



High tech versus high hype: Avoiding the pitfalls of black-box technologies

An E Source white paper

By Essie Snell

August 27, 2014

Contents

[Why care about black boxes?](#)

[A different approach](#)

[Technical considerations](#)

[Behavioral and psychological considerations](#)

[Final thoughts](#)

[Acknowledgements](#)

[Notes](#)

At E Source, we define black-box technologies, in part, as things that are bolted on, wired in, or otherwise added to existing technologies in order to yield energy savings. Although some black-box technologies can achieve savings under a wide variety of conditions, some achieve savings under limited conditions, and others provide no savings whatsoever. To figure out which is which, we look at the claimed mechanism of savings and only review independent test data to see whether it validates that mechanism. Our strategy can be in stark contrast to the approach used by many energy managers, utility program managers, and energy

consultants, who may review product test reports or conduct tests themselves to see what energy savings result without necessarily understanding the reason for those savings.

How E Source can help with technology assessment

Fill out this short form to start a conversation about your needs and how E Source can help.

The methodology we use to evaluate black-box technologies is a comparatively quick and cost-effective way to review products, and it avoids many of the pitfalls that “commonsense” methods of evaluating technologies—including field tests and vendor guarantees—may fall into when used to assess potential energy savings. In addition to developing the overarching methodology itself, we’ve also identified a number of specific technical and psychological considerations that can lead to confusion about a product’s likely efficacy, but which are easily overcome if evaluators know what to look for. Overall, we expect readers to come away with a variety of new ideas to incorporate into their own evaluations, making them less likely to adopt or promote products that don’t work as well as expected.

Why care about black boxes?

Black-box products are sold everywhere there’s an interest in energy efficiency. As a result, anyone whose job entails identifying cost-effective facility upgrades, qualifying efficiency technologies for demand-side management programs, picking solid “green” investments, or deciding what types of new efficient technologies should undergo laboratory or field testing may eventually be approached by an enthusiastic representative of a black-box vendor who will invariably have a laundry list of reasons why purchasing his company’s product will be one of the best decisions a person could ever make. Although this sales pitch could indeed hit the mark and the technology could be a great success, it’s also entirely possible that the product may not perform as expected (if at all), wasting both time and money. Having a strategy in place to gauge the likely potential of a given product to meet expectations before a vendor comes knocking can reduce the time and resources spent dealing with products that may end in disappointment—it can also allow for the adoption of new, previously unknown technologies with great potential.

A different approach

Because E Source relies on secondary research and doesn’t have a laboratory in which to perform independent tests, we developed a unique process to evaluate unfamiliar technologies based on several decades of experience assessing hundreds of black-box products. This approach is a relatively quick and inexpensive process that can help weed out ineffective or underperforming technologies early on, and, in some cases, it may even preclude the need for additional testing. Beyond just cutting costs, however, it can also help evaluators avoid numerous potential pitfalls that testing alone may not overcome. For example, a

test under a narrow band of conditions may suggest that a product has enormous savings potential, but it may miss the fact that the operating conditions under which the product is likely to be installed may be quite different from the test conditions. If the product can save energy only under a very limited set of conditions (which happen to match those of the test), average savings in the field may be much lower than expected. Alternatively, otherwise competent evaluators may neglect to account for key variables in their tests if they don't fully understand the theoretical principles behind the device they're testing. Even beyond these technical considerations, evaluators who jump straight to testing a product to see how well it works can easily fall prey to confirmation bias, in which they unconsciously favor information that agrees with the vendors' claims while discounting conflicting information.¹ In essence, our approach can be boiled down to two fundamental questions: What's the theory of operation, and does test data substantiate that theory?

What's the theory of operation?

When we evaluate energy-savings claims, we first check for a theory of operation that aligns with the laws of physics and chemistry. Only after such a theory is established do we look for independent performance evaluations that verify the theory and its efficacy. We take this approach because without understanding how the underlying technology works, we have no way of assessing whether test conditions may be representative of the operating conditions where the device is likely to be installed in the field, whether the test accounts for all relevant variables, or whether the test results can be generalized to other scenarios or applications. Given all these unknown factors and the kinds of potential test-related problems we've described, it's possible to greatly overestimate the applicability of such products unless an attempt is made to first understand the underlying methodology.

To check for a viable theory of operation, we first verify that the energy waste the product claims to reduce actually exists and that it's significant enough to yield benefits if it's somehow reduced. To do this, we may look for existing robust, independent studies that quantify the extent of energy waste and the conditions under which that waste is most likely to exist. Or, particularly when this type of research is unavailable, we may reach out to impartial experts in the field for their insight. We're generally skeptical of vendor-provided case studies and nonindependent test data because there's always a chance that biases—both intentional and unintentional—may shade the results.

Next, we collect information on how the vendor claims the product is able to reduce this energy waste. Sources for this information include vendor websites, conversations with company representatives, articles in outside publications, or experts in the field. Once a plausible method of savings has been identified, we check for any incongruities with either physical laws or real-world experience to verify that the product can—at least in theory—reduce energy loss to the extent the vendors claim.

A common problem with vendor-provided theories of operation is an apparent confusion about basic technical concepts such as the distinction between correlation and causation, the different types of heat transfer, real

versus apparent power, or the interaction of various components in a particular system. In some cases, the supposed theory of operation may appear to contradict known laws of physics and lead to concerns about a product's viability (more on this later). Because this confusion can stem from a variety of sources—ranging from a simple misunderstanding of the technology to a deliberate attempt to confuse potential customers—we generally try to clarify the source of these seeming incongruities with company representatives before coming to a decision regarding a given product. Only after we've successfully established that it's theoretically possible for the product to save energy will we move on to an investigation of any available independent test data.

Does test data substantiate vendor claims?

When examining test data, the first thing we look for is an indication that the testing process actively investigates the reported savings mechanism. Simply plugging in a black-box product and observing an apparent reduction in power draw isn't rigorous enough, since factors such as statistical variance or uncontrolled variables may provide false positives. In our view, testing should not only show that savings were achieved, but it also should reveal how.

The test conditions should also accurately reflect the wide variety of conditions under which the product is likely to be installed. Because the energy-saving potential of black-box technologies may vary significantly in different applications, realistic test conditions can help indicate where the product is likely to be more or less successful in yielding benefits to users.

As with any testing, it's vitally important that sources of error are accounted for and minimized wherever possible. In particular, the best test results will account for all the controlling variables and include a detailed description of the setup and measurement procedures, a reasonable sample size of both control and experimental measurements, an appropriate normalization of any changes in controlled variables between cases, and statistical error analysis to indicate the significance of the results. In addition, it's essential that the report be internally consistent. For example, if a value is claimed to be the average of a set of measured data, it should actually be the average of the data described, not a different and seemingly unrelated number. (Although this may seem like an obvious example, it's often not clear that numbers are incorrect, or they may change throughout a document unless one carefully combs through all of the calculations.)

Technical considerations

In addition to the more general considerations we've described, there are a number of specific areas where technical information provided by a vendor may break down upon closer scrutiny, including claims that seem to violate the laws of physics, incorrect units, or a lack of error analysis. Although it's always possible that unlikely claims or apparent flaws in vendor-provided information may simply be the result of typos or a misunderstanding in the company's marketing department—and it's worth double-checking with the organization to clarify that its claims are stated as intended before making a judgment of its products—all of

these problems should raise red flags for evaluators.

Laws of physics

Simply put, any new device or technology must behave according to the laws of physics, and if vendor-provided claims about a product run counter to currently accepted scientific theories or posit the existence of new or unknown scientific principles, the burden of proof should be on the vendor to back up those claims with peer-reviewed theoretical and experimental data. In our experience, some discrepancies come up more often than others, and several such examples are provided below.

Basic thermodynamics. The four laws of thermodynamics are fundamental to modern physics and establish (among other things) that energy can't be created or destroyed, and that the entropy of a system tends to increase over time. Nonetheless, a surprising number of vendors make claims or implications that necessitate their product's violation of these laws. Some specific examples that are representative of extreme vendor-provided claims we've previously reviewed include:

Every photon produced by our heater creates another photon with the same energy and wavelength without requiring additional energy.

Independent testing proves that the energy output is greater than the input.

Such claims should serve as a clear indicator that skepticism about the potential validity of a company's products is warranted.

Heat transfer. Heat can be transferred from one location to another through one of three basic mechanisms:²

- *Conduction*, in which heat is exchanged through physical contact (for example, burning yourself by touching a hot pan)
- *Convection*, in which heat is moved from one place to another due to the motion of a fluid (one of the main drivers behind ocean currents)
- *Radiation*, in which heat is transferred from one object to another via the emission or absorption of electromagnetic radiation (such as moving into a sunny area to get warmer)

With regard to energy efficiency, the two most important types of heat transfer generally tend to be conduction and radiation. Building insulation (including fiberglass batting and foam) is designed primarily to reduce heat transfer via conduction into or out of a building, and the level of insulation generally increases as

materials get thicker. In contrast, radiant barriers and light-colored “cool roof” products are designed to reduce heat transfer via radiation and can be essentially flat. The problems come when vendors conflate different modes of heat transfer, as illustrated in this representative vendor claim:

Our paint product reflects nearly all of the sun’s radiation and can replace 6 to 8 inches of traditional insulation.

Even when coatings can provide benefits by limiting radiative heat transfer, they typically aren’t thick enough to affect conductive heat transfer in any appreciable way. Thus, it’s incorrect to view them as a replacement for traditional insulation that’s capable of limiting conductive heat transfer (but which may not block radiative heat transfer). Simply put, both approaches can indeed be effective at reducing heat transfer, but one is not a substitute for the other.

Magnetic monopoles. Some products, such as magnetic fuel conditioners, claim to use magnetic monopoles (magnets with only one pole—for example, only north or only south) to create substantial energy savings in a variety of systems. Although the possibility of magnetic monopoles was first posited in 1931 by physicist Paul Dirac,³ at the time of this writing, no known naturally occurring magnetic monopoles have yet been discovered. Some groundbreaking experiments have recently been able to study “synthetic” magnetic monopoles created by manipulating magnetic fields into patterns consistent with those that a real magnetic monopole is likely to create if it exists.⁴ But that’s still a far cry from a real magnetic monopole that could be produced on command to serve the interests of energy efficiency. Barring major future developments, any mention of the use of magnetic monopoles should cause concern about the potential validity of a product.

Unit consistency

In some cases, vendor-provided materials may include calculations designed to support their claims that may intentionally or unintentionally use flawed math to reach their conclusions. In such cases, performing a simple unit check on the results (and/or steps of the mathematical calculations) can be a quick way to identify potential errors. For example, in imperial units, R-value (a measure of the insulating ability of a given material) should have units of square feet-hours-degrees Fahrenheit per British thermal unit [(ft² x h x °F)/Btu],⁵ or (h x ft² x °F)/(Btu x in) if measuring R-value per inch of material thickness. As a result, if a vendor-provided calculation yields results with different units on the numerator or denominator [for example, (h x ft²)/(Btu x °F)], it’s likely that the value purported to be R-value is something else entirely and isn’t representative of the insulating value of the material.

Statistical analysis

It's impossible to make any real-world measurements with 100 percent accuracy. The only reliable way to verify that measured results actually mean anything (that is, that they're statistically significant) is to do some level of error analysis. Say, for example, an additive is introduced to a refrigeration system and the overall system efficiency is measured to have increased by 2 percentage points. This may or may not say anything about the additive. If the statistical error is plus or minus 2 percentage points (or greater), it's possible that the additive had no effect at all and that the perceived "improvement" is simply a result of inevitable variability in the measurement process. Unfortunately, although it's often taught in many undergraduate science courses as a fundamental part of the scientific process, error analysis seems to be widely neglected in the energy-efficiency industry. Even reports from otherwise reputable sources may neglect to include such analysis. Nonetheless, statistical analysis can be a huge boon to technology evaluators when it exists in research reports, and it can be well worth the effort to ask for such data to ensure that results aren't misleading.

Reality checks

Taking a step back to consider the big picture—such as whether vendor claims seem plausible in a general sense or whether claimed savings are even in the right order of magnitude—can often help quickly identify aspects of vendors' claims that may require further information and research or that are clearly nonsensical.

For example, one vendor with a product designed to reduce heat transfer claims that it offers an R-value (in imperial units) of 400 per inch of thickness. However, a quick Internet search reveals that the material currently considered to be the least dense and most insulating solid in the world is called aerogel. Silica aerogels in particular can have very low thermal conductivities (with associated R-values of around 10 per inch of thickness) and are so light and translucent that they are often called "solid smoke."⁶ Given that the vendor's claimed level of insulation is 40 times larger than one of the most insulating solids currently known to science, it seems fair to say that this claim is highly unrealistic at best.

As another example, a number of vendors profess to be able to provide overall building electricity savings in the range of 20 to 30 percent by improving power quality through strategies like power factor correction. Although we continue to search for studies of customer-side electrical distribution system energy losses, to date we've found very few broad, publicly available, statistically valid studies of potential energy savings via improvements to the electrical distribution systems. In the absence of such data, E Source has discussed the potential for energy savings via strategies like power factor correction with a wide range of power quality experts across the US. The response has been consistent: Savings are usually below 1 percent, but they can rise to as much as 3 percent in extreme cases. The discrepancy between the opinions of these experts and the claims made by vendors is so large that it casts doubt on the ability of such products to come anywhere near the claimed savings levels (though it's certainly true that improving power quality can sometimes

provide a number of other benefits).

Behavioral and psychological considerations

Vendors can use myriad persuasive psychological strategies to encourage acceptance of their products, particularly when additional investigation on the part of a potential customer may reveal lackluster results. Many of these strategies are relatively common marketing practices because they're so effective at convincing people to buy products. The problem with black-box products in particular is that vendors can rely on these tactics almost exclusively and use them as persuasive distractions from the kinds of technical details we've mentioned that are vital to the assessment of a product's potential performance. For this reason, evaluators who are aware of these general strategies are less likely to be swayed by subjective factors when assessing the potential of a given product, regardless of whether the product ultimately works as claimed.

Red herrings

Red herrings can include any piece of information that sounds credible and convincing but isn't independently verifiable or doesn't directly address the claims a vendor is making. Red herrings are so successful that it's possible to find examples of their use in both dubious and well-established products going back centuries (**Figure 1**). These can take various forms, including:

- *Testimonials or personal opinions*, which, even if taken from real customers, may not be representative of all customer experiences, may stem from customers' feelings about the company, or may even be due to a misunderstanding of the actual benefits of the product. Nonetheless, testimonials can be a very persuasive technique that makes good use of the bandwagon effect—a well-established cognitive bias in which people are more likely to do something just because they believe others are doing it.⁷
- *"Phantom" test data* that's summarized on a vendor's website or cited by a representative but is not available for reference anywhere (and, in some cases, may not even exist). Some vendors may go so far as to promise to send along robust independent research that they claim will address all skepticism about a product, but then never provide that research or offer a list of excuses as to why it's not available to share.
- *A list of the number of installations*, which doesn't indicate how well a product works, regardless of whether the figure is accurate. Because people often make purchases based on a variety of factors unrelated to the value of the product itself, it can be fallacious to imply the efficacy or intrinsic value of a product by sales alone (as an example, consider the success of fad products like the Pet Rock, which made its inventor a millionaire). Listing the number of installations is another strategy that uses the bandwagon effect described earlier.
- *Deferral to authority figures* to imply that another organization has thoroughly investigated and vetted a product even when that isn't necessarily true. One example is the claim that a product was developed by NASA. Although this claim may be indirectly true—NASA's research has indeed produced a host of spin-off technologies like Velcro and aerogel—NASA generally doesn't develop products for saving energy and likely hasn't evaluated the energy-saving potential of such products. Another example is the statement that the US Department of Energy has funded research into the product or installed it at a federal facility (which, again, may be true, but doesn't provide any representative information about the

performance of the product). Yet another example is when a company becomes an Energy Star partner (meaning that the company has agreed to improve the performance of its own facilities), then includes the Energy Star logo frequently in its product portfolio to indirectly imply that ENERGY STAR has qualified or approved the product.

FIGURE 1: Historical red herrings

This 1868 advertisement for a cure-all product of dubious efficacy demonstrates several red herrings—including number of “successes,” personal experience, and deferral to authority figures—similar to those used by many modern vendors of purportedly energy-saving black-box products.



Each of these pieces of information can be persuasive individually—and the product can seem even more compelling when multiple red herrings are simultaneously thrown at an evaluator—but none of them actually provides any concrete indication of how well a product really works.

Obfuscation

In some cases, vendors may offer overcomplicated technical descriptions that make products sound more impressive and cutting-edge than they actually are. They may also give the impression that a lot of research and science have gone into developing the product, even if that’s not necessarily the case. Furthermore, technical evaluators might be discouraged from asking hard questions about the method of savings if it seems like they might not understand the answer. In the most egregious cases, vendors may provide reports that are many pages long and use a plethora of scientific jargon but that don’t include much, if any, information that’s pertinent to a would-be evaluator. Here’s a short example of an obfuscating explanation about how a product works:

Through the magic of nanotechnology, our product is able to change the molecular structure of hydrocarbon fuels to release never-before-tapped energy.

At first glance, this sentence may sound impressive, but it doesn’t provide any substantive information about how the device functions. Buzzwords like quantum mechanics or nanotechnology seem to crop up in marketing materials with increasing frequency, but because terms like these are essentially just categories for certain areas of study and don’t necessarily refer to specific processes or techniques, they don’t ultimately reveal anything about the underlying methodology used. From a technology assessment standpoint, the most valuable explanations about how unfamiliar devices work tend to be simple and straightforward, with supplemental details about the underlying scientific principles if a reader is interested in learning them.

The “poor inventor” archetype

Another strategy that can help build empathy and goodwill while also explaining a lack of robust independent research is to frame the manufacturing company as a small organization (potentially built around one person’s genius) that offers a wholly unique product that could change the world but that currently has limited resources available for testing. Although this could, in some cases, be a completely accurate description of the company, it may also simply be a convenient excuse to provide limited supporting information about a product. In our experience, truly unique products are rare, and we expect that even small companies should be able to explain (at least at a high level) how their products work.

Guarantees

On their surface, performance guarantees may seem like a great way to ensure that a product is likely to deliver on its promises. But vendors can use a wide range of approaches to offer a guarantee while also ensuring that it will rarely, if ever, be fulfilled. For this reason, we don’t find guarantees to be reliable indicators of a product’s efficacy. We’ve outlined some specific techniques that can undermine the value of guarantees.

Vendor provides proof of savings. A simple way to “prove” that a device saved a customer energy, even if it didn’t, is to require the customer to provide several months of utility billing data from before and after the product was installed. The vendor will then “analyze the data.” If, for whatever reason, energy bills for a month or two after the installation were lower than some of those before the installation, the vendor can claim that the product saved the customer energy and deny a customer’s request for a refund. Because energy bills tend to vary from month to month, based on many factors, a vendor is likely to be able to find at least some selective data that can be used to apparently confirm its claims.

Burden of proof is on the customer. If a customer decides he wants his money back, a company may require him to provide extensive and detailed testing results over multiple years before and after the product was installed. He may also need to submit a plethora of other information (including weather data, utility bills, service invoices, documentation of any and all changes during the guarantee period, occupancy records, and sales data) in a short time frame. Unless, from the moment of installation (or possibly even earlier), the customer had the foresight to spend the time and funds needed to get this level of analysis and maintain extensive records about everything that might possibly relate to the product, he is unlikely to be able to make the case to the vendor’s satisfaction that the product failed to deliver as promised.

Fine print. Some manufacturers or vendors include fine print in their guarantees that can make it difficult for a refund request to even be processed or reviewed. For example, one manufacturer we’re aware of states that the “guarantee is not valid unless received by customer through the US Postal Service directly from [the manufacturer].” If, instead, a distributor or vendor provides the customer with a performance guarantee, or that guarantee is not delivered through the US Postal Service, then the manufacturer could claim that the

guarantee is null and void. Alternatively, some vendors require that their product be installed in a very specific way (which may not be obvious unless the customer has read through all the fine print) in order for the guarantee to be valid.

Sense of urgency

To make all of the aforementioned behavioral and psychological strategies even more effective, vendors may urge evaluators or customers to make a hasty decision (or, in the case of guarantees, to get everything submitted in a very short amount of time). When trying to come to a decision quickly, it becomes more difficult to gather and analyze the kinds of information needed to make an informed judgment. It also makes it more likely that evaluators and customers alike will rely more on subjective factors than empirical data, increasing the likelihood that they will side with the vendor's claims.

Final thoughts

Assessing the energy-savings potential of black-box technologies can require analysis from a variety of different angles, but having a single overarching approach to start with can help evaluators get on the right track quickly and relatively easily. In our experience, the general strategy we've outlined (asking for a logical theory of operation and independent test data to substantiate that theory) is an effective and low-cost way to get a sense of the viability of an unfamiliar energy-efficiency device before expending more time and energy to dig deeper into potential savings. The specific criteria described in the preceding two sections can help raise red flags about individual vendor claims.

Ultimately, all of these steps go toward assessing the legitimacy of a vendor's energy-savings claims. A lot more work goes into determining whether a product is appropriate for a given application—including consideration of such factors as cost-effectiveness, reliability, quality control, the availability of a support network, and how the product compares with competing technologies. Nonetheless, this initial assessment is invaluable for establishing whether the underlying technology can, in fact, provide beneficial results if used properly and whether further investigation may be warranted. Although it would be unreasonable to expect readers to adopt the exact methodology used by E Source, we expect that incorporating some of the strategies and information outlined in this report into their own evaluations will improve the likelihood of finding successful new technologies that will work as promised and yield substantial energy savings.

Acknowledgements

Over the past several decades, a number of people have helped to develop and refine E Source's approach to analyzing technologies. In particular, we'd like to thank Bill Howe, Jay Stein, and Michael Shepard for laying the groundwork of our methodology and providing guidance on how to better address the challenges of technology assessment.

Notes

- 1 Jonathan J. Koehler, "The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality," *Organizational Behavior & Human Decision Processes*, v. 56 (September 1, 1993), pp. 28-55.
- 2 Daniel V. Schroeder, *Introduction to Thermal Physics* (Addison-Wesley, 1999).
- 3 P.A.M. Dirac, "Quantised Singularities in the Electromagnetic Field," *Proceedings of the Royal Society A*, v. 133 (1931), pp. 60-72.
- 4 M.W. Ray, E. Ruokokoski, S. Kandel, M. Möttönen, and D.S. Hall, "Observation of Dirac Monopoles in a Synthetic Magnetic Field," *Nature*, v. 505 (January 30, 2014), p. 657-660.
- 5 "Seal and Insulate with Energy Star Program for Residential Insulation Manufacturers: Partner Commitments," US Environmental Protection Agency (March 2013), [www.energystar.gov/products/specs/sites/products/files/SI%20w%20ENERGY%20STAR_Def%20and%20Testing%20Req_Partner%20Commitments%20V1%20\(R%20rev%20Mar-2013\)%20\(2013-03-28\).pdf](http://www.energystar.gov/products/specs/sites/products/files/SI%20w%20ENERGY%20STAR_Def%20and%20Testing%20Req_Partner%20Commitments%20V1%20(R%20rev%20Mar-2013)%20(2013-03-28).pdf).
- 6 David R. Lide, *CRC Handbook of Chemistry and Physics*, 86th edition, (CRC Press, 2005), p. 227.
- 7 Andrew Colman, *Oxford Dictionary of Psychology* (Oxford University Press, 2003).