Beyond kWh: myths and fixes for energy competition game design

Philip M. Johnson, Yongwen Xu, Robert S. Brewer,
George E. Lee, Michelle Katchuck, Carleton A. Moore
Collaborative Software Development Laboratory
Information and Computer Sciences
University of Hawaii
Honolulu, HI USA
johnson@hawaii.edu

The Kukui Cup project investigates the use of "meaningful play" to facilitate energy awareness, conservation and behavioral change. Each Kukui Cup Challenge combines real world and online environments in an attempt to combine information technology, game mechanics, educational pedagogy, and incentives in a synergistic and engaging fashion. We challenge players to: (1) acquire more sophistication about energy concepts and (2) experiment with new behaviors ranging from micro (such as turning off the lights or installing a CFL) to macro (such as taking energy-related courses, joining environmental groups, and political/social advocacy.)
To inform the design of the inaugural 2011 Kukui Cup, we relied heavily on prior collegiate energy competitions, of which there have been over 150 in the past few years. Published accounts of these competitions indicate that they achieve dramatic reductions in energy usage (a median of 22%) and cost savings of tens of thousands of dollars. In our case, the data collected from the 2011 Kukui Cup was generally in agreement, with observed energy reductions of up to 16% when using data collection and analysis techniques typical to these competitions. However, our analysis process caused us to look more closely at the methods employed to produce outcome data for energy competitions, with unexpected results.
We now believe that energy competitions make significant unwarranted assumptions about the data they collect and the way they analyze them, which calls into question both the accuracy of published results from this literature and their effectiveness as serious games. We believe a closer examination of these issues by the community can help improve the design not only of future energy challenges, but other similar forms of serious games for sustainability.
In this paper, we describe the Kukui Cup, the design myths it uncovered, and the fixes we propose to improve future forms of meaningful play with respect to energy in particular and sustainability in general.

Introduction

The rising cost, increasing scarcity, and environmental impact of fossil fuels as an energy source makes a transition to cleaner, renewable energy sources an international imperative. Barriers to this transition include the historical success of electrical utilities in making energy low-cost, ubiquitous, reliable, and easy to access, thus enabling widespread ignorance about basic energy principles and trade-offs. In Hawaii, the need for transition is especially acute, as in our state the price of energy is the highest in the nation and we are the most reliant on oil as an energy source.

Moving away from petroleum involves technological, political, and social changes, requiring citizens to not only think differently, but behave differently with respect to energy policies, methods of generation, and their own consumption. Unfortunately, there is no tradition of teaching "energy" as a core subject area in the United States, even though this subject appears to be one of the most important emergent issues of the 21st century. Anecdotal reports indicate the lack of basic energy literacy at the secondary school level (Ammons, 2010). A survey of the energy literacy of New York State middle and high school students found energy knowledge was poor, with mean scores of only 42% correct (DeWaters & Powers, 2011).

One of the most widespread and successful approaches to raising the profile of energy use is the collegiate dorm energy competition, which has been held on over 150 cam-

The Kukui Cup is supported in part by grant IIS-1017126 from the National Science Foundation, by the University of Hawaii (Facilities Management, Housing, and Information and Computer Sciences), by the Hawaii State Department of Business, Economic Development, and Tourism, and by the HEI Charitable Foundation. We gratefully acknowledge the 418 players of the 2011 Kukui Cup and the members of the project team in addition to the authors who made the vision a reality: Kaveh Abhari, Hana Bowers, Greg Burgess, Caterina Desiato, Risa Khamsi, Alex Young, and Chris Zorn.

puses in the past few years (Hodge, 2010). Published reports claim that these competitions are extremely successful. Hodge finds median reductions in energy use of 22% and a maximum reduction of 80% in energy use for the competitions she studied. A case study of Elon University claims that a seven week competition reduced energy consumption by 231,454 kWh and produced $2,000 in electricity cost savings per week (Durr, 2010).

In an attempt to build upon these promising initial results, we began the Kukui Cup project in 2010. Our goal was to expand the scope from a relatively simple "competition" where the primary outcome measure was energy consumption in kilowatt-hours (kWh) to a more elaborate "challenge" in which we combined information technology, community-based social marketing, serious games, incentives, and educational pedagogy to support sustained change in energy-related behaviors. In addition to measuring kWh consumption, the Kukui Cup also implemented a point system intended to measure and encourage player involvement with educational materials, workshops, excursions, social media, and group participation. Through a series of challenges held in residence halls at the University of Hawaii and elsewhere, we are attempting to gain deeper insight into how these various factors contribute to positive behavioral change.

Our first Kukui Cup challenge was held at the University of Hawaii in Fall, 2011 for the 1,000 students living in the Hale Aloha residence halls. The challenge divided the students into 20 teams (called "lounges") of approximately 50 students each. We started measuring energy consumption for most teams at five weeks prior to the competition, although several teams did not have operational energy meters until shortly before the competition started. We followed the traditional approach of using this pre-challenge data to produce baselines as a way of assessing whether and to what extent energy reductions occurred as a result of the challenge.

Our initial analyses of the data we collected during the challenge appeared promising. Over 400 of the students participated in the three week challenge, spending over 850 hours in the online environment. Student feedback regarding both real world and online aspects of the challenge was uniformly positive. Several teams appeared to achieve a 10-16% reduction in their energy usage during the challenge. Participating students made over 1,000 commitments and earned over 80,000 points.

As we continued to look at the outcomes from the 2011 challenge in order to understand how to better support sustainable behavioral change, we began to be troubled by the way the Kukui Cup and other dorm energy competitions measure outcomes, and the variety of unwarranted assumptions underlying the measurements and results. These design problems are not merely theoretical or scientific: they have a direct impact on the experience of participants and the effectiveness of these competitions as "meaningful play." For example, the baseline calculation method used during the 2010 Campus Conservation Nationals energy competition led some students at Oberlin College to stop participating in the competition "out of frustration" (Willens, 2010).

Making matters worse, published reports concerning energy competitions rarely document how, for example, baseline energy consumption is calculated, or the extent to which energy reductions are sustained after the competition ends. Without this information, it is hard to assess the true "meaningfulness" of the "play." For example, if incorrectly defined baselines enable a team to "coast to victory" with minimal behavior change, then is the play fair? If team energy consumption returns immediately to pre-competition levels, no matter what level of reduction is achieved during the competition, then is the play meaningful?

This paper presents our findings to date about how to better understand the impact of energy competitions and challenges as meaningful play, and how to improve this impact over time. We believe this process must start with a better understanding of the limitations of *kWh consumed* as an outcome measure, as tempting a metric as it might be. We argue that behavior must be interpreted and measured more broadly and that games must be designed to promote and reward a much more diverse spectrum of behavior change. Rather than reward students for unsustainable, temporary behavior changes such as unplugging vending machines (as at Oberlin College (Petersen, Murray, Platt, & Shunturov, 2007)) or camping outside during the competition (as at Carleton College (Hodge, 2010)), future games should reward students for more sustainable, lasting changes such as enrolling in a class on energy in an upcoming semester, or joining an environmental group that promotes renewable energy.

The next section of this paper provides a brief overview of the Kukui Cup. The following section presents the myths (unwarranted assumptions) that we believe to be widespread in the current design and reporting of energy competitions. We conclude with our fixes (recommendations) for how we and others should design future energy challenges to be more effective as meaningful play.

The Kukui Cup

A defining feature of Kukui Cup challenges is a blend of real world and online activities, all utilizing game mechanics. In the real world, players participate in workshops and excursions, win prizes, and most importantly, learn about their current lifestyle and its impact on energy consumption. In the online world of the Kukui Cup web application, players earn points, achieve badges, increase their sustainability "literacy" through readings and videos, and use social networking mechanisms to engage with friends and family about the issues raised. The challenge is designed to make real world and online activities complementary and synergistic.

Figure 1 illustrates the home page of the web application.

Each Kukui Cup Challenge is typically designed with the following goals for its participants:

- Increase **knowledge** about energy issues;
- Gain **insight** about the impact of one's current behaviors.
- **Motivate** them to change their behaviors for the better;
- Build **community**, through awareness of local and national sustainability organizations and initiatives;
- Create **commitment**, from minor (turn off the lights

Figure 1. An example Kukui Cup home page

when not in use) to major (pursue a profession related to sustainability).

The design of the Kukui Cup game was inspired in part by the Community-Based Social Marketing (CBSM) process (McKenzie-Mohr, 2009), which provides a toolbox of behavior change techniques demonstrated to be effective in changing environmental behaviors. In particular, the Kukui Cup allows users to make public commitments to pro-environmental behavior such as turning off the lights when leaving a room, which has been shown to be effective in past studies (Freedman & Fraser, 1966; Pallak, Cook, & Sullivan, 1980). The Kukui Cup also uses energy conservation goals set for each team to encourage them to reduce their energy use, which has also been shown to be helpful in energy conservation (Becker, 1978).

To create sophisticated games based upon energy consumption, it is helpful to integrate real-time energy data from meters into the challenge through goal tracking and energy visualizations. We developed WattDepot (Brewer & Johnson, 2010) to provide an open source, vendor-neutral framework for energy data collection, storage, analysis, and visualization. WattDepot is useful not only as technology infrastructure for the Kukui Cup, but as infrastructure for other energy-related initiatives such as the Smart Grid.

Implementation of game mechanics is provided by another system we developed called Makahiki (Lee, Xu, Brewer, & Johnson, 2012). It provides an open source, component-based, extensible environment for developing sustainability challenges such as the Kukui Cup and tailoring them to the needs of different organizations. One configures the Makahiki framework to produce a "challenge instance" with a specific set of game mechanics, user interface features, and experimental goals. Makahiki provides sophisticated instrumentation to support evaluation of how well the game mechanics supported the organization's goals for the challenge.

Myths and misperceptions in energy challenge game design

The preceding sections present the design and implementation of the 2011 Kukui Cup along with some basic outcome data in a manner consistent with the way many other energy competitions have been presented. Perhaps the most significant outcome from the 2011 Kukui Cup is the insight that virtually all energy challenges, when viewed as game designs, contain significant flaws including one or more of the following: (1) they do not create a fair game, (2) they do not measure what they think they are measuring, (3) they do not measure the right things, and (4) they do not promote the right outcomes. In this section we discuss how we came to these realizations.

To begin: a fundamental property of energy challenges is competition, and a fundamental property of competitions is the ability to creating rankings. The first myth involves the assumption that the traditional ranking method produces valid (i.e. fair) energy competitions.

Table 1
Total energy consumption ranking method. Winning number in bold.

Team	Actual (kWh)
A	**1,140**
B	1,239

Table 2
Reduction from baseline ranking method. Winning numbers in bold.

Team	Baseline (kWh)	Actual (kWh)	Reduction (kWh)	Percent Reduction
A	1,200	1,140	60	**5.0 %**
B	1,300	1,239	**61**	4.6 %

Myth #1: Percentage reduction from baseline is a fair ranking method

There are two traditional methods for creating rankings in energy competitions: (1) by total energy consumption or (2) by reduction from a baseline.

The first ranking method is simple: order the teams according to their total energy consumption (from least to most) during the competition interval. As illustrated in Table 1, if Team A consumes 1,140 kWh during the competition period, and Team B consumes 1,239 kWh, then Team A wins. This is the simplest method, but produces a fair ranking only if every team is equivalent with respect to the factors affecting their energy consumption. For example, to use this ranking method, every team should have the same number of members. If Team A has 50 members and Team B has 60 members, then Team A's victory might be due to its unfair advantage of having fewer energy consumers. In addition, to use total energy consumption as the ranking method, every team should have the same energy infrastructure. If, for example, Team A lives in a well-insulated building and Team B lives in a poorly insulated building, then Team A again has an unfair advantage. Because it is so difficult to obtain equality among teams with respect to all significant energy consumption factors, and because the unfairness of a competition using this ranking method can be so obvious, the use of total energy consumption is relatively uncommon.

In an attempt to address the fairness problems that arise with the total energy consumption ranking method, the more common approach used by almost all energy competitions is to compute a "baseline" for each team based upon historical usage and then produce rankings based upon reductions from this baseline. The hope is to effectively normalize energy consumption so that teams with differing factors affecting their energy consumption can be fairly ranked. Table 2 illustrates this approach.

As in the prior example, Team A consumes 1,140 kWh and Team B consumes 1,239 kWh. Using energy data collected prior to the competition, a baseline of 1,200 kWh is calculated for Team A and a baseline of 1,300 kWh is calculated for Team B.

Table 3
Adjusted absolute consumption ranking method. Winning numbers in bold.

Team	Baseline (kWh)	Actual (kWh)	Adjusted Actual (kWh)	Percent Reduction
A	1,000	900	900	**10 %**
B	1,300	1,180	**880**	9 %

What is interesting about this example is that there are two winning numbers. Team A wins if *percentage reduction* from baseline is used as the ranking method, while Team B wins if *absolute reduction* from baseline is used as the ranking method. In all of the energy competitions we have seen, it is percentage reduction from baseline, not absolute reduction from baseline, that is used as the ranking method.

The uniform use of percentage reduction from baseline as the ranking method reveals one problem with the use of baselines, because there is no *a priori* reason that percentage reduction from the baseline produces a more fair ranking than absolute reduction from the baseline. To see why this is so, consider a (very plausible) scenario where Team A lives on one floor and Team B lives on an adjacent floor. The two floors are similar in structure and occupancy except that Team B's floor includes a shared laundry room. In this case, the most fair way to produce a ranking is to calculate the energy consumption of the shared laundry room, subtract that consumption from Team B's data, then use absolute values to rank the floors. Put simply, the two teams should be considered equal once the load from the shared laundry room is factored out.

Table 3 illustrates this scenario. Team A reduces their consumption by 100 kWh from the baseline, for a winning percentage reduction of 10%. Team B reduces their consumption by 120 kWh from the baseline, for losing percentage reduction of 9%. But the fairer way to compare these two teams is to simply subtract out the 300 kWh of energy consumed by the laundry room, then compare the two consumptions directly. Under this "adjusted" ranking scheme, Team B wins because their adjusted actual consumption of 880 kWh is less than Team A's actual consumption of 900 kWh.

Of course, the adjusted absolute consumption ranking method is not a panacea for all energy competition situations. If Team A has 50 players and Team B has 70 players, then some sort of percentage-based normalization to create a per capita consumption is required to provide a fair ranking. In many cases, some combination of absolute adjustment (to compensate for structural differences like laundry rooms) and percentage-based adjustment (to compensate for differences in number of players) might be required to produce a fair ranking.

Our conclusion is the following: to produce a fair ranking based upon historical energy consumption, you must determine the reasons behind the differences in consumption by teams, because those reasons are crucial to creating a fair ranking.

Myth #2: Representative energy data can be collected immediately prior to the competition

The preceding section assumes that representative, or *normal* energy usage during the competition can be determined from historical data, and then discusses the problems that arise in using such data to create a fair ranking. Let's now step back and consider the question: is it possible to gather historical energy data immediately prior to the competition and be confident that it reflects representative use (i.e. normal use, as would occur in the absence of the competition).

This question is important because many energy competitions create baselines using the average energy consumption from several weeks immediately prior to the start of competition. Using this recent data (as opposed to data from a year or more in the past) has two significant advantages with respect to obtaining representative data. First, this data is collected from the players who will actually participate in the competition. Second, this data is based upon the state of the infrastructure (buildings, appliances, etc.) as it will be during the competition.

We discovered the problem with this approach after collecting and evaluating energy data from immediately before the 2011 Kukui Cup. Figure 2 shows data from the weeks prior to the Kukui Cup 2011 energy challenge for five of the 20 teams.

What makes this data problematic for predicting normal conditions during the competition are the variety of trends that can be observed in the different teams. The energy consumption of Mokihana-A, Mokihana-B, and Mokihana-D is trending upwards, while the energy consumption of Mokihana-C and Mokihana-E is trending downwards. The change can be substantial: in the case of Mokihana-A, weekly energy consumption varies by approximately 30% during the five weeks prior to the start of the challenge.

If we are to use this data to predict normal conditions during the challenge, we must decide whether the trends represent persistent or transitory changes in consumption. If we assume the trends represent persistent changes, then the most representative data is that collected immediately prior to the start of the challenge. For example, given this assumption, Mokihana-E would compete based upon a baseline consumption of approximately 1300 kWh per week.

If, on the other hand, we assume the trends represent transitory changes in consumption, then a better choice is to compute the baseline from the average energy usage over several weeks prior to the competition. Under this assumption, Mokihana-E would compete based upon a baseline consumption of approximately 1100 kWh. Thus, the choice between persistent and transitory assumptions can make a difference of 15-20% in the baseline value in real world conditions.

All energy competitions that we have seen that have actually published their baseline calculation approach use the averaging method, which implicitly assumes that observed trends are transitory, not persistent. But there is no *a priori* reason to assume that trends are transitory. For example, perhaps a half dozen additional residents moved in to Mokihana-E in late September, creating a persistent increase in energy consumption. In this case, using the average of several weeks to compute the baseline unfairly penalizes Mokihana-E by producing a low baseline value that will not reflect normal conditions during October. Similarly, using the average approach when consumption is trending down in a persistent manner (as might be the case if residents move out or structural improvements are introduced) gives that team an unfair advantage, as their baseline will be abnormally high.

This same problem can occur when seasonal change is taken into account: the amount of energy required to heat a team's residence might be significantly less in the month preceding the competition if the competition is held in the Fall. Similarly, heating-related energy consumption might be significantly more in the weeks preceding the competition if the competition is held in the Spring.

This problem of assuming data collected immediately prior to the competition is representative caused a significant problem with the first Campus Conservation Nationals event (Willens, 2010). According to John Peterson, the "baseline period... was, in some cases, resulting in percentage changes for individual buildings... that were more attributable to changes in weather and other factors than to the choices that students were making in their dorms."

Our conclusion is the following: Use of data collected immediately prior to the competition can predict normal conditions during the competition with limited accuracy. Depending upon the choice of assumptions, baseline values can vary by at least 20% under real world conditions. This issue alone is enough to wipe out (or create) the median energy reductions found in published energy competition results.

Myth #3: Representative energy data can be collected from years prior to the competition.

Given the problems with obtaining representative data immediately prior to the competition for the purpose of creating baselines, an alternative is to use data from prior years. This was in fact the choice made during the first Campus Conservation Nationals event (Willens, 2010), as it was felt that such data would take into account seasonal changes more effectively than data collected immediately prior to the competition. Unfortunately, using data from prior years has significant validity problems when used to represent normal energy consumption during a future year, including:

- The residents of the building are typically different from year to year. Individuals can vary significantly in their energy consumption, and there is no *a priori* reason to believe that these differences simply average out.
- Building infrastructure can change from year to year. HVAC and other energy systems can degrade (leading to more energy consumption in future years) or be updated

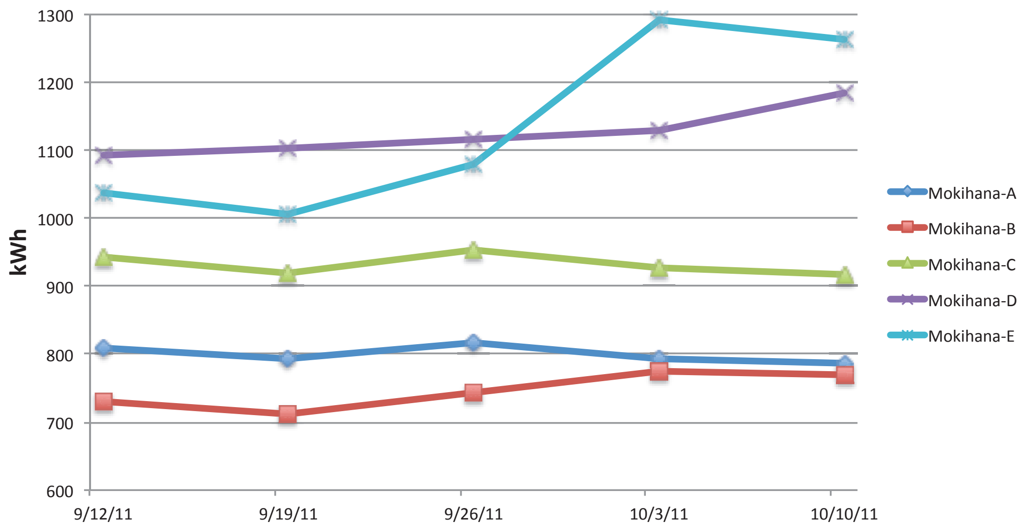

Figure 2. Energy data for five teams for the weeks immediately preceding the 2011 Kukui Cup

(leading to significantly less energy consumption in future years). Similarly, building maintenance (such as weather stripping, lighting upgrades, etc.) can all change the energy consumption needs significantly.

- Weather can vary considerably from year to year. The number of heating degree days and cooling degree days in a given month can change considerably from one year to the next, leading to volatility in energy demand.
- The devices residents use in their rooms can change over time. The trend of students switching from desktop computers to more energy efficient notebook computers might reduce energy use, while the increasing affordability of flat-screen TVs might lead to increased energy use.

As one example, we examined the energy consumption data for the month of October for the past 12 years in the Hale Aloha residence halls, and discovered variability of over 30% from one year to the next. In Hawaii, this variability cannot be attributed to seasonal variation (the Hale Aloha residence halls have neither centralized heating nor cooling) but due to some combination of changes in the energy habits of the residents and ongoing building infrastructure changes.

As another example, consider the outcome data graphic regarding the 2011 Green Cup at University of California, Berkeley (Dhong et al., 2011), shown in Figure 3.

This table shows the reported changes in both electricity and natural gas consumption for 18 fraternity houses during a two month competition. The houses "are ranked based upon reductions in per-capita electricity and natural gas relative to individual chapter baselines". The baselines were based on the "house's previous energy bills for the same two months."

What this report fails to discuss is the huge amount of variation in results among houses. With respect to electricity consumption, results varied from -39% to +24%. With respect to natural gas consumption, results varied from -31% to +71%. The aggregate outcome measure (sum of percentage change in both electricity and gas) varies a staggering amount: from -63% to +66%.

What is the cause of this extreme variability in results? One possibility is that the behaviors of the occupants of houses changed in wildly different manners in response to the competition, with some houses deciding to conserve, and others deciding to increase their consumption. This explanation seems quite implausible. An alternative, more plausible explanation is that their use of baseline data from prior years was not a valid predictor of the normal consumption to be expected in the absence of the competition. If so, one cannot say with any confidence who won and how much was saved.

Our conclusion is the following: Similar to data collected immediately prior to a competition, data collected from prior years can predict normal conditions during the competition with only limited accuracy. The Kukui Cup data suggests that the margin of error can easily exceed 20%. The Green Cup data suggests that the margin of error could be even greater.

Myth #4: Competition data can be used to estimate actual savings

Most energy competitions report on the cost savings generated due to changed behaviors. For example, the Campus Conservation Nationals reported that "savings nationwide totaled 509,000 kilowatt hours, $50,200, and 816,00 pounds of carbon dioxide" (Willens, 2010). The UC Berkeley Green Cup report states that "29,221.17 kWh of electricity and 667.33 Therms of natural gas were saved" (Dhong et al., 2011). The Oberlin competition described in (Petersen et al., 2007) reported savings of $5,107 and 148,000 pounds of carbon dioxide. The 2008 Go For the Green challenge at Western Washington University reported savings of $7100 dollars and an emissions reduction of 100,000 pounds of carbon dioxide (Mauney, 2008).

Unfortunately, our analysis calls into question the assump-

Competition Rank	Council	Chapter	Electricity use % change	Natural gas use % change	Sum % change
1	IFC	Phi Gamma Delta (Fiji)	-39.66	-23.81	-63.47
2	IFC	Kappa Alpha (KA)	-38.38	-23.14	-61.52
3	IFC	Tau Kappa Epsilon (TKE)	-36.04	-19.55	-55.59
4	PHC	Chi Omega	-19.94	-28.79	-48.73
5	IFC	Delta Upsilon (DU)	-3.73	-31.02	-34.75
6	IFC	Alpha Epsilon Pi (AEPi)	-30.30	-4.24	-34.53
7	IFC	Theta Delta Chi (TDX)	-14.67	-18.57	-33.24
8	IFC	Sigma Alpha Mu (Sammy's)	-13.37	-18.56	-31.94
9	PHC	Kappa Alpha Theta	-13.70	-17.96	-31.66
10	PHC	Kappa Kappa Gamma	-9.02	-17.81	-26.83
11	PHC	Delta Delta Delta	-9.52	-6.76	-16.28
12	IFC	Alpha Tau Omega (ATO)	-8.99	3.21	-5.78
13	PHC	Alpha Phi	0.22	-1.36	-1.13
14	MCGC	Alpha Delta Chi	21.29	-7.24	14.05
15	PHC	Alpha Chi Omega	24.75	2.78	27.53
16	PHC	Delta Gamma	3.19	26.42	29.60
17	PHC	Alpha Delta Pi	-14.17	52.81	38.64
18	PHC	Gamma Phi Beta	-4.20	71.04	66.84

Figure 3. Outcome data for the 2011 UC Berkeley Green Cup

tion underlying these claims, that it is possible to gather energy data that can serve as representative of the consumption to be expected during the competition.

Unfortunately, if it is not possible to generate representative consumption with any accuracy, then it is not possible to estimate actual savings with any accuracy. This is because estimated savings are always calculated by assuming that the baseline data not only puts teams on an equal footing with each other for the purposes of the competition, but also represents an accurate prediction of what each team would have consumed during the time interval of the competition, had the competition not taken place.

It is interesting to note that these two applications of baseline data are independent: it is plausible to design baselines that succeed in creating a level playing field among teams but that do not predict what the energy consumption would have been in the absence of the competition.

Our conclusion is the following: virtually all published claims for dollar savings due to energy competitions could not withstand a detailed analysis of the assumptions upon which these claims were made.

Myth #5: The "good guys" win

One intuitive assumption concerning energy competitions is that the winners will be those who are most energy conscious. In reality, under typical competition conditions (i.e., collection of baseline data immediately prior to the competition, and the use of percentage reduction from baseline for ranking), those teams who are most energy conscious prior to the competition might be at a distinct disadvantage.

This result is due to energy conscious teams being more likely to have have been practicing conservation oriented behaviors prior to the competition period, and thus their baseline energy use will be lower than their "energy hog" peers. This energy efficient behavior makes it harder for them to compete on the basis of percentage reduction, since they have less "fat" (from an energy perspective) to cut from their consumption. As a concrete example, we discovered that it is not uncommon for residence hall rooms to have two refrigerators and a few even had three refrigerators (Brewer, 2012). For those residents, simply turning off one refrigerator for the duration of the competition can achieve significant reductions in consumption, whereas more ecologically conscious residents who chose to use only one refrigerator would have to forego the refrigerator entirely to obtain the same reduction.

Our conclusion is the following: although the good guys do not necessarily win, this is not necessarily a bad thing. From a game design perspective this might seem unintuitive and unfair, but from a serious game design perspective is it potentially positive. It is possible that giving positive reinforcement to profligate energy users might be more effective than rewarding those who have already bought into energy conservation principles. What is missing is an understanding of whether this phenomenon is occurring and what the effectiveness of rewarding this group actually is.

Myth #6: Competitions encourage sustainable behavior change

It is an implicit assumption of all collegiate energy competitions that students will acquire sustainable, energy conscious behaviors as a result of participation. If this assumption did not exist and the goal was only to reduce energy consumption during the competition period, then far easier approaches could be implemented (for example, rolling blackouts.)

What is remarkable about published reports of energy competitions is the lack of evidence for sustained, positive behavior change. Instead, reports indicate unsustainable changes, from unplugging vending machines (Petersen et al., 2007) to camping outside (Hodge, 2010) to *(cite other un-*

sustainable behaviors here). With few exceptions, reports do not track or analyze anything about what happens after the competition ends, even though this is perhaps the most important behavioral question of all.

Our conclusion: We believe this absence of post-competition information is mostly due to the student-run nature of collegiate energy challenges, where all time and energy is focused on the event itself, and once the event is over, there appears to be nothing more to be done. To fix this problem, student organizers must become aware of the importance of post-competition monitoring and assessment and allocate their time and resources accordingly.

Myth #7: Energy competitions measure the right thing

We believe that measuring *percentage change from baseline* is so widespread because it appears to have so many appealing properties. First, measuring in-competition energy consumption energy use is extremely easy, as is the collection of pre-competition data. Second, using percentage change from baseline as a ranking method appears more valid than ranking according to absolute energy use. Third, it is straightforward to convert reduction values into dollars saved by multiplying by the local cost per kilowatt-hour.

The fact that this approach is easy does not make it appropriate for at least two reasons. First, as the preceding myths have discussed in detail, there is considerable evidence that the numbers produced are not accurate: baseline data as collected for energy competitions do not appear to be valid as predictors of normal consumption in the absence of the competition during the time intervals of interest.

Second, measuring and reporting only consumption data during the competition misses the most meaningful data of all: the behaviors and energy consumption that occurs after the competition is over. For example, if consumption simply returns to its precompetition level, then the impact of the competition is unclear. Furthermore, it incentivizes unsustainable, short-term, radical changes in behaviors which students undertake only because they know they can cease to follow them within a short period of time.

Our conclusion: the almost exclusive focus of collegiate energy competitions on measuring percentage reduction from baseline during the competition leads to (1) invalid claims regarding savings and (2) an incentive to engage in unsustainable behaviors.

Supporting more meaningful play through improved energy challenge game design

We claim that the effectiveness of energy competitions can be improved if they provide better opportunities for both meaning (in the sense of sophisticated insight as well as impact upon future behavior) and play (in the sense of providing fair competitions as well as providing other, non-competitive forms of play). We conclude with our recommendations for game design "fixes" resulting from our experiences with the 2011 Kukui Cup and analysis of other energy competitions.

Fix #1: Don't compete on kWh

Perhaps the single most important fix for future energy challenge game design is to abandon *kWh consumed* as the primary outcome measure. As an example of an alternative, the Kukui Cup implements a competition in which players and teams compete based upon a point system, where points can be earned for successfully answering questions about energy videos, attending workshops and excursions, and other sustainability-related activities. The 2011 Kukui Cup also had a competition based upon kWh consumed, but because the infrastructure for the Hale Aloha residence hall floors is so similar, we measured absolute energy consumption rather than a percentage reduction from baseline.

Avoiding competitions based upon kWh consumed helps to address at least the first four myths.

While we advocate against competition based solely upon kWh, we do not imply that measurement of consumption has no use. In fact, measurements of kWh consumed still have a very important role: as empirical feedback to help players understand and assess the impact of structural and behavioral change. Energy competition may also lead to introspection about energy use, even if the quantitative results of the competition are not particularly meaningful.

Fix #2: Use long time intervals

Hodge states that one of the five core components of collegiate energy competitions is "a short timeframe", and that this "can be a catalyst for students to go all out, which in turn creates hype and energy" (Hodge, 2010). Based upon our analyses, we believe that a short time frame encourages exactly the wrong behaviors: short-term, unsustainable changes that the players have no intention of retaining after the competition ends. Research on the effectiveness of energy feedback devices on energy conservation indicates that changing habits can require effort over three months (Darby, 2006).

Instead, we recommend that future energy challenge games last a minimum of four months: too long a time for students to camp outside, unplug their vending machines, and so forth. By analogy, the strategies for winning an energy "marathon" are very different, and much more likely to be sustainable than the strategies for winning a "sprint".

We believe this fix helps to address Myth #6: competitions encourage sustainable behavior change.

Fix #3: Use dynamic baselines

As discussed in depth above, we believe that current approaches to baseline computation are, in most cases, irreparably flawed. Yet providing an objective measure of progress is very useful to creating effective meaningful play.

We claim that the problem with current approaches to baselines is that they are used to both measure progress with respect to behavior change as well as an estimate of what normal consumption would have been in the absence of the challenge.

We recommend that future energy challenges address these two issues independently. For example, an alternative type of baseline that can be used to measure progress is a dynamic baseline based upon a "sliding window" of average consumption for each team over the preceding few weeks, and the goal is to simply stay under this baseline of the recent past, where the recent past can and will include at least a portion of the competition interval itself.

The use of a sliding window means that the baseline is no longer static, but rather changes throughout the competition, and thus incentivizes sustainable changes (because each reduction in consumption is incorporated into the baseline measurement and thus impacts on the goal behavior.) This aspect helps to address Myth #6: competitions encourage sustainable behavior change.

The second impact of a sliding baseline is its obvious inapplicability as an estimator for normal consumption in the absence of the challenge. This aspect helps to address Myth #4: competition data can be used to estimate actual savings.

Fix #4: Measure both micro and macro behavioral change

Traditional energy competitions tend to focus on what we call *micro* behavioral changes, such as turning off the lights when you leave the room, enabling the screen saver on your computer, and so forth. These are meaningful changes, but there exists another level that we call *macro* behavioral change. Examples of macro behavioral change include: choosing to put off buying a car until a hybrid or EV can be obtained; choosing to enroll in a course on sustainability or energy issues; choosing a degree program related to energy or sustainability; choosing political candidates to vote for based (in part) on their energy views; and choosing a job (during or after school) based upon the work's impact on energy or sustainability. Micro behaviors tend to involve habits, while macro behaviors tend to involve decision making.

Considering macro behavioral change precipitates a significant shift in thinking about almost every aspect of an energy competition. For example, in the 2011 Kukui Cup, we held a workshop entitled "Your Sustainable Future" in which the first year students were able to meet and talk with professors of Electrical Engineering, Environmental Studies, and Computer Science, as well as with the Vice Chancellor in charge of sustainability regarding degree opportunities in renewable energy at the University of Hawaii. Instead of monitoring energy consumption over a period of a few weeks or months, a focus on macro behavioral change leads to the design of longitudinal research to investigate the impact of the program on students over a period of years.

Creating macro behavioral change will be extremely difficult, and we do not claim to have made much significant progress toward it within the Kukui Cup. But we believe that any consideration of macro behavioral change within the design of the competition improves the potential impact of the event, and will help address Myths #6 (competitions encourage sustainable behavior change) and #7 (energy competitions measure the right thing).

Fix #5: Do challenges, not competitions

We call the Kukui Cup a "challenge" rather than a "competition" in order to emphasize change within individuals and teams as opposed to triumphing over others. As discussed above, to enable a competition, one must provide a ranking method. But rankings and competitions become relatively meaningless when considering, for example, macro behavioral change: how can one create a ranking method that decides whether or not "joining an environmental group" is better or worse than "voting for a candidate with progressive energy views"?

This is not to say that competitions aren't fun and engaging, but we believe they should be just one component of the overall challenge. This fix will address Myths #7 (energy competitions measure the right thing).

Fix #6: The most meaningful data become available later

When planning future energy competitions or challenges, we recommend allocating time and resources to observing the impact of the event on participants after the event is over. For example, continue to monitor energy consumption to see if usage rebounds to its pre-competition level. Interview and/or survey students at some time period following the event to learn about any persistent impact of the event on their behavior. If possible, do not depend only on self-reported data, but attempt to independently verify if changes have occurred. If participants self-report that they are now washing loads of laundry in cold water, see if that can be validated by looking at records of hot water usage in the laundry room (or even spot checks of the washing machine settings).

Once again, we are advocating a broadening of focus. While in-competition behavior is important and interesting (because the competition is where *initial* behavioral change occurs), it is equally if not more important to learn about post-competition behavior (because that is where behavioral change is *sustained*. This will help to address Myth #7 (energy competitions measure the right thing).

Fix #7: Do science; be skeptical

In the majority of the energy competitions that we reviewed, we were struck by the lack of rigor with which data and analyses were performed and the unsupportable claims regarding energy reductions and cost savings. We believe this arises from a situation in which well intentioned people (volunteer students, usually) put tremendous time and effort into an activity with clearly desirable social outcomes. Although everyone (ourselves included) wants a positive, successful outcome from all of these efforts, progress depends upon a thoughtful, critical analysis of how each instance of this particular type of meaningful play is designed and what can be said with confidence about the outcome. For example, instead of focusing only on the winners, it is important to also consider the losers, and what their experiences and data reveal about the competition. Rather than simply report correlations (we held a competition and at the same time energy consumption appeared to decrease for certain teams and increase for others), it will be far more useful to the community if organizers would attempt to discover why those results were obtained. As just one example, it would be very helpful to understand why some of the UC Berkeley fraternity houses appeared to decrease consumption by 63% and others appeared to increase consumption by 66%.

It is in this spirit that we present this paper: not to discredit or embarrass past efforts, but rather to help build a firmer foundation for future designs. There is much good to be learned from the many collegiate energy competitions that have been held to date, and we can progress even faster in future with a more scientific, skeptical attitude towards our outcomes. This fix helps to address all of the Myths.

References

Ammons, J. (2010, July). *In the classroom, a lack of energy education.* Indy Week. Available from http://www.indyweek.com/indyweek/in-the-classroom-a-lack-of-energy-education/Content?oid=1547796

Becker, L. J. (1978). Joint effect of feedback and goal setting on performance: A field study of residential energy conservation. *Journal of Applied Psychology*, *63*(4).

Brewer, R. S. (2012, Jan). *Results from energy audit of Hale Aloha* (Tech. Rep. No. CSDL-11-12). Department of Information and Computer Sciences, University of Hawaii, Honolulu, Hawaii 96822. Available from http://csdl.ics.hawaii.edu/techreports/11-12/11-12.pdf

Brewer, R. S., & Johnson, P. M. (2010, October). WattDepot: An open source software ecosystem for enterprise-scale energy data collection, storage, analysis, and visualization. In *Proceedings of the first international conference on smart grid communications.* Gaithersburg, MD. Available from http://csdl.ics.hawaii.edu/techreports/10-05/10-05.pdf

Darby, S. (2006). *The effectiveness of feedback on energy consumption* (Tech. Rep.). Environmental Change Institute, University of Oxford.

DeWaters, J. E., & Powers, S. E. (2011). Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior. *Energy Policy*, *39*(3), 1699–1710. Available from http://www.sciencedirect.com/science/article/pii/S0301421511000073

Dhong, C., Walker, A., Seacat, S., Kaenel, M. von, Doyle, K., & Smith, P. (2011, December). *Greek energy savings competition and outreach campaign– "green cup".* The Green Initiative Fund. Available from http://tgif.berkeley.edu/index.php/grants/projects/2011-projects/28-greek-energy-savings

Durr, E. (2010, July). *Elon university strives to reduce energy consumption through behavior change.* Lucid Design Group Case Study. Available from http://www.luciddesigngroup.com/casestudies/elon.php

Freedman, J. L., & Fraser, S. C. (1966, August). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195–202. Available from http://psycnet.apa.org/journals/psp/4/2/195/

Hodge, C. (2010, November). *Dorm energy competitions: Passing fad or powerful behavior modification tool?* Presentation at the 2010 Behavior Energy and Climate Change conference. Available from http://www.stanford.edu/group/peec/cgi-bin/docs/events/2010/becc/presentations/2C_ChelseaHodge.pdf

Lee, G. E., Xu, Y., Brewer, R. S., & Johnson, P. M. (2012, January). *Makahiki: An open source game engine for energy education and conservation* (Tech. Rep. No. CSDL-11-07). Department of Information and Computer Sciences, University of Hawaii, Honolulu, Hawaii 96822. Available from http://csdl.ics.hawaii.edu/techreports/11-07/11-07.pdf

Mauney, K. L. (2008). *The effects of the Go for the Green challenge on electricity use, behaviors, and attitudes of Western Washington University residents*. Unpublished master's thesis, Western Washington University.

McKenzie-Mohr, D. (2009). *Fostering sustainable behavior: Community-based social marketing*. McKenzie-Mohr & Associates, Inc. Available from http://www.cbsm.com/pages/guide/preface

Pallak, M. S., Cook, D. A., & Sullivan, J. J. (1980). Commitment and energy conservation. In (Vol. 4, pp. 352–370). Sage Publications. Available from http://books.google.com/books?hl=en&lr=&id=TFsnqmk97csC&oi=fnd&pg=PA352&ots=KDxy1IG6AJ&sig=9pkY_mTwoyDE7tLkJx_Ory6nzE8

Petersen, J., Murray, M. E., Platt, G., & Shunturov, V. (2007, September). Using buildings to teach environmental stewardship: real-time display of environmental performance as a mechanism for educating, motivating, and empowering the student body. In *Proceedings of greening the campus vi.* Muncie, Indiana. (http://www.oberlin.edu/faculty/petersen/ColorPrint/Petersen2007BuildingsToTeachStewardship.pdf)

Willens, A. (2010, December). *Dorm energy resource competition succeeds despite challenges.* The Oberlin Review. Available from http://www.oberlinreview.org/article/dorm-energy-resource-competition-succeeds-despite/