CITE Suitability: An Exploration of Product Evaluation Methodologies for Developing World Technologies

by

Christopher Anthony Pombrol

B.S. Mechanical Engineering
California Institute of Technology, 2012

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author ....
Department of Mechanical Engineering
May 9, 2014

Signature redacted

Certified by ....
Daniel D. Frey
Professor of Mechanical Engineering and Engineering Systems
Thesis Supervisor

Signature redacted

Accepted by ....
David E. Hardt
Ralph E. and Eloise F. Cross Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Theses

CITE Suitability: An Exploration of Product Evaluation Methodologies for Developing World Technologies

by

Christopher Anthony Pombrol

Submitted to the Department of Mechanical Engineering
on May 9, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

There are a multitude of technological products that have been developed to improve the lives of bottom of pyramid consumers in the developing world. Unfortunately, many of these products fail to have the desired impact and there is a serious gap in knowledge of what works and what does not work. It is the goal of the Comprehensive Initiative for Technology Evaluation to fill this gap by creating a methodology for evaluating such products. This thesis documents the first efforts of the Suitability team, as part of the Comprehensive Initiative for Technology Evaluation, to develop these product evaluation methodologies. The Suitability team is focused on evaluating the technical performance of these products in a comparative manner with respect to user expectations and use patterns. Two different product evaluation frameworks were developed, implemented, and compared. The first is an attribute-based product evaluation, in the style of *Consumer Reports*, in which all products under evaluation are subjected to a standardized set of laboratory tests designed to differentiate among products in key performance areas. The second is a problem-based evaluation in which user feedback is gathered in order to guide product testing. Product evaluations were produced using both methods. Both user groups ranked the importance of solar lantern characteristics similarly, with a Kendall's τ coefficient of 0.4545. Each method also had unique advantages, revealing different kinds of information on solar lantern performance. This suggests that it may be useful to employ both methods simultaneously, with the problem-based evaluation informing the attribute-based evaluation.

Thesis Supervisor: Daniel D. Frey
Title: Professor of Mechanical Engineering and Engineering Systems

Acknowledgments

First, I would like to acknowedge USAID and the CITE program for providing me with the opportunity to work on this project as well as financial support for this research as well as my studies at MIT.

Next, I would like thank my thesis advisor, Dan Frey. He provided insightful and thoughtful feedback and guidance throughout the project. He was always made himself readily available to me for meetings and consultations.

I would also like to thank Derek Brine for helping to guide this research project and providing day to day support. I would also like to Dr. Jeff Asher for engaging with us and sharing his detailed knowledge with us.

I am also thankful to all those in the CITE program and in D-Lab who have helped along the way. I have had many informative and thoought provoking converstations with many members of the CITE and D-Lab family. I would also like to thank Jack Whipple for opening his shop to me and helping with building various things for the project. I would like to thank Amit Gandhi and Victor Lesniewski for their continued support and friendship. Finally, I would like to thank all those who took the time to participate in the pilot crowd-sourced evaluation.

Lastly, I would like to thank my family for their support and love.

Contents

List of Figures

List of Tables

Chapter 1

Background and Purpose

This section will provide the reader with a brief background of the Comprehensive Initiative on Technology Evaluation (CITE), focusing specifically on the Suitability team. The text in this section will explain the history, main goals, and structure of the program.

1.1 Overview of CITE

CITE was formed as part of USAID's Higher Education Solutions Network (HESN). HESN is a network seven of universities that have partnered with USAID in order to create a system of international development labs. The purpose of HESN is to facilitate the development and deployment of technology-based solutions to development problems.

Within HESN, CITE was formed in response to a gap in current technology evaluation methods which has impeded innovation and diffusion of technology [1]. There is no shortage of solutions being created to solve developing world problems. The problem lies in the fact that despite this abundance of available technologies, there is very little assessment of what actually works. This has resulted in a multitude of failed products and products that have not had the desired impact. For example, many solar lanterns continue to perform below expectations, despite the presence of standards organizations such as Lighting Africa [2].

CITE's purpose is to develop evaluation protocols to assess the efficacy of various technological products designed for the developing world. These evaluation protocols will be used to produce technology evaluation reports, which will be distributed to NGO's, donors, and partner organizations in order to help them make decisions about appropriate products to invest in. In addition, in the course of producing these technology evaluation reports, CITE will identify the weaknesses and gaps in existing technologies as well as areas for improvement, allowing development agencies to target these areas for investment.

In order to perform complete technology evaluations, CITE is using what we call a "3S framework." This framework is designed to take into account all the complexities of designing appropriate and effective technologies by assessing products in three different areas: Suitability, Sustainability, and Scalability. The driving principle behind this framework is that a product must satisfy a multitude of criteria in order to be successful, including technical performance and user acceptability, the ability to reach consumers on a large scale, and ability to affect positive change over the product life-cycle. It is not enough to simply develop a technology that performs well or a technology that is highly affordable, the technology must be appropriate in many different ways.

The Suitability team is focused on the technical performance with respect to user expectations and use patterns. Technical performance will be established by testing products in a laboratory setting. In contrast to industry standards, however, the Suitability team will seek to test products in way that aligns with the actual use patterns of the product. User acceptability will be evaluated by performing field evaluations in which data are gathered on how users perceive and use these products. The data from these two efforts are combined to determine which products have adequate technical performance and are acceptable to users.

The Scalability team is focused on a product's ability to reach sufficient scale to affect a large number of users in the developing world. To this end, the Scalability team will map and analyze a product's supply chain. This will allow the Scalability team to determine the potential of certain products to be scaled up. Furthermore, the

Scalability team will be able to identify strengths and weaknesses in various supply chains and suggest improvements for future products.

The Sustainability team is focused on product adoption and use over the long term. The Sustainability team will utilize complex systems analysis in order to characterize the diffusion of new technologies in the developing world. This understanding of the dynamics of technology adoption and continued use will be used to identify products and programs that are most likely to succeed in the marketplace. In addition, future technologies can be designed with adoption in mind.

The results of the three team's evaluations will be combined into a single technology evaluation report, combining a wide array of factors relevant to the success and impact of technologies designed for the developing world. These reports will identify technologies that work and should be invested in. They will also reveal weaknesses and gaps in existing technologies and suggest specific areas in which they can be improved. Lastly, requirements for future technologies can be discovered in the process of understanding the state of current technologies and why they succeed or fail.

1.2 Context: Developing World Products

Evaluating products designed for the developing world is very different, and in some ways much more difficult, than evaluating products designed for consumers in the developed world. Yet this is a very important task and one that has the potential to impact a large number of people. Development projects are focused on people at the bottom of the pyramid. According to the United Nations, there are currently approximately 1.2 billion people worldwide living in extreme poverty, defined as those making $1.25 per day or less.

Those living in poverty often do not have access to a vast array of information in the way a consumer in a developed country would. This makes it harder for them to make informed purchase decisions. Making a bad purchase decision then has much larger consequences for these vulnerable consumers. For example, a consumer is living on $1.25 per day and decides to buy a solar lantern, which may cost in the range of

$10-$20. If this consumer makes a purchasing mistake and buys a lantern does not suit his or her needs or does not work well, this can lead to economic hardship.

There are other challenges unique to the developing world. In the developed world, it is often very easy to obtain information regarding consumers and products. In the developing, this may be much more difficult. Consumer surveys cannot be conducted as easily because of the difficulty of reaching a large number of consumers. Additionally, consumer characteristics and preferences and product use patterns may vary across geographic areas, such as from country to country or from village to village.

Despite these hurdles, however, these new technologies have the potential to make a very large impact on the lives on those at the bottom of the pyramid. They are often exposed to hazards which are very uncommon in the developed world. For example, many families in rural areas may rely on burning charcoal in order to cook food, which can result in exposure to air pollution. Improved cookstove designs can potentially alleviate this problem at very little cost.

Technology evaluations of products designed to help those at the bottom of the pyramid in the developing world pose unique challenges when compared with the same task in the developed world. However, performing this work is very important and has the potential to bring large benefits to these vulnerable populations. These issues are at the forefront of CITE's effort to create technology evaluation methodology.

1.3 CITE Goals: Impact on the Developing World

The goal of CITE is improve the lives of those consumers at the bottom of the pyramid in the developing world. CITE will achieve this goal by influencing the market of technologies available to these consumers. First, CITE will seek to understand the current market of available technologies by performing technology evaluations. CITE will evaluate products and determine which products work and which do not. This information will be used to build a database of technologies and the contexts in which they are most effective.

This information will be passed on to NGO's, partner organizations, and procurement agencies. CITE's technology evaluation reports will serve as guide to these organizations about which technologies are most promising and are deserving of additional investment. This will result in an increase in the distribution of products that are effective in the relevant context. Furthermore, a better understanding of consumer needs will drive the development of improved products and technologies in the future.

CITE's end goal is that through these efforts, the array of products and technologies available to bottom of the pyramid consumers is improved. Products and technologies are developed that are well aligned with their needs and preferences. Part if this goal is also to improve the efficiency of donor organizations and aid agencies. Development dollars will be spent more effectively if they are focused on effective technologies and interventions.

1.4 Theoretical Underpinnings of Product Evaluation

The topic of product evaluation has been previously explored. The goal of such evaluation is usually to uncover the "quality" of product, as consumers prefer "high quality" products. These "high quality" products are precisely what CITE would like to identify and promote in the developing world. This notion of quality is, however, difficult to define. There five major approaches to quality: the transcendent approach, the product-based approach, the user-based approach, the manufacturing-based approach, and the value-based approach [3].

First, the transcendent approach defines quality as "innate excellence" that can be universally recognized, yet cannot be precisely defined and is an "unanalyzable property that we learn to recognize only through experience" [3]. Quality in this sense reflects a certain, unspecified standard or achievement and would appear to be difficult to quantify or even measure. Given the vagueness of this definition and the difficulty of measuring "innate excellence," CITE did not pursue this definition further.

Second, according to the product-based approach, quality can be measured precisely [3]. Quality is derived from some attributes of a product, each of which a product possesses in varying and measurable quantities [3]. This very rigid definition allows products to be ranked according the amounts of each desirable attribute they possess. For example, solar lanterns could be ranked based on how many LED's they had. However, this definition does not specify a method of combining values for multiple attributes and assumes that all consumers find all attributes desirable to the same extent.

Third, the user-based approach defines quality from the perspective of a product's end user. Each user is assumed to have specific needs and preferences. Products that best meet the needs of each specific user will then be regarded as the "highest quality" product by that user. This approach has an aggregation problem. This approach makes it difficult to generalize results to a large group of users. In reality, many users, especially users that share common characteristics such as geographical location or income level, probably have similar needs and preferences.

Fourth, the manufacturing-based approach is rooted in quality control. That is, the most important attribute in determining quality is conformance to specifications [3]. This approach has an emphasis on production and engineering, appropriately enough, but it may not be representative of what a consumer might regard as quality. The manufacturing-based approach recognizes that consumers do indeed have an interest in well-made products; however, this is just one of many attributes that a consumer may be interested. Because this definition of quality seems to lend itself more towards industry standards organizations than consumer focused product evaluations, CITE did not pursue this definition further.

Finally, the value-based approach combines two distinct ideas: performance and price [3]. A "high quality" product is thus one that can provide the highest performance at the lowest price. CITE did not incorporate this approach into its initial product evaluation because of the difficulty of determining price. For example, in some parts of the developing world, haggling is commonplace, and two different customers might pay different prices for the same product from the same seller. Additionally, some products are sold at higher prices in the developed world, prices that are unrepresentative of what a developing end user might actually pay. In the future, this may be an important area for exploration, due to the low income levels associated with the bottom of pyramid users that CITE is focused on.

The Suitability team chose to use a combination of the product-based approach and the user-based approach to quality. This has precedent in the marketing community, which represents a product as a bundle of attributes [4]. Product users derive utility from a product through some combination of these product attributes. This takes into account variability in needs and preferences that individual users might have, but also assumes that a group of users in a similar context will have sufficiently similar needs and preferences such that a large majority of the group will consider the same set of products to be of "high quality."

In order to properly evaluate a product, two steps must be completed. First, consumer needs and preferences must be identified. Second, product attributes, and the extent to which they satisfy consumer needs and preferences, must be measured. These steps can be accomplished through either user-centered design methods or expert-based evaluation.

In user-centered design, the user is actively involved in the design process of the product [5]. During the design process, actual users are asked to evaluate a product or prototype. These users are given representative tasks to perform with a product. They are observed and quantitative performance data, such as the amount of time required to perform a task, are recorded. Users are also interviewed to obtain qualitative data on the product. The information gained from this type of evaluation can positively influence product design, because consumer preferences and performance metrics that may not be considered by the product designer are revealed. User evaluation of products has the disadvantage of being resource intensive. Users must be recruited, evaluations must be scheduled, and qualitative data must be mined. Additionally, there may be some doubt as to the realism of the tasks presented to users.

Expert-based evaluations can include guideline reviews, cognitive walkthroughs, and heuristic evaluations. In guideline reviews, aspects of a product's design are evaluated individually in the absence of a task scenario by an expert according to established standards. On the other hand, cognitive walkthroughs focus on learnability and ask experts to walk through a series of actions with the user interface and to evaluate if a user will be able to complete the task. Heuristic evaluation engages a small number of evaluators in assessing a product according to understood design principles. In all of these analytic methods, experts must accurately predict the problems that users might experiences. If experts are not successful in their predictions, the results may lack generality [6].

In addition, expert evaluations may not be relevant for all classes of products. Users often prefer direct observation or experience with the product over secondary information [7]. In addition, consumers only utilize impersonal, independent sources of information, such as *Consumer Reports* when there is a high technical performance risk associated with the product [7]. Since CITE will have similar characteristics as *Consumer Reports* in this sense, product selection will be very important to the success of CITE evaluations.

In summary, the Suitability team was influenced by the product-based and user-based approaches to product quality. The Suitability team would attempt to uncover both user preferences and needs, as well as product attributes and the extent to which products satisfied users. The evaluation methods employed would be a combination of user-based evaluation through solicitation of information from users and expert-based evaluation through lab testing. Product selection would also be important because of the way consumers perceive organizations like CITE.

1.5 Purpose of This Thesis

CITE has now completed evaluations of its first product family: solar lighting devices in Uganda. At the time of this writing, CITE is now proceeding with evaluation of its second product family: water treatment devices in India. A large, diverse, multi-disciplinary team at MIT contributed a large amount of time and effort in order to make this project a reality. The CITE team includes faculty, research staff, and graduate and undergraduate students from different laboratories, departments, and organizations across campus.

This thesis will focus on the work of the Suitability team, of which the author is a member. The Suitability team specifically consisted of a number of faculty and staff, as well as three graduate students. As such, a large portion of the team is transient. Graduate students may only be on the team for two years or less, after which time they may graduate or move to another project. Additionally, although specific test procedures and results were documented in great detail in reports submitted to the sponsor, internal discussions and decision making frameworks have not been well documented. Thus, it may be difficult for future students to trace the development of the Suitability methodology and understand the framework on which it has been based.

This thesis has two main purposes. The first purpose of this thesis is to document the first efforts of the CITE Suitability team to establish a framework for evaluating developing world technologies. This will expand on the detailed test procedures and results that have already been documented in detail and present the decision-making frameworks and assumptions that were employed to arrive at these specific methodologies. The second purpose of this thesis is to examine the assumptions that have been utilized in applying a developed world process to the developing world. Separating those assumptions that remain applicable from those that do not will result in better technology evaluations in the future.

Chapter 2

The Suitability Team and Consumer Product Testing

This section provides additional background on the CITE Suitability team. In addition, it examines product evaluation methods in the developed world.

2.1 The Suitability Team

The CITE Suitability consisted of a team drawn from the faculty, staff, and students of the Mechanical Engineering Department at MIT. For the solar lantern evaluation team members included three graduate students: Amit Gandhi, Victor Lesniewski, and the author; one faculty: Dan Frey; and program manager: Derek Brine. In addition, the team had access to a consultant from *Consumer Reports*, Dr. Jeffrey Asher, who had worked extensively on consumer product evaluation. This group worked closely as a team to produce the Suitability portion of the first CITE product evaluation.

The focus of the Suitability team is on the technical performance and user acceptability aspects of a product. The end goal of the Suitability team was to compare the technical performance and user acceptability attributes of similar products. This required the Suitability team to make decisions about the methodology that would be used to determine which products to evaluate, which attributes to test, and how to test.

For the initial evaluation, the methodology utilized was informed and influenced by *Consumer Reports* method. This method focuses on testing a product based on its observed use patterns. Thus, the *Consumer Reports* method is user-centric, and its focuses on user needs and preferences rather than on manufacturer claims and industry standards. The *Consumer Reports* method of product evaluation is discussed in more detail in the next two sections.

In order to perform laboratory testing of the technical aspects of product performance, the Suitability team was given access to laboratory space in MIT's D-lab. For user acceptability aspects, Suitability team members traveled to the field in order to gather information on end users and product use patterns.

2.2 Overview of Consumer Product Testing in the Developed World

Consumer product testing is not a new concept in the developed world. There are a number of prominent organizations that perform this service. Examples include *Which?* Magazine and *Good Housekeeping*. Perhaps the most well-known of these organizations is *Consumer Reports*, which was founded in 1936. Since its inception, *Consumer Reports* has grown into a large and well-funded organization. *Consumer Reports* is headquartered in Yonkers, New York and has Consumer Advocacy Offices in Washington, DC; Austin, Texas; and San Francisco, California. In addition, there is an associated 320 acre Auto Test Track in Eastern Connecticut. *Consumer Reports* has yearly expenditures of around $250 million [8] for its activities and employs approximately 600 staff. It currently has over 7.3 million subscribers [9] and tests approximately 100 products every year.

Thus, it can be seen that consumer product testing is a large and well-developed multi-million dollar industry in the developed world. A large amount of resources is devoted to this testing, with large staff and extensive facilities. Despite this however, the information resulting from these efforts is available at relatively low cost when compared to the income of the average consumer in the developed world. Furthermore, Information from these sources is very easy to obtain either through a magazine subscription or on the web.

Consumer Reports and other consumer organizations have had great success performing product testing in developed countries. Many consumers rely on these organizations for information to make their purchase decisions. There are also many examples of these organizations identifying unsafe products and influencing profound changes in product markets.

Consumer product testing organizations generally practice user-focused, attribute-based, comparative product evaluations. This product testing framework will be discussed in the next section.

2.3 The Attribute-based, Comparative Product Testing Process

This section describes the user-focused, attribute based, comparative product evaluation method practiced in the developed world by organizations such as *Consumer Reports*. The Suitability team chose to use this framework as a starting point in developing world product evaluations because of its effectiveness in driving the market towards better products in the developed world. In addition, this method of product evaluation is already well understood in the developed world context.

This chapter is substantially based on the white paper (Brine, et al, 2014) submitted to USAID. There are significant additions and modifications based on the author's involvement in the work.

2.3.1 Overview of Attribute-based, Comparative Product Testing

The attributed-based, comparative product testing method is used to understand the differences in technical performance between a set of similar products, called a **product family**. This set of similar products generally shares many common characteristics and generally solves a challenge that a consumer may have in a similar way. The individual products within product family are referred to as **models**. Each model may be made by a different manufacturer but is available to consumers on the mass market. The mix of models available in region as well as their respective market shares constitute that region's **marketplace**.

Within each product family, a set of product attributes can be identified. **Product attributes** are a set of testable characteristics that are common to all models within the product family, and are directly relevant to the function the product is designed to perform. In contrast, **product features** are characteristics that may not be common to all members of the product family and are not essential to the product's function. Nonetheless, these product features are important because if well-designed, they may increase the usefulness or desirability of a product. On the other hand, a poorly designed feature may make a product harder to use and less desirable.

Once the relevant product attributes and features have been identified, a **test protocol** must be developed. Product performance is generally tested in all attribute categories identified. In feature categories, there is usually no testing. Features are simply identified as being present or not. Exceptions to this are features that are deemed especially important, for which test protocols are developed and carried out just as for attributes. These protocols specify the test procedure to be used and how the results will be compared across models. Every model is evaluated with respect to every attribute and feature using the relevant protocol. This allows an apples-to-apples comparison between each model within the product family.

This type of evaluation is **user-focused**, rather than manufacturer focused. As such, test protocols are developed with the needs and preferences of the end users in mind. Test protocols are based on either **reasonable use cases** or **extreme use cases**. Reasonable use cases are based on the most common ways a product may be used and may be determined through consumer interviews or observations. Extreme use cases are based on the most severe ways in which a product may be used or the most severe conditions it may be subjected to by the user. Extreme uses cases are generally considered less likely to occur than reasonable use cases, but they can be used to identify top performing models. Test protocols are designed such that laboratory experiments simulate the effect such reasonable and extreme use cases may have on each model.

These test protocols are designed and carried out by an **expert evaluation team**. The results of a comparative product evaluation are often shared widely, as this is required for them to make an impact. This impact may be large and affect sales of a product or consumer safety. Thus, these evaluations must be performed rigorously by a knowledgeable and experienced product evaluation team.

The end product of an attribute-based comparative product evaluation is a **ratings matrix** and a composite score. The ratings matrix has each product model one axis and each attribute on the other axis. Each cell displays an icon showing each model's performance in each attribute category. This matrix is a visual representation of the evaluation results and can easily be used to compare and contrast each model's performance against other models. An example of a matrix tool is shown in Table 2.1.

Model	Attribute A	Attibute B	Attribute C
A	2	3	4
B	3	3	2
C	2	5	3

Table 2.1: An example ratings matrix. Three models are rated with regards to three attributes on a scale of 1 to 5 in each attribute.

In addition to the ratings matrix, **weightings** are assigned to each attribute. These weightings are used to indicate the relative importance of each attribute. Weightings should be aligned with consumer preferences and are generally informed by user surveys and interviews, as well as observations of use patterns. These weightings are then used to combine the attribute scores into a single **composite score**. The composite score represents the overall performance of that product model. The composite scores from all models are then used to establish a **product ranking**. In this way, an apples-to-apples comparison between all product models is obtained and gradations in overall performance are revealed.

2.3.2 Assumptions

There a number of assumptions inherent in the attribute-based product testing process. This section will detail these assumptions.

First, it is possible to obtain reliable data on user needs, tastes, and preferences, as well as product use patterns. In order to perform product evaluation, it is necessary to understand the users themselves. This requires gathering various data points about end users. This can be done via interviews or surveys by phone, mail, or electronically. However, it must be possible to reach a sufficiently large number of users and users must be willing and able to provide this information. This information is necessary to identify relevant product attributes, design tests, and assign weightings.

Second, all target consumers of products in a certain product family are relatively homogenous in their tastes, needs, and preferences. For the ratings matrix and rankings produced by this testing method to be impactful on a sufficiently large scale, they must accurately reflect the preferences and needs of a large number of users. In other words, once a sufficient amount of consumer data has been gathered, it is a necessary condition that a relatively uncontroversial set of important attributes and weightings describing the relative importance of each attribute emerge.

Third, a single objective number can represent the entire performance of a product with respect to user needs and preferences. It must be possible to use the attribute ratings and weightings to arrive at a composite score in an uncontroversial way. In other words, the performance of a product model across all its performance categories can be reduced to a single metric. This score then represents the degree to which the product meets a user's needs. The score can then be used to rank all models in a product family in a meaningful way that is consistent across a large number of consumers.

Fourth, there is a well-developed market with multiple models to test. By definition, comparative testing requires that more than a single model is available to test within a product family. If there is only a single product, no gradations in performance can be uncovered. Evaluations are generally only performed on products that are sufficiently developed such that there are a number of product models that are similar enough to be considered a product family

Fifth, testing and publishing of results can have an impact on the marketplace. This type of consumer product testing is predicated on providing useful information to consumers so that they can make informed product purchasing decisions. The ultimate utility is derived from steering consumers to better products, which will eventually result in improvements in the quality of products available in the marketplace. However, consumers must have access to a marketplace where many different models in a product family are available. That is, consumers must actually be able to make choice between products.

Sixth, there is a need to maintain independence from manufacturers and commercial interests. In order to gain the trust of consumers, the audience for this kind of consumer product evaluation, it is necessary to maintain a certain distance from commercial interests. Product samples are therefore purchased through consumer outlets and manufacturers are not notified beforehand that their products are being tested. This ensures that the product models tested are the same as those actually available to the consumer and that manufacturers have not provided samples with superior performance. Furthermore, the evaluating body does not become involved in the design process or give advice regarding design improvements in order to prevent a conflict of interest.

2.3.3 The Attribute-based, Comparative Product Testing Process

The attribute-based, comparative product testing process can generally be broken down into ten steps. This section will detail each of these steps in the evaluation process.

Step 1: Product Family Definition

The first step in performing an attribute-based, comparative product evaluation is defining several possible product families to evaluate. Potential product families of interest should meet certain criteria, although specific criteria may vary for the specific organization performing the evaluation and may even very from evaluation to evaluation. Examples of criteria that may be used include widespread use with target user population, safety concerns, or frequent user complaints. At this stage, information is typically gathered through secondary sources, such as market reports.

Step 2: Market Intelligence Gathering, Background Research, and Product Family Selection

The next step is to select the actual product family to be evaluated. This is an important step as it represents a large commitment and time and resources to a single product evaluation. Furthermore, product families that are evaluated should be relevant to consumers. Thus, product family selection must be done with care. Product families are generally selected so as to maximize the impact of the evaluation.

In order to guide this process, further research is gathered on each prospective product family. This includes the gathering of market intelligence on each product family, such as which and how many models are available, the market share of each, and the overall size of the market. This is primarily done through secondary research, although some primary research may be done at this stage. In addition, the product family definitions are refined as more information is gathered.

The final selection of the product family is done by comparing each prospective product family along a certain set of criteria. These criteria may vary depending on the interests of the testing team, but is generally centered on maximizing the impact of the evaluation. Some examples of criteria that may be used include areas of in-house expertise, ease of obtaining test samples, or size of market for the product.

Step 3: User/Consumer Data Gathering

Next is the gathering of consumer data in order to understand the product family's end users. This step can be accomplished through consumer surveys or interviews, market research reports, or interviews with industry experts. The goal of this step is to identify the needs of the product end user and understand the actual product use patterns. This information is vital for identifying key attributes and features, designing testing protocols, and assigning weightings.

Step 4: Attribute and Feature Identification

From the information obtained in the consumer data gathering step, the key attributes and features of the product family should emerge. These should be aligned with common product use patterns and consumer preferences on the important characteristics of the products. This may result in small alterations in the definition of the product family. This attributes and features then form the basis for designing testing protocols.

Step 5: Product Testing

The fifth step is to develop testing protocols and perform the product testing. In this stage, the evaluating team gathers the necessary expertise, designs the test protocols, and conducts the actual product testing. The testing protocol design and testing process is informed by use cases inferred from the use data that was previously gathered. In addition, these protocols and procedures may also be influenced by existing product evaluation reports and industry standards.

Step 6: Results and Scoring

The next step is the analysis of results and the conversion of these results into scores. Results first needed to be converted into a raw score that describes the model's performance in relation to the performance attribute being tested. Next, these raw scores must be converted into a scaled score. This scaled score serves two purposes. First, it allows all results to be compared across all attributes by converting the raw scores into values on a common scale that can be related to one another. Second, for all models tested in each attribute category, it rescales the quantitative difference between the raw score values into differences between scaled score values that are qualitatively relevant to product end users.

Step 7: Weightings

The next step is to assign weightings to each attribute. These weightings determine heavily each attribute category is considered in calculating the final composite score. The final composite score combines the product attribute scores in each category and are used to establish a comparative ranking of the models tested. The weightings should represent the importance of each attribute category to consumers. This information can be gathered from user surveys and interviews.

Step 8: Composite Weightings Tool

Step 8 is to compile the results into a composite ratings tool. The purpose of this tool is to convey the results of the product testing stage in a compact and easy-to-use form. This tool usually takes the form of a comparative ratings chart. The comparative ratings chart graphical depicts all of the test results as well as the rankings derived from the composite scores that were calculated. The final manifestation of the chart may depend on the target audience or consumer group of the product evaluation.

Step 9: Publish Results

Once the results have been compiled into a useful ratings chart, this chart and accompanying information must be published. In order for a product evaluation to have a meaningful impact, the results must be made available to those making purchase decisions. In practice, this generally means that results are published and made as widely available as possible. As in the last step, the delivery mechanism depends on the target audience.

Step 10: Follow-up Evaluation and Results

The final step is to perform follow-up evaluations as necessary and publish the results. Follow-up evaluations are necessary because marketplaces are dynamic environments. Over time, products are generally updated with new designs and completely new products may be introduced as well. Follow-up evaluations should be performed periodically in order to keep the information up to date and add new products to the ratings.

Chapter 3

Pilot Attribute-based Product Evaluation: Solar Lanterns in the Ugandan Market

In this chapter, the Suitability team's first efforts at implementing an attribute-based product evaluation are chronicled. This evaluation was intended to be a proof-of-concept evaluation, establishing that the same general evaluation principles can be applied to a variety of products in the developing world. In addition, the team hoped to learn what changes to the process might be necessary to better adapt this framework to the developing world. The product family chosen was solar lanterns in Uganda. This product family was chosen due to the presence of a USAID partner organization, Solar Sister, which sold solar lanterns in Uganda using a social enterprise business model. In addition, the CITE team felt that solar lanterns were a good starting technology to learn about the evaluation process because it presented a low safety-risk to users, was a commonly available and at relatively low cost, and had significant technological aspects to its performance. In the future, a more robust product family selection process will be used that includes other factors as well, such as potential impact, previous evaluation work, interest from NGO's and procurement organizations, alignment with USAID goals, and research interests and expertise of the evaluating teams.

It is also worth noting that in performing this product evaluation, the Suitability team deviated from the framework of product evaluation in the developed world in a key way. In the developed world, product evaluations are performed with a user focus and the target audience are the users themselves. For the developing world, the Suitability retains this user focus, but the target audience is not solely the product end users. NGO's, procurement agencies, and partner organizations are the also an audience, as they often make purchasing decisions for these types of products.

This chapter is substantially based on the technical report (Brine, et al, 2014) submitted to USAID. There are significant additions and modifications based on the author's involvement in the work.

3.1 Step 1: Product Family Definition

The initial product family definition was performed in collaboration with other CITE teams. There were several criteria that were important to CITE in this first group of candidate product families. First, the product family should have a large user base so as to maximize the impact of the resulting evaluation. Second, the product family should be well-developed and have multiple models available for testing. In addition, the products should be sufficiently well-developed that the product family is easy to define and the products available on the market should be similar enough that they can reasonably be assigned to the same relevant product family.

For this first product family survey, CITE used the internal knowledge of its team members to tentatively identify several potential product families. These product families included cookstoves, solar lighting devices, water treatment technology, and information and communications technology. For the pilot evaluation, CITE relied primarily on internal knowledge and desk research in order to expedite the first product evaluation.

In the future the CITE, will combine the internal knowledge of its team members with surveys and interviews of implementing partners in a more rigorous product family definition process. This will allow CITE to consider the interests of partner organizations in product family definitions. In addition, this could possibly streamline the process by revealing the most impactful and widespread product families early on in the process.

3.2 Step 2: Market Intelligence Gathering, Background Research, and Product Famility Selection

Once product families have been defined, the next step was to gather additional information and select a single product family for evaluation. For the first project, the CITE had six criteria, primarily based on strategic and organizational factors, for selecting the product family to be evaluated. Solar lighting devices in the Ugandan market matched the desired product characteristics most closely and were chosen as the first product family to be evaluated. The criteria used for selection and the characteristics of this first product family are described below.

First, CITE preferred to evaluate a technology where CITE already had in-house expertise. A member of the Suitability team, Amit Gandhi, was in the process of completing a thesis on solar lighting devices in Ghana [10]. CITE sought to leverage this knowledge of the technology in the first product evaluation.

Second, the presence of implementing partners was crucial. In the area of solar lanterns, USAID had an implementing partner in Uganda, an NGO called Solar Sister. Solar Sister is a social enterprise that sells solar lighting devices through local entrepreneurs. The presence of Solar Sister in Uganda would be instrumental in understanding the Ugandan market and in reaching local consumers of solar lighting devices.

Third, in order to maximize impact, CITE sought to evaluate products which had not been evaluated before. There were organizations that performed solar lantern evaluations in existence. For example, Lighting Africa has extensive testing standards and labs to perform testing. However, these were generally industry-focused, standards based testing.

Fourth, for the pilot evaluation, CITE preferred to evaluate a product with minimal safety risks to users. CITE sought to avoid products that could potentially be dangerous for the evaluation team and users for this first proof-of-concept evaluation. In the future, it is possible that CITE would seek out products with safety concerns in order to effect changes in the market and to maximize the positive impact of evaluations.

Fifth, a relatively large number of models for testing and ease of acquisition was desired. The solar lantern market has grown quickly in the past few years and a number of models were quickly identified. In addition, solar lighting devices prices are generally in the tens of dollars and they are small and light-weight. This made the acquisition of a large number of solar lighting devices at low cost easy.

Sixth, CITE preferred to evaluate a product that would be relatively easy to test. Many of the factors that made solar lanterns easy to acquire also made them easy to test. They are small and light-weight and do not require any special tools to handle. A large number of samples could be easily acquired, allowing for many tests to be performed.

Once the product family was selected, it was necessary to gather market intelligence and refine the product family definition. This was done with a combination of desk work and field work. Available solar lantern models were identified through interviews with Solar Sister leadership in the United States and internet research. In addition, Amit Gandhi and Victor Lesniewski traveled to Uganda to ascertain market conditions in country.

Through these efforts, the Suitability team was able to identify eleven solar lantern models that were suitable for testing. These models were either readily available in Uganda, were about to be released onto the Ugandan market, or were readily available in neighboring countries. Models available from neighboring countries were included so that potential high performing models could be identified, and if not available in the Ugandan market, they could be recommended to expand into the market.

Throughout this thesis, lantern models will be referred to by a letter designation, starting at Model A. The corresponding lantern names and manufacturers can be found in Appendix D. As this is a proof-of-concept evaluation, however, the ratings and rankings in this thesis should not be regarded as definitive, nor should it be regarded as consumer advice. The main purpose of this pilot evaluation was to understand the evaluation process.

3.3 Step 3: User/Consumer Data Gathering

In addition to performing detailed market research while in Uganda, Gandhi and Lesniewski also gathered user information. They worked with Solar Sister to identify solar lantern users and conducted user surveys and studies. They interviewed a total of 38 solar lantern users who were owners of either lantern Model B or lantern Model D. The Suitability team chose to focus on these two lantern models because they were Solar Sister's top-selling solar lantern models in Uganda. In addition to these interviews, instrumented lanterns were distributed to these users to track their actual user patterns for one to two weeks.

Before the interview, users were first asked to consent to the two stage interview process and to agree exchange their solar lantern for an instrumented version of the same model for about two weeks. This ensured that users would use the instrumented lantern that they were provided. The instrumented lanterns were fitted with an Arduino-based data logger that could read the lantern's battery voltage and battery current. This would allow the Suitability team to determine when the lantern was on or off, when it was charging, and what brightness setting was selected when the lantern was on. In addition, an accelerometer was included so that they number of times the lantern was dropped could be measured. The data logger was designed to fit inside the existing lantern housing so that it would not change the aesthetic appearance of the solar lantern. The data logger also ran on a separate battery so as the maintain the solar lantern's performance.

In the first stage of interviews, users were asked about general solar lantern use patterns and the decision making process they went through when deciding to purchase a solar lantern. Demographic information was also collected. After the first stage interview, lanterns were exchanged, with instrumented lanterns left with users. One to two weeks later, interviewers returned for the second stage of interviews. In this stage, users were asked to specifically describe their lantern usage in the last week. They were also asked to rank the most and least important lantern characteristics using a MaxDiff survey.

MaxDiff analysis is a method used to show relative preference from a defined collection of choices. Interviewees were showed subsets of these characteristics and asked to identify the most and least important one from each subset. The results are then converted into a chart with shows the relative importance of each characteristic. The results obtained by Gandhi and Lesniewski are shown below. These findings will be used to influence the lantern testing process and the assigning of weightings later in the evaluation.

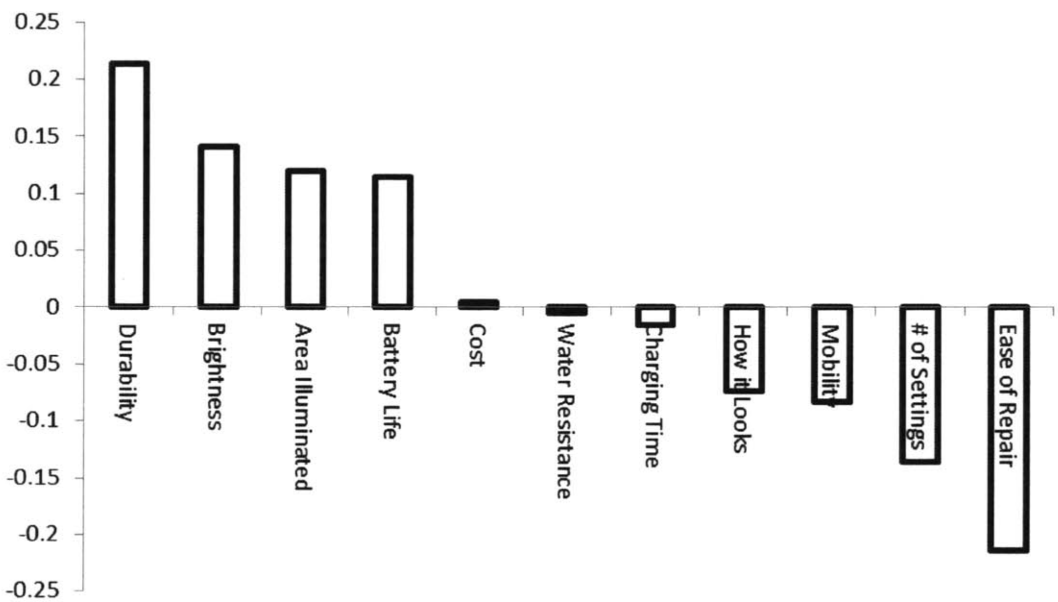

Figure 3-1: MaxDiff survey of lantern characteristics administered in Uganda. This plot shows the relative importance of each solar lantern characteristic. Results obtained by Amit Gandhi and Victor Lesniewsk from interviewing 38 solar lantern users in Uganda.

The user survey results regarding the specific usage while users were in possession of instrumented lanterns was also compared to usage information obtained from the data loggers. In was found that on average users over-reported their usage by a factor of two. This implies that consumer surveys may not give an accurate picture of actual consumer use patterns.

The user survey and data logging instrumentation were designed primarily by Amit Gandhi. The surveys were administered to users by Amit Gandhi and Victor Lensiewski. For more information regarding survey design, instrumentation design, and methods used, refer to Amit Gandhi's thesis [10].

3.4 Step 4: Attribute and Feature Identification

From the consumer data gather in the previous step, the Suitability team was able to identify several attributes relevant to solar lantern performance. They are listed below.

1. Time to achieve full charge from a fully discharged state
2. Time to fully discharge from a state of full charge
3. Sensitivity of charging performance to light levels
4. Brightness
5. Number and range of brightness settings
6. Color of light output
7. Task lighting performance
8. Ambient lighting performance
9. Water resistance
10. Impact resistance
11. Dust resistance

After further consideration and information gathering, the CITE team elected to perform tests of all attributes except for impact resistance and dust resistance. Impact resistance and dust resistance were not tested because of the existence of previous testing and uncertainty in test parameters, respectively. The rationale and test design for each attribute tested is discussed in Section 3.5.

From the consumer data, the Suitability team was also able to identify the following important features of the solar lantern product family.

1. Battery charging indicator

2. Battery charged indicator

3. Ability to charge from A/C power

4. Cell phone charging capability

For features, the Suitability team decided that indicating the presence of absence of features was sufficient for the product evaluation. Testing of these features was not needed because they did not affect the main function of these solar lanterns and did significantly change any other characteristics.

The exception to this was the cell phone charging capability. Information gathered from the field indicated that this was an especially important feature. Cell phone use is very widespread in the developing world, although charging can be problematic. Cell phone users often have intermittent access to power sources or must pay to have their cell phone charged. Thus, the ability of a solar lantern to double as a charging station for cell phones has the potential to save users money or even be a source of income.

3.5 Step 5: Product Testing

Creating product testing protocols and performing product testing were the next step in the product evaluation. The experimental design and rationale behind each aspect of testing is described briefly in this section. For more details, refer to Appendix A.

Test samples were obtained in multiple ways. Some were ordered directly from the manufacturer, some were ordered online through resellers, and others were purchased in-country in Uganda and brought back to Cambridge by Amit Gandhi and Victor Lesniewski.

3.5.1 Spectral Analysis

Light color is an important factor to consider when evaluating lighting sources. Working with Dr. Peter Bex, a researcher at Harvard Medical School, the Suitability team learned that the color of light can affect the ability of the human eye to distinguish colors and contrast. Thus, the Suitability team decided to measure the spectral power density of each solar lantern to determine the color of its light output.

The spectral power density of each solar lantern was measured using a spectroradiometer. Lanterns were taken to a dark room and the spectral output of the lanterns was measured on each brightness setting. Some lanterns had multiple LED's that corresponded to different settings. These were measured as well.

3.5.2 Charge Time

Time to charge is an important variable in solar lantern performance. Many manufacturers make headline claims about the charge time and place these prominently on the packaging. Furthermore, being able to charge the lantern quickly has many benefits for consumers. It minimizes the time a lantern has to be left outside and exposed to damage from the elements and to theft. It also allows flexibility in the decision of when to charge the lantern in order to take advantage of favorable weather conditions.

In order to measure this aspect of performance, solar lanterns were instrumented so that their battery voltage and charging current could be measured. Lanterns were fully discharged and then set out to charge in natural sunlight for two days while their battery voltage and charging current were continuously logged. Natural sunlight was used to charge the lanterns rather than a solar simulator or power supply in order to test the efficiency of the complete system in conditions similar to what a user would experience. In addition, a pyranometer was used to measure the solar insolation during this test.

3.5.3 Discharge Time

Discharge time is a second important variable in solar lantern performance. Manufacturers also make various claims about discharge time that are displayed prominently on product packaging. The length of time a lantern can remain lit on a single charge determines its usefulness. Discharge time performance is, however, very closely related to charge time. Some lanterns may take a long time to charge but have a long discharge time because they have a large battery. Others may have a short discharge time but charge quickly because they have a small battery. Thus, it is not enough to only evaluate charge time or discharge time alone. These two aspects of performance must be considered together.

Measuring discharge time was accomplished using the same test equipment as was used to measure charging time. Lanterns were fully charged and then their solar panels were unplugged or covered with a black polyurethane cover in order to prevent further charging. They were then turned on to their highest setting and left to discharge over the course of thirty hours while their battery current and voltage were continuously logged.

3.5.4 Solar Sensitivity

Solar sensitivity refers to the ability of a lantern to charge in varying levels of light. This test was not originally part of the evaluation but was added after the Suitability team noticed that some lantern models did not charge on cloudy days. This information is potentially useful to consumers, as it might affect their purchasing decisions or use of the product. For example, during the rainy season in Uganda, the sky is overcast for much of the time. In this case, a lantern that does not charge in cloudy conditions may be a poor investment. Furthermore, this may be an area in which product performance can be improved in the future.

No additional test was conducted for solar sensitivity. Rather, all the necessary data were already gathered during the charge time test. The Suitability team utilized the battery charge current measurements collected from the solar lanterns and the solar insolation data collected from the pyranometer. These data were used to plot each lantern model's charge rate with respect to the solar insolation over the course of the charge time test. This could then reveal each lantern's charging performance in various lighting conditions since the charge time test was run over the course of two days, with insolation ranging from zero at night to a maximum at around noon.

3.5.5 Task Lighting

Solar lanterns are commonly used to illuminate a small work area for specific tasks such as cooking, reading, and studying. This task lighting often requires a solar lantern to brightly illuminate a relatively small work area. The lanterns are usually placed very close to work area to provide task lighting. Some lanterns appeared to be designed for this type of use and take the form of a desk lamp. Regardless, of designed however, it is reasonable to expect that solar lanterns will be employed in this way.

In order to test a lantern's ability to perform task lighting, the Suitability team measured the area which each lantern could illuminate relatively uniformly to a predetermined brightness level. For this test, a photometer was placed on a flat surface. Next, a solar lantern was placed over it and its height adjusted until the photometer was illuminated at 25 lux. The photometer was then moved away from the lantern along the horizontal surface until it was illuminated at 12.5 lux. This was repeated in two other angularly equidistance directions. These measurements were then used to calculate the area illuminated to the threshold brightness level.

3.5.6 Ambient Lighting

Another common use for solar lanterns is to provide general room lighting. This allows all the occupants of a room to share a single light source in order to perform various tasks. For this use case, the solar lantern would generally be mounted high in the room or even suspended from the ceiling. Many of the models tested included design elements to allow exactly this type of lantern positioning. Thus, this was considered an important test of performance for a common usage mode.

Ambient lighting performance was tested by simulating the use of a lantern to provide light to a small room when suspended from the ceiling. This was done by first suspending a solar lantern six feet from the ground in a dark room. The ability of the lantern to illuminate objects at three different heights was determined by using a photometer to measure light levels. At each of the specified heights, light level readings were taken at various points, forming a cylinder.

3.5.7 Brightness

After durability, brightness was ranked as the second most important characteristic of solar lantern lanterns by the consumers surveyed in Uganda. This is an intuitive result, as the purpose of a solar lantern is to provide light. The more light a solar lantern can provide, the more beneficial is its to a consumer.

This test was performed by placing each solar lantern 18 inches away from a photometer. The lantern was then turned onto its highest setting and the brightness level was recorded.

3.5.8 Setting Versatility

Setting versatility refers to the range of different brightness levels that a lantern is capable of outputting. Having a range of different brightness settings allows a consumer to select the correct light levels for the task they are doing. For example, walking outside at night might require a very high level of light, while a night light might require a much dimmer light. Having multiple settings available also allows users to conserve battery when maximum brightness is not needed.

This test was performed by placing each solar lantern 18 inches away from a photometer. The lantern was then cycled through all its brightness settings. The light level at each brightness setting was recorded. This was then used to calculate the ratio of the light level at the lowest setting to the light level at the highest setting.

3.5.9 Water Resistance

Water resistance was considered to be a important test because exposure to water is often damaging for electronic devices. Furthermore, in the case of solar lanterns, it is likely that they will be exposed to water in the course of regular use. Solar lanterns that have an integrated solar panel must be placed outside to charge. This makes it likely that they will be left out in the rain. Even when being used for tasks such as cooking, it is not unlikely that a lantern will be inadvertently exposed to water.

Water resistance was evaluated by performing a three-tiered series of tests. The first test was the extreme case. Lanterns were completely submerged in water for one minute. The second test was a reasonable worst case. Lanterns were placed in a vulnerable orientation and subjected to a simulated rainstorm. The vulnerable orientation was determined by the Suitability team and generally depended on factors such as the location of open ports, fabric handles that could wick, or other potential weaknesses. The third case was a mild case. Lanterns were subject to a simulated rainstorm while place in charging orientation. In this case, lanterns were placed standing up, as they would be placed if a user was charging them. Lanterns were evaluated based on the highest tier water resistance test they were able to pass and the resulting degradation in performance.

3.5.10 Features

Features are characteristics or functions that are beyond the basic set that defines the product family. Some of these are features are important and affect the desirability and functionality of certain models. For example, in Uganda, the Suitability team found that the ability to charge cell phones with a solar lantern was important and could affect purchasing decisions. The Suitability thus decided to develop a test for this important feature.

The lantern charging indicator, the lantern charged indicator, and the A/C charging capability were not tested. These features were simply identified as present or absent on each lantern model. For the phone charging feature, each lantern model equipped with this feature was used to charge a smart phone and a feature phone. The rate at which lanterns could charge the phones, and the amount of charge that a lantern could provide to the phones before exhausting the battery were measured. Lanterns were evaluated on these two metrics.

3.5.11 Note on Durability

The Suitability team divided durability into three subcategories: water resistance, impact resistance, and dust resistance. These are the ways in which the Suitability team felt solar lanterns were most likely to fail based on field surveys and observations of consumers in Uganda. Additionally, durability was ranked as the most important factor in lantern purchases according to the MaxDiff survey administered in Uganda. Thus, it was important to evaluate these aspects of lantern performance.

The Suitability team devised a test for water resistance, which was described in Section 3.5.9. The Suitability team, however, decided to exclude impact resistance and dust resistance from testing despite the indicated importance of durability due to the existence of previous testing and lack of sufficiently detailed information, respectively.

In the case of impact resistance, there was an existing Lighting Africa Standard. In this test, six samples of each lantern are dropped from a height of one meter onto a concrete surface in six different orientations. Lanterns models pass this test if five of the six samples continue to function [11]. The Suitability team felt that the Lighting Africa test was sufficient for most users and the Suitability team would not be able to provide more valuable information. Thus, the Suitability team elected to not develop and perform a new test for impact resistance.

There was also a Lighting Africa standard for dust resistance. The standard specifies that most lanterns are subjected to a 12.5mm probe and pass the test if the probe does not enter the lantern enclosure. There is also a more rigorous test that some lanterns are subject to which requires the sample to be kept in an environmental chamber in which talcum powder is circulated in the air for four hours. Lanterns pass if powder does not enter the enclosure [11].

In both cases, the Suitability team felt that the existing tests were not sufficient. A 12.5mm probe will give little information on the effect of dust on lantern performance. Talcum powder may indicate if a lantern has a permeable enclosure, but talcum power is very soft. It would not have the same effect as dirt or sand. However, due the large variations possible in the types of dirt and dust a solar lantern may be exposed to depending on geographic location and other factors, the Suitability team felt that it would be difficult to create a representative test for aspect of durability. For example, desert sand may be much more damaging to a solar lantern than household dust. The performance effect on a lantern would then depend on many factors other than simple exposure. Thus, this aspect of solar lantern performance was not evaluated because of the uncertainty that would result from such a test.

3.6 Step 6: Results and Scoring

The results and interpretation of each test are briefly described here. For more details, refer to Appendix A.

After the raw score for each test had been calculated, it is then converted into a scaled score. The scaled score ranges between 0.5 for the lowest score and 5.49 for the highest score. Raw scores are converted in scaled scores by setting the lowest raw score to 0.5 and the highest raw score to 5.49. A linear interpolation is then used to map the remaining scores onto this number line.

This conversion method maps the entire range of raw scores onto a defined range of scaled scores. Because all attribute scores are converted to a 0.5 to 5.49 scaled score, performance on different attributes can then be related to one another. This is necessary to create a single composite score for each model evaluated.

3.6.1 Spectral Analysis

All solar lantern models in this evaluation had a similar spectral power density. They all had spectral peaks at about 450 nm and 570 nm, which is typical of white LED's. In addition, the relative magnitudes of the peaks were also similar. In consultation with Dr. Peter Bex, a vision researcher at Harvard Medical School, the Suitability team concluded that the color of the light was sufficiently similar across lantern models that it would not materially affect a user's ability to read or perform other tasks. Therefore, the results were not used to score the lanterns.

3.6.2 Charge Time

For this test, lanterns were scored on the amount of time it took them to fully charge. This was determined by observing the battery current trace and identifying the time at which the charging current dropped to zero despite the presence of sunlight, which was measured using the pyranometer. A shorter charge time was considered better.

A large variation in performance was observed in this test, with some lantern models charging twice as fast as others. This was likely due mostly to differences in battery capacity. A lantern that had a high capacity smart phone battery would probably take longer to charge than a lantern with a smaller battery. Thus, the results of this test are not definitive and must be considered in conjunction with the results of the discharge time test.

3.6.3 Discharge Time

Lanterns were scored on the amount of time they were able to stay lit. Determining this time was more difficult than for the charge time test. Some lanterns simply discharged and dimmed as the battery voltage dropped. Other lanterns had more sophisticated circuit which maintained a constant current until a preset battery level had reached. The lantern would then switch to a lower brightness setting. This would continue until the battery level dropped below another preset level and the lantern would power off. This behavior drove the Suitability team to develop a separate test to measure light output with respect to battery output current in order to develop a method for scoring this test, since there is no obvious point at which lanterns were discharged.

The team observed that for the lantern with the largest difference in output between brightness settings, the ratio of brightness in the lowest setting to highest setting was approximately 0.05. This was selected at the cut-off point at which lanterns would be considered discharged. Using the data correlating brightness level to current, the time at which a lantern dropped below 0.05 of its original brightness was determined by examining the current trace obtained from the discharge test. This was then used to determine the time the lantern had taken to discharge. A longer discharge time was considered better.

A large variation in performance was observed in this test as well, with some lanterns able to remain on for more than twice as long as others. As in the charging test, this was probably due mostly to differences in battery capacity. This test therefore needs to be considered in conjunction with the charging test.

3.6.4 Solar Sensitivity

All solar lanterns models generally had relatively linear charge rate vs. solar insolation curve above 100 - 200 W/m^2. Below this level, however, most lanterns had a steep drop-off in charge rate. Because performance was similar across most lantern models in high light conditions, the Suitability team decided to focus on low light performance for scoring purposes. Lanterns were scored based on the lowest insolation at which they were able to charge. Being able to charge at a lower insolation level was considered better.

There was some variation observed in this test. Some lanterns, such as Model D, were able to charge with less solar insolation than others. This may be due to solar panel chemistry and size, as the lantern models tested had mono-crystalline and poly-crystalline panels of various sizes.. Lanterns that are able to charge in low light conditions have a distinct advantage because they can continue to function even if there are a string of cloudy days.

3.6.5 Task Lighting

The task lighting test was scored by using the three measurements of distance at which a lantern could illuminate the photometer to 12.5 lux. Assuming that the illuminated area is a circle, these three measurements are used to calculate the total area that a lantern can illuminate to 12.5 lux. Being able to illuminate a larger area was considered better.

In this test, the highest performing lantern, Model J, was able to illuminate over five times the area of the lowest performing lantern, Model K. This test does not distinguish between lantern form factors, however. A floodlight-style lantern will probably score better on this test than a flashlight-style lantern. Form factors are called out on the final ratings chart and allowing comparisons to be made across lanterns with similar designs. The choice of lantern design is left up to the consumer as a qualitative factor.

3.6.6 Ambient Lighting

Lanterns were scored in this test by their ability to provide light to an entire room. This was done by weighting the light level measurements taken by height and then summing these values. A higher level of summed illumination was considered better.

No lanterns were able to illuminate the points a ground level to a anything more than a very dim level. At 30 inches, Models A, E, F, and J were able to provide slightly more illumination. At 66 inches, Models A, E, and F were able to provide a very high level of illumination. There were, however, some lantern models that were not able to provide an adequate level of illumination at any height.

3.6.7 Brightness

Lanterns were simply ranked ordered by the brightness level measured on the high setting for this test. Brighter was considered better for this test.

A larger than usual spread in performance was observed in this test. The highest performing lantern, Model E, was four times brighter than the lowest performing lantern, Model D. As in the task lighting test, a model's score on this test may be influenced by its design. Form factors are called out in the final ratings chart and are left as a qualitative factor for the consumer.

3.6.8 Setting Versatility

In this test, lanterns were scored simply by calculating the ratio of illumination measured at the lowest brightness setting and the highest brightness setting. A larger ratio was considered better.

There were some lanterns models in this evaluation that only had one brightness setting. In this case they offered the user no flexibility in this aspect of lantern use. Some lanterns had up to four brightness settings. The highest performing of these lanterns, Model E, had a ratio of illumination at the lowest brightness setting to the highest setting of approximately 0.03.

3.6.9 Water Resistance

For the water resistance test, lanterns were scored based on the highest tier of water resistance test that they were able to pass. In addition, they were also scored on any performance degradation that they might have suffered. Passing a higher tier, and therefore more extreme, water resistance test was considered better. In addition, less performance degradation was considered better, with no loss in performance the ideal case.

The results of the water resistance tests were surprising. Models B, C, D, F, G, H, I, and K were able to pass the tier 1 complete submersion test with little or no loss in performance. Because so many lanterns were able to survive this test, the Suitability team decided to use this as the benchmark for performance in this test. Only two lanterns, Model A and Model E, failed all three water resistance tests. Model J was only able to pass the tier 3 shower test, however, it was still rated poorly in the context of the outstanding performance of a majority of solar lantern models.

3.6.10 Features

The lantern charging indicator, lantern charged indicator, and A/C charging capability were simply identified as present or not present. For cell phone charging capability, lanterns were scored on the rate at which they could charge a sample and the amount of charge they were able to transfer to the cell phone before the lantern battery became depleted. Lanterns that could transfer more charge at a higher rate scored higher.

Almost all lanterns tested had a lantern charging indicator, the exception being model K. About half of the lantern models had a lantern charged indicator. Similarly about half the lantern models could be charged from A/C power. Finally, about half of the lantern models were able to charge a cell phone, all there was much variation in the performance of this feature. Some lanterns were able to fully charge a smartphone very quickly, while others were only able to charge a feature phone over the course of a few hours. Model B required use of the solar panel to charge the cell phone, during which time the lantern itself could not be charged.

3.7 Step 7: Weightings

Once all product evaluation tests had been completed and scaled scores were calculated, the next step was to assign weightings to each product attribute so that a composite score and ranking could be created. For this pilot evaluation, the Suitability team based the weightings on the results of the MaxDiff survey that was administered in Uganda and the intuition of the evaluation team. First, each member of the Suitability team proposed a set of attribute weightings. These proposed weightings were aggregated and averaged, producing a consolidated set of weightings. Adjustments were then made to these consolidated weightings based on the consumer survey results. The weightings values that the Suitability selected are shown in Table 3.1.

Category	Weight (%)
Discharge Time	20
Charge Time	10
Solar Sensitivity	10
Brightness	10
Task Lighting	15
Ambient Lighting	10
Setting Versatility	5
Water Resistance	10
Features	10

Table 3.1: Weighting table. This table contains the weighting assigned to each attribute. Each weight is represented as a percentage of the overall score.

For this study, the consumer survey had a sample size of 38. In addition, because solar lantern users were identified through CITE's partner organization in Uganda, Solar Sister, rather than randomly chosen, these users are most likely not representative of the general population. For these reasons, the Suitability team did not rely solely on the survey data in order to determine the weightings.

It should be noted that although durability was selected as the most important solar lantern characteristic in the MaxDiff survey given to Ugandan users, water resistance was weighted lower than charge time and task lighting performance. This is because water resistance comprises just one aspect of durability; the others that were of interest to the Suitability team but were not tested being dust resistance and impact resistance. Thus, water resistance was given a lower weighting to reflect the fact that it is only partially representative of the broader category of durability.

In the future, the Suitability team will rely on more consumer data in order to determine attribute weightings. More in-depth market intelligence research and consumer data gathering will precede future evaluations. More robust survey methods will be developed and surveys will be applied to a larger and more representative group of consumers.

3.8 Step 8: Composite Rating Tool

The scaled scores and weightings from the previous two steps are then combined in order to create the composite rating tool and rank order all the models evaluated. The composite score aggregates each solar lantern model's performance over all tested attributes. The composite score of the i^{th} lantern model is calculated using the following formula:

$$Composite\ Score_i = 100 * (Y_i - MIN)/(MAX - MIN)$$

where Y_1, MAX, and MIN are defined as follows:

$$Y_i = FMAX + A_{i,1}W_1 + A_{i,2}W_2 + \ldots + A_{i,n}W_n$$

$$MAX = FMAX + A_{max,1}W_1 + A_{max,2}W_2 + \ldots + A_{max,n}W_n$$

$$MIN = A_{min,1}W_1 + A_{min,2}W_2 + \ldots + A_{min,n}W_n$$

In these equations, $FMAX$ is the sum of the highest scaled scores from each attribute tested. W_j is the weighting assigned to attribute j. $A_{i,j}$ is the scaled score of the i^{th} lantern for attribute j. $A_{max,j}$ is the maximum attribute score observed over all lanterns for attribute j and similar $A_{min,j}$ is the minimum attribute score observed over all lanterns for attribute j.

Y_i is the aggregated score for lantern model i. This is computed by first multiplying all attribute scores for lantern i with the appropriate attribute weightings. The resulting values are then summed and $FMAX$ is then added to this total. MAX is the aggregated maximum score, computed by first summing the product of the highest score observed over all lanterns for each attribute and the weighting assigned to each attribute. $FMAX$ is then added to this total. MIN is the aggregated minimum score, computed by summing the product of the lowest score observed over all lanterns for each attribute and the weighting assigned to each attribute.

The final composite score is then computed by subtracting MIN from Y_i and then dividing by the difference between MIN and MAX. Calculating the composite score in this way results in a number ranging between 0 and 100. The difference between the minimum and maximum observed scores is used to normalize the composite score so that in falls within this range, while ensuring most of the range is used. The addition of $FMAX$ to the aggregated scaled scores and MAX compresses the scale slightly, so that extreme values such as 0 and 100 are not actually used. This results in composite scores that generally range from the 20's to the 70's.

Composite scores were calculated for each lantern model evaluated and these were used to rank order them. The resulting evaluation matrix is shown in Figure 3-2.

3.9 Step 9: Publish Results

For this evaluation, results were submitted to USAID. However, because this was a pilot evaluation, the full results will not be published widely. The main purpose of this evaluation was to understand the process of attribute-based, comparative product evaluation and how it can be applied to the developing world. Adjustments to the process will be made based on the lessons learned from this pilot evaluation.

In the future, CITE's product evaluation results will be shared with USAID, implementing partners, and published more widely. Publishing results is necessary in order to achieve CITE's long term goals of influencing the market for technological solutions in the developing world and improving products for consumers. Before this can happen, however, CITE needs to more fully develop the product evaluation process and develop protocols for this endeavor.

3.10 Step 10: Follow-up Evaluation and Results

This evaluation was just completed recently. Therefore, there has not been a follow-up evaluation. In the future, however, follow-up evaluations will be necessary in order to maintain timely and relevant data. For example, while this evaluation was in progress, a manufacturer released an updated version of a lantern model that was under evaluation. As other lantern models are updated, eventually the ratings chart will become outdated.

Product Information				Product Attributes										Product Features			
MODEL	100	Relative Cost	Type	Discharge Time		Charge Time		Solar Sensitivity	Brightness	Task Lighting	Ambient Lighting	Setting Versatility	Water Resistance	Battery Charging Indicator	Battery Charged Indicator	Device Charges from AC	Mobile Phone Charger
				Hours	Score	Hours	Score	Score									
Model J	70	5.65	Handheld/Desk	21.2	●	17.7	○	●	●	●	●	○	○	Yes	Yes	Yes	Yes
Model G	61	2.11	Handheld/Desk	22.1	●	13.4	◑	●	◑	◔	◔	○	●	Yes	Yes	No	No
Model F	60	2.82	Handheld/Desk	13.1	◑	8.7	●	○	●	●	●	○	◑	Yes	Yes	No	Yes
Model I	53	2.82	Handheld/Desk	18.1	◕	19.6	○	●	◔	◔	◕	○	●	Yes	Yes	No	No
Model E	47	3.53	Handheld	6.1	○	13.3	◑	○	●	●	●	●	○	Yes	Yes	Yes	Yes
Model H	47	1.41	Handheld/Desk	19.1	●	8.2	●	◑	○	○	○	○	●	Yes	Yes	No	No
Model B	46	2.61	Desktop	7.5	○	6.7	●	◑	◕	◑	○	◔	●	Yes	No	No	Yes
Model D	43	1.24	Handheld/Desk	11.8	◔	9.8	◕	●	○	○	◔	○	●	Yes	No	Yes	No
Model C	40	1.00	Desktop	12.0	◔	13.7	◑	●	◔	○	◑	◔	◕	Yes	No	Yes	No
Model A	34	4.23	Handheld	12.5	◑	19.0	○	○	◔	◑	●	◔	○	Yes	No	Yes	Yes
Model K	32	1.41	Desktop	10.4	◔	17.9	○	●	○	○	○	◔	●	No	No	No	No

Legend	
●	Outstanding
◕	Very Good
◑	Average
◔	Marginal
○	Poor

Figure 3-2: Comparative ratings chart. Lanterns are ranked ordered by composite score. Note that relative cost is listed rather than actual price.

In the future, CITE will perform a follow-up solar lantern evaluation in order reflect new and updated products that may have been introduced. When a follow-up evaluation is performed may depend on factors such as the number of models that have been updated, the number of new models introduced to the market, and major changes in the market, product family, or technology since the previous evaluation. Follow-up evaluations will update the data contained in the previous evaluation and may also include updated procedures to account for such factors as changes in technology or use patterns.

Chapter 4

Discussion of Pilot Evaluation

In performing the pilot solar lantern product evaluation, the Suitability team learned many lessons and encountered many challenges in applying the attribute-based, comparative testing process to the developing world. This chapter will discusses the challenges encountered and strategies that the Suitability team may employ on the future to improve evaluations.

4.1 Limitations of Attribute-based Comparative Product Testing

Product evaluation is performed with great success in the developed world. Organizations like *Consumer Reports* are able to raise hundreds of millions of dollars to fund their operations and have millions of subscribers who depend on these organizations for information. They are able to perform sophisticated evaluations of a large number of products and also have a large influence on markets in developed countries.

Consumer product evaluations also have the potential to be very useful in the developing world. However, there are a number of challenges that must be overcome. Many of these are unique to the developing world, such as disseminating results to consumers and obtaining funding. Others are inherent to the product evaluation framework, such as scoring method. The major challenges CITE faces are discussed below.

4.1.1 Resource Intensiveness

A major challenge in performing attribute-based, comparative product evaluations in the developing world is the resource intensiveness of this evaluation method. Large amounts of time and financial resources are required. For the pilot solar lantern evaluation, the Suitability team employed three graduate students, along with other program staff and consultants. It took this team approximately eight months to complete the pilot evaluation.

In addition to the time required, there was also a large investment in lab space and equipment for the laboratory testing portion of the product evaluation. A dedicated lab space was maintained and equipment such as data acquisition hardware and software were purchased. Furthermore, the test samples themselves had to be purchased, with up to ten samples of each solar lantern model acquired by the Suitability team. Finally, Suitability team members traveled to Uganda for one month in order to perform market research and gather consumer data. All of this adds up to a significant investment of time and money to perform a single product evaluation.

Furthermore, once an evaluation has been completed, it will periodically need to be updated. As products are updated and new models are released, the evaluations will eventually become out of date. This requires performing additional market intelligence gathering to understand how the market has changed as well as laboratory testing on any new and updated models so that they can be added to evaluation results. This will require that equipment be stored and maintained even after the initial evaluation has been concluded. Personnel will have to devote time to performing this follow-up testing. Thus, maintaining existing evaluations in and of itself also demands significant time and money resources.

Organizations like *Consumer Reports* operate in high income countries and are able to raise money from subscription fees. In the developing world, however, consumers have much less disposable income, especially the bottom of the pyramid consumers targeted by CITE and other development organizations. Consumers are therefore less likely to be able pay for the information provided by consumer product testing organizations. Furthermore, CITE's goals are inherently somewhat different than a consumer product testing organization in the developed world. CITE's goal is not only to influence the market and improve the market, but also to improve the lives of the poorest consumers. This means CITE must disseminate its results differently, perhaps providing them at no cost to consumers.

The resource intensiveness of attribute-based, comparative product evaluations and the limited ability of CITE to raise fees from subscribers means that CITE cannot be scaled up in the same way as *Consumer Reports*. This inability to scale up limits the number of product families that can be evaluated, thus making the choice of product family to evaluate a very important one because it is a large allocation of limited resources. Choosing one product family to evaluate means another product family will not be evaluated. In this case, CITE must carefully select product families to evaluate that have the largest potential impact on developing world users.

The inability to evaluate large numbers of product families might also limit CITE's potential long term impact. In order for a CITE to have a significant impact on a large number of users, CITE must complete many evaluations of different products in different contexts. If CITE only produces a few product evaluations focused on a certain region or technology, CITE's impact is likely to limited to that region or technology. Additionally, evaluating a large number of products will be important in establishing CITE as a credible and authoritative resource on consumer products. Indeed, part of the reason users in developed world turn to organizations like *Consumer Reports*, which evaluates 100 product families each year, is the fact they can find information of a wide variety of product families from these sources.

As a possible solution to these issues, the Suitability team may develop cost-effective testing methods, just as cost-effective technologies have been developed. The Suitability team may also choose to focus on developing test methodologies, leaving the actual product testing to NGO's and partner organizations. Alternative evaluation frameworks could also be developed, with an eye towards decreasing the resource intensiveness of product evaluations.

4.1.2 Assumptions about Products, Markets, and Users

Products, markets, and users are very different in the developing world from what organizations like *Consumer Reports* are familiar with, and this presents unique challenges for CITE. Some of the assumptions that are made by consumer product testing organizations in the developed world may simply not apply in the developing world. These assumptions were detailed in Chapter 2 and will now be discussed in more detail in the context of the developing world.

First, assumptions were made that consumers are a relatively homogeneous group and therefore product performance could be expressed in a single, widely accepted metric. In the developed world, the assumption that consumers are relatively homogeneous is probably approximately accurate. For example, when *Consumer Reports* tests light bulbs, this is probably a good assumption. The average home in Massachusetts is similar to a home in California, with a similar light fixtures and reliable access to electricity. In addition, the occupants of both homes use their light bulbs for similar purposes: to illuminate activities such as reading, cooking, and studying. Thus, it is possible for *Consumer Reports* to create a battery of tests that is reflective of the use patterns of the majority of users.

In the developing world, these assumptions about the similarity of consumers may not be valid. Use conditions may vary greatly from country to country or even village to village. For example, use patterns and user preferences for solar lanterns may be completely different for a user from a dry, sunny climate and a user from a wet, rainy climate. The user who lives in a dry, sunny area may not care at all about a lantern's water resistance or ability to charge in cloudy weather. These attributes would be very important to a user who lives in a wet, rainy area. These differences mean that any calculation of an overall performance score or performance ranking is inherently dependent on context. It would therefore not be possible to reduce product performance to a single composite score and create a single product ranking that would be relevant to a large portion of developing world users.

Second, an assumption was made about that product families are well-developed and consumers can easily obtain many members of the product family through retail outlets. The assumption that there are well-developed product families is also probably acceptable for the developed world. Many consumer products that we are familiar with and use on a daily basis are in fact from well-developed product families. For example, televisions, automobiles, light bulbs, and dishwashers are product families that developed consumers might be interested in and all are from well-developed product families. There are many different manufacturers that produce many different models of each product. These products are also easily obtained by consumers through a variety of outlets and thus, consumers are able to choose which model they purchase.

These assumptions may not hold in the developing world. Many technological interventions deployed in the developing world are not part of well-developed product families. Rather, they are prototype technologies still in development or are one-off technologies developed to solve a specific problem. These present a problem when trying to perform a comparative because there may not be a suitably similar product to evaluate. Additionally, there may be some interventions that are not products, but rather methods of providing the same service. For example, in the context of the solar lantern evaluation, Liter of Light [13] can be used provide indoor light during daytime hours, but it is an open source method rather than a product provided by a manufacturer. This is also challenging to include in a comparative product evaluation because it is different from a manufactured product and its performance depends on how the consumer assembles it.

Consumers in the developing world also do not have the same access to products in the developed world. In developed countries many models within a product family are generally easily obtainable through many retail outlets and online. In the developing world, availability of products may be reduced. Consumers may only have access to products that they can obtain locally or they may receive products from NGO's or other donors. In this case, the consumer may not have much choice in which product they obtain. If consumers are unable to make purchase decisions, performing a comparative product evaluation and publishing the results to use may not have any market impact. Alternative audiences may need to be considered, such as NGO's and purchasing organizations. This presents its own set of challenges, as these groups may have different preferences than end users.

Third, it was assumed that consumers would be easy to reach. In the developed world, this assumption generally holds true, as consumer can be reached through many means, including by mail, phone, or electronic means. In the developing world, reaching consumers is much more difficult, especially the bottom of pyramid users that CITE is targeting. Internet access is not as widespread as in the developed world [12] and even SMS communications is not always reliable and may still be too expensive for bottom of pyramid users. Performing evaluations focused on this user group is difficult when information cannot be obtained regarding the use patterns and preferences of consumers in this group.

Finally, in order to properly conduct an unbiased product evaluation, it was assumed that CITE would maintain independence from product manufacturers. This may interfere with some of CITE's other activities. For example, the Scalability team focuses on supply chain research, but this type of information is generally only obtainable by developing a relationship with product manufacturers. In addition, CITE may be able to provide value by engaging with product manufacturers to develop and improve future products. It is not clear that it is within CITE's best interest to maintain complete independence from all product manufacturers.

4.1.3 Manipulation of Scores

The purpose of a comparative product evaluation is to influence the marketplace, eventually resulting in the availability of superior products. This is done through the scoring system, which rates products on each attribute and encourages manufacturers improve their products which will then score better on a subsequent evaluation. This requires that that scoring system is designed very carefully. Even then, with the most well-intentioned scoring system in place, it is possible that product manufacturers will redesign their products such that the score is improved in the product evaluation, but there is no benefit to the consumer. In fact, it is even possible that the product may be worse from a consumer standpoint, or even become dangerous.

For example, consider the brightness test performed as part of CITE's solar lantern evaluation. Lanterns were scored based on how bright they were and brighter was considered better. Product manufacturers could then redesign their solar lanterns to be brighter and thus score better in a future evaluation. While a small improvement in brightness might be desirable, very large increases in brightness may not be. If a solar lantern manufacturer were to design a lantern with an extreme increase in brightness, it may at best be a useless modification if no one ever uses the brightest setting because it is too bright. At worst, it has the potential to hurt rather than help users, as the light may be bright enough that is uncomfortable to use.

Thus, it is important to carefully design the scoring system to prevent manufacturers from manipulating scores with design changes that are not useful to consumers. Even with a carefully thought out scoring system, it is possible that unintended consequences will occur. For this type of testing, market response to the evaluation has to be carefully monitored, and the scoring system may have to be modified to accommodate for score manipulation or other changes in the market or user preferences.

4.1.4 Meaning of Scores

A comparative product evaluation must differentiate between product models in order to rank order them. Many product tests are designed such that differences in performance can be detected and quantified. It is important, however, to ensure that any differences in performance that are detected are great enough that they would be significant to a consumer. Furthermore, the resulting scores must be presented in such a way that the magnitude of these differences in accurately conveyed to the consumer.

This can sometimes be difficult to do when all raw test scores must be converted into scaled scores, which are in a common range. For example, if one lantern achieves a score of outstanding in the brightness category and another achieves a score of average, it may not be clear to the user how large the performance difference actually is. All the consumer can conclude is that one was brighter and the other one fell somewhere near the middle of the range of brightness levels observed in the test. This issue is also manifested in the single composite score. The highest scoring lantern in the solar lantern achieved a score of 70, while the lowest scoring lantern achieved a score of 32. From this, it is clear that one lantern performed better than the other. However, what it means for one lantern to be twice as good as another is unclear.

Another issue is that the rankings may mask the shortcomings of a model in a specific area. For example, Model J was the highest ranked lantern despite failing the water resistance test. The fact that a solar lantern model could be achieve the highest score despite failing a durability test, especially when durability was ranked as the most important characteristic by consumers, illustrates a weakness of this rating system. Because of the way scores are calculated weighted, a solar lantern can make up for a shortcoming in this area by performing well in other tests. This is an area that requires future work, perhaps by adjusting the weightings or the way in which scores are calculated.

Thus, although products models become differentiated and ranking emerges, care must be taken to ensure that these differentiations are meaningful. Furthermore, consumers need to understand the way these results are conveyed and how the scores and rankings are relevant to them.

4.2 Suggestions for Improvement

In performing the pilot solar lantern evaluation, the Suitability team also noticed many areas in which improvements could be made. These improvements are discussed here.

4.2.1 Organizational Development

In the pilot evaluation, it was difficult for the Suitability team to gather accurate market intelligence and user data. Developing organizational partners is a potential solution to this problem. Local partners would have much more detailed knowledge about the local context, as well as the ability to easily connect with consumers in-country. In addition, they would be able to procure product samples in-country for laboratory testing, which would be helpful because some products are not available in the United States. Furthermore, the Suitability team noticed that products purchased in the United States sometimes had different packaging and included peripherals.

The presence of local partners in-country has the potential to streamline the process of performing product evaluations in the future. The quality of evaluations would also be improved to better information from the field and access to a larger number of consumers. Organizational partners may also be able to provide assistance in disseminating evaluation results in a way that is meaningful for local consumers.

4.2.2 Standardized Protocol

Another challenge that the Suitability team faced was the ability to scale up in order to perform a large number of product evaluations and the required follow-up to keep data current. A possible solution to this is to develop a set of standard testing protocols. These protocols could then be followed by any testing organization, effectively outsourcing the actual laboratory product testing. Lab testing could then be run by various testing organizations locally within various countries.

In this case the Suitability team would become developers of test methodologies, rather than actual product testers. This would allow larger numbers of product evaluations on many product families in many locations. Running product evaluations in-country would also significantly reduce the cost of such evaluations, as well as the large travel burden on the Suitability team. Additionally, the local knowledge inherent in such arrangements would make context relevant evaluations easier to produce.

The main concern arising from this approach is maintaining the quality of testing and reporting performed by these independent testing labs. There may be some oversight or more stringent reporting required. Independent testing organizations could also be screened beforehand. However, this idea is worth exploring in the future as the potential benefits are very large.

4.2.3 Sensor Improvements

The Suitability team learned through consumer surveys and the deployment of data loggers in Uganda that many users do not accurately report their user patterns. This creates a challenge in accurately understanding the needs of the consumer. The data logger deployed to Uganda was a prototype model and had severe limitations. It was not able to continuously log data and had a limited battery life of approximately two weeks. It was therefore only able to log discontinuous data for two weeks in the field. Furthermore, all data was stored locally on an SD card and required the Suitability team to physically retrieve the data from the solar lantern.

In the future, CITE will develop improved sensors and data loggers in order to track use more accurately and over longer periods of time. An improved data logger would incorporate a lower power design and possibly a larger battery in order to allow it record data continuously and remain in the field for a longer period of time. A Bluetooth or mobile communications chip will also be added to allow data to be acquired remotely or wirelessly. Lastly, a standard data logger will be used that can interface with a variety of sensors in order to allow this data logger to be used in a wide variety of products.

These sensor and data loggers improvements will allow CITE to collect much more accurate user data. In addition, use patterns can be monitored over time. More accurate use patterns over time would allow the Suitability team to develop more representative and relevant laboratory tests. Product performance could also be tracked over time and this could be part of future product evaluations.

Chapter 5

Reconceptualization and Future Evaluations

In this chapter, an alternative product evaluation method proposed by the Suitability team is discussed. The problem-based evaluation method is designed to address many of the challenges facing the Suitability team in performing the attribute-based product evaluation. In the future, problem-based evaluations may be run concurrently with attribute-based evaluations. This co-evolution of testing frameworks has the potential to provide better product evaluations based on the advantages of each method.

5.1 Problem-based Product Evaluation

After the completion of the pilot attribute-based, comparative solar lantern evaluation, an alternative evaluation model, the problem-based evaluation was proposed. The problem-based evaluation method is inspired by the success of crowd-sourcing enterprises such as *Wikipedia.* This type of evaluation emphasizes obtaining data from large amount of users regarding the behavior of the product under evaluation in order to inform laboratory testing, rather than relying on a team of experts to identify various reasonable and extreme use cases. This section will discuss the formulation, implementation, and advantages of this model of product evaluation.

5.1.1 Overview

In problem-based product evaluations, the goal is to identify, validate, and document phenomena that are relevant to users instead of comparing all models across a pre-determined set of attributes. The output of such an evaluation would be an organized, searchable catalog of verified observations regarding product performance and failure modes, rather than a scoring matrix.

Problem-based product evaluations are based on gathering large of amount of user feedback regarding the product under evaluation. This user feedback takes the form of user anecdotes regarding any phenomena observed regarding product performance and the conditions under which it occurred. A large number of anecdotes are desired in order to cover as many use cases and contexts as possible. This data may be crowdsourced from users through a variety of means, including internet or mobile-device based platforms. There will likely be an incentive system to encourage users to report their observations and provide useful and detailed feedback.

Next, the user feedback must be mined for relevant and verifiable user anecdotes. Factors that determine which anecdotes are selected may include the severity or frequency of the phenomena that is observed. In addition, anecdotes must provide enough specific details such that the phenomena and conditions under which it occurred can be recreated in the laboratory.

These relevant and verifiable user anecdotes are then used to guide laboratory testing. The conditions of each anecdote are recreated and the occurrence of the observed phenomena is either verified or refuted. If a phenomenon is verified, it is added to an organized, searchable database. If it is refuted, it is removed from the CITE resource. The Suitability team may also test related aspects of product performance that are relevant to each observed phenomena.

The resulting database of verified user anecdotes is then made widely available. This information then allows consumers to make purchase decisions based on phenomena other users have encountered. In this way, consumers will have a access to a large amount of relevant information, but there will still be a filtering mechanism to ensure the quality of the information. This allows the Suitability team to provide large amounts of useful information through targeted testing.

5.1.2 Implementation Issues

The implementation of such a problem-based product evaluation presents some challenges to the Suitability team. These include developing a platform to gather user information, creating a method to identify user anecdotes of interest, and disseminating results.

First, developing a platform to gather user anecdotes is a challenge. Internet or mobile-device based mechanisms are obvious choices. However, for the bottom of pyramid consumers that CITE is targeting, even these methods may not be able to easily and reliably reach them. A data gathering system that is not prohibitive in the amount of labor required will need to be developed. Furthermore, it is likely that there will need to be an incentive system to encourage users to submit anecdotes. This will need to be carefully designed so as to encourage not only submissions of anecdotes, but submission of high quality anecdotes.

Second, a method to identify anecdotes of interest needs to be developed. This product evaluation model relies on gathering a large number of user anecdotes; however, the Suitability team will only be able to test a limited number of anecdotes. Anecdotes must be screened on two levels. On the first level, anecdotes must be screened for their specificity and testability. On the second level, anecdotes must be screened on the basis of relevance and importance. The Suitability team must have a way of narrowing down a large number of anecdotes to only those that are relevant and testable.

Finally, there remain challenges to disseminating results with problem-based evaluations. Part of this is related to the data gathering problem. If consumers do not have ready access to the internet, accessing this information will be quite difficult. In addition, there is the issue of familiarity with such types of data structures. Users in the developed world are generally familiar with databases, but users who have not interacted with such methods of storing information may not be able to easily access and use the information stored within. Furthermore, there is a language issue. Users may speak a variety of different languages so it may be difficult to find a common language which all consumers can use.

5.1.3 Advantages

Despite the presence of some challenges, the problem-based evaluation has many advantages. These include the ability account for non-homogeneous user groups, the ability to test product families that are not well-developed, the ability to test non-products, substantial resource savings, and identification of many and unexpected performance issues.

First, the problem-based evaluation method is able to produce relevant results even for non-homogeneous user groups. Users submit anecdotes that inherently take into account the users context. For example, in the case of the shower test for water resistance, the user might specify that a rain shower of 2 inches per hour caused their lantern to fail. The Suitability team would then simply have to verify that this level of rain can cause a lantern to fail, rather than determining what rainfall rate is a reasonable standard. Furthermore, there is no need to create a composite score and ranking system for products. Consumers who use this information will understand what it means in their own contexts and will be able to make buying decision based on anecdotes that are most relevant to them.

Second, products without well-developed product families can be tested using this method. Since problem-based evaluations are not comparative, there does not need to be a direct competitor. The problem-based evaluation method would simply identify the failure modes and deficiencies in the product. This information, without any comparison to other products, can be useful for consumers making buying decisions. Thus, the Suitability team can deliver value to consumers much early in a product's design cycle than with attribute-based testing. Furthermore, CITE can actively work with product designers and manufacturers to make better products by providing feedback early in the design process.

Third, the problem-based evaluation method can be used to test non-products or aspects of products that manufacturers may not have control over. For example, for the solar lantern evaluation, it would have been possible to include Liter of Light [13], even though it is not a product and not directly comparable to solar lanterns. Users would still be able to report their experiences with it. This delivers value to other consumers by providing them with information regarding a product alternative. Consumers would then be able to decide if it fit their needs or if purchasing a solar lantern was necessary.

Additionally, products that are sensitive to such factors as installation, assembly, or user training can be evaluated. This cannot be done with attribute-based comparative product testing, because results would not be easy to compare. With problem-based testing, however, useful user anecdotes regarding failure modes and other phenomena could be gathered and verified. This information may be useful to consumers in making a buying decision or avoiding problems other users have encountered.

Fourth, use of this method can potentially save time and money resources by directing testing to only the most important areas. In attribute-based, comparative product testing, all products have to undergo standardized testing in each of the attributes being evaluated. For example, in the solar lantern evaluation, all models were subjected to all three tiers of water resistance testing, even though approximately half of them appeared to impervious to water damage. In problem-based evaluation, testing is only performed in areas where there may be performance issues based on user experiences. For example, in the context of solar lantern water resistance, only lanterns that users encountered problems with would be tested. Lanterns that were highly resistant to water damage would probably not be tested because it is unlikely that users would observe failures. These resource savings would allow more products and product families to be tested.

Finally, problem-based testing has the potential to identify many more and unexpected product performance issues than attribute-based testing. Attribute-based testing requires the evaluating team to define a set of important attributes to be tested. It is possible that there will be relevant performance issues that are not captured. Problem-based testing can identify many more of these issues because it depends on on users to report these issues when they are encountered. Furthermore, problem-based evaluations can be run on an ongoing basis, with information being continually added, while attribute-based evaluations are finite and provide a snapshot of the market at a specific point in time, requiring follow-up evaluations at regular intervals.

Chapter 6

Pilot Problem-based Evaluation

After the completion of the pilot attribute-based, comparative product evaluation, the Suitability team decided to further explore the concept of crowd-sourced, problem-based evaluation. The author designed and implemented a pilot crowd-sourced, problem-based evaluation of solar lanterns. This chapter will discuss the details of this alternative pilot evaluation.

6.1 Implementation of Pilot Problem-based Evaluation

A pilot problem-based evaluation was conducted at MIT. The pilot evaluation had three main goals. The first goal was to show the viability of the problem-based evaluation method. The Suitability team hoped to produce useful confirmed anecdotes about lantern performance issues. A secondary goal of was to demonstrate the advantages of the problem-based evaluation method by identifying performance issues that were not detected in the attribute-based product evaluation. The third goal was to test some of the assumptions required for the attribute-based evaluation. The two assumptions of interest were the assumption that users are relatively homogeneous in their tastes and preferences and the assumption that products can be ranked ordered in terms of performance in way that is relevant for a large number of users.

This pilot problem-based evaluation was carried out over a three week period with seventeen volunteers at MIT. The pilot evaluation was a simplified version of what might be deployed in the future. MIT staff and students were substituted as users because of ease of access compared to developing world users. They also had access to the internet and it was easy to communicate and exchange information. A small group was selected in order to keep the number of responses small enough that data mining would not be difficult.

Volunteers were given three solar lantern models, which they were asked to use and provide feedback on over this period. Users were asked to subject lanterns to conditions similar to what they might face in the developing world. The three lanterns were chosen from the models that were tested in the attribute-based evaluation. Models B, D, and F were picked for the problem-based evaluation. Models B and D were chosen because they were used in the consumer data gathering stage of the attribute-based evaluation in Uganda. Both of these models were rated slightly below average in the attribute-based evaluation, scoring 46 and 43 respectively. Lantern F was chosen because it was a relatively high scoring lantern in the attribute based evaluation, achieving a score of 60.

A data gathering mechanism in the form of a forum-style website was set up to collect feedback from this group of users. Users were encouraged to submit reports detailing and failure modes or performance issues they may have observed. User submissions were visible to all other users and user interaction was encouraged. During the study, user submissions were mined to identify relevant and testable observations. These were then used to develop laboratory tests to confirm or refute observed phenomena. Test results were disseminated through the data gathering mechanism and users were allowed to comment on them.

At the end of the trial period, evaluation volunteers were given a questionnaire which asked volunteers to rank the three lanterns in the study in order of performance. Additionally, volunteers were asked to complete the MaxDiff survey that was administered to solar lantern users in Uganda. This allowed preferences of MIT study volunteers to be compared with solar lantern users in Uganda. Furthermore, the comparative ranking obtained from the attribute-based testing can be compared with the ranking given by the MIT volunteers.

As stated previously, this pilot problem-based evaluation is simpler version of what may be employed in the future. In order to employ this in a useful manner for future evaluations, the sample size must be much larger and consist of product users in the field rather than surrogate users. A new data gathering mechanism also has to be developed that can be accessed and used by consumers in the field. Finally, more efficient data mining method must be created that can process large numbers of user submissions.

6.1.1 Data Gathering Mechanism

The data gathering mechanism for this pilot evaluation was a forum-style website as shown in Figure 6-1 below. The forum was hosted local on a CITE computer. Study volunteers were each given a username, password, and URL which allowed them to log onto the website and post user submissions without revealing their identities. Volunteers were allowed to log in and post submissions at any time during the study period.

This data gathering mechanism was chosen because it allowed allow user submissions to be available to all study participants. Users were encouraged to interact with each other in order to improve user submissions. For example, when one user identified a problem, others could identify additional conditions under which it occurred. Additionally, this platform allowed laboratory verified results to be communicated easily back to the users. Test results could simply be posted on the forum under a separate sub-forum, effectively creating a cataloged and searchable database of verified results. This also allowed all users to view verified results and provide further comments on them.

A data gathering mechanism employed in the field would be significantly different than that employed in this pilot evaluation. It would have to provide much of the same functionality in terms of collecting data from users and allowing users to view data, but with a much larger number of users. It must be easily accessible to bottom

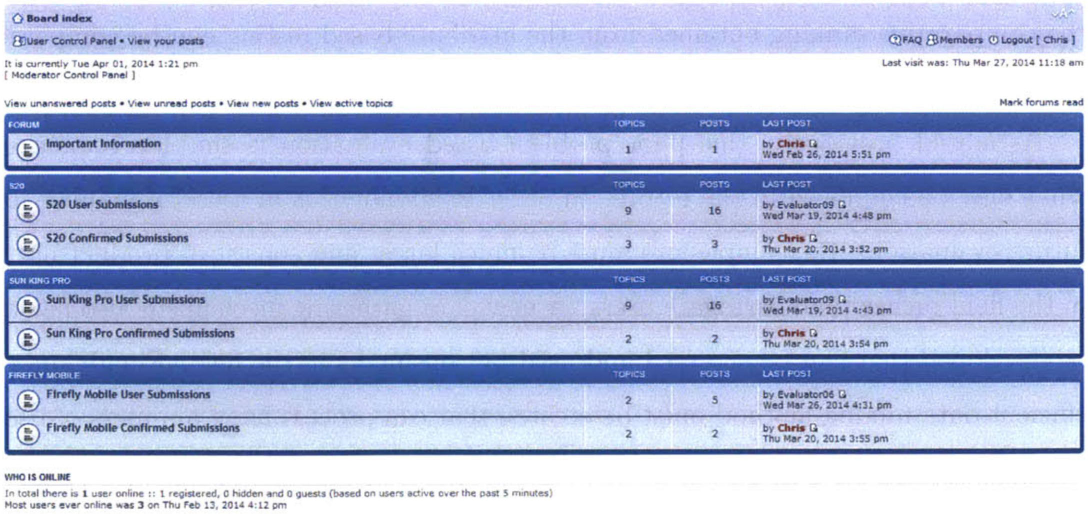

Figure 6-1: Problem-based evaluation data gathering mechanism. The data gathering mechanism was an online forum-style website. The website was divided into subsections for each lantern and users were encouraged to submit feedback by posting on the forum.

of pyramid users. In addition, data needs to be aggregated in such a way that privacy is preserved.

6.1.2 Methodology for Identifying "Problems"

Initially, the methodology for identifying phenomena of interest was simple. The evaluating team mined the data by manually inspecting user submissions and identifying those that were of interest. The factors that determined whether submissions were flagged included severity of phenomena reported, number of submissions documenting similar phenomena, and testability of observed phenomena.

For example, multiple users submitted reports stating that one of the lantern models did not charge well when it was cloudy. This phenomenon was chosen for testing because of the frequency with which users noticed it. One user noticed that another lantern model did not charge at all when the A/C charging cable was plugged in. This phenomenon was tested because of its relative severity. Finally, another user reported that all three lanterns product a light that was too blue and therefore uncomfortable to use. This phenomenon was tested because of the ease with which it could be verified.

For future crowd-sourced evaluations, a more sophisticated method for identifying interesting user submissions will need to be employed. There will be a much larger number of users and therefore user reports. Mining these reports to identify phenomena for testing will need to be done automatically. Additionally, more precise definitions of what observed phenomena are suitable for laboratory testing are needed.

6.1.3 Pilot Users

The users for this pilot problem-based evaluation consisted from volunteers at MIT. Within MIT, these volunteers were selected from staff and students from D-lab and CITE. Volunteers were chosen from D-lab and CITE because they had experience traveling in the developing world and working with bottom of pyramid consumers. With this knowledge and experience, these volunteers would be suitable surrogate users for the pilot problem-based evaluation.

As surrogate users, study volunteers were asked to simulate use conditions that solar lanterns may be subjected to in the field. Surrogate users were desired because the Suitability team felt that they would generate more relevant and realistic results than developed world users. Developed world users are unlikely to encounter the same issues and developed world users. For example, failing to charge a lantern would not result in any significant consequences because it is not their only source of light. Furthermore, issues such as poor charging in cloudy conditions may not be noticed because of easy accessibility of A/C power. Thus, although volunteers from CITE and D-lab are developed world users themselves and therefore not perfect surrogate users, the Suitability team felt that they would nonetheless be able to identify some lantern performance issues that would be relevant to developing world users.

Future crowd-sourced would ideally employ the product user to provide feedback directly rather than surrogate users. Product end users will be able to provide the most relevant and context sensitive information regarding the products that they are using. The challenge for the Suitability team going forward is identifying and reaching these users and developing an appropriate system for gathering product feedback from them.

6.1.4 Results

User responses collected through the data gathering website revealed many interesting aspects of lantern performance that are relevant to users but were either not detected or evaluated during the attribute-based, comparative evaluation.

First, users reported that two of the solar lanterns, Model B and Model F, produced light that was unpleasant to use. A user said that Model B produced light that was too blue and seemed like a fluorescent lamp, while another user simply reported that Model F's light was unpleasant. In the attribute-based evaluation, the Suitability team had, in fact, evaluated the color of light produced by solar lanterns by measuring the spectral power density of the light, shown in Figure 6-2. In consultation with Dr. Peter Bex, a vision researcher at Harvard Medical School, the Suitability team had concluded that all lanterns produced light with similar spectral properties which would not significantly affect a user's ability to read and perform other tasks with the solar lanterns. Therefore, this aspect of performance was not considered in the attribute-based evaluation. Despite this however, this property of solar lanterns appears to relevant to users and may warrant further investigations. Even if it is not possible to differentiate between lantern models, it may benefit future designs to change the spectral output of the LED.

Second, a user discovered that Model D could not be charged using the solar panel if the A/C charging adapter was left plugged into the solar lantern. This issue was not identified or tested in the attribute-based evaluation because the ability to charge from A/C power was simply identifies as present or not. Even if this feature had been fully tested, it is unlikely that the Suitability team would have detected this phenomenon due to the structure of the attribute-based testing process. All products are subjected to a standard testing procedure, so phenomena which occur under very specific use conditions will not be revealed. Nonetheless, phenomena such as this may be relevant to end users as it may affect their use of the product.

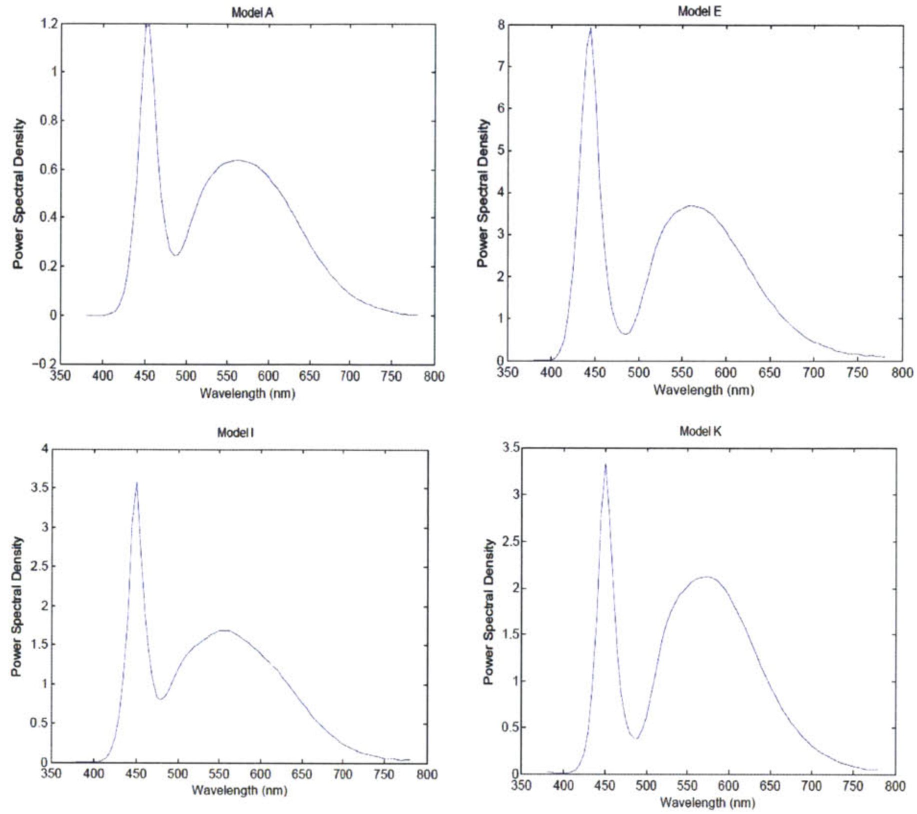

Figure 6-2: Spectral power density of select solar lantern models. The solar lanterns in this evaluation had LED's which all produced a very similar spectral power density. The Suitability therefore chose to not include these results in the final evaluation.

Third, many users reported that Model F had particularly poor charging performance as compared to Model D when there is no direct solar insolation. That is, Model F appears to charge more slowly than Model D on a cloudy day or when placed in a shaded area. This is consistent with solar sensitivity results from the attribute-based evaluation. In this case, users identified an issue that was detected by the Suitability team and tested in detail. Solar sensitivity testing was not, however, initially among the attributes to be tested and was only added when the Suitability team noticed similar performance issues. In the future, this user feedback could be used to accelerate the identification of performance issues such as this.

In addition to gathering feedback through the forum-style website, users were given a survey which asked them to indicate which of the three lanterns under evaluation were the best and worst performing models. In addition, there were asked to fill out the same MaxDiff survey as Ugandan solar lantern users were during the attribute-based evaluation. Only 14 of 17 problem-based evaluation participants returned the survey. Despite this, however, the results of the survey administered also revealed interesting information about user preferences and the ranking of solar lantern models in the attribute-based evaluation.

The MaxDiff survey results, shown in Figure 6-3 revealed that Ugandan users and surrogate users had different priorities regarding the importance of specific solar lantern characteristics. For example, Ugandan users rated durability as the most important characteristic, while surrogate users rated durability as the fifth most important characteristic. Water resistance was in the middle for Ugandan users, while it was the second least important attribute for surrogate users. Thus, the two different groups of users placed widely differing weights on specific solar lantern attributes.

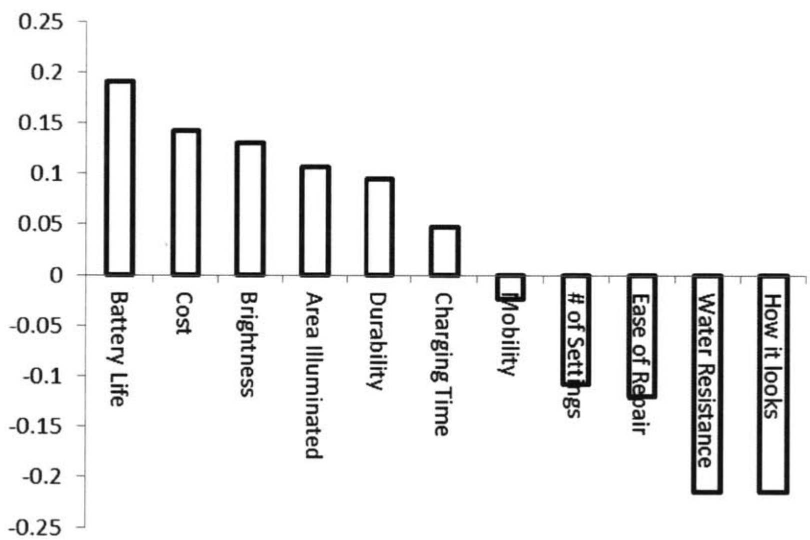

Figure 6-3: MaxDiff survey of lantern characteristics administered to surrogate users. This plot shows the relative importance of each solar lantern characteristic. Results obtained by surveying surrogate users of the pilot problem-based evaluation. Note that these are results from the 14 participants who responded to the survey.

Taking a broad view, however, both user groups still have a degree of similarity, with the two MaxDiff surveys having a Kendall's τ coefficient of 0.4545. More specifically, looking at the four most important attributes as group, it is apparent that there is a large overlap. Brightness, area illuminated, and battery life are in the top four for both user groups. The four least important attributes show a similar pattern, with number of settings, ease of repair, and aesthetics common to both user groups. Thus, although users from each group may place different levels of importance on each specific attribute, they were more similar in broadly defining groups of more and less important attributes.

These initial findings suggest that more work is needed in order to understand to what extent different user groups in varying contexts are homogeneous or heterogeneous in their preferences. If different user groups have broadly similar preferences as observed here, it may be possible to create more general product evaluations that are relevant to a large cross-section of users. Thus, future work identifying the preferences of various user groups in differing contexts could be useful for future evaluations.

Users were also asked to identify the solar lantern model that they felt performed the best and the worst. These survey results are shown in Figure 6-4 for the best performing lanterns and Figure 6-5 for the worst performing lanterns.

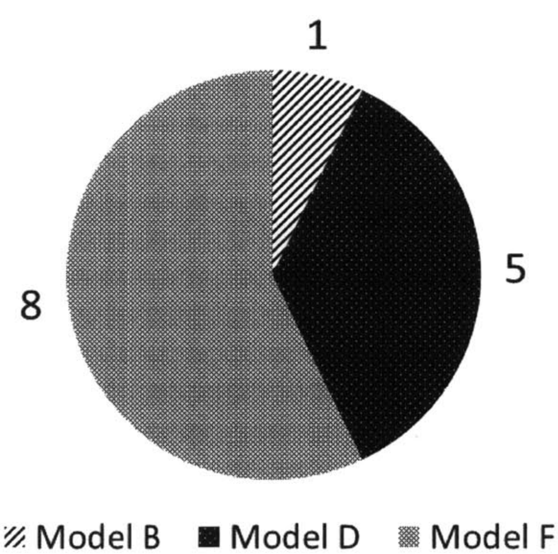

Figure 6-4: Highest performing lantern models in pilot problem-based evaluation according to surrogate users. This plot shows the number of users that selected each lantern model as the highest performing lantern.

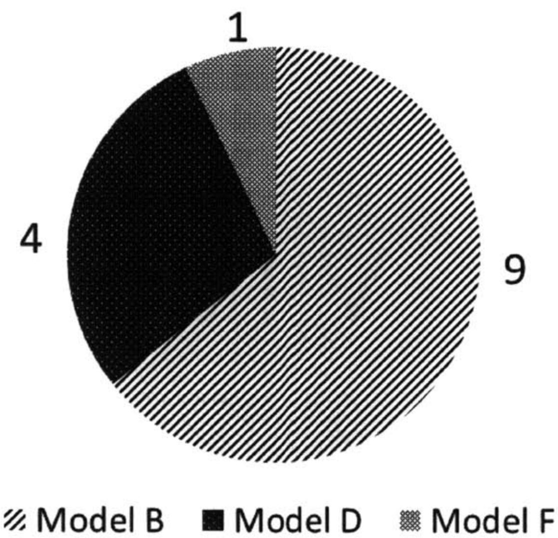

Figure 6-5: Lowest performing lantern models in pilot problem-based evaluation according to surrogate users. This plot shows the number of users that selected each lantern model as the lowest performing lantern.

From these results, a clear ranking emerges. One user selected model B as the best performing lantern while 9 users identified as the worst performing lantern. For Model D, 5 users selected it as the best performing lantern and 4 users identified it as the worst performing lantern. Lastly, 8 users selected Model F as the best performing lantern while 1 user identified is as the worst performing lantern. Thus, according to the surrogate users in this study, Model F is the best performing lantern, Model D is in the middle, and Model B is the worst performing lantern.

This compares somewhat favorably with the rankings from the attribute-based evaluation, in which Model B had a score of 46, Model D has a score of 43, and Model F had a score of 60. Model F was selected as the best performing lantern in the problem-based evaluation, which is consistent with that fact that it scored much higher than the other two models in the attribute-based evaluation. Models B and D, which scored lower in the attribute-based testing are also selected by users as lower performing lanterns. Thus, the two evaluations agree on the rankings to a large extent.

For Model B and Model D, however, their scores were close in the attribute-based evaluation, while surrogate users clearly selected Model B as the worst performing lantern. This discrepancy suggests that some minor adjustments may need to be made to the methodology of the attribute-based evaluation method to bring its results in line with user preferences. Adjustments may be as simple as changing the weightings used to calculate the composite scores. This may require further research into user preferences, some of which may involve non-technical factors.

Overall, however, there seems to be a some amount of agreement between the two evaluations methods, suggesting that with further refinement, it will be possible to create product evaluations that are indeed reflective of user preferences and therefore relevant to consumers.

6.2 Discussion of Pilot Problem-based Evaluation

The pilot problem-based evaluation was able to demonstrate many of the advantages of this testing methodology. The Suitability team was able to identify many performance issues that were relevant to solar lantern users and address them with specific, directed testing. While some of these issues were discovered by the evaluating team in attribute-based testing, others were not and are unlikely to be detected in a rigid testing regime. Furthermore, the pilot problem-based evaluation required a lower level of resource commitment to perform the evaluation. The problem based evaluation was performed over a three month period by the author. In comparison, the attribute-based evaluation was performed over an eight month period by a team consisting of three graduate students, in addition to other staff.

In the future, it may be beneficial for CITE to conduct both types of evaluations, in order to not only generate complete performance data, but also discover and provide useful information to consumers that may be difficult to obtain with a standardized testing process. A problem-based evaluation can also be used to inform an attribute-based evaluation, saving resources that would otherwise be spent identifying relevant product attributes and on tests that may not be relevant to consumers.

Although the initial results of the pilot problem-based evaluation are promising, there remain many challenges to implementing this type of methodology to evaluate actual products in the field. These challenges include developing a feedback gathering mechanism that is accessible for large number of users with limited access to the internet and who may speak many different languages and developing a method for processing the large amounts of data that would be acquired. Solving these challenges would allow for more product evaluations in the future, by allowing these two methodologies to be used in conjunction, providing the advantages of each.

It may also be possible to improve the usefulness of the problem-based evaluation methods. For example, it is possible that surrogate users may be better able to identify issues relevant to actual users if they are shown the MaxDiff survey results of actual users. Surrogate users might then be able to probe lantern performance in a way that is more representative of the preferences of actual users. This and other modifications to the problem-based evaluation methods are an area of possible future work for CITE.

Chapter 7

Conclusions

The Suitability team designed and implemented two different product evaluation methodologies, an attribute-based comparative methodology and a problem-based methodology. Each methodology produced useful results and had different strengths.

Using the attribute-based comparative methodology, the Suitability team was able to combine market intelligence, user data gathered through consumer surveys and dataloggers, and lab tests in order to produce a comparative product evaluation and ranking. Products were evaluated on eight different attributes deemed to be relevant to Ugandan users and in which products had observable differences in performance. The overall ranking was determined by weighting using input as well as the judgment of the evaluating team.

Using the problem-based methodology, the Suitability team was able to identify relevant product performance issues that were not detected in the attribute-based evaluation. In addition, the surrogate user group produced a similar product ranking and indicated similar weightings on the importance of product attributes. This suggests that diverse user groups may have similar preferences and that product rankings may be at least somewhat generalizable.

Taken together, use of these two methodologies in parallel may be able to generate relevant product evaluation reports with a reasonable time and resource commitment. The more frugal problem-based evaluation can be used to direct resources in the more resource intensive and thorough attribute-based evaluation. This is, in fact, the hybrid product evaluation model initially proposed by CITE, with detailed *Consumer Reports*-style laboratory testing supplemented by *in situ* data gathering and interaction with end users. More work is needed, however, in determining to what extent user groups are homogeneous across varying contexts, how well do product rankings reflect actual user preferences, and what is the best way to disseminate the results of these product evaluations.

Appendix A

Solar Lantern Testing Protocols, Results and Scoring

This appendix provides more detailed information about the test protocols and scoring methods for each of the tests performed in the attribute-based, comparative solar lantern evaluation. This section is excerpted from the technical report (Brine, et al) submitted to USAID.

A.1 Spectral Analysis

A.1.1 Motivation

CITE worked with a researcher, Dr. Peter Bex, from Harvard Medical School to determine the relative importance of certain characteristics of light including intensity and spectrum according to their impact on human vision. Through these discussions, it was determined that if the spectrum of light emitted by each lantern could be eliminated as a variable, then cross-comparison based on lighting performance could be solely determined based on the intensity of light. Thus, CITE chose to conduct a spectral analysis of the product family.

A.1.2 Description

Spectral analysis was performed by characterizing each lantern's light on multiple settings with a PR-655 Spectrascan spectroradiometer.

A.1.3 Results and Scoring

Spectral analysis yielded the important result that each of the lanterns displayed similar spectral characteristics. This is expected since each lantern is illuminated with a white light emitting diode (LED). A typical result from the spectroradiometer is shown in Figure A-1 below.

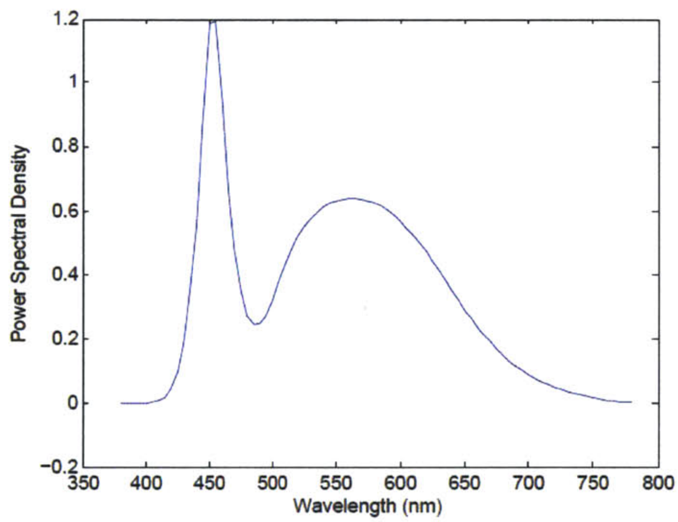

Figure A-1: Sample solar lantern spectral power density. This plot shows spectral power vs frequency for a single solar lantern model. Note that all solar lantern models exhibited a similar spectral power density.

The detailed physics of LEDs are beyond the scope of this paper; however, notice that there are two spectral peaks. The first is the excitement wavelength, an initial burst of light between 400 and 450nm, and the second is the duty wavelength. It is this second peak that determined the spectrum of light visible to humans when the unit is switched on. Figure A.1 below shows the results from spectral analysis by lantern. As is evident from the data, the lanterns displayed a similar spectral power density. The peaks are at roughly the same frequency and the relative magnitudes are also similar. This result simplified comparing each lanterns performance in the subsequent task and ambient lighting tests as each could be compared based on intensity alone.

Lantern	Primary Peak (nm)	Secondary Peak (nm)
Model A	~450	~560
Model B	~450	~550
Model D	~450	~570
Model E	~450	~560
Model F	~450	~580
Model G	~450	~580
Model I	~450	~560
Model J	~450	~560
Model K	~450	~570

Table A.1: Table of spectral peaks. This table shows the frequency at which each lantern model has a spectral peak in light output.

Because the lanterns produced light at very similar wavelengths, they would appear very similar, if not indistinguishable to the human eye in terms of the color of the light. This particular test was thus used as a screening mechanism and, as such, did not affect the overall rating score.

A.2 Charge Time

A.2.1 Motivation

Time to charge is an important variable in a portable solar lighting device. In order to charge a solar lighting device, users must leave it exposed to the sun for an extended period of time. The shorter this period of time is, the less likely the unit is to be exposed to damage or theft. Furthermore, shorter time to charge gives a user greater flexibility in taking advantage of favorable charging conditions.

A.2.2 Description

In order to best simulate conditions that a lantern would be subject to when being used by consumers, the evaluation team elected to perform charge testing outdoors using natural sunlight, rather than with a solar simulator. In this way, the efficiency of the entire lantern system would be measured, rather than just the solar panel as is usually the case. Additionally, the sensitivity of the lantern to changing lighting conditions could be measured.

Because the sun was to be as the light source for this test, it was performed on the roof of D-lab (Building N51 at MIT). A lantern charging station was constructed, which held the lanterns in secure manner. The lanterns were attached to platforms which were removable from the charging stations in order to allow for the lanterns to be moved indoors in case of inclement weather. The charging station can be seen in the Figure A-2. The charging test was primarily performed during dry weather in order to prevent damage to the lanterns. A mix of sunny and cloudy days were utilized to obtain best case charging times, as well as to observe degradation in charging performance due to clouds.

Figure A-2: Solar lantern charging station. This charging station is located on the roof of building N51 at MIT. Solar lanterns were placed on these platforms for the charging tests.

Battery voltage and current measurements were taken in a manner similar to that used for the field study. A voltmeter was connected across the battery terminals in order to measure battery voltage. A small resistor was connected between the battery positive terminal and the rest of the lantern circuit. This allowed for measurements of current flow to be taken. These two parameters were used to measure battery state and charge rate of the lanterns. A wiring diagram can be seen in Figure A-3 below. In contrast to the field study, however, the data was read by a National Instruments cDAQ data acquisition system rather than an Arduino. This allowed for data to be taken very reliably over a long period of time.

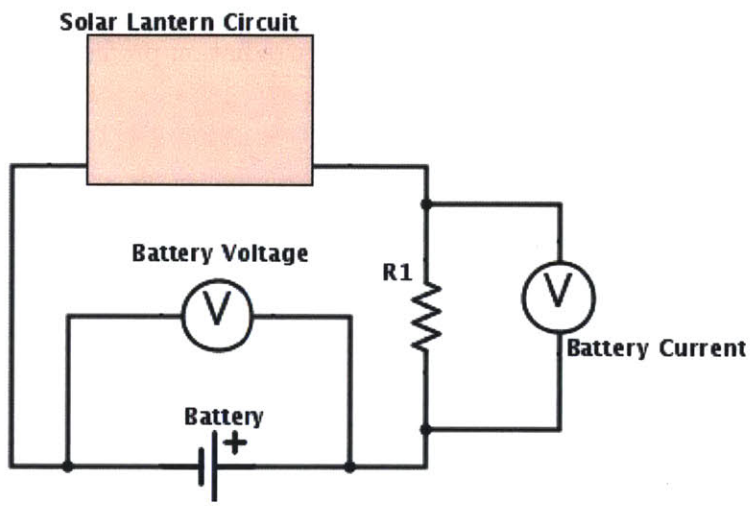

Figure A-3: Solar lantern instrumentation diagram. Voltmeters were used because an ammeter with suitable rang and resolution was not available. R1 was 0.1 or 0.5 Ohms depending on the lantern model.

In addition to monitoring the lanterns, a pyranometer was mounted on the roof adjacent to the charging station that records both incident solar radiation and diffuse radiation. The pyranometer used was a model SPN-1 Sunshine Pyranometer from Delta-T Devices.

The test was performed by first fully discharging the lanterns. They were monitored using a method similar to the discharge test described in the next section to ensure they were, in fact, fully discharged. The lanterns were then connected to the data acquisition system and left to charge over the course of two days. Two days was selected as the test duration because some lanterns had large batteries and would not become fully charge in one day. Two days allowed all lanterns to become fully charged.

The amount of time each lantern took to charge was determined by examining the current and voltage traces obtained from the data acquisition system. Time to charge was determined by identifying the point at which the charge current dropped to zero, despite the continued presence of solar irradiation as measured by the pyranometer. Most lanterns also exhibited a concurrent spike in voltage measured in the circuit.

In order to obtain consistent results, two or three samples of each lantern model were used in this test. In addition, this test was performed twice over different consecutive two day periods with similar solar irradiation profiles.

A.2.3 Results and Scoring

Although all lanterns were able to fully charge within the two day test, there were significant variations in the charging times. This is likely due the different battery capacities and differently sized solar panels. Some lanterns provide batteries with smartphone-level capacities but a small solar panel and these lanterns take a long time to charge. Other lanterns have smaller capacities and charge much more quickly. There is also likely some variation due to internal efficiencies, battery chemistry, and charge control.

Because of the fact that much of the charge time variation is due to differently sized batteries, it is not enough to evaluate these lanterns on charge times alone. They will undoubtedly also have highly variable discharge times. For example, a lantern that takes a very long time to charge may still be acceptable if the resulting full charge allows the lantern to stay on for a very long time. Thus, a discharge time test is also required to fully determine lantern performance and is described in the next section.

Most lanterns exhibit a charge control feature in which the current entering the battery will be regulated as the battery becomes full. This prevents overcharge of the battery. Figure A-4 that shows this behavior.

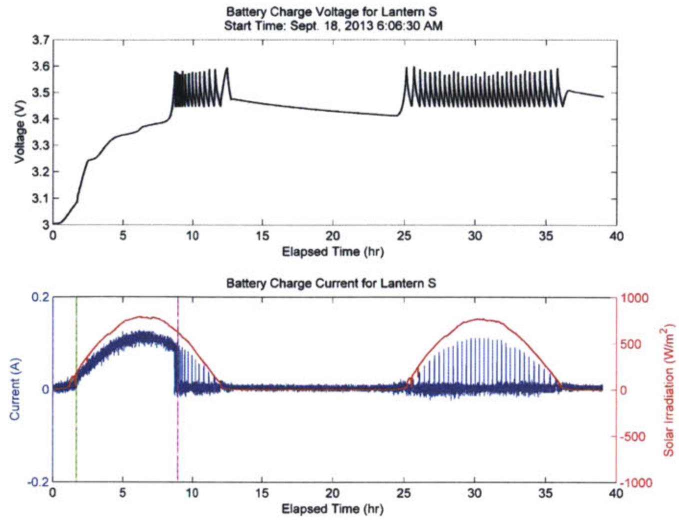

Figure A-4: Charging behavior of a solar lantern. This plot shows the charging current for a single solar lantern model in blue. This lantern was fully charged in the first day of the test.

To score this test lantern were rank ordered by their time to charge as determined by identifying the time at which the charging current dropped to zero despite the continued presence of sunlight. Each lantern had 3 models (in some cases 2 models) subjected to the same conditions on two different days in which the max insolation and the insolation patterns was approximately the same.

A.3 Discharge Time

A.3.1 Motivation

Lantern discharge testing is perhaps the most important test that CITE undertook. The amount of time a lantern can stay on is very important attribute from the perspective of consumers. This determines how useful a lantern is. In addition, this is an area in which manufactures make many of their headline claims. This piece of information is generally prominently featured on the product packaging and in marketing materials. Determining the actual performance of these lanterns is thus crucial.

A.3.2 Description

Time to full discharge is described as the time a lantern will remain illuminated at a usable intensity starting from a full battery charge. This attribute was measured using the same lantern solar charging and data recording station used for the charge time test described in the previous section. This was performed primarily at night due to the ease of performing discharge only, due to the lack of sunlight. However, because some lanterns were able to remain on for longer than one night, testing sometimes continued into the morning. In these cases, the lanterns were disconnected from there solar panels where possible. For lanterns with an integrated solar panel, a black polyurethane cover was applied to the solar panel. In this way, lanterns were prevented from charging. Similar to the charging test, lantern battery voltage and battery current were logged throughout the test. In addition, a time-lapse camera was installed to record the lanterns as they discharged to note any visible phenomena occurring during the discharge period.

In order to perform this test, lanterns were first fully charged. This test was performed after the charging test. It is thus possible to verify the charge state of the lantern prior to the discharge test. The lanterns were then prepared for discharge as described above. Lanterns were then turned on to the highest setting and allowed to discharge over a period of 30 hours. 30 hours was selected as the test period because based in previous observations, the longest on time of the lanterns was in the range of 20-25 hours. This duration allowed all lanterns to be completely discharged by the end of the test.

In order to obtain consistent results, two or three samples of each lantern model were used in this test. In addition, this test was performed twice over two different nights.

A.3.3 Results and Scoring

The data showed that lanterns demonstrated two distinctive discharge patterns when switched on and left to fully discharge. One class of lanterns began at 100%

of full brightness intensity, slowly diminishing in intensity until the battery could no longer run the LED and they were extinguished. This behavior is shown in Figure A-5 below.

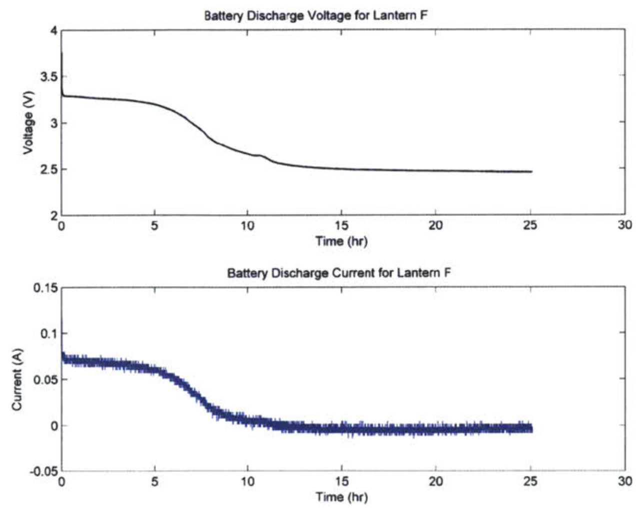

Figure A-5: Discharge behavior of solar lanterns with no constant current circuit. The LED dims over time, with the brightness asymptotically approaching zero.

Other lanterns employed a more sophisticated circuit to extend the amount of usable light for the user. These lanterns began at 100% of brightness and continued at this brightness until they reached a battery voltage level predetermined by the lantern manufacturer, at which time they stepped down to a lower brightness setting and continued to discharge at the lower brightness level. Some lanterns continued this process through all the possible brightness settings until finally the battery did not hold enough charge to run the unit at even its lowest setting. At this time they would shut off. This behavior is shown in Figure A-6 below. These results were corroborated by time lapse photography taken during discharge.

These two behaviors presented an interesting problem when trying to determine the end of charge. For lanterns with a charge controller, the end of charge was obvious. For lanterns that dimmed over time, however, there is not a single obvious point where this occurs. Furthermore, as the lantern dimmed, it was not clear that a small level of light output would still be useful to a consumer. This ultimately drove

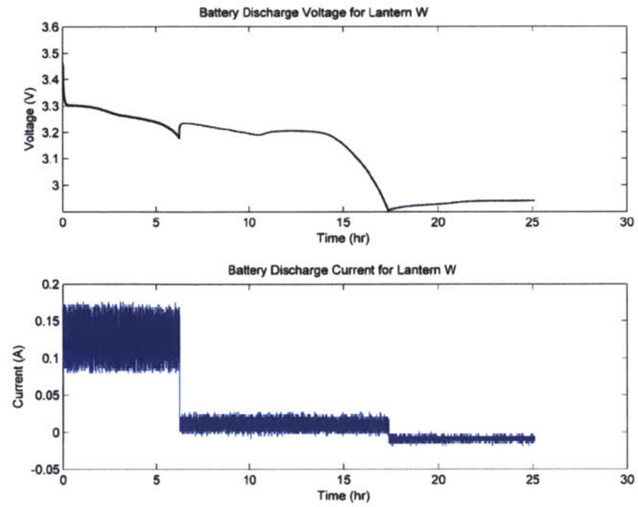

Figure A-6: Discharge behavior of solar lanterns with constant current circuit. Smart circuit monitors charge state of battery and switches lantern to lower setting at predetermined charge level.

the team to develop another laboratory test to characterize the current-luminance curve for lanterns that exhibit dimming behavior.

This test allowed the team to correlate the lantern battery current with its light output. Observing that the largest ratio of brightness at lowest setting to brightness at highest setting was approximately 0.05 among lanterns with more than one setting, this was selected as the point at which all lanterns were considered discharged. Using the data correlating brightness to current, the time at which the lantern was discharged was determined by examining the current trace taken from the lantern during the discharge test.

A.4 Solar Sensitivity

A.4.1 Motivation

Solar sensitivity is defined as the amount of current entering the battery per unit flux of incident sunlight. This was not initially planned in the suite of tests CITE was to perform. Rather, this test evolved as it became apparent that several models exhibited behavior indicating that overcast skies might impede the rate at which the unit could charge. Since oftentimes in a tropical climate like Uganda the rainy season means overcast skies for a long period of time, the results of this test may prove important in the buying decision. Further, it alerts the manufacturer to an issue that they may want to investigate to remain competitive in the marketplace.

A.4.2 Description

No additional testing was required for completing this analysis. Instead, plots of current entering the lanterns battery versus solar insolation were prepared for each model.

A.4.3 Results and Scoring

Most solar lanterns exhibited a roughly linear relationship between charging current and solar insolation above 100 - 200 W/m^2. Below this range of solar insolation, however, most solar lanterns exhibited a large drop-off in the charging rate. This behavior is shown in Figure A-7.

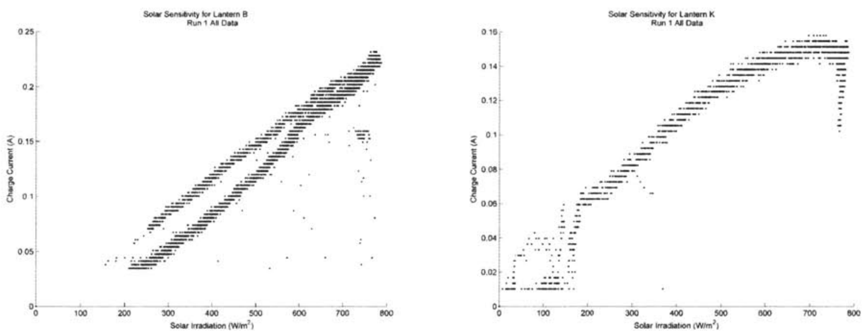

Figure A-7: Solar sensitivity of solar lanterns. These plots show charging current as a function of solar insolation for two different models of solar lanterns. Note that they appear to be fairly linear with a drop-off in low light conditions.

Because all lanterns are similar in there linear behavior in high light conditions and drop-off in performance in low light conditions, it was decided that lanterns would be rated on the lowest solar insolation at which they could charge. Scoring the linearity or charging rates at higher solar insolation levels probably would not be very useful to consumers because the performance was so similar that differences would not be meaningful. Knowing the lowest lighting level in which a lantern could charge, however, could be useful to consumers in deciding whether or not to put a lantern out to charge on a cloudy day. This could prevent users from risking damage to their lantern from attempting to charge in suboptimal lighting conditions. Thus, lanterns were scored based on the minimal light level required to be able to charge.

A.5 Task Lighting

A.5.1 Motivation

Task lighting is an important use category for solar lighting devices. Common tasks for which solar lighting devices can be used include studying, reading and cooking. In fact, lighting with traditional kerosene lamps can be dangerous for these activities, which often require close proximity to the light source. Solar lanterns have the potential to provide lighting much more safely in these conditions. However, it is necessary to determine how well the various solar lanterns models provide task lighting.

A.5.2 Description

In order to comparatively analyze each lights performance for task lighting the team subjected each lantern to the following protocol. Lights were secured to an adjustable tripod placed directly above a photometer placed on a flat, dark surface in a dark room. Each light was then turned to its highest setting and adjusted such that the photometer read 25 lux, a predetermined standard set by Lighting Africa [11]. The distance between the photometer and the light source was then recorded. The photometer was then moved radially in three angularly equidistant directions on the same plane until the photometer read 12.5 lux. The distance at each of these locations was then recorded. The same procedure was then followed until the photometer read

6 lux and the distance recorded.

After recording these values, the lantern was turned onto its next setting and the procedure was repeated until all settings had been evaluated.

A.5.3 Results and Scoring

Lanterns were scored based on the area that they could illuminate at a level of 12.5 lux. The area was determined by taking the average of the three distances measured between the lantern and the photometer at 12.5 lux as described above. The area was then assumed to be circular and the average distance was used as the radius in order to calculate the total area illuminated at 12.5 lux. The area was assumed to be circular because as the distance from the lantern to a surface increases beyond a few inches, the light from pattern is fairly uniform and approaches a circular shape. The 12.5 lux level was selected because it is roughly twice the output of a kerosene lantern [14]. Additionally, from our observations, many lanterns were able to illuminate a fairly large area at this light level. A large area illuminated at 12.5 lux would be much better than a kerosene lantern.

A significant range in performances was observed across the various lantern models. For example, the highest performing lantern in this category was able to illuminate an area of 2,260 square inches, approximately five times the area illuminated by the lowest performing lantern. It is important to note, however, that performance in this test is determined in significant part by the embodiment of the lantern. For example, a lantern designed as a flashlight, which has a highly directed beam, will perform worse on this test than a lantern designed as a floodlight, which has a highly diffuse light pattern.

This test does not distinguish between different lantern types. This may be a qualitative factor for the consumer to consider in deciding which lantern to select. In the ratings chart, different lantern forms are distinguished. This will allow the consumer to compare different lanterns within the same embodiment category and also allow them to compare performance across categories.

A.6 Ambient Lighting

A.6.1 Motivation

Ambient lighting, or the ability to light an entire room for general purpose, proved to be a common and important characteristic to many interviewees in Uganda. Oftentimes each room in a home will be occupied and used by several family members at once, all of whom are completing different tasks. In this case, it was observed that a solar lantern might be mounted on the ceiling. Family members would use this diffuse light, since they had to share the light source.

A.6.2 Description

In order to test the ability to provide general-purpose light, each lantern was suspended from the ground at 6 feet inside a dark room. The point directly under the photometer will be called the origin. A photometer was used to record brightness readings at four different points at ground level. These points are each located three feet from the origin and are 90 degrees apart, forming a cross. This is then repeated at a height of 30 inches and 66 inches, for a total of twelve data points.

A.6.3 Results and Scoring

The brightness readings from this test were weighted by location and then summed to produce a raw score. There was a wide range of performance in this test, with some lanterns only able to provide minimal illumination at each of these points, while others were able to illuminate the points at greater heights reasonably well. Almost all lanterns had difficulty illuminating the points at ground level.

A.7 Setting Versatility

A.7.1 Motivation

This simple test is designed to give a user a sense of each lanterns luminance range, or the relative brightness of the lowest setting to the highest setting. Many users in Uganda use their solar lights for several purposes including reading, ambient light and security light. Each of these tasks requires a different intensity of light. For instance a search task might require a very bright light while a nightlight or security light may only require a dim glow. Lanterns that are able to operate in a range of brightness settings prove to be more flexible than those that only offer one brightness setting or a very limited brightness range.

A.7.2 Description

This test is designed to comparatively evaluate each lanterns lighting range; the range of illumination that it can produce from its lowest to highest setting. This is a rather simple test in which the lantern is placed at 18 inches from the photometer and cycled through its various settings while the intensity at each setting is recorded in turn. The test occurs in a dark room.

A.7.3 Results and Scoring

Lanterns were scored based on the relative of brightness of their highest setting compared to their lowest setting. Lanterns with a larger ratio receive a higher score and vice versa. Larger variations are viewed as favorable because they allow users more flexibility in the usage of their lantern. For example, they may be able to use the highest setting for ambient light for an entire room. A medium setting could then be used for reading. Finally, a low setting could be used as a nightlight, allowing users to conserve power to they have light for an entire night.

In this test, a wide range of performances was observed. Several lanterns have only one brightness setting and thus, are not very versatile in terms of lighting flexibility.

There were many other lanterns with multiple settings but still did not provide a very large range of lighting levels. One lantern did particularly well in this category, providing a ratio of approximately 0.03 of lowest brightness to highest brightness.

A.8 Water Resistance

A.8.1 Motivation

Exposure to moisture can present serious problems for electronic devices. For those living in extreme poverty, investing in what for them may be an expensive product that is easily ruined by moisture and not easily repaired can cause undue financial hardship. Further, existing testing organizations have failed to develop standardized moisture exposure tests that take into consideration the real operating conditions under which many of these devices might find themselves. As such, CITE developed a rigorous a three-tiered testing system in order to rate the lighting units based on their resistance to water.

A.8.2 Description

The following battery of tests was developed in order to determine each unit's ability to resist water damage. Importantly, in order to determine the effect of each of the following tests on the performance of the lantern, each sample tested was subjected to the Task Lighting procedure described above prior to and after being tested for water resistance. Any variation in these tests constituted a measurable effect of the water test on the unit. For each of the three tiers of testing given below a different test sample was used for each of at least 3 trials. These test samples were only used for water resistance testing and each was fully charged before this testing began.

Tier 1 Resistance: Complete submersion

In this test a sample of each model of lantern and its associated solar panel was completely submerged in a tank of water such that its electronics were beneath

6 inches of head. At this depth the sample being tested was rotated to find the orientation that allowed the maximum amount of water to penetrate the housing as determined by visually observing the rate of air leaving the unit. Once air ceased leaving the unit the sample was left underwater for a period of 60 seconds at which time it was removed from the water. Clearly, this test represents a worst-case test for water resistance.

Tier 2 Resistance: Exposure to shower in the most vulnerable orientation

In this test a fresh sample of each unit was subjected to a uniform, constant flow rate of 4 inches/hour impacting the entire exposed surface of the unit for a period of 15 minutes. A 4inch/hour rainfall is typical of a heavy tropical storm [15]. The surface exposed was the most vulnerable surface as determined visually by the testing team. Characteristics of the most vulnerable position varied by lantern but included such attributes as open ports, fabric handles that could wick, or other penetrations in the housing that may allow water to contact the electronics of the unit.

Tier 3 Resistance: Exposure to shower in as charging orientation

In the final water resistance test, each lantern was subjected to the same treatment as in the most vulnerable test, except the lanterns orientation was set up as if it were charging. This means that the solar panel was oriented such that it was perpendicular to the sky and thus to the flow of water. In the case of lantern models that did not have an integral panel, the lantern was placed in its normal desktop use position.

Draining and Drying and Performance Testing for Tier 1 to 3 Water Resistance Testing

The team sought to recreate the manner in which a user might realistically try to drain and dry a unit after exposure to water. These procedures remained the same across all three tiers of testing. Once each unit had been removed from further contact with water, it was rotated to obtain the maximum rate of water drainage from the housing, determined visually, and left in that position until water had had ceased to exit. The unit was then shaken off and dried with a towel.

After draining and drying each lantern was observed for noticeable changes in behavior, such as spontaneous discharge. Any changes were noted. Subsequently the unit was then turned on and any changes in behavior were noted and recorded.

Afterward, if the unit could be easily disassembled without the use of tools, then those parts that could be removed for the purpose of drying were removed. Each unit was then placed in a fume hood to dry. After 1 hour had elapsed each lantern was again turned on and any changes in behavior were noted. The procedure was again repeated at 24 hours, at which point the lantern would once again be run through the Task Lighting procedure.

A.8.3 Results and Scoring

Lanterns were scored based on the highest tier of water resistance testing which they passed and the extent to which their performance was degraded.

The most interesting result from the water resistance tests was the fact that more than half the models, Models B, C, D, F, G, H, I, and K, passed the worst-case scenario, Tier 1 Total Submersion Testing. All of these models achieved the highest score for water resistance. Furthermore, the fact that so many lanterns passed the test gave credibility to its inclusion as a gold standard for manufacturers to meet.

This particular test presented an interesting situation which comparative product testing organizations often face. Some lanterns had user manuals which specified that the unit should be exposed to water in certain areas and certain orientations. However, CITE evaluators felt as though given the environments and use patterns discovered in the fieldwork, it was not reasonable to assume that water would never contact the vulnerable area of the device. In this type of situation in which the manufacturer has warned the consumer that vulnerability exists but the reasonable use case suggests that vulnerability may be exposed, it is important to take into consideration several factors in determining if one should test and subsequently report on the vulnerability. This includes the likelihood that the vulnerability would be exposed in everyday use, the performance of, the presence of the vulnerability. In this case, many lanterns failed permanently and completely. Model A and Model E, failed all three water resistance tests, while Model J was only able to pass the tier 3 shower test. Due to the nature of these failures, this test was deemed important.

A.9 Features

A.9.1 Motivation

Features are generally considered to be those characteristics that exceed the set of traits that define the product family. Sometimes these features are simply bells and whistles that are present to increase the price, but not provide greater convenience or function. Nonetheless, during this study, the testing team identified several key features, each of which is highlighted below. Key features are those that were either commonly found within the product family, identified through consumer interviews in Uganda, or identified by the evaluation team to be potentially useful to a user.

A.9.2 Description and Results

Battery Charging Indicator: During solar charging, it is convenient for the lantern to in some way indicate that it is charging. This allows a user to know that their unit is operational and receiving a charge. The best of these was an LCD that showed a variety of functions including the state of charge. Only one lantern did not have this indicator.

Battery Charged Indicator: This feedback is quite helpful for the user to be sure that they have attained a full battery charge. As noted in the time to charge testing, some lanterns took two days or more to completely charge, in which case this feature becomes even more relevant. About half of the lantern models tested had this feature.

Ability to Charge from A/C Power: On occasions when the sun is obscured by cloud cover, one might want to have the option to charge their lantern through alternative means such as an A/C outlet. About half of the lantern models tested had this feature.

Mobile Phone Charging: The Suitability team discovered through field interviews that cellphone charging was a major driver in the decision to purchase a solar lighting device. During the testing process, however, the testing team discovered that each lantern that provided this capability had varying degrees of performance and convenience issues. The most common findings centered around the ability to charge ones mobile device while also charging the lantern, the mode of mobile phone charging, and the extent to which a charged lantern could charge a mobile device with reserves left over for its own power.

Appendix B

Attibute-based, Comparative Evaluation Raw Scores

This appendix contains raw scores obtained from the testing phase of the attribute-based comparative product evaluation. For more details on the testing procedure, see Appendix A.

Model	Time to Charge (hr)
Model A Sample 1	No Data
Model A Sample 2	6.9
Model A Sample 3	14.4
Model B Sample 1	7.4
Model B Sample 2	6.9
Model B Sample 3	7.3
Model C Sample 1	13.9
Model C Sample 2	13.0
Model C Sample 3	13.0
Model D Sample 1	No Data
Model D Sample 2	10.0
Model D Sample 3	9.1
Model E Sample 1	No Data
Model E Sample 2	13.5
Model E Sample 3	13.0
Model F Sample 1	8.6
Model F Sample 2	7.4
Model G Sample 1	12.5
Model G Sample 2	13.7
Model G Sample 3	12.9
Model H Sample 1	7.3
Model H Sample 2	8.5
Model H Sample 3	7.8
Model I Sample 1	10.5
Model I Sample 2	20.0
Model I Sample 3	19.3
Model J Sample 1	17.9
Model J Sample 2	17.5
Model J Sample 3	17.8
Model K Sample 1	15.4
Model K Sample 2	19.4
Model K Sample 3	20.2

Table B.1: Run 1 time to charge data. This table contains the charging time observed in the first run of charge time testing.

Model	Time to Charge (hr)
Model A Sample 1	20.9
Model A Sample 2	22.7
Model A Sample 3	18.0
Model B Sample 1	6.7
Model B Sample 2	6.2
Model B Sample 3	6.4
Model C Sample 1	14.7
Model C Sample 2	13.1
Model C Sample 3	14.3
Model D Sample 1	10.1
Model D Sample 2	10.1
Model D Sample 3	9.7
Model E Sample 1	0.0
Model E Sample 2	13.8
Model E Sample 3	13.0
Model F Sample 1	9.5
Model F Sample 2	9.5
Model G Sample 1	13.6
Model G Sample 2	14.3
Model G Sample 3	13.5
Model H Sample 1	7.8
Model H Sample 2	9.8
Model H Sample 3	8.4
Model I Sample 1	11.1
Model I Sample 2	19.7
Model I Sample 3	19.3
Model J Sample 1	17.8
Model J Sample 2	17.4
Model J Sample 3	17.7
Model K Sample 1	15.4
Model K Sample 2	19.2
Model K Sample 3	17.5

Table B.2: Run 2 time to charge data. This table contains the charging time observed in the second run of charge time testing.

Model	Average Time to Charge (hr)	Standard Deviation
Model A	19.0	3.6
Model B	6.7	0.5
Model C	13.7	0.7
Model D	9.8	0.4
Model E	13.3	0.4
Model F	8.7	1.0
Model G	13.4	0.6
Model H	8.2	0.9
Model I	19.6	0.3
Model J	17.7	0.2
Model K	17.9	2.1

Table B.3: Charge time data summary. This table contains the statistics for charge time testing over both runs.

Model	Time to Discharge (hr)
Model A Sample 1	18.0
Model A Sample 2	19.0
Model A Sample 3	17.0
Model B Sample 1	6.0
Model B Sample 2	7.9
Model B Sample 3	6.0
Model C Sample 1	10.4
Model C Sample 2	10.5
Model C Sample 3	9.5
Model D Sample 1	11.7
Model D Sample 2	9.5
Model D Sample 3	9.9
Model E Sample 1	1.5
Model E Sample 2	6.6
Model E Sample 3	6.3
Model F Sample 1	12.4
Model F Sample 2	16.4
Model G Sample 1	18.0
Model G sample 2	16.5
Model G Sample 3	17.4
Model H Sample 1	16.6
Model H Sample 2	16.9
Model H Sample 3	16.8
Model I Sample 1	15.7
Model I Sample 2	16.6
Model I Sample 3	19.8
Model J Sample 1	22.8
Model J Sample 2	23.1
Model J Sample 3	22.9
Model K Sample 1	7.8
Model K Sample 2	8.5
Model K Sample 3	9.3

Table B.4: Run 1 time to discharge data. This table contains the discharging time observed in the first run of charge time testing.

Model	Time to Discharge (hr)
Model A Sample 1	9.7
Model A Sample 2	8.9
Model A Sample 3	16.9
Model B Sample 1	6.5
Model B Sample 2	7.1
Model B Sample 3	6.6
Model C Sample 1	12.1
Model C Sample 2	11.7
Model C Sample 3	10.6
Model D Sample 1	12.9
Model D Sample 2	10.9
Model D Sample 3	10.8
Model E Sample 1	1.5
Model E Sample 2	6.4
Model E Sample 3	6.0
Model F Sample 1	11.8
Model F Sample 2	12.1
Model G Sample 1	19.7
Model G sample 2	17.8
Model G Sample 3	19.1
Model H Sample 1	17.3
Model H Sample 2	17.8
Model H Sample 3	17.4
Model I Sample 1	16.9
Model I Sample 2	18.1
Model I Sample 3	19.4
Model J Sample 1	22.6
Model J Sample 2	22.8
Model J Sample 3	22.7
Model K Sample 1	9.0
Model K Sample 2	9.3
Model K Sample 3	11.6

Table B.5: Run 2 time to discharge data. This table contains the discharging time observed in the second run of charge time testing.

Model	Time to Discharge (hr)
Model A Sample 1	5.5
Model A Sample 2	1.5
Model A Sample 3	5.2
Model B Sample 1	8.8
Model B Sample 2	10.1
Model B Sample 3	8.6
Model C Sample 1	14.2
Model C Sample 2	14.2
Model C Sample 3	14.4
Model D Sample 1	14.6
Model D Sample 2	13.6
Model D Sample 3	12.6
Model E Sample 1	5.8
Model E Sample 2	5.8
Model E Sample 3	5.5
Model F Sample 1	3.2
Model F Sample 2	12.6
Model G Sample 1	31.4
Model G sample 2	28.6
Model G Sample 3	30.5
Model H Sample 1	21.2
Model H Sample 2	24.1
Model H Sample 3	23.3
Model I Sample 1	1.5
Model I Sample 2	1.5
Model I Sample 3	20.0
Model J Sample 1	17.5
Model J Sample 2	18.7
Model J Sample 3	17.8
Model K Sample 1	9.4
Model K Sample 2	10.8
Model K Sample 3	18.4

Table B.6: Run 3 time to discharge data. This table contains the discharging time observed in the third run of charge time testing.

Lantern	Average Time to Discharge (hr)	Standard Deviation
Model A	12.5	5.8
Model B	7.5	1.4
Model C	12.0	1.9
Model D	11.8	1.7
Model E	6.1	0.4
Model F	13.1	1.9
Model G	22.1	6.1
Model H	19.1	3.0
Model I	18.1	1.7
Model J	21.2	2.4
Model K	10.4	3.2

Table B.7: Discharge time data summary. This table contains the statistics for discharge time testing over all three runs.

12.5 Lux Area			
Model	Mode	Sq. in.	Sq. cm
Model A	On	1205	7771
Model B	High	1299	8378
Model C	On	446	2878
Model D	High	668	4309
Model E	High	1963	12665
Model F	High	1963	12665
Model G	High	990	6384
Model H	High	524	3381
Model I	High	917	5914
Model J	High	2262	14590
Model K	On	440	2837

Table B.8: Task lighting data. This table contains the measured area which each lantern model could illuminate to a level of 12.5 lux.

Model	H0, Origin	H0, 3 ft Average	H0, 3 ft SD	H1, Origin	H1, 3ft Average	H1, 3 ft SD	H2, Origin	H2, 3ft Average	H2, 3 ft SD
Model A	5	3	1	10	5	2	71	9	6
Model B	1	0	0	2	1	0	14	2	1
Model C	2	1	0	4	2	0	44	2	1
Model D	2	1	0	4	3	0	29	5	1
Model E	5	4	0	12	7	0	75	8	2
Model F	5	4	0	10	7	0	70	12	0
Model G	2	1	0	5	3	0	36	5	0
Model H	1	0	0	2	1	0	21	2	0
Model I	3	2	0	8	5	0	59	7	0
Model J	4	4	0	10	8	1	64	2	1
Model K	1	0	0	2	1	0	22	0	0

Table B.9: Ambient lighting data. This table contains brightness measurements in lux taken during the ambient lighting test. H0 is at ground level, H1 is at a height of 30 inches, and H2 is at a height of 66 inches.

Location	Weighting
H0 Origin Weighting	0.20
H0 3ft Weighting	0.10
H1 Origin Weighting	0.20
H1 3ft Weighting	0.10
H2 Origin Weighting	0.20
H2 3ft Weighting	0.10

Table B.10: Ambient lighting weightings. This table contains the weightings used to calculate the processed raw score for the ambient lighting test.

Model	Ambient Lighting Processed Raw Score
Model A	18.7475
Model B	3.65
Model C	10.2975
Model D	7.88
Model E	20.125
Model F	19.2575
Model G	9.3725
Model H	5.055
Model I	15.435
Model J	16.9
Model K	5.1175

Table B.11: Processed ambient lighting data. This table contains the processed scores for the ambient lighting test. These were calculated by first multiplying the average reading recorded at each location and the weighting assigned to that location. These values were then summed.

Model	High	Medium High	Medium	Medium Low	Low	Spot
Model A			52.9			
Model B	79		27		5	
Model C			44.5			
Model D	22				11	
Model E	110	62		31	3	
Model F	99		42		21	
Model G	59				9	68
Model H	26		11		3	
Model I	48.5		27.7		10.2	44.6
Model J	105	57		38	19	
Model K			28.7			

Table B.12: Setting versatility data. This table contains brightness levels in lux measured at each brightness level for all lantern models.

Model	Number of Settings	Ratio
Model A	1	N/A
Model B	3	16
Model C	1	N/A
Model D	2	2
Model E	4	37
Model F	3	5
Model G	3	8
Model H	3	9
Model I	4	5
Model J	4	6
Model K	1	N/A

Table B.13: Setting versatility summary. This table contains the ratio of the brightness at the highest setting to the brightness level at the lowest setting for each lantern.

MODEL	Shower: Charge Position	Shower: Most Vulnerable Position	Submersion
Model A	FAIL	FAIL	FAIL
Model B	PASS	PASS	PASS
Model C	PASS	PASS	PASS
Model D	PASS	PASS	PASS
Model E	FAIL	FAIL	FAIL
Model F	PASS	PASS	PASS
Model G	PASS	PASS	PASS
Model H	PASS	PASS	PASS
Model I	PASS	PASS	PASS
Model J	PASS	FAIL	FAIL
Model K	PASS	PASS	PASS

Table B.14: Water resistance data. This table shows the results of each tier of water resistance test for all lantern models.

Appendix C

MaxDiff Survey

This appendix contains the survey administered to participants of the pilot problem-based solar evaluation. The survey was administered at the end of the evaluation period, after participants had become family with the solar lanterns. Participants are asked to select the best and worst performing lanterns. They are also asked to complete a MaxDiff survey. This survey was developed by Amit Gandhi and is the same survey used during Amit Gandhi and Victor Lesniewski's field research trip to Uganda. This survey was used in order to compare the preferences of developed world users to those of the developing world users. For more information on the design of this survey, see Amit Gandhi's Master's thesis [10].

Solar Lantern Evaluation Survey

Please fill in the following table:

Study ID number (Your username for the website, EvaluatorXX)	
Which was the best performing lantern?	
Which worst performing lantern?	

MaxDiff Section

In the following tables, please indicate which attributes are MOST important to you and which attributes are LEAST important to you when you consider owning a solar lantern. Note this survey section is general and does not refer to any specific lantern.

Check ONLY ONE issue for each of the most and least columns, in each table. Each table will have one item ticked for the MOST preferred and one item for the LEAST preferred.

EXAMPLE: About Characteristics of Rice

LEAST IMPORTANT	CHARACTERISTIC	MOST IMPORTANT
X	Softness	
	Color	
	Size of Grain	
	Saltiness	X

[MD1].

LEAST	CHARACTERISTIC	MOST
	Charging Time	
	Battery Life	
	Ease of Repair	
	Water Resistance	

[MD2].

LEAST	CHARACTERISTIC	MOST
	# of Settings	
	How It Looks	
	Ease of Repair	
	Cost	

[MD3].

LEAST	CHARACTERISTIC	MOST
	Durability	
	Battery Life	
	Brightness	
	How It Looks	

[MD4].

LEAST	CHARACTERISTIC	MOST
	Charging Time	
	Mobility	
	How It Looks	
	Area Illuminated	

[MD5].

LEAST	CHARACTERISTIC	MOST
	# of Settings	
	Brightness	
	Area Illuminated	
	Water Resistance	

[MD6].

LEAST	CHARACTERISTIC	MOST
	Durability	
	Mobility	
	Water Resistance	
	Cost	

[MD7].

LEAST	CHARACTERISTIC	MOST
	Durability	
	# Of Settings	
	Battery Life	
	Mobility	
	Ease of Repair	
	Area Illuminated	

[MD8].

LEAST	CHARACTERISTIC	MOST
	Durability	
	Charging Time	
	Brightness	
	Ease of Repair	
	Area Illuminated	
	Cost	

[MD9].

LEAST	CHARACTERISTIC	MOST
	Charging Time	
	# Of Settings	
	Battery Life	
	Brightness	
	Mobility	
	Cost	

[MD10].

LEAST	CHARACTERISTIC	MOST
	Durability	
	Charging Time	
	# of Settings	
	Battery Life	
	How It Looks	
	Area Illuminated	
	Water Resistance	
	Cost	

[MD11].

LEAST	CHARACTERISTIC	MOST
	Battery Life	
	Brightness	
	Mobility	
	How It Looks	
	Ease of Repair	
	Area Illuminated	
	Water Resistance	
	Cost	

[MD12].

LEAST	CHARACTERISTIC	MOST
	Durability	
	Charging Time	
	# Of Settings	
	Mobility	
	How It Looks	
	Brightness	
	Ease of Repair	
	Water Resistance	

Appendix D

Table of Solar Lantern Models

This appendix contains a table of models and manufacturers of the solar lanterns evaluated by CITE in the pilot product evaluation.

Evaluation Name	Manufacturer	Model
Model A	Brennenstuhl	SCL 24
Model B	Barefoot Power	Firefly Mobile
Model C	d.light	S2
Model D	d.light	S20
Model E	d.light	S300
Model F	Greenlight Planet	SunKing Pro
Model G	Greenlight Planet	SunKing Solo
Model H	Greenlight Planet	SunKing Eco
Model I	WakaWaka	WakaWaka Light
Model J	WakaWaka	WakaWaka Power
Model K	Unite to Light	UTL-1

Table D.1: Table of Solar Lantern Models. This table shows the number of users that selected each lantern model as the lowest performing lantern.

Bibliography

[1] Kebede, K. Y., Mulder, K. F. (2008) *Needs Assessment and Technology Assessment: Crucial Steps in Technology Transfer.* Revista Internacional Sostenibilidad, Tecnologia y Humanismo, 85-104.

[2] Mills, E., Jacobson, A. (2008) *The Need for Independent Quality and Performance Testing of Emerging Off-Grid White-LED Illumination Systems for Developing Countries.* Light and Engineering 16(2):5-24.

[3] Garvin, David A. (1984) *What Does "Product Quality" Really Mean?.* Sloan Management Review 26(1):25-43

[4] Krishnan, V., Ulrich, Karl T. (2001) *Product Development Decisions: A Review of the Literature.* Management Science 47(1):1-21. `http://dx.doi.org/10.1287/mnsc.47.1.1.10668` Accessed on April 28, 2014

[5] Abras, C., Maloney-Krichmar, D., Preece, J. (2004) *User-Centered Design.* In Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications.

[6] Lavery, D., Cockton, G., Atkinson, MP. (1997) *Comparison of Evaluation Method Using Structured Usability Problem Reports.* Behavior & Information Technology 16(4-5):246-266.

[7] Richard J. Lutz and Patrick J. Reilly. (1974) *An Exploration of the Effects of Perceived Social and Performance Risk on Consumer Information Acquisition.*

Advances in Consumer Research Volume 01, eds. Scott Ward and Peter Wright, Ann Abor, MI : Association for Consumer Research, Pages: 393-405.

[8] Consumer Reports. (2013) *Consolidated Financial Statements: May 31, 2013 and 2012* http://www.consumerreports.org/content/dam/cro/corporate/about/Consumers_Union_Financial_Statements_2013.pdf. Accessed on May 2, 2014

[9] Bounds, G. (2010) *Meet the Sticklers: New Demands Test Consumer Reports; Flying in Cat Fur.* http://online.wsj.com/news/articles/SB10001424052748703866704575224093017379202. Accessed on May 2, 2014

[10] Gandhi, Amit. (2014) *Development of Methodologies for the Testing and Evaluation of Solar Lanterns.* SM Thesis, Massachusetts Institute of Technology.

[11] International Electrotechnical Commission. (2013) *Recommendations for small renewable energy and hybrid systems for rural electrification - Part 9-5: Integrated system - Selection of stand-alone lighting kits for rural electrification.* IEC/TS 62257-9-5 ed2.0.

[12] International Telecommunications Union. (2014) *ICT Facts and Figures: The World in 2013.* http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf Accessed on May 2, 2014

[13] Liter of Light. http://aliteroflight.org/. Accessed on May 2, 2014

[14] Mills, Evan. (2003) *Technical and Economic Performance Analysis of Kerosene Lamps and Alternative Approaches to Illumination in Developing Countries.* Lawrence Berkeley National Laboratory.

[15] State of Virginia (2006) *State of Viginia Plumbing Code.* https://www2.iccsafe.org/states/Virginia/Plumbing/Plumbing_Frameset.html. Accessed on April 1, 2014