Energy for Sustainable Development 48 (2019) 154–163

Contents lists available at ScienceDirect

Energy for Sustainable Development

# Development of a practical evaluation for cookstove usability

Nicholas D. Moses, Mohammad H. Pakravan, Nordica A. MacCarty *

*Oregon State University, School of Mechanical, Industrial, and Manufacturing Engineering, 204 Rogers Hall, Corvallis, OR 97331, USA*

ARTICLE INFO

*Article history:*
Received 20 June 2018
Revised 10 December 2018
Accepted 10 December 2018
Available online 9 January 2019

*Keywords:*
Cookstove testing
User preference
Interdisciplinary
Ethnography

ABSTRACT

While improved cookstoves have been designed and distributed for decades with the goal of addressing the human health and environmental issues caused by traditional biomass cooking, many have not achieved the impact intended. One of the main reasons for this shortcoming is that engineers tend to focus on technical attributes of cookstove designs, such as improved fuel and combustion efficiency, but neglect usability. If a stove design does not meet a cook's needs and preferences, however, the stove will likely be used only as a supplement to a traditional stove, or not used at all. To help close this gap, a testing protocol for cookstove usability was created. The development process and resulting protocol are described in this article. The proposed protocol is based on established usability practices from fields such as software and consumer product design, as well as usability criteria taken from existing cookstove research and interviews with subject experts. Ethnographic testing methods from the field of anthropology have also been incorporated to make the protocol more appropriate for cross-cultural applications, as well as adaptive to a wide range of testing scenarios. The protocol includes objective measurements and observation, as well as subjective survey and semi-structured interview questions. Usability criteria are generally assessed with paired Likert scale survey questions that elicit user perceptions of criteria, as well as the relative importance of each. These results are supplemented by interview results and objective measurements, wherever possible, for comparison and to identify bias or uncertainty in the results. This protocol may be useful to stove designers as a way to better understand users and validate designs, to implementers as a way to assist with the selection of the most appropriate stove for a given project, and to researchers as a tool to assess cookstoves and cookstove programs.



## Introduction

Every year, 4 million people die prematurely from indoor air pollution due to cooking with biomass and other polluting fuels (WHO, 2016). Fuel collection and smoke emissions have also been shown to degrade environmental quality and contribute to regional and global climate change (Bond et al., 2007; Ramanathan & Carmichael, 2008). While up to a billion dollars is spent on improved cookstoves each year (Putti, Tsan, Mehta, & Srilata, 2015) due to their potential to address all of these problems simultaneously, some studies report that the tens of millions of stoves that have been distributed have had little measurable effect (Mobarak, Dwivedi, Bailis, Hildemann, & Miller, 2012; Thacker & Mattson, 2014). This is often because improved stove designs do not meet the user's cooking needs, so users instead supplement or forgo cleaner technologies in favor of inefficient, polluting, traditional methods even when improved cookstoves are available (Johnson, 2012; Mobarak et al., 2012; Thacker, Barger, & Mattson, 2015; Thacker & Mattson, 2014; Victor, 2011).

Usability testing, or the study of how well a product can be used for a given purpose, has become an integral and well-studied aspect of modern product design across many sectors in industrialized nations, including healthcare systems, web design and software, and consumer products (Wood & Mattson, 2016). Usability testing is especially applicable when designers have little inherent understanding of user needs and cannot easily draw on their own experience to make appropriate design decisions, such as in international development settings where users and designers often come from different cultures and contexts (Wood & Mattson, 2016).

Despite the success and ubiquity of usability in design applications in high-income countries, as well as its appropriateness for product design in an international development context, there has historically been little research into usability as a part of design for global development. This may be attributed to factors including the competition between usability and high technical performance in cookstove design, which is more familiar to many designers (MacCarty & Bryden, 2015; Thacker et al., 2015), as well as the limited resources and expertise available in the sector, which does not often have access to the professional, multi-disciplinary design teams that have developed usability practices in other industries.

* Corresponding author.
*E-mail address:* nordica.maccarty@oregonstate.edu (N.A. MacCarty).

https://doi.org/10.1016/j.esd.2018.12.003

This work seeks to apply existing usability knowledge and practices to cookstoves in a way that is accessible to designers and implementers, and which offers a practical method to evaluate and understand usability in the regions and contexts of their work. This information may then be used to better balance user needs with technical performance, emissions, and other objectives in the hopes of increasing the rate of uptake and the impact of improved cookstoves. Elements of this work are incorporated into the ISO/TC 285 international standard for cookstove testing, currently under development, to improve awareness of the need for usable designs. The protocol may also be referenced directly by the completed standard (International Organization for Standardization, 2018).

The proposed protocol organizes cookstove usability into six main criteria, each with multiple sub-criteria, to provide both a high-level and more detailed framework for assessing usability. These criteria are evaluated with subjective survey and interview-based testing methods, as well as objective, quantitative methods when possible, and require minimal testing equipment. A weighted average score for each main usability criterion is calculated from corresponding sub-criteria results and a relative importance assigned to each sub-criterion by respondents. Margins of error are also calculated to assist with interpretation.

Tests are organized into four separate sections by testing method and are meant to be conducted together, ideally in a kitchen during normal cooking activities, for the most representative results. However, test sections may be omitted and/or conducted in less representative settings to accommodate limited testing resources or in cases where less detailed results may be acceptable. Tests are designed to be applicable to common cooking technologies, including traditional and improved biomass, solar, and modern-fuel cookstoves, though not all usability criteria and testing methods apply equally well to all technologies and cooking contexts. Some discretion is required on the part of the test administrator to adapt the protocol to the testing scenario.

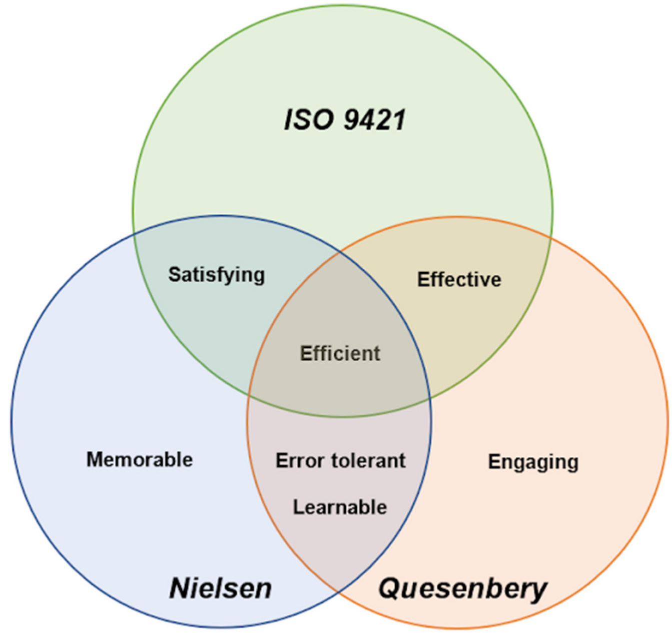


**Fig. 1.** Common definitions of usability.
Source: Moses and MacCarty (2018a).

## Background

### *Definitions of usability*

While the concept of usability has been a part of design fields for several decades, there is no single accepted definition of the term, but instead multiple, sometimes overlapping or competing definitions. Existing definitions include a variety of concepts regarding how users think about, approach, and judge the success of their interaction with a product (Quesenbery, 2004), and are often tailored to specific applications, such as the usability of websites or consumer electronics (Bevan, 1991; Hertzum, 2010; McNamara & Kirakowski, 2005). Notable definitions include ISO 9241: Ergonomics of Human System Interaction, which describes usability as the effectiveness, efficiency, and satisfaction with which a system meets the needs of its users (International Organization for Standardization, 1998), and Nielsen's well-known book, “Usability Engineering,” which expands the definition of usability to include memorability, learnability, and the consideration of user errors (Nielsen, 1994). A graphic depicting the similarities and differences of these definitions of usability is shown in Fig. 1. Broader definitions of usability may also include factors such as ease of access, learning curve, aesthetics, and safety (Bevan, 2008; Fulton Suri & Marsh, 2000), though all definitions share the core idea that the needs of the intended users in a given context must somehow guide the design of a product [15].

### *Usability testing within the field of design*

Within the broader field of design research, usability is one of many approaches for characterizing user behavior and product efficacy. Other methodologies, such as applied ethnography and contextual inquiry, have also received attention for their value in similar applications (Daae & Boks, 2014). While usability testing has traditionally been a relatively top-down, expert-led approach with less direct inclusion of user preferences and opinions than more participatory approaches (Sanders & Stappers, 2008), it does not necessarily require the same level of expertise and investment of time in qualitative research, which is less familiar to many designers and engineers. The proposed usability testing protocol is not meant to replace or exclude the use of other research approaches, but to serve as an accessible starting point for practitioners to understand how a user relates to and interacts with a cookstove. Depending on the needs of the test implementer, supplemental user or design research and testing methods may be appropriate.

### *Design of usability evaluations*

There is no one established method for the development of usability evaluations or their implementation, but instead many related approaches that may be adapted to create the most appropriate evaluation for a specific context (Dumas, 2003; Kwahk & Han, 2002). The development of any evaluation requires a clear identification of the purpose of the test, as well as the test administrator and user of the resulting data, in order to select appropriate testing methods and derive thorough usability criteria (Rettger, 2010). Depending on these factors, an evaluation may be designed for formative testing, which is meant to improve a product or service during its development, or summative testing, which provides validation after the design process is finished, or a combination of the two (Bevan, 2008).

An evaluation may also be designed for use either in a controlled, laboratory setting, or in the field, and may include the collection of quantitative and/or qualitative data. Many evaluations for conventional product design are done in the laboratory where detailed, quantitative data on a user's interaction with a specific product can be recorded (Bevan, 1991). While this may be preferred by many designers and engineers, such a controlled setting is not always available, as is often the case with design for global development. Qualitative data can also provide more context and insight into users' perceptions (Bernard, 2011), which are critical in the decision to purchase or use a product (Han, Hwan Yun, Kim, & Kwahk, 2000).

Examples of applicable evaluation design methods include a systematic method to evaluate 48 common dimensions of consumer product usability, divided into performance-based and perception-based criteria (Han et al., 2000), a quantitative approach based on the interface features of a product and the context of the user, product, and activity

(Kwahk & Han, 2002), and a holistic product assessment model that also includes "safety, wellbeing, satisfaction, health, effectiveness, efficiency, and other aspects" to expand the breadth and potential effectiveness of usability evaluations (Merino, Teixeira, Schoenardie, Merino, & Gontijo, 2012). Many methods and theories have also been incorporated into ISO 9241:11, an international standard for ergonomics and computer user interface design developed by industry experts to promote best-practices (International Organization for Standardization, 1998). While much of the standard is specific to human-computer interface design, it provides foundational guidance for usability evaluation; specifically, the notion that not only a product, but also a user and task, must be a part of evaluation to maximize validity.

Though these methods for evaluation design differ in their content and application, they have in common a systematic approach, which requires the following basic steps:

- The identification of relevant usability criteria for a product or service based on a predetermined definition of usability.
- The identification of the test administrator and user of the resulting data, as well as the scope and purpose of the test.
- The assignment of appropriate testing methods to optimize the effectiveness of the evaluation given the results of the previous two steps.
- The validation of the testing protocol.

## *Cookstove usability research*

Much of the body of research on biomass cookstoves focuses on technical performance, which is easier to measure and more familiar to designers with engineering backgrounds (MacCarty & Bryden, 2017; Thacker et al., 2015). Research into cookstove usability has only recently begun to receive significant attention; a 2013 review found two usability-focused works in existence (Kumar, Kumar, & Tyagi, 2013). While several relevant studies have been completed since that time (Kshirsagar & Kalamkar, 2014; Thacker et al., 2015; Thacker & Mattson, 2014; Urmee & Gyamfi, 2014), no research exists up until this point which compares the usability of a wide range of stove designs or different usage contexts.

The inability of high technical performance to drive cookstove adoption and sustained use has been established, however, as has the importance of balancing technical goals and user needs to maximize adoption, sustained use, and impact (MacCarty & Bryden, 2015; MacCarty & Bryden, 2016a; MacCarty & Bryden, 2017; Mobarak et al., 2012; Thacker et al., 2015; Thacker & Mattson, 2014). Several researchers have also identified the need for additional usability research for biomass cookstoves, as well as the development of standards and tests to allow practitioners to effectively measure and communicate their findings on cookstove usability (MacCarty & Bryden, 2016b; Mobarak et al., 2012; Thacker et al., 2015). Usability has been highlighted as a critical factor in designing or selecting products for users from a different culture (Wood & Mattson, 2016), as well, especially when products are intended to create behavior change (Daae & Boks, 2014), which is the case for many improved cookstove projects.

This small body of cookstove usability literature has also highlighted the existence of common cookstove user needs and lists of key usability criteria. These include cooking speed, firepower, tending frequency, the ability to use different cooking vessels, visibility of the fire, and a variety of context-specific needs, such providing light and heat (García-Frapolli et al., 2010; Ruiz-Mercado, Masera, Zamora, & Smith, 2011; Thacker et al., 2015; Urmee & Gyamfi, 2014). While these lists overlap and offer many insights into user needs, none claim to be complete, nor have they been organized into repeatable, standard testing formats. This has been identified as a necessary step in enabling effective usability evaluation and generating awareness of the need for usability within the cookstove sector (Adkins, Tyler, Wang, Siriri, & Modi, 2010).

To better understand and compare the priorities of cookstove practitioners, a survey of 32 attendees at the InStove Stove Summit, an international development conference focused on cookstoves, was conducted in 2016. Practitioners were asked to rate the importance of a range of criteria related to cookstove design and programming from zero to five in order to measure what they feel is most important to the effectiveness of their work. This research was conducted with oversight from the Oregon State University Institutional Review Board under study number 7603.

These results were compared to end user priorities identified in a review of 13 studies related to cookstove adoption and user preferences (Masera, Díaz, & Berrueta, 2018; Jeuland et al., 2015; Rhodes et al., 2014; Lambe & Atteridge, 2012; Thacker & Mattson, 2014; J. Rosenbaum, Derby, & Dutta, 2015; Adkins et al., 2010; Pilishvili et al., 2016; Nathan G. Johnson & Bryden, 2012; Thacker et al., 2015; Pattanayak et al., 2016; Namagembe et al., 2015; Troncoso, Armendáriz, & Alatorre, 2013). Note that Thacker & Mattson's, 2014 work is itself a review of over 150 published articles. Fig. 2 shows the percentage of practitioners who rated each criterion a 5 out of 5 ("very important") in the Stove Summit survey is compared to the percentage of existing studies that identified criteria as important to users. Though many criteria are partially related to usability, technical, and other factors, they have been roughly organized depending on whether their strongest association is with usability, technical performance, or equally with both. Note that responses from cooks were either not collected or not reported as important by cooks for approximately half of the criteria.

While the language and methods used to solicit input in the practitioner survey and in each of these studies varies significantly, clear gaps can be identified between user and practitioner priorities. Users tend to favor usability-related criteria, such as the ability to use different pots and to cook common dishes effectively, while practitioners rank technical criteria, including thermal efficiency, more highly. There is also a wide dispersion in practitioner priorities; no single attribute was identified as very important by more than 50% of respondents. This demonstrates that there are significant differences in priorities, as well as likely understanding of users, amongst practitioners. Finally, while there is some overlap between the two groups, this indicates that though practitioners may have an awareness of user priorities, they do not necessarily understand which are most important to users. These results reinforce the work of MacCarty and Bryden, Mobarak, Thacker, and others in emphasizing that cookstove usability is generally undervalued by practitioners (N. A. MacCarty & Bryden, 2016b; Mobarak et al., 2012; Thacker et al., 2015).

It should also be noted that while cookstove usability and user satisfaction have been recognized as a critical part of the decision to purchase or use a stove, there are many other important factors influencing technology diffusion and adoption. These include cost, cultural norms and gender roles, education, relationships with the stove vendor or distributing organization, and others, many of which are also commonly neglected in favor of technical stove performance and other objectives (MacCarty & Bryden, 2016b; Mobarak et al., 2012; Pakravan & MacCarty, 2018; Puzzolo, Pope, Stanistreet, Rehfuess, & Bruce, 2016; Rosenbaum et al., 2015). These aspects deserve consideration and additional research, but are beyond the scope of this work.

## *Special considerations for cookstove usability evaluation*

### *Multi-disciplinary and mixed-method evaluation design*

Though the study of usability is generally included within the domain of engineering, it can be implemented most effectively when a variety of disciplines are included to help capture the range of technical, cultural, and other human factors involved in product usability (Johnson, Salvo, & Zoetewey, 2007; Loo et al., 2016; Redish, 2007; Rettger, 2010). Ethnographic methods including interviews, participant observation, and the documentation of physical artifacts (Kiewe, 2008), are sometimes employed by usability practitioners because of their effectiveness in eliciting user opinions, thought processes, and needs, especially when they are not initially known or obvious to the researcher

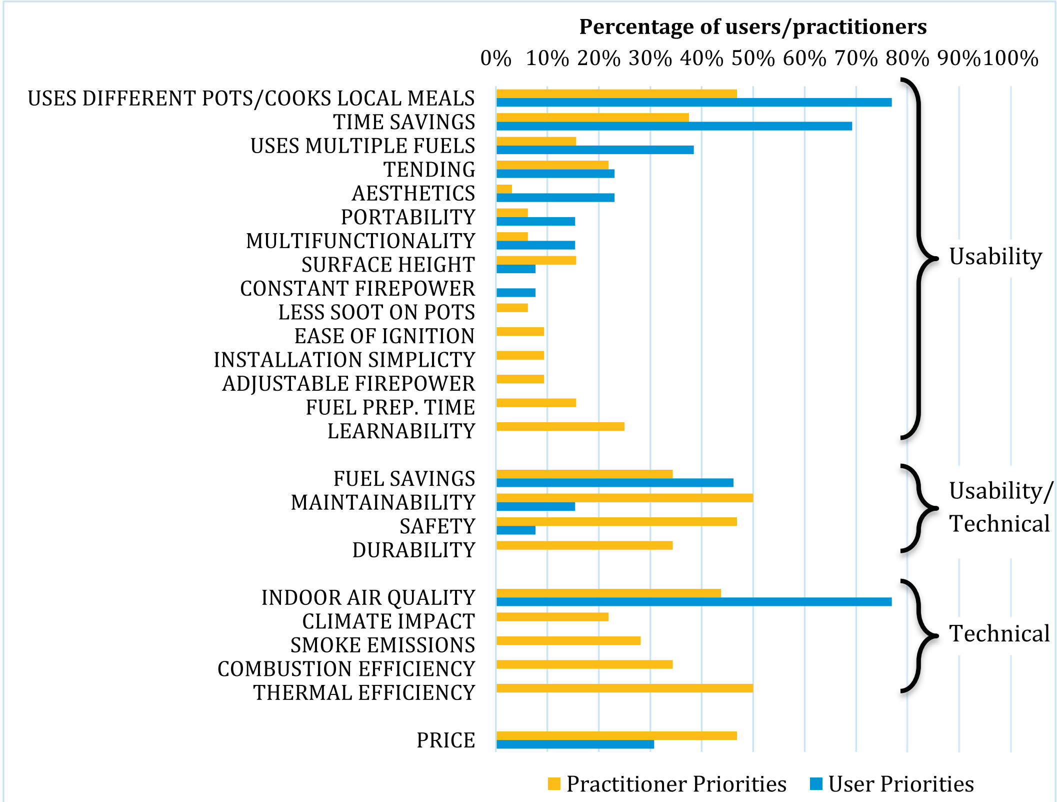


**Fig. 2.** Cookstove user vs. practitioner priorities.

(Loo et al., 2016; S. Rosenbaum, 2008). The use of mixed-methods research has also been suggested as more effective than either quantitative or qualitative methods, alone, in cookstove research (Stanistreet et al., 2015).

Rapid ethnography is another related area of practice that adapts ethnographic methods to the constraints of industry applications such as usability testing (Millen, 2000). Unlike traditional ethnographic studies, which may require months or years and begin without a structured research plan, rapid ethnography often includes a prescribed research approach and may be conducted over a period of days or hours to accommodate the limited resources and time available to an organization, yet still provide valuable information that would not be available without the use of ethnographic methods (Isaacs, 2012).

*Effects of culture*

While usability evaluation design outside of high-income nations has not been as thoroughly studied, multiple researchers have observed that culture affects the process of usability testing, and that when the user and evaluator do not belong to the same culture, the risk of misunderstanding and miscommunication is increased (Clemmensen, Hertzum, Hornbaek, Shi, & Yammiyavar, 2009; Vatrapu & Perez-Quinones, 2006; Wood & Mattson, 2016). In the context of cookstove testing, culture may be defined as the way a user relates to their stoves, cooking practices, and to foreign or local researchers. Considerations for cross-cultural testing and test design have been explored for applications such as computer user interface design, however. This research has suggests that to maximize effectiveness, tests must account for the user's testing experience as well as their relationship with the evaluator, and ideally an evaluator from the same culture should be chosen so that they are more likely to be aware of and able to interpret culture-specific communication queues and styles (Clemmensen et al., 2009; Vatrapu & Perez-Quinones, 2006).

Certain types of testing methods, such as questionnaires and observation, also tend to yield different results in different cultures (Clemmensen et al., 2009; Vatrapu & Perez-Quinones, 2006). Guidance for creating tests for users in India and Africa, specifically, where the majority of the cookstove industry is focused, has also been provided by Oyugi et al., who suggest that Western usability testing methods often do not directly transfer well into these regions and that special care must be taken to create culturally appropriate, relevant evaluations (Oyugi, Dunckley, & Smith, 2008).

*Field testing*

While most available usability protocols are intended for a laboratory, there is a class of evaluations described collectively as "rapid usability tests," which are designed for less controlled scenarios with limited time and resources for testing. These are likely to overlap with international development applications, where practical limitations of field work and constrained resources may limit time and access to participants. One method, named "Extremely Rapid Usability Testing," was designed for trade shows and makes up for limited contact time with each user by engaging them in a real task with a product and recording their actions and comments (Pawson, 2009). Another method, "Rapid Assessment of Product Usability & Universal Design," incorporates universal design principles to help overcome any unfamiliarity or lack of understanding between the user and test administrator, and is therefore a relevant precedent for a cookstove usability protocol designed for the diverse global population of biomass cookstove users, as well as for an evaluation that is likely to be implemented by foreigners (Lenker, Nasarwanji, Paquet, & Feathers, 2011).

Remote usability testing, or the remote collection of real-world usage data from computer applications and websites, is also relevant to the field testing of consumer products. In these tests real-world usage may be evaluated, which can help to avoid unwanted variables introduced by potentially less representative usage scenarios created in a laboratory (Hartson & Castillo, 1998). While there are limitations to the types of measurements that can be conducted outside of a laboratory, remote evaluations have been shown to produce results that are of comparable value and have the potential to elicit complimentary information not captured in the laboratory (Tullis, Fleischman, McNulty, Cianchette, & Bergel, 2002).

## Cookstove usability protocol development

### *Cookstove usability criteria selection*

To judge cookstove usability, an appropriate set of criteria was identified. Criteria were compiled from a systematic review of existing international development literature (García-Frapolli et al., 2010; Kshirsagar & Kalamkar, 2014; Ruiz-Mercado et al., 2011; Thacker & Mattson, 2014; Urmee & Gyamfi, 2014), as well as discussions with industry experts. Most criteria were mentioned by multiple literature sources and experts. Cookstove durability and safety, which by some definitions could be included in or overlap with usability, are incorporated into separate testing protocols (Center for Energy Development; Energy Institute at Colorado State University, 2014; Nathan G Johnson & Bryden, 2015) and in the upcoming ISO/TC 285 cookstove testing standard (International Organization for Standardization, 2018). To avoid duplicating these efforts, only subjective user impressions of safety and durability are considered in this protocol, which are likely to be independent of and complementary to the existing technical assessments and the ISO standard.

The resulting compilation of criteria, shown in Table 1, has been grouped into six categories corresponding to major areas of cookstove usability, each with multiple sub-criteria. Note that the location-specific needs category includes space for test administrators to write in additional non-cooking needs found in their region or cooking context, as all auxiliary uses of cookstoves cannot be captured in this list.

### *Characterization of the users of the test and data*

The backgrounds and values of the cookstove practitioners and researchers likely to apply the protocol and incorporate the results into their work were considered in order to design an appropriate evaluation. From the author's experience, they tend to be international development or engineering professionals, or may be missionaries or other humanitarians from a variety of professional backgrounds. With few exceptions, however, cookstove practitioners are not usability or user experience experts, nor are they social scientists, and are not likely to have deep experience with social or user research. Similarly, practitioners often work under constrained timelines with limited human and financial resources at their disposal. Finally, as demonstrated in the comparison of practitioner and user priorities in the Cookstove usability research section, practitioners may not enter into testing work with a strong understanding of the needs of the users they will be evaluating.

### *Testing methodology*

Given backgrounds of the cookstove practitioners described above, as well the diversity in the design, research, and stove selection applications that practitioners may have for usability testing, the testing methods included in this protocol have been designed to be:

- Broadly applicable to encompass the majority of cooking technologies and contexts
- Adaptive, with the use of multiple and mixed methods to overcome potential shortcomings of a method in a given testing scenario.
- Used for either a formative or summative evaluation, appropriate for new stove design, design iteration, or the selection or evaluation of existing designs.
- Self-contained and self-explanatory to accommodate varying levels of prior testing experience, especially with qualitative methods and data analysis. References to outside testing and study design resources are included.
- Flexible; evaluators may use or modify the test to fit a range of testing needs.

Existing usability evaluation methods were used as a foundation for the protocol, with the addition of anthropological testing methods to help bridge cross-cultural communication gaps and create a deeper understanding of user interactions with their stoves. In addition to characterizing the effectiveness of user interaction with a cooking system, the methods and testing processes used in this protocol are meant to elicit new user needs and preferences that may not already be included in the protocol or existing cookstove literature. While this is not typical of conventional usability tests, it has been identified as a necessary step in the context of improved cookstoves to allow practitioners to deliver effective products (Loo et al., 2016), and is consistent with the expansion of the scope and methods of usability testing over the years since its inception several decades ago (Redish, 2007; Rosenbaum, 2008).

Quantitative measurements and objective observations of stoves and cooking process were chosen to evaluate as many criteria as possible to help reduce the potential for bias, though some criteria cannot be easily measured objectively and are based on subjective user responses, only. Appropriate testing methods for these criteria, in the form of Likert scale survey questions to elicit user perceptions, were modeled on rapid usability tests (Lenker et al., 2011; Pawson, 2009). These questions have a set number of qualitative answers (usually 5, in this protocol), which correspond to a numeric value. Sections for supplemental field notes are also provided for each question to note any additional information provided by a respondent, or any perceived bias or uncertainty in the validity of the response. This approach serves as a simple way to quantify qualitative data, and is a compromise between richer, purely qualitative testing methods, and the amount of time and expertise needed to effectively

**Table 1**
Cookstove usability criteria
Source: Moses and MacCarty (2018a)

| Fuel Convenience | Cooking Performance | Operability | Maintenance | Comfort | Location-Specific Needs |
|---|---|---|---|---|---|
| Fuel availability | Firepower Range | Tending/refueling frequency | Routine maintenance | Cooking area soot deposits | Space heating |
| Fuel preparation | Firepower control | Tending/refueling effort | Long-term maintenance | Perceived smoke exposure | Insect repellant |
| | Cooking speed | Fuel feed entry size | | Perceived durability | Lighting |
| | Versatility | Visibility of fire | | Pot soot deposits | Portability |
| | | Ease of lighting | | Cooking height | Food drying/smoking |
| | | Fire start-up delay | | Stove aesthetics | Water heating |
| | | User error | | Perceived safety | *(Additional needs may be added by the test administrator)* |
| | | User instruction | | Perceived value | |
| | | | | Taste | |

interpret qualitative results. Quantitative methods, on the other hand, would have been unlikely to elicit user perceptions effectively in this context, alone (Loo et al., 2016); without a detailed prior knowledge of a particular group of cooks and cooking technologies, designing an effective, purely quantitative evaluation is not practical. Survey questions have also been applied to the criteria assessed by objective measurements and observations, with the purpose of enabling the test administrator to more easily identify error or bias from the results from one method or the other.

A small set of semi-structured interview questions has also been included to give the test administrator a chance to clarify and identify error in other portions of the test, and the cook an opportunity to share additional information not explicitly asked for by the test administrator. This method was selected to elicit as much detail as possible from the cook, while still generally directing responses towards cooking and usability. Semi-structured interviews have also been suggested as an effective method when follow-up with the participant at a later date for additional testing or clarification is not preferable or possible (Bernard, 2011), as is likely to be the case in international testing situations.

### *Validation and improvement*

The protocol has been tested and improved through preliminary field trials in 10 households and 2 institutional kitchens near Lira, Uganda, 20 households near Gulu, Uganda, and 4 households in Guatemala, as well as a laboratory trial on 11 stove models in the United State. In general, the protocol was shown to elicit valuable and relevant information, though more work is needed to improve its effectiveness and to correlate usability results to effects on the adoption and sustained use of cookstoves, health outcomes, etc. Additional details about the methods used in this work and conclusions regarding validity of the protocol as a tool to measure cookstove usability are discussed in a separate article submitted for publication in *Energy Research and Social Science* (Moses & MacCarty, 2018c).

## Cookstove usability protocol overview

The cookstove usability testing protocol is described here to give the reader an understanding of the methods, criteria, and testing and data analysis processes used. The full protocol, which includes an instructional guide covering study planning, test administration, the scoring for each criterion, as well as a template for recording test data and an Excel spreadsheet to analyze and summarize results, are not included here in the interest of space. The entire protocol document and the Excel spreadsheet are available online from Oregon State University (Moses & MacCarty, 2018b).

### *Instructional guide*

#### *Study design and test administration considerations*

General guidance for study design and test administration is given within the protocol. While no field testing scenario is perfect, factors that should be considered in planning for testing and evaluating results include, but are not limited to, the information in the following sections.

*Sample selection and testing saturation.* Factors to consider when choosing test participants include familiarity with the cookstove, representation of intended users, and relevance of the sample to testing goals. Cooks should have used a stove for at least several days, and ideally several weeks or longer, prior to usability testing to ensure a basic level of familiarity with the stove and representative cooking usage (International Organization for Standardization, 2018). Similarly, cooks should be representative of the range of expected users of the stove (in terms of age, income, proximity to urban areas, etc.).

At least three tests with three different cooks are recommended per stove model to provide a reasonable, qualitative understanding of usability, which may provide significant value to designers, especially in early stages of design and validation (Moses & MacCarty, 2018c). A minimum of 15 tests are suggested to produce significant quantitative results for most, but not all of the usability sub criteria, which may be approximately defined by a margin of error of less than 0.5 points for scores rated from 0 to 4 with a 95% confidence level (Moses & MacCarty, 2018c).

While this represents a low level of statistical significance compared to technical cookstove tests, it also reflects real variations in opinions and needs between individual cooks within the same community for some criteria. All criteria cannot necessarily be characterized effectively with purely quantitative methods in a given context, and may not have results with truly normal distributions that allow for validity in common statistical tests. Uncertain quantitative results in these cases act as indicators that criteria must be assessed qualitatively, as well, to be understood thoroughly, as no amount of quantitative testing is likely to produce adequate results to inform stove design or selection. These criteria may be identified by preliminary qualitative assessments of cooking practices, or frequent assessments of test results during a study, and may be assessed in more detail during the interview portions of subsequent tests or by separate follow-up work. Improved survey questions or measurements created to address these criteria could increase the validity of quantitative results and reduce the need for qualitative data and interpretation to clarify uncertainty.

With very large sample sizes, the number of criteria identified as candidates for further qualitative assessment may not increase significantly past a certain point, leading to diminishing returns from additional tests. Even if high statistical significance is achieved for these criteria, the results may in effect mask a more complicate reality. Differences within sample populations may produce bimodal or other distributions in real user needs and perceptions, which may require multiple stove designs and program solutions, but will not be reflected in a single average score.

The suggestion of a minimum of 15 test repetitions and an acceptable margin of error of 0.5 points (with a 95% confidence level) is proposed as a practical compromise between the time and expense required for larger studies, and the likely value of the testing results for stove design, selection, or research. Higher or lower standards for statistical significance and different usability testing needs may require larger or smaller sample sizes than prescribed here, however. More detailed guidance on sampling and statistical significance may be found in the UNFCCC CDM Guidelines for Sampling and Surveys, which is referenced in the protocol (UNFCC CDM Executive Board, 2012). It should also be noted that testing may be most valuable when done on baseline and current or potential improved stove designs to provide a comparison. Understanding the usability of existing stoves will give context to a cook's perception of new stoves, which are likely to be judged against familiar cooking technologies.

*Test administrator skill-set and identity.* The portions of the test related to stove measurements and objective observations should be done by a person familiar with common cookstove designs. It is critical that the survey and semi-structured interview portions of the test be administered by a person who is proficient in a language spoken by the cook, familiar enough with the cook's culture to recognize culture-specific communication cues, and whose presence in the kitchen is as unobtrusive as possible to allow for representative cooking activity. Past experience with surveys or with related work may also be helpful, especially to administer of the survey in an informal, conversational format, which can keep the respondent more engaged and allow for more natural and valid answers to Likert scale questions.

The objective and subjective portions of the tests may be done by the same person, or by two different people simultaneously. A second test administrator frees the surveyor/interviewer from the distractions of taking measurements, and also allows for the added benefit of a second perspective on cooking behaviors and the cook's responses to the questions (Pawson, 2009). Two administrators with different backgrounds may enhance this effect.

There are many other ways that the test administrator's identity may impact test results and should be considered. For example, because it is unusual for a man to spend time in the kitchen in some regions, a male test administrator is not likely be as welcome or receive the same quality of responses. While no test administrator is likely to receive perfectly accurate responses from all participants, in many cases, a local woman who is an expert in local cooking practices, and who also has with relevant survey or interview experience, may make a high-quality test administrator, especially for the interview and survey portions of the protocol. Similarly, the more familiar a cook is with the test administrator, the more likely they are to behave normally and give direct answers. However, a cook may also prepare a more complex meal, use different stoves or fuels, etc. if they consider the test administrator(s) to be guests (even if they are asked to prepare a meal normally). It should also be noted that cooking practices can also depend on weather, harvest or seasonal employment schedules, etc. In addition, if an evaluator has a relationship with an NGO, government agency, or other source of authority, a cook may be more likely to bias their behavior or responses to meet the expectations of the evaluator, or to gain something in return. More broadly, the histories of colonialism and foreign-led development programs in many of the countries where cookstove work is focused may shape the relationships between cooks and foreign test administrators, which could also impact the validity of testing results.

*Construct validity.* New usability criteria that are not included in the protocol may be present in many regions, and various aspects of the testing methods and usability criteria included in the protocol may be more or less valid in a given context due to cultural factors and variety in cooking technologies and practices. Preliminary survey- or interview-based qualitative studies and/or consultation with a local expert, who is familiar with the cooking culture and context, may be able to identify any shortcomings or omissions in the protocol methods, criteria, or study design, so that adjustments may be made before testing begins. A local test administrator may be able to advise whether a cook has deviated from a typical meal, how and when seasonal patterns or other factors may impact test results, etc. Similarly, frequent, reflexive assessments of test results during a study, especially the triangulation of subjective and objective results, as well as critical observations by test administrators, may allow for the identification and correction of issues with the construct validity of the protocol.

*Testing location*

The protocol was designed to be conducted in a kitchen during a normal cooking event to increase the relevance of the results and allow for the test administrator to observe potential discrepancies between participant cooking practices and verbal responses. Bringing a cook into a laboratory setting creates the risk of placing them in an unfamiliar setting and significantly affecting outcomes, especially if they are not already familiar with the stove design being tested (Lenker et al., 2011). Testing in a familiar location also makes it easier to conduct the survey and interview portions conversationally, as opposed to formally, which may help to keep the cook engaged and encourage natural responses throughout the duration of the evaluation, increasing the quality of the results (Pawson, 2009).

*Test format and methods*

The protocol relies on four separate test sections organized by testing methods, as shown in Table 2. The use of multiple and mixed testing methods provides overlapping assessments of usability criteria wherever possible, allowing for the identification of conflicting responses and likely miscommunications or misunderstandings. Each test section also contains space for field notes, as well as basic guidance for test administrators. Supplemental field notes, photography, and audio and video recordings are also encouraged (with consent, and consideration for the potential impact on the participant and their responses) to elicit additional details and maximize value from the effort invested in testing (Bernard, 2011). If more than one stove is used in the preparation of the same meal, the tests may be conducted in parallel for each stove, although it should be noted that in this case each stove may be used to its strengths by the cook, and the usability results are likely to be different than if the stove was assessed in isolation.

*Alternative testing procedures*

The protocol is designed to measure cookstove usability as thoroughly as possible in the field for common cookstove designs. However, variations may be appropriate for specific testing needs.

*Rapid field testing.* When time is limited or less thorough usability data are required, such as when choosing between a limited number of cookstove models as opposed to improving a stove design, the User Cooking Event Observation section may be omitted. In this case, the remaining physical measurement, survey, and interview portions of the test may be carried out in 20 min or less per household without a cooking event taking place. Though much contextual information may be lost without observing a cooking event, the remaining tests may still provide valuable information.

*Laboratory testing.* Laboratory testing can be used to collect preliminary or basic data before field testing or if field testing is not feasible, and may also be able to provide additional data that compliments field testing results. This testing can be done in one of three ways:

1. A Cookstove Characteristics Evaluation can be done without lighting a stove. The physical measurements and observation of stove features included in this section provide basic information about likely usability performance. This information is most valuable with a prior understanding of cooking needs and habits in the region of interest.
2. In addition to a Cookstove Characteristics Evaluation, a User Cooking Event Observation can be simulated by evaluators or other surrogate cooks. This provides additional information about usability performance and can offer valuable first-hand experience to the stove test administrator, although the results are likely to be less valid than testing by a representative stove user.
3. A User Cooking Event Evaluation can also be approximated in a lab with a representative cook operating the stove. This may offer a higher level of validity than is possible with a foreign or inexperienced stove operator, though asking a user to cook in an unfamiliar laboratory setting instead of their personal kitchen may introduce many variables and limit the validity of the test.

*Additional case-specific testing scenarios.* This test may be done concurrently with the Controlled Cooking Test (CCT) or Uncontrolled Cooking Test (UCT) (Household Energy and Health Programme, 2004; Robinson & Ibraimo, 2011). Some fuel and time measurements are shared between this usability protocol and the CCT and UCT, and conducting both protocols at once may save time and effort.

Questions regarding personal and cultural perceptions towards a stove will have different significance to many institutional cooks, as well, who often have less personal investment and input in the selection and use of their cookstove. These aspects may be skipped or assigned the highest or lowest rating for a given sub-criteria, as appropriate.

*Questions and scoring*

A field data collection template for recording test results is also included in the protocol. It contains directions for all observational- and

**Table 2**
Protocol test sections.
(Source: Moses and MacCarty (2018a)).

| | Test section name | Test methods | Purpose |
|---|---|---|---|
| 1. | Cookstove characteristics evaluation | Quantitative measurements and observations | To measure stove dimensions and features. |
| 2. | User cooking event evaluation | Quantitative measurements and observations | To measure fuel use, cooking event duration, and cooking practices and patterns during cooking activity. |
| 3. | User survey | Quantified survey with primarily Likert scale questions | To elicit perceptions about, and the relative importance of, each criterion from the cook's perspective. |
| 4. | Semi-structured interview | Qualitative interview | To clarify results from other test sections, as well as give participants the opportunity to share additional information they feel is important. |

measurement-based tests, as well as all survey and interview question needed by the test administrator. The template provides additional in-line guidance for the administration of each test or question, as well as sections for field notes to supplement the results. The full field data collection template is included in the supplemental materials for this article.

Most usability criteria are first scored, then assigned a relative weight reflecting how important a criterion is to the cook, through paired sets of Likert scale survey questions. This has been done because the relative importance of many usability criteria may vary significantly in different contexts. For example, in an area where wood is freely and easily accessible, fuel consumption may not be as much of a concern as in arid regions. Results from different testing method for the same criterion are reported in groups so they may be more easily compared, and potential bias or error more easily identified. The full scoring methodology for each usability sub-criteria is given in the supplemental materials.

While the protocol is meant to assess many types of technology, not all questions and measurements apply to household cooking technologies other than wood-burning and charcoal stoves. These include solar stoves, gas and liquid fuel stoves, and electric stoves. Aspects of the test that do not apply may be skipped or assigned the highest or lowest rating for a given sub-criteria, as appropriate, to make for a fair comparison with other stove types.

*Data analysis and reporting*

A Microsoft Excel spreadsheet is included with the protocol to simplify the entry, storage, and analysis of test data. This spreadsheet calculates numerical scores for each usability sub-criteria from 0 to 4, generally based on Likert scale questions. This has been done to align with the range of the Likert scale survey questions used in the test, and is the same range used by the ISO-IWA 11:2012 tiers of performance for improved cookstoves (International Organization for Standardization, 2012). Since many cookstove practitioners are familiar with the 0–4 scale used by the ISO-IWA guidance for technical stove performance criteria, this is meant to facilitate easier communication and understanding. A margin of error for each score is also calculated, similar to the Kitchen Performance Test (KPT) commonly used to assess fuel consumption and other factors (Rob Bailis & Edwards, 2007). Margin of error is calculated with a Student's t distribution given the sample size, standard deviation of the sample, and a confidence level specified by the spreadsheet user. The spreadsheet also includes a qualitative data analysis tool to assist with the coding and interpretation of interview results. Qualitative data do not factor into numerical scores, but may provide additional context to results and help to identify potential biases or errors. If small test sample sizes are used, these data may be more valid and meaningful than the numerical scores.

An overall score for each of the six main usability criteria is calculated from a weighted average of sub-criteria scores to provide a concise results summary, as well as to account for the relative importance of each criterion to the users. The highest margin of error from the corresponding sub-criteria is also reported alongside each main criterion score to indicate potential uncertainty, and whether the sub-criteria should be investigated in more detail to better understand the validity of the overall score.

As with each sub-criterion, scores for the main criteria range from 0 to 4. These numeric scores are not necessarily meant to be a direct prediction of adoption or usage behavior, especially without large test sample sizes, but are meant to highlight areas of potential concern that should be investigated further, or to be used as a means of comparison between two or more stove designs used in the same context. Similarly, no single overall score is calculated for a stove, as small differences in scores between stoves would likely reflect imperfections in the protocol and study designs as much or more than real differences in usability with implications for adoption and usage. A template for reporting final results is included in the supplemental materials.

*Known limitations*

This protocol has several limitations and differences compared to the existing cookstove field tests, such as the Uncontrolled Cooking Test or Controlled Cooking Test (Bailis, 2004; Robinson & Ibraimo, 2011). For example, validity may be impacted by the cook's level of familiarity and comfort with survey and interview-style questioning, as well as various other factors that impact their responses and relationship with the test administrator. A relatively large number of tests may also be needed to achieve a statistically valid comparison between stoves, or the characterization of a single stove model, though it should be noted that smaller sample sizes may still elicit many important, if qualitative or anecdotal, usability results for a given stove and context. In addition, variation should be expected between regions and cultures; results from one study may or may not be applicable to other contexts.

Significant differences in opinion and cooking habits may also be expected from person to person within one sample population, though this may indicate a real diversity in cooking needs and preferences in addition to or instead of limitations to the repeatability or validity of the protocol. The universality of protocol also comes at the expense of sensitivity to regional cooking needs and cultural factors, which may be accounted for by the test administrator through reflexive testing practices or through preliminary, qualitative work to improve the construct validity of the protocol for a given context.

## Discussion

While more validation is needed, this protocol is intended to offer practitioners an accessible method to increase their understanding of user needs and how effectively a cookstove meets them in a given context. Based in both engineering and anthropological methods, the protocol includes a mix of quantitative and objective tests, as well as subjective survey and interview questions, to provide overlapping evaluations of fuel processing and collection habits, cooking performance, stove operability, maintenance, comfort and aesthetic considerations, and location-specific needs. The inclusion of multiple, mixed methods is meant to provide more valuable context to the test administrator, help overcome communication barriers inherent in often cross-cultural cookstove testing, and to identify and clarify potential bias, error, and uncertainty in test results.

The resulting understanding of usability should allow designers and implementers to better balance technical and user needs in cooking systems. This will potentially lead to higher rates of adoption, sustained use, and impact. More generally, the experience of spending an extended amount of time with users and the use of ethnographic methods may also allow practitioners to develop a deeper understanding of the cooks they serve, the underlying health and other issues at hand, and the international development environment in which they work. This research and the interdisciplinary approach underlying the design of the protocol will also hopefully be of value for additional user research and standardization in other areas of international development.

Future research could be designed to validate and improve the protocol through testing in more contexts, on different technologies, and for different purposes. Additional research could also be designed to correlate the results of this protocol to adoption, purchasing, sustained usage, and impact, as well as to create new versions of the protocol that are optimized for specific cultures or technologies, as are most conventional usability tests. The protocol has been distributed to partners of the Global Alliance for Clean Cookstoves to solicit feedback for these purposes. This work to further increase the standardization and prevalence of cookstove usability testing may also allow for an increase in accountability for design or selection of stoves for a given context, as well as clearer relationships between usability and health and environmental outcomes.

The full protocol is available online for use and evaluation from the Oregon State University website: https://humanitarian.engineering.oregonstate.edu/project-page/usability-testing-protocol-cookstoves.

## Acknowledgments

The authors would like to thank International Lifeline Fund – especially the staff in Lira, Uganda – and Aid Africa for facilitating the first field trials of this protocol, as well as the US Technical Advisory Group to the International Standards Organization for helping to incorporate this work into the ISO/TC 285 standard for cookstove testing. We would also like to thank Paul Means, Christa Roth, Vahid Jahangiri, Dale Andreatta, Shaozeng Zhang, and Alan Bigelow for their contributions to this work, as well as the Aprovecho Research Center for allowing us access to cookstoves in their laboratory, and Sam Bentson for assisting in the selection of stoves for laboratory testing. We are grateful to the many individuals, organizations, and Regional Testing and Knowledge Centers currently evaluating and providing feedback on the protocol to help us make improvements. Finally, we would like to thank the National Science Foundation (CMMI grant # 162485) and the Oregon State University School of Mechanical and Industrial Engineering for supporting this research.

## References

Adkins, E., Tyler, E., Wang, J., Siriri, D., & Modi, V. (2010). Field testing and survey evaluation of household biomass cookstoves in rural sub-Saharan Africa. *Energy for Sustainable Development*, *14*(3), 172–185. https://doi.org/10.1016/j.esd.2010.07.003.

Bailis, R. (2004). *Controlled Cooking Test (CCT)*. Shell Foundation Retrieved from http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/80-1.pdf.

Bailis, R., & Edwards, R. (2007). *Kitchen Performance Test (KPT)*. Shell Foundation Retrieved from https://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/83-1.pdf.

Bernard, H. (2011). *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed.). Lanham: AltaMira Press Retrieved from https://books.google.com/books?hl=en&lr=&id=WhKYqATAySwC&oi=fnd&pg=PR1&dq=bernard+2011+inductive+approach&ots=6zAlzyO0vH&sig=_5A-4JA9MjJRlIAxy0pb0aupwt0.

Bevan, N. (1991). What is usability. *Proceedings of the 4th International Conference on HCI, Stuttgart* (pp. 651–655)https://doi.org/10.1.1.83.8558.

Bevan, N. (2008). Combining quantitative and qualitative data for measuring user experience of an educational game. *International Workshop on Meaningful Measure: Valid Useful User Experience Measurement (VUUM)* (pp. 27). https://doi.org/10.1145/2070481.2070498.

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., ... Trautmann, N. M. (2007). Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000. *Global Biogeochemical Cycles*, *21*(2). https://doi.org/10.1029/2006GB002840.

CDM Executive Board (2012). Guidelines for sampling and surveys for CDM project activities and programme of activities. Retrieved from https://cdm.unfccc.int/Reference/Guidclarif/meth/meth_guid48.pdf.

Center for Energy Development; Energy Institute at Colorado State University (2014). Cookstove Durability Protocol (CDP). Retrieved from https://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/89-1.pdf.

Clemmensen, T., Hertzum, M., Hornbaek, K., Shi, Q., & Yammiyavar, P. (2009). Cultural cognition in usability evaluation. *Interacting with Computers*, *21*(3), 212–220 Retrieved from http://mortenhertzum.dk/publ/IWC2009_preprint.pdf.

Daae, J., & Boks, C. (2014). A classification of user research methods for design for sustainable behaviour. *Journal of Cleaner Production*, *106*, 680–689. https://doi.org/10.1016/j.jclepro.2014.04.056.

Dumas, J. S. (2003). User-based evaluations. In J. A. Jacko, & A. Sears (Eds.), *The human-computer interaction handbook* (pp. 1093–1117). Lawrence Erlbaum Associates Retrieved from http://dl.acm.org/citation.cfm?id=772141.

Fulton Suri, J., & Marsh, M. (2000). Scenario building as an ergonomics method in consumer product design. *Applied Ergonomics*, *31*(2), 151–157. https://doi.org/10.1016/S0003-6870(99)00035-6.

García-Frapolli, E., Schilmann, A., Berrueta, V. M., Riojas-Rodríguez, H., Edwards, R. D., Johnson, M., & Masera, O. (2010). Beyond fuelwood savings: Valuing the economic benefits of introducing improved biomass cookstoves in the Purépecha region of Mexico. *Ecological Economics*, *69*, 2598–2605. https://doi.org/10.1016/j.ecolecon.2010.08.004.

Han, S. H., Hwan Yun, M., Kim, K. J., & Kwahk, J. (2000). Evaluation of product usability: Development and validation of usability dimensions and design elements based on empirical models. *International Journal of Industrial Ergonomics*, *26*(4), 477–488. https://doi.org/10.1016/S0169-8141(00)00019-6.

Hartson, H. R., & Castillo, J. C. (1998). Remote evaluation for post-deployment usability improvement. *Proceedings of the working conference on Advanced visual interfaces - AVI '98* (pp. 22). https://doi.org/10.1145/948496.948499.

Hertzum, M. (2010). Images of usability. *International Journal of Human Computer Interaction*, *26*(6), 567–600. https://doi.org/10.1080/10447311003781300.

Household Energy and Health Programme, SF (2004). Controlled Cooking Test (CCT). Retrieved from https://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/80-1.pdf.

International Organization for Standardization (1998). *ISO 9241:11: Ergonomics of human system interaction.*

International Organization for Standardization (2012). *ISO/IWA 11: 2012 guidelines for evaluating cookstove performance.*

International Organization for Standardization (2018). *ISO/TC 285: Clean cookstoves and clean cooking solutions.* Geneva: Switzerland Retrieved from https://www.iso.org/committee/4857971.html.

Isaacs, E. (2012). The value of rapid ethnography. *Advancing ethnography in corporate environments: Challenges and emerging opportunities. Vol. 1.* (pp. 92–107) Retrieved from http://www.izix.com/pubs/Isaacs-RapidEthnography-2013.pdf.

Jeuland, M. A., Bhojvaid, V., Kar, A., Lewis, J. J., Patange, O., Pattanayak, S. K., ... Ramanathan, V. (2015). Preferences for improved cook stoves: Evidence from rural villages in north India. *Energy Economics*, *52*, 287–298. https://doi.org/10.1016/j.eneco.2015.11.010.

Johnson, N. G. (2012). *Village energy system dynamics of an isolated rural West African village.* Digital Repository, Graduate Theses and Dissertation *1*. (pp. 1–239). Iowa State University, 1–239 Retrieved from http://lib.dr.iastate.edu/etd.

Johnson, N. G., & Bryden, K. M. (2012). Energy supply and use in a rural West African village. *Energy*, *43*(1), 283–292. https://doi.org/10.1016/J.ENERGY.2012.04.028.

Johnson, N. G., & Bryden, K. M. (2015). Field-based safety guidelines for solid fuel household cookstoves in developing countries. *Energy for Sustainable Development*, *25*, 56–66. https://doi.org/10.1016/j.esd.2015.01.002.

Johnson, R. R., Salvo, M. J., & Zoetewey, M. W. (2007). User-centered technology in participatory culture: Two decades "beyond a narrow conception of usability testing". *IEEE Transactions on Professional Communication*, *50*(4), 320–332. https://doi.org/10.1109/TPC.2007.908730.

Kiewe, H. (2008). How may I help you? An ethnographic view of contact-center HCI. *Journal of Usability Studies*, *3*(2), 74–89.

Kshirsagar, M. P., & Kalamkar, V. R. (2014). A comprehensive review on biomass cookstoves and a systematic approach for modern cookstove design. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2013.10.039.

Kumar, M., Kumar, S., & Tyagi, S. K. (2013). Design, development and technological advancement in the biomass cookstoves: A review. *Renewable and Sustainable Energy Reviews*, *26*, 265–285. https://doi.org/10.1016/j.rser.2013.05.010.

Kwahk, J., & Han, S. H. (2002). A methodology for evaluating the usability of audiovisual consumer electronic products. *Applied Ergonomics*, *33*(5), 419–431. https://doi.org/10.1016/S0003-6870(02)00034-0.

Lambe, F., & Atteridge, A. (2012). *Putting the cook before the stove: A user-centred approach to understanding household energy decision-making.* , 30 Retrieved from http://www.sei-international.org.

Lenker, J. A., Nasarwanji, M., Paquet, V., & Feathers, D. (2011). A tool for rapid assessment of product usability and universal design: Development and preliminary psychometric testing. *Work*, *39*(2), 141–150. https://doi.org/10.3233/WOR-2011-1160.

Loo, J., Hyseni, L., Ouda, R., Koske, S., Nyagol, R., Sadumah, I., ... Stanistreet, D. (2016). User perspectives of characteristics of improved cookstoves from a field evaluation in western Kenya. *International Journal of Environmental Research and Public Health*, *13*(2), 167. https://doi.org/10.3390/ijerph13020167.

MacCarty, N., & Bryden, K. M. (2015). Modeling technology strategies for thermal energy services in rural developing communities. *Proceedings of the ASME IDETC/CIE 2015* (pp. 1–13). https://doi.org/10.1115/DETC201546806.

MacCarty, N. A., & Bryden, K. M. (2016a). A generalized heat-transfer model for shielded-fire household cookstoves. *Energy for Sustainable Development*, *33*, 96–107. https://doi.org/10.1016/j.esd.2016.03.003.

MacCarty, N. A., & Bryden, K. M. (2016b). An integrated systems model for energy services in rural developing communities. *Energy*, *113*, 536–557. https://doi.org/10.1016/j.energy.2016.06.145.

MacCarty, N. A., & Bryden, K. M. (2017). Costs and impacts of potential energy strategies for rural households in developing communities. *Energy*, *138*, 1157–1174. https://doi.org/10.1016/j.energy.2017.07.051.

Masera, O. R., Díaz, R., & Berrueta, V. (2018). From cookstoves to cooking systems: the integrated program on sustainable household energy use in Mexico. Retrieved from http://www.bioenergylists.org/stovesdoc/Masera/ESDVol9No1%255Cmexicoprogram.pdf.

McNamara, N., & Kirakowski, J. (2005). Defining usability: Quality of use or quality of experience? *IPCC 2005. Proceedings. International Professional Communication Conference, 2005* (pp. 200–204). https://doi.org/10.1109/IPCC.2005.1494178.

Merino, G. S. A. D., Teixeira, C. S., Schoenardie, R. P., Merino, E. A. D., & Gontijo, L. A. (2012). Usability in product design - The importance and need for systematic assessment models in product development - Usa-Design Model (U-D)©. *Work*, *41*(SUPPL.1), 1045–1052. https://doi.org/10.3233/WOR-2012-1011-1045.

Millen, D. R. (2000). Rapid ethnography: Time deepening strategies for HCI field research. Retrieved from http://onemvweb.com/sources/ethnography/rapid_ethnography.pdf.

Mobarak, A. M., Dwivedi, P., Bailis, R., Hildemann, L., & Miller, G. (2012). Low demand for nontraditional cookstove technologies. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27), 10815–10820 https://doi.org/DOIhttps://doi.org/10.1073/pnas.1115571109.

Moses, N. D., & MacCarty, N. A. (2018a). A practical evaluation for cookstove usability. *Proceedings to ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

Moses, N. D., & MacCarty, N. A. (2018b). *Usability testing protocol for cookstoves.* Oregon State University Retrieved from https://humanitarian.engineering.oregonstate.edu/project-page/usability-testing-protocol-cookstoves.

Moses, N. D., & MacCarty, N. A. (2018c). What makes a cookstove usable? Trials of a of cookstove usability testing protocol in Uganda, Guatemala, and the United States. *Energy Res. Soc. Sci.* (in review).

Namagembe, A., Muller, N., Scott, L. M., Zwisler, G., Johnson, M., Arney, J., ... Mugisha, E. (2015). Factors influencing the acquisition and correct and consistent use of the top-lit updraft cookstove in Uganda. *Journal of Health Communication*, *20*(sup1), 76–83. https://doi.org/10.1080/10810730.2014.994245.

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann.

Oyugi, C., Dunckley, L., & Smith, A. (2008). Evaluation methods and cultural differences: studies across three continents. *NordiCHI '08: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges* (pp. 318–325). https://doi.org/10.1145/1463160.1463195.

Pakravan, M. H., & MacCarty, N. A. (2018). Evaluating user intention for uptake of clean technologies using the theory of planned behavior. *Proceedings to ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

Pattanayak, S. K., Jeuland, M. A., Lewis, J. J., Bhojvaid, V., Brooks, N., Kar, A., & Philippone, L. (2016). *Cooking up change in the Himalayas: Experimental evidence on cookstove promotion.*

Pawson, M. (2009). Extremely rapid usability testing. *Journal of Usability Studies*, *4*(3), 124–135.

Pilishvili, T., Loo, J. D., Schrag, S., Stanistreet, D., Christensen, B., Yip, F., ... Bruce, N. (2016). Effectiveness of six improved cookstoves in reducing household air pollution and their acceptability in rural western Kenya. *PLoS One*, *11*(11), e0165529. https://doi.org/10.1371/journal.pone.0165529.

Putti, V. R., Tsan, M., Mehta, S., & Srilata, K. (2015). The state of the global clean and improved cooking sector. *ESMAP and GACC* (pp. 1–179) https://doi.org/007/15.

Puzzolo, E., Pope, D., Stanistreet, D., Rehfuess, E. A., & Bruce, N. G. (2016). Clean fuels for resource-poor settings: A systematic review of barriers and enablers to adoption and sustained use. *Environmental Research*, *146*, 218–234. https://doi.org/10.1016/j.envres.2016.01.002.

Quesenbery, W. (2004). Balancing the 5Es of usability. *Cutter IT Journal*, *17*(2), 4–11.

Ramanathan, V., & Carmichael, G. (2008). Global and regional climate changes due to black carbon. *Nature Geoscience*, *1*(4), 221–227. https://doi.org/10.1038/ngeo156.

Redish, J. (2007). Expanding usability testing to evaluate complex systems. *Journal of Usability Studies*, *2*(3), 102–111.

Rettger, M. B. (2010). The seven habits of highly effective usability people. *Journal of Usability Studies*, *5*(1), 76–80.

Rhodes, E. L., Dreibelbis, R., Klasen, E., Naithani, N., Baliddawa, J., Menya, D., ... Jaime Miranda, J. (2014). Behavioral attitudes and preferences in cooking practices with traditional open-fire stoves in Peru, Nepal, and Kenya: Implications for improved cookstove interventions. *International Journal of Environmental Research and Public Health*, *11*(10). https://doi.org/10.3390/ijerph111010310.

Robinson, J., & Ibraimo, M. (2011). The uncontrolled cooking test: Measuring three-stone fire performance in northern Mozambique. *The 19th International Conference Domestic Use of Energy* (pp. 2–7) Retrieved from http://unf.mediapolis.com/binary-data/DOCUMENT/file/000/000/82-1.pdf.

Rosenbaum, J., Derby, E., & Dutta, K. (2015). Understanding consumer preference and willingness to pay for improved cookstoves in Bangladesh. *Journal of Health Communication*, *20*(1), 20–27. https://doi.org/10.1080/10810730.2014.989345.

Rosenbaum, S. (2008). The future of usability evaluation: Increasing impact on value. *Human-Computer Interaction Series. 2008*. (pp. 344–378). London: Springer. https://doi.org/10.1007/978-1-84628-941-5_15.

Ruiz-Mercado, I., Masera, O., Zamora, H., & Smith, K. R. (2011). Adoption and sustained use of improved cookstoves. *Energy Policy*, *39*(12), 7557–7566. https://doi.org/10.1016/j.enpol.2011.03.028.

Sanders, E. B. -N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068.

Stanistreet, D., Hyseni, L., Bashin, M., Sadumah, I., Pope, D., Sage, M., & Bruce, N. (2015). The role of mixed methods in improved cookstove research. *Journal of Health Communication*, *20*(S1), 84–93. https://doi.org/10.1080/10810730.2014.999896.

Thacker, K. S., Barger, K. M. C., & Mattson, C. A. (2015). Balancing technical and user objectives in the redesign of a peruvian cookstove. *Development Engineering*, *2*, 12–19. https://doi.org/10.1016/j.deveng.2016.05.001.

Thacker, K. S., & Mattson, C. A. (2014). A global review of end user needs: Establishing the need for adaptable cookstoves. *Proceedings of the 4th IEEE Global Humanitarian Technology Conference, GHTC* (pp. 649–658).

Troncoso, K., Armendáriz, C., & Alatorre, S. (2013). Improved cook stove adoption and impact assessment: A proposed methodology. *Energy Policy*, *62*, 637–645. https://doi.org/10.1016/J.ENPOL.2013.07.074.

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., & Bergel, M. (2002). An empirical comparison of lab and remote usability testing of web sites. *Usability Professionals Association Conference* Retrieved from https://www.researchgate.net/publication/228540469.

Urmee, T., & Gyamfi, S. (2014). A review of improved cookstove technologies and programs. *Renewable and Sustainable Energy Reviews*, *33*, 625–635. https://doi.org/10.1016/j.rser.2014.02.019.

Vatrapu, R., & Perez-Quinones, M. (2006). Culture and usability evaluation: The effects of culture in structured interviews. *Journal of Usability Studies*, *1*(4), 156–170 Retrieved from https://arxiv.org/ftp/cs/papers/0405/0405045.pdf.

Victor, B. (2011). Sustaining culture with sustainable stoves: The role of tradition in providing clean-burning stoves to developing countries. *Consilience-The Journal of Sustainable Development*, *5*(1), 71–95. https://doi.org/10.7916/D86973X0.

WHO (2016). Household air pollution and health. *Fact sheet*http://www.who.int/mediacentre/factsheets/fs292/en/.

Wood, A. E., & Mattson, C. A. (2016). An experiment in engineering ethnography in the developing world. *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE* (pp. 1–10).