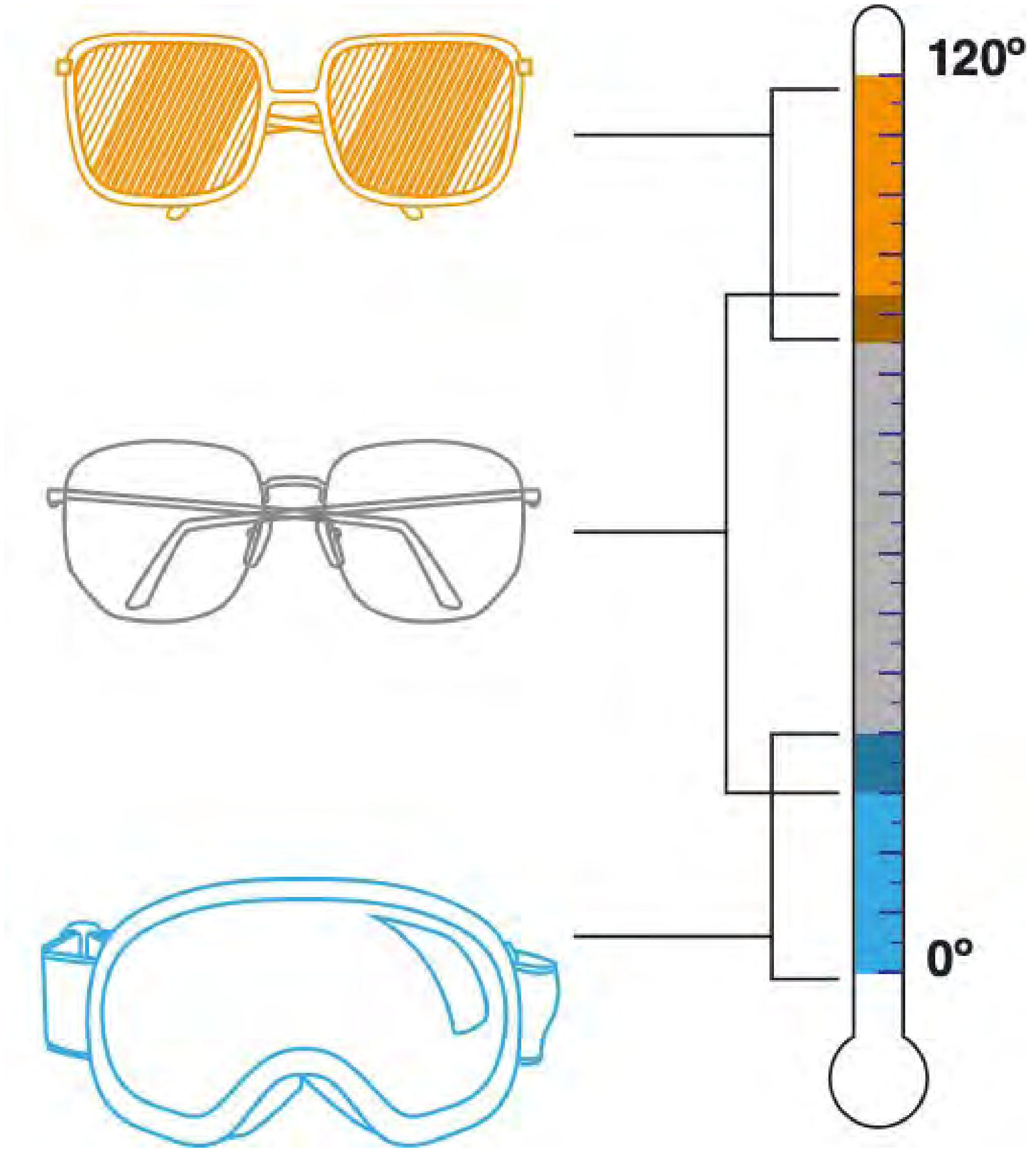
– Not too expensive



A GOOD RATING IS:

- Representative

This means it is an accurate predictor of in-field performance



 |

Why change current heat pump rating tests?

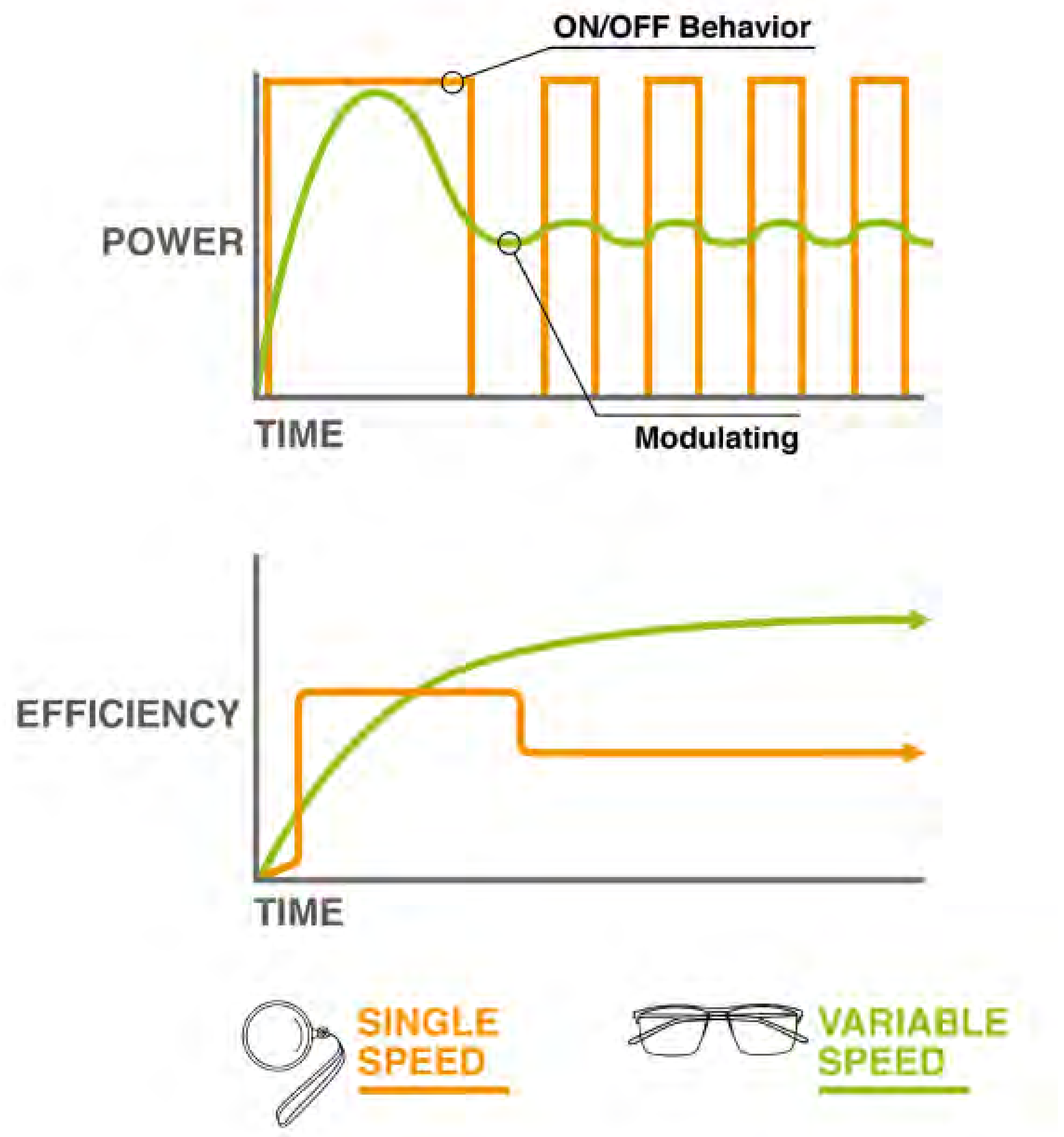
Current ratings do not predict in-field performance accurately

- ✓ Easy
- ✓ Repeatable
- ? Reproduceable
- ✗ Representative

Old Style vs. New Style

The performance of new style variable speed heat pumps is dynamic

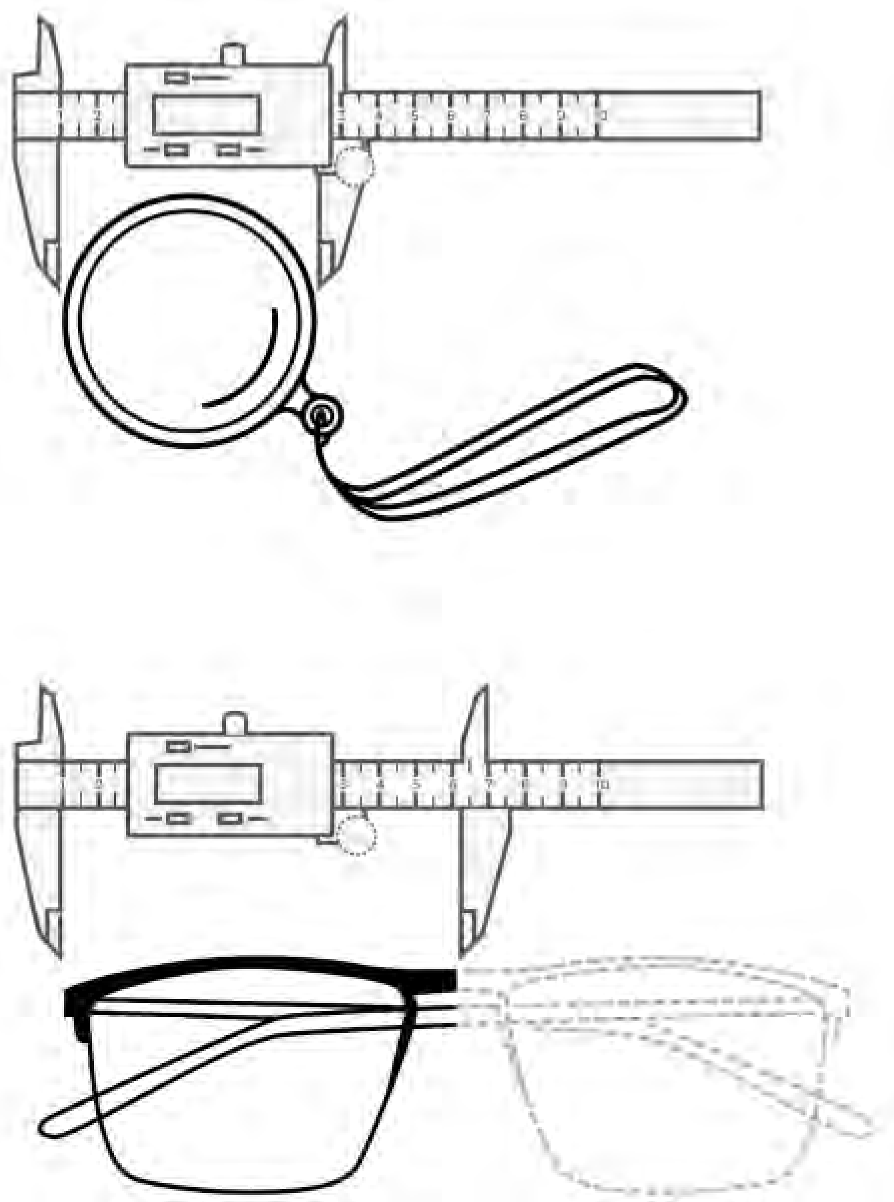
Current tests works well for simple on/off operation



 |

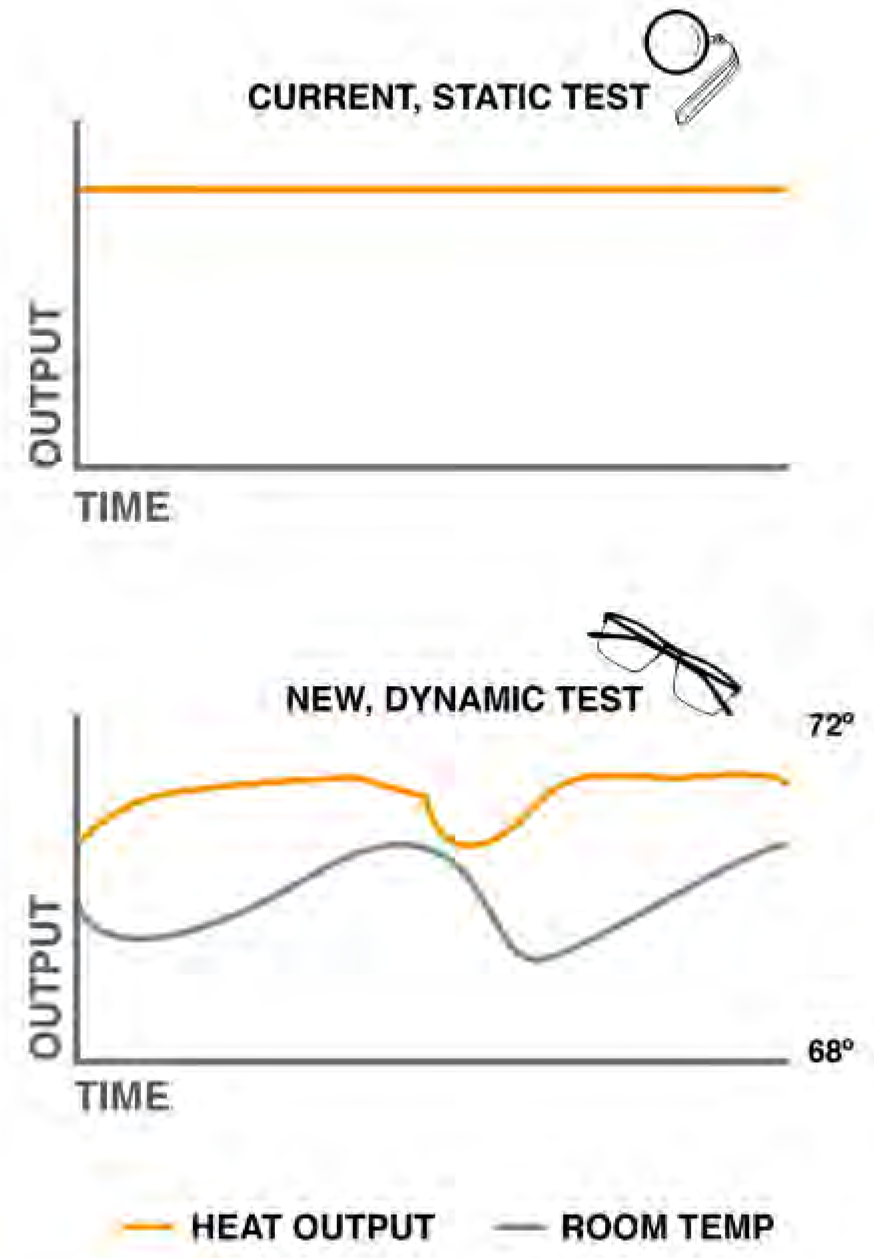
Current tests were designed for old style heat pumps

…this only measures part of the heat pump performance



Current tests are locked into static test modes

…new test works under realistic, dynamic conditions, using the heat pump's own controls



 |

Current tests apply to only one climate

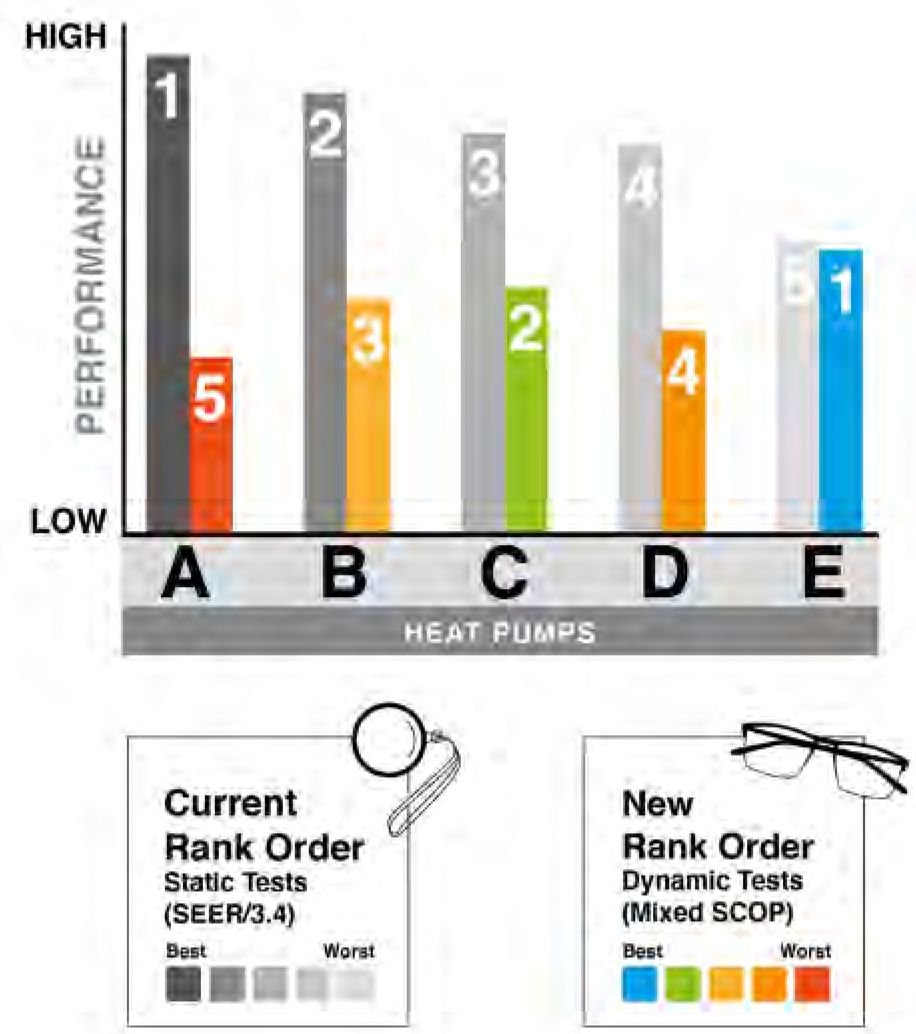
…new tests will predict how the heat pump will perform in different climates

In a lab test comparison
(shown here)

…new tests revealed big changes in ranking & performance

(Reference: Why Metrics Matter paper)

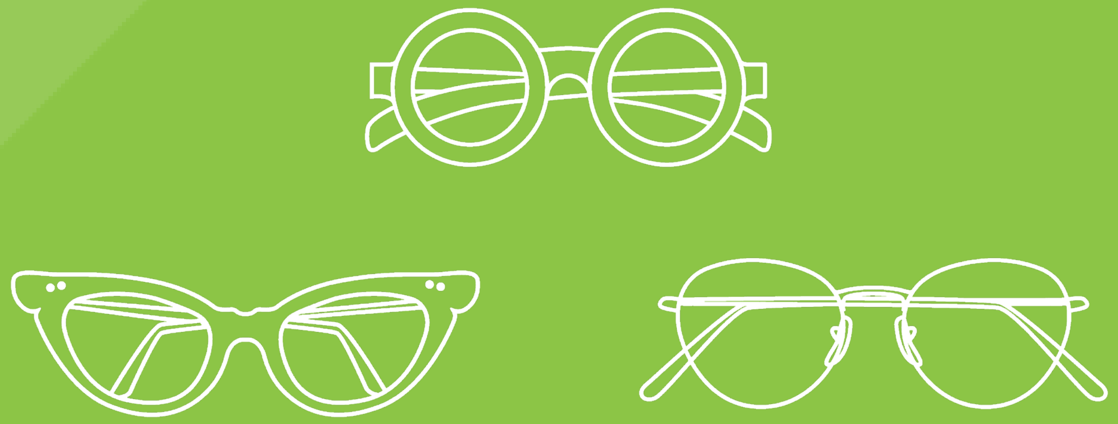
RANKING ORDER CHANGED



 |

How does this benefit you?

Good quality ratings help everyone to make better decisions.



Satisfied customers

Machines that are sold are correct for the climate

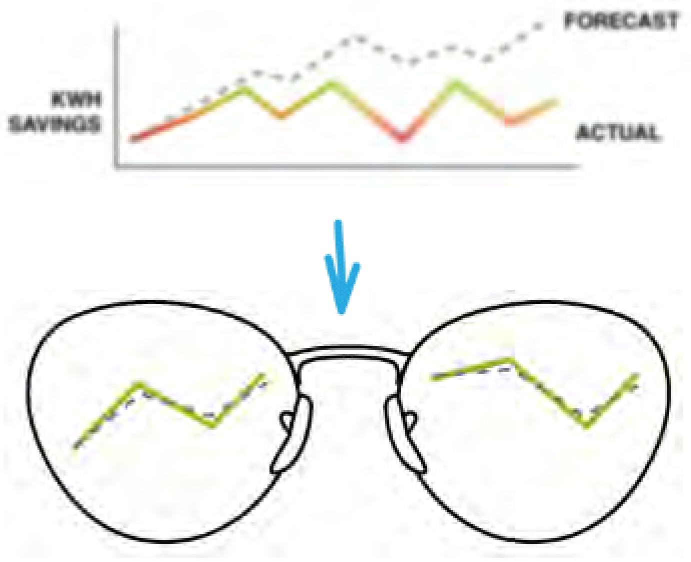
Low performing manufacturers are encouraged to improve.

Consumers get the comfort, quality and savings they were expecting.

 |

Program Goals are Met

More Accurate Carbon and Energy Forecasting



Program Savings are REAL

What can you do?

1

Read "Why Metrics Matter" Full Report

[NEEA reports]

2

Use climate specific, representative metrics
(don't rely on HSPF and SEER)

3

Join the Advanced Heat Pump Coalition

[Click here]

Thank you

Christopher Dymond, Sr. Product Manager
cdymond@neea.org