Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
IDETC/CIE 2018
August 26-29, 2018, Quebec City, Canada

DETC2018-85728

A PRACTICAL EVALUATION FOR COOKSTOVE USABILITY

Nicholas D. Moses
Oregon State University
Corvallis, Oregon, USA

Nordica A. MacCarty
Oregon State University
Corvallis, Oregon, USA

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 methods, they often have not achieved the intended impact. One of the main reasons for this shortcoming is that engineers often 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, 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 developed. The proposed protocol is based on established usability practices from fields such as software and consumer product design, and includes usability criteria taken from existing cookstove research and interviews with subject experts. The protocol includes objective and subjective testing methods, is designed to elicit user perceptions of and the relative importance of each usability criterion in a given context, and incorporates ethnographic methods to improve validity in cross-cultural applications and in diverse testing scenarios. This protocol may be useful to stove designers as a way to better understand users and validate or improve designs, to implementers as a method to assist with the selection of the most appropriate stove for a project, and to researchers as a tool to assess cookstoves and cookstove programs. Preliminary field and laboratory work to test the validity of the protocol demonstrated a mixture of meaningful and uncertain results, indicating that while it is a reasonable tool to assess cookstove usability, the protocol requires interpretation of qualitative data and assessment of uncertainty to be most effective.

INTRODUCTION

Every year, over 4 million people die prematurely from exposure to smoke from biomass cooking and heating fires [1]. Fuel collection and smoke emissions have also been shown to degrade environmental quality and contribute to regional and global climate change [2,3]. While up to a billion dollars is spent on improved cookstoves each year 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 [4,5]. 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 [4–8].

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 [9]. Usability 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 [9]. 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 on usability as a part of humanitarian product design. 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 [6,10], 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.

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 context of their region and work. This information may then be used to better balance user needs with technical performance, emissions, and other objectives to increase the overall uptake and impact of improved cookstoves. Elements of this work are incorporated into the ISO/TC 285 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 standard, once the standard is published [11].

The proposed protocol organizes cookstove usability into six main criteria, each with multiple sub-criteria, to provide 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. The margin of error for each criterion, calculated for a confidence level specified by the test administrator, is also reported 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 valid 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 may be required on the part of the test administrator to adapt the protocol to uncommon testing scenarios.

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. These are often tailored to specific applications, such as the usability of websites or electronics [12–14].

Existing definitions include a variety of concepts regarding how users think about, approach, and judge the success of their interaction with a product [15]. 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 [16], and Nielsen's well-known book, "Usability Engineering," which expands the definition of usability to include memorability, learnability, and the consideration of user errors

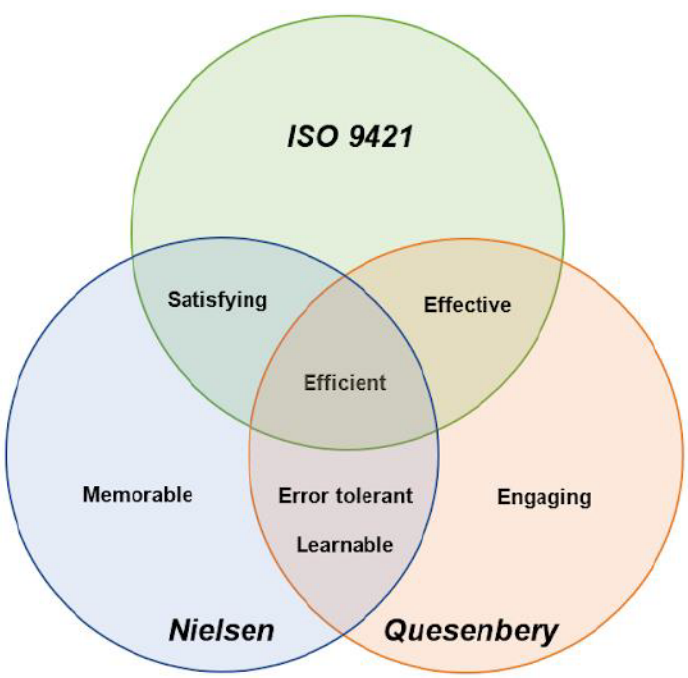

Figure 1: COMMON DEFINITIONS OF USABILITY

[17]. A graphic depicting the similarities and differences of these definitions of usability is shown in Figure 1. Broader definitions of usability may also include factors such as ease of access, learning curve, aesthetics, and safety [18,19], 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 [20]. 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 [21], it does not necessarily require the same level of expertise and time investment in qualitative research methods, which are less familiar to many designers and engineers. The proposed usability testing protocol is not meant to replace or exclude the use of other research methods, but to serve as an accessible starting point 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 evaluations or their implementation, but instead many related approaches that may be adapted to create the most appropriate evaluation for a specific context [22,23]. 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 [24]. 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 [18].

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 [14]. While this may be preferred by many designers and engineers, such a controlled setting is not always available, as is often the case with humanitarian engineering design. Qualitative data can also provide more context and insight into users' perceptions [25], which are critical in the decision to purchase or use a product [26].

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 [26], a quantitative approach based on the interface features of a product and the context of the user, product, and activity [22], 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 [27]. 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 [16]. 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 differ in their content and application, they have in common a systematic approach to evaluation design, 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 within the given parameters.
- The validation of the testing protocol.

Existing Cookstoves Usability Research

Much of the body of research on biomass cookstoves and existing testing protocols focus on technical performance, which is easier to measure and more familiar to designers with engineering backgrounds [6,28]. Research into cookstove usability has only recently begun to receive attention; a 2013 review found two studies of usability in existence [29]. Conducted by Adkins et al., these studies considered usability alongside technical performance tests as factors in adoption, concluding that a variety of factors, including usability, influenced adoption decisions [30]. While two studies focused on understanding global cookstove user needs and designing usable stoves have been completed by Thacker since that time [5,6], no published research exists which examines the usability of a wide range of stove designs or different usage contexts.

The inability of high technical performance to drive cookstove adoption has been established, however, as has the importance of balancing technical goals and user needs to maximize adoption, sustained use, and impact [4–6,10,28,31]. 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 [4,6,32]. It has also been established that usability is a critical factor in designing or selecting products for users from a different culture, especially when products are intended to create behavior change [20], which is the case for many improved cookstove projects.

This small body of cookstove usability literature has 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 [6,33–35]. While these lists overlap and offer many insights into user needs, none claim to be complete, nor have they been organized into repeatable, practical 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 [36].

It should also be noted that while cookstove usability and user satisfaction has 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 have also been neglected in favor of technical stove performance and other objectives [4,32,37–39]. These aspects also 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 [24,40–42]. Anthropological methods, such as ethnographic interviews and participant observation, are sometimes employed by usability practitioners because of their effectiveness in eliciting user opinions, thought processes, and needs, especially when these are not initially known or obvious to the researcher [41,43]. The use of mixed-methods research has also been suggested as more effective than either quantitative or qualitative methods, alone, in cookstove research [44].

Effects of culture. While usability evaluation design outside of industrialized nations has not been as thoroughly researched, 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 increases [9,45,46]. Considerations for cross-cultural testing and test design have been explored for applications such as computer user interface design. To maximize effectiveness, these tests must account for the user's testing culture and experience as well as their relationship with the evaluator. 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 queues and styles of communication [45,46].

Certain types of testing methods, such as questionnaires and observation, also tend to yield different results in different cultures [45,46]. 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 [47].

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 [48]. 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 [49].

COOKSTOVE USABILTY PROTOCOL DEVELOPMENT

Cookstove Usability Criteria Selection

To judge cookstove usability, an appropriate set of criteria from usability literature was identified. Criteria were compiled from a systematic review of existing international development literature [5,33–35], as well as discussions with industry experts. Most criteria were mentioned by multiple literature sources and experts. Cookstove safety and durability, which by some definitions could be included in or overlap with usability, are incorporated into separate technical testing protocols within the upcoming ISO/TC 285 cookstove testing standard [11]. 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 objective metrics evaluated by the existing protocols.

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 expected users of this protocol are designers, implementers, and researchers working with cooking systems in low- and middle-income countries. From the authors' experience, these practitioners 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.

Testing Methodology

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

- Broadly applicable to encompass the majority of usability testing needs and scenarios.
- Used as 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 build from or modify the test to fit their specific needs.

Testing methods were developed with existing usability evaluation methods as a foundation, and were supplemented with 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 user needs and preferences. While this is not typical of many traditional usability tests, it has been identified as a necessary step in the context of improved cookstoves to allow practitioners to deliver effective products [41], and is consistent with recent trends in the expansion of the scope and methods of usability testing [42,43].

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 are based on subjective user opinions and cannot be easily measured objectively. Appropriate testing methods for these criteria were modeled on rapid usability tests [48,49], with the inclusion of Likert scale survey questions to elicit user perceptions. These questions have a set number of qualitative answers (usually 5, in this protocol), which correspond to a numerical value. This method 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

Table 1. COOKSTOVE USABILITY CRITERIA

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 safety	Lighting
	Versatility	Visibility of fire		Pot soot deposits	Portability
		Ease of lighting		Cooking height	Water heating
		Fire start-up delay		Stove aesthetics	Food drying/smoking
		User error		Perceived durability	*(Additional needs may be added by test administrator)*
		User instruction		Perceived value	
				Taste	

expertise needed to effectively interpret qualitative results. Purely quantitative methods, on the other hand, are unlikely to elicit user perceptions effectively in this context [41]; without a detailed prior knowledge of a particular group of cooks and cooking technologies, designing an effective quantitative evaluation is not practical. Likert scale questions were chosen for the bulk of the evaluation to offer insight into a cook's perceptions and priorities while allowing for relatively simple testing and analysis compared to more open-ended or qualitative methods. Survey questions have also been applied to the criteria assessed by objective measurements and observations for 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 allow test administrators to clarify and identify error in other portions of the test, and to give the cooks an opportunity to share additional information not explicitly asked for by the test administrators. This method was selected to elicit as much detail as possible, 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, as is likely to be the case in international testing situations [25].

COOKSTOVE USABILITY PROTOCOL OVERVIEW

In the interest of page length, this paper summarizes and provides examples for each portion of the protocol. A more detailed description of the protocol design methodology has been submitted for publication in *Energy for Sustainable Development* [50]. The most recent version of the stand-alone protocol is also available at:
https://humanitarian.engineering.oregonstate.edu/usability-testing-protocol-cookstoves

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 without room for bias and error, factors that should be considered in planning testing and evaluating results include, but are not limited to, the following:

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 [11]. Similarly, cooks should be representative of the range of expected users of the stove (by age, income, proximity to urban areas, etc.). Finally, sampling methods and sample size should be adequate for the desired level of detail and significance of the results. Detailed guidance on sampling and statistical significance in this context may be found in the UNFCCC CDM Guidelines for Sampling and Surveys [51].

Based on the field testing described in this document, at least three tests with three different cooks are recommended per stove model to provide a minimum, qualitative understanding of usability, and approximately 10 tests are needed to produce quantitative results with a margin of error of 0.5 points or less for three quarters of the usability criteria in the protocol (which are scored from 0 – 4 points) with a 95% confidence level. 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. These criteria cannot be characterized effectively with purely quantitative methods, 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 adequately understood, as no amount of quantitative testing is likely to produce adequate results to inform stove design or selection.

Frequent data analysis in between tests may allow for the identification of such criteria during a study, which can then be targeted with qualitative methods during subsequent tests to achieve a more complete understanding. Higher or lower standards for statistical significance and different usability testing needs may require larger or smaller sample sizes than prescribed, here, as well. Ten repetitions is suggested as a practical compromise by the author, beyond which the number of criteria identified as candidates for further qualitative assessment may not increase, leading to diminishing returns from additional tests.

It should also be noted that testing may be most valuable when done on both baseline and current or potential improved stove designs. Understanding the usability of existing stoves will give context to a cook's perception of new stoves, which are likely to be judged against what the cook is familiar with.

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 subtle, 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.

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 [48].

There are many other ways that the test administrator's identity may impact test results and should be considered. For example, because it may be unusual for a man to spend time in the kitchen in some regions, a male test administrator may not always be as welcome or receive the same quality of responses. In many cases, a local woman with relevant survey or interview experience may make the ideal test administrator. Similarly, the more familiar a cook is with the test administrator, the more likely they are to behave normally and give direct answers in most cases. It should be noted that if the evaluator has a relationship with an NGO, government agency, or other source of authority, a cook may be more likely to bias her behavior or

Table 2. PROTOCOL TEST SECTIONS

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.

responses to meet the expectations of the evaluator, or in the hopes of gaining something in return.

Additional considerations. A cook may prepare a more complex meal, use different stoves or fuels, etc. if she considers the test administrator(s) to be guests (even if she is asked to prepare a meal normally). Cooking practices can also depend on weather, harvest or seasonal employment schedules, etc. A local test administrator can advise whether a cook has deviated from a typical meal, or if and when seasonal patterns or other factors may impact test results.

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 behavior and verbal responses. Bringing a cook into a laboratory setting would create 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 [49].

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 [48].

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 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 [25]. Examples of each major type of question/test included in the protocol are shown in Figure 2.

Figure 2: EXAMPLE TESTS AND QUESTIONS

2.1 Example Observation Test

Ease of lighting	Kindling or accelerants are required: **Y / N** Fire must be lit in an enclosed space within the stove: **Y / N**

2.2 Example Quantitative Measurement Test

Cooking time	Start time: ____ : ____ Completion time: ____ : ____ (Starts when food, water, etc. is first heated on the fire. Completed when the last dish is removed from the fire)

2.3 Example Paired Likert-Scale Survey Question Set

	Question	Response Options				
A	"Does your stove cook quickly or slowly?"	"Very slowly"	"Slowly"	"Neither quickly nor slowly"	"Quickly"	"Very quickly"
B	"How important is it that a stove cook quickly?"	"Very unimportant"	"Unimportant"	"Somewhat important"	"Important"	"Very important"

2.4 Example Semi-Structured Interview Question

"Can you list a few of your least favorite things about your stove? Why are they important to you?"
(Response paraphrased here by test administrator)

Alternative Testing Procedures

The protocol is designed to measure cookstove usability 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 is required, such as when choosing between a limited number of cookstove models instead of changing 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 minutes or less per household (without a cooking event taking place). Though much contextual information is 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. This may be done in one of three ways:

1. A Cookstove Characteristics Evaluation can be done without lighting a stove. This provides basic information about likely usability performance, and is most valuable with a thorough understanding of local cooking needs and habits.
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 local 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). Some fuel and time measurements are shared between this Usability Protocol and the CCT and UCT [52,53], and doing multiple types of tests 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 may 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.

Scoring

Most usability criteria are first assessed and evaluated before being assigned a relative weight, to determine how important a criterion is to the cook, through paired sets of survey questions. Results from each testing method are reported in groups for each criterion so they may be easily compared. This has been done because the relative importance of many usability criteria varies significantly between contexts. For example, in an area where wood is freely and easily accessible, fuel consumption may not be as large of a concern as in arid regions.

While the protocol is meant to assess many types of technology, not all questions and measurements apply to household cooking technologies other than wood 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

A Microsoft Excel spreadsheet is included as a part of the protocol to simplify the entry, storage, and analysis of test data. This spreadsheet calculates numerical scores for each usability sub-criteria (from 0-4, generally based on Likert scale questions), as well as a margin of error for each score for a certain confidence level, similar to the Kitchen Performance Test (KPT) commonly used to assess fuel consumption and other factors [54]. The spreadsheet also includes a qualitative data analysis tool to assist with the interpretation of survey results. Qualitative data from field notes does not factor into numerical scores, but may provide additional context to results and help to identify potential biases or errors. In the case of small sample sizes, this data may be more valuable than the numerical scores.

Data Reporting and Interpretation

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, and to account for the relative importance of each criterion to the users. Averages are weighted based on the perceived importance of each criterion indicated by the cook in the survey portion of the test. If no survey is conducted, all sub-criteria are weighted equally. The widest sub-criterion margin of error is also reported for the specified confidence level, to indicate whether there may be significant uncertainty in each overall score value. As with each sub-criterion, overall scores range from 0-4. This has been done to coincide 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 familiar to many cookstove practitioners, which may serve to facilitate easier communication and understanding [55].

These numerical scores are not meant to be a direct prediction of adoption or usage behavior, but to highlight areas of potential concern that should be investigated further. Examples of the scoring results and format for the six main usability criteria, as well as more detailed sub-criteria results, may be seen in Figure 3 and Figure 4, respectively.

Known Limitations

This protocol has several limitations and differences compared to the existing cookstove tests familiar to stove designers and implementers, such as the CCT [52] or UCT [53]. 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 cultural factors. A relatively large number of tests may also be needed to achieve a statistically valid comparison between stoves, or understanding of a single stove model, though it should be noted that smaller sample sizes may still elicit many key qualitative aspects of usability for a particular stove and context. Variation should also be expected between regions and cultures, so results may or may not be applicable to other contexts. Relatively large differences in opinion and cooking habits can also be expected from person to person within one sample group, though this may indicate an actual diversity in cooking needs, in addition to any limits to the repeatability or validity of the test methods. Finally, the universality of protocol comes at the expense of some sensitivity to regional cooking needs and cultural factors.

FIELD TRIAL AND LABORATORY TESTING RESULTS

This protocol was trialed in Lira, Uganda and nearby rural communities in the summer of 2017 with assistance from International Lifeline Fund (ILF), an international NGO operating cookstove and other development programs near Lira. This human subject research was conducted with oversight from the Oregon State University Institutional Review Board under study number 7257. The quantitative and objective stove measurements and observations portions of the test were also trialed on 11 stove models in the possession of the Aprovecho Research Center in Cottage Grove, Oregon, in the US. The purpose of the field study and laboratory measurements were not to evaluate the usability of particular stove models, but to test the application of the protocol itself in representative households and on a variety of cookstove designs. More detailed data and analysis from this work have been submitted for publication in the journal *Energy Research and Social Science* [56].

Field Trial Results

A description of the kitchens and cooking technologies evaluated is shown in Table 2. Note that total number of stoves evaluated is greater than the number of kitchens, as some cooks used multiple stoves of the same or different designs at the same time.

Table 2. TYPES AND QUANTITIES OF COOKSTOVES EVALUATED IN UGANDA

Kitchen Description	Cooking Technologies Used
8 - Rural household	2 - Traditional mud stove 2 - Traditional three stone fire 9 - Improved clay rocket stove
2 - Urban household	1 - Traditional mud charcoal stove 2 - Improved charcoal stove
2 - Urban institutional	1 - Improved charcoal/wood rocket 1 - Modern 4 burner LPG

Two ILF staff members, both Ugandan women from the region where testing was conducted, alternated administering tests in each household. The author was present to observe and guide testing as well as to complete the physical measurement and observation portions of the test. Over the course of the field study, the author and evaluators collaborated on improving question language and the testing procedures. At the end of the testing period, the ILF test administrators were interviewed about their opinions and experience administering the protocol. They agreed that the results of the test were in-line with their experience working with and around cookstoves in these locations over a three-year period.

While the role of ILF as a source of free or subsidized goods and services – including subsidized ILF improved clay rocket stoves to most of the study participants in 2015 – certainly introduced a potential for bias in the test results, the established relationships between the ILF staff and the cooks, as well as the staff's past experience with cookstove implementation, allowed for the advantage of previously established trust and personal relationships and made ILF staff well-qualified to judge the process of administering the protocol.

It became clear by the end of the field trial that insufficient data had been collected to effectively assess the usability of the stove models tested, with the possible exception of the improved clay rocket stove. This stove was evaluated in five households, which was sufficient to reach approximate saturation of qualitative usability insights where no significant new concepts were elicited in the last two evaluations. The numerical scores for approximately half of the usability criteria had a margin of error of greater than 25% of the mean with a 95% confidence level, however. For the remaining stove types, with sample sizes of two or fewer households there was not enough data to calculate the standard error and confidence intervals for quantitative data, nor was a saturation of qualitative data reached. A summary of the overall results for the five households using improved clay rocket stoves, output by the data processing Excel spreadsheet, is shown in Figure 3. These preliminary results are given not to provide a conclusive assessment of usability, but to demonstrate the type and meaning of results output by the protocol.

Figure 3. OVERALL USABILITY RESULTS TABLE: IMPROVED CLAY ROCKET STOVE

Results Summary	Score	Highest Sub-criteria MoE (95% CI)
I. Fuel cost and convenience	2.2	1.89
II. Cooking performance	3.0	2.72
III. Operability	2.2	0.68
IV. Maintenance	2.9	2.07
V. Comfort	3.2	0.68
VI. Location-specific needs	N/A	

While there was at least one sub-criterion with a relatively large margin of error for each overall category score, these values were in alignment with the ILF staff's general understanding of the stove's strengths and weaknesses. This high variation does not necessarily indicate that the overall score is invalid, but that the results for each sub-criterion should be examined in more detail to arrive at an effective interpretation. Note that the location-specific needs score was assigned an N/A because the evaluation and scoring process was changed significantly as a result of the field trial. The data collected was insufficient to rate the stove according to the revised protocol.

The cooking performance results for the improved clay rocket stove is shown in Figure 4 as an example of the more detailed sub-criteria results output by the data processing spreadsheet, again for the five households using improved clay rocket stoves. The margins of error were relatively small for the

Figure 4. EXAMPLE SUB-CRITERIA RESULTS FROM UGANDA

II. Cooking performance	**Result**	**Relative Weight**	**MoE (95% CI)**
A. Cooking speed (percieved)	3.8	4.0	0.56
Cooking duration (measured)	*02:59*		*1:05*
B. Firepower control	2.2	3.8	2.04
C. Firepower range	1.6	2.4	2.72
D. Use of all pots and pans	3.8	4.0	0.56
		Overall performance score:	3.0
		Highest subcategory margin of error:	2.72

"cooking speed" and "use of all pots and pans" categories, indicating a strong agreement between cooks surveyed on the scores reported, as well as a high confidence that these results are representative of the responses that would be obtained from others in their community who also received subsidized ILF stoves. The wide margins of error for the remaining criteria could indicate that there is either high variation in opinion between cooks, or that the test results for these criteria were not consistently valid. Additional qualitative and quantitative testing would have been needed to clarify the likely cause of this variation, as well as the significance of the scores for these criteria. It may be that these results indicate high variation in perspectives between cooks, or perhaps variation in honesty, where some cooks responded in a way that would please ILF staff and others responded with real opinions. In cases like these, qualitative data, either taken during the tests or during follow-up interviews, is necessary to understand cooks' differing perspectives and any implications for stove design or selection.

Also note that quantitative measurements, such as cooking duration, are presented alongside subjective scores when possible to allow for comparison, as well as to avoid making value judgements on numbers that may have different significance to different cooks and cooking cultures.

Laboratory Testing Results

The quantitative measurement portion of the protocol (the *Cookstove Characteristics Evaluation*, described previously) has also been tested on 11 different cookstove models at the Aprovecho Research Center in Cottage Grove, Oregon, to help refine and calibrate these tests. These stoves included a variety of wood-burning rocket stoves, top lit updraft pellet stoves (TLUD's), and charcoal stoves, all of which are in production and use in at least one location in Africa, Asia, or Central America. Key results and observations from these tests included:

- The improvement and expansion of several tests in the *Cookstove Characteristics Evaluation* portion of the protocol.
- Verification that all protocol tests and questions are relevant to the common biomass cookstove designs tested.
- The understanding that amount of information, quantitative or qualitative, that can be collected from a cookstove without observing a cooking event or engaging a cook, is relatively small.

DISCUSSION

This protocol is designed to offer practitioners an accessible method to increase their understanding of user needs, and how effectively a cookstove meets those needs in a given context. Based in both engineering and anthropological methods, the protocol includes a mix of quantitative and objective tests, as well as qualitative 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 methods is meant to add context to the data, allow for easier identification of error or bias, and to help overcome communication barriers inherent in the cross-cultural scenarios common in cookstove programs. The calculation of the margin of error for each criterion allows for the identification of criteria that may be assessed effectively with comparatively simple quantitative data, as well as indicates which criteria require more thorough qualitative assessment to be adequately understood, allowing for a more efficient use of both quantitative and qualitative methods.

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 use of ethnographic methods such as participant observation and direct interviews with cooks may also encourage practitioners to develop a deeper understanding of the people they serve, the underlying health and other issues at hand, and the international development environment in which they work.

This work to increase the standardization and prevalence of cookstove usability testing may allow for an increase in accountability for design or selection of stoves in a given context, as well as clearer relationships between usability and health and environmental outcomes. This work will also hopefully serve as a step towards future user and design research and standardization in other areas of international development. Future research could be designed to connect the results of this protocol to adoption and sustained usage, further validate and refine testing processes and criteria, and create new versions of the protocol that are optimized for specific regions, cultures, and testing applications, and require less qualitative testing and data analysis.

The full protocol is available online for use and evaluation at: *https://humanitarian.engineering.oregonstate.edu/usability-testing-protocol-cookstoves*

ACKNOWLEDGMENTS

The authors would like to thank International Lifeline Fund – especially the staff in Lira, Uganda – for facilitating the field trial 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, and Dale Andreatta for their contributions to this work, as well as the Aprovecho Research Center for allowing us access to cookstoves in their laboratory. 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, Industrial, and Manufacturing Engineering for supporting this research.

REFERENCES

[1] World Health Organization, 2016, "Household Air Pollution and Health" [Fact sheet].

[2] T. C. Bond *et al.*, "Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850-2000," *Global Biogeochem. Cycles*, vol. 21, no. 2, Jun. 2007.

[3] V. Ramanathan and G. Carmichael, "Global and regional climate changes due to black carbon," *Nat. Geosci.*, vol. 1, no. 4, pp. 221–227, 2008.

[4] A. M. Mobarak, P. Dwivedi, R. Bailis, L. Hildemann, and G. Miller, "Low demand for nontraditional cookstove technologies," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 27, pp. 10815–10820, 2012.

[5] K. S. Thacker and C. A. Mattson, "A Global Review of End User Needs : Establishing the Need for Adaptable Cookstoves," in *Proceedings of the 4th IEEE Global Humanitarian Technology Conference, GHTC*, 2014, pp. 649–658.

[6] K. S. Thacker, K. M. C. Barger, and C. A. Mattson, "Balancing technical and user objectives in the redesign of a peruvian cookstove," *Dev. Eng.*, vol. 2, pp. 12–19, 2015.

[7] N. G. Johnson, "Village energy system dynamics of an isolated rural West African village," PhD dissertation, College of Engr., Iowa State Univ., Ames, 2012.

[8] B. Victor, "Sustaining Culture with Sustainable Stoves: The Role of Tradition in Providing Clean-Burning Stoves to Developing Countries," *Cons. J. Sustain. Dev.*, vol. 5, no. 1, pp. 71–95, 2011.

[9] A. E. Wood and C. A. Mattson, "An Experiment in Engineering Ethnography in the Developing World," in *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE*, 2016, pp. 1–10.

[10] N. MacCarty and K. M. Bryden, "Modeling Technology Strategies for Thermal Energy Services in Rural Developing Communities," *Proc. ASME IDETC/CIE 2015*, pp. 1–13, 2015.

[11] *Clean cookstoves and clean cooking solution,* ISO/TC WD 285: Geneva, Switzerland, 2018. https://www.iso.org/committee/4857971.html

[12] M. Hertzum, "Images of Usability," *Int. J. Hum. Comput. Interact.*, vol. 26, no. 6, pp. 567–600, 2010.

[13] N. McNamara and J. Kirakowski, "Defining usability: quality of use or quality of experience?," *IPCC 2005. Proceedings. Int. Prof. Commun. Conf. 2005.*, pp. 200–204, 2005.

[14] N. Bevan, "What is usability," in *Proceedings of the 4th International Conference on HCI, Stuttgart*, 1991, no. September, pp. 651–655.

[15] W. Quesenbery, "Balancing the 5Es of Usability," *Cut. IT J.*, vol. 17, no. 2, pp. 4–11, 2004.

[16] *Ergonomics of Human System Interaction*. ISO 9241:11. Geneva, Switzerland, 1998. https://www.iso.org/standard/16883.html

[17] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1994.

[18] N. Bevan, "Combining Quantitative and Qualitative Data for Measuring User Experience of an Educational Game," in *International Workshop on Meaningful measure: Valid Useful User Experience Measurement (VUUM)*, 2008, p. 27.

[19] J. Fulton Suri and M. Marsh, "Scenario building as an ergonomics method in consumer product design," *Appl. Ergon.*, vol. 31, no. 2, pp. 151–157, 2000.

[20] J. Daae and C. Boks, "A classification of user research methods for design for sustainable behaviour," *J. Clean. Prod.*, vol. 106, pp. 680–689, 2014.

[21] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design," *Co-Design*, vol. 4, no. 1, pp. 5–18, 2008.

[22] J. Kwahk and S. H. Han, "A methodology for evaluating the usability of audiovisual consumer electronic products," *Appl. Ergon.*, vol. 33, no. 5, pp. 419–431, 2002.

[23] J. S. Dumas, "User-based evaluations," in *The human-computer interaction handbook*, J. A. Jacko and A. Sears, Eds. Lawrence Erlbaum Associates, 2003, pp. 1093–1117.

[24] M. B. Rettger, "The Seven Habits of Highly Effective Usability People," *J. Usability Stud.*, vol. 5, no. 1, pp. 76–80, 2010.

[25] H. Bernard, *Research methods in anthropology: Qualitative and quantitative approaches*, 4th ed. Lanham: AltaMira Press, 2011.

[26] S. H. Han, M. Hwan Yun, K. J. Kim, and J. Kwahk, "Evaluation of product usability: Development and validation of usability dimensions and design elements based on empirical models," *Int. J. Ind. Ergon.*, vol. 26, no. 4, pp. 477–488, 2000.

[27] G. S. A. D. Merino, C. S. Teixeira, R. P. Schoenardie, E. A. D. Merino, and L. A. Gontijo, "Usability in product design - The importance and need for systematic assessment models in product development - Usa-Design Model (U-D) ??," *Work*, vol. 41, no. SUPPL.1, pp. 1045–1052, 2012.

[28] N. A. MacCarty and K. M. Bryden, "Costs and impacts of potential energy strategies for rural households in developing communities," *Energy*, vol. 138, pp. 1157–1174, 2017.

[29] M. Kumar, S. Kumar, and S. K. Tyagi, "Design, development and technological advancement in the biomass cookstoves: A review," *Renewable and Sustainable Energy Reviews*, vol. 26. pp. 265–285, 2013.

[30] E. Adkins, E. Tyler, J. Wang, D. Siriri, and V. Modi, "Field testing and survey evaluation of household biomass cookstoves in rural sub-Saharan Africa," *Energy Sustain. Dev.*, vol. 14, no. 3, pp. 172–185, 2010.

[31] N. A. MacCarty and K. M. Bryden, "A generalized heat-transfer model for shielded-fire household cookstoves," *Energy Sustain. Dev.*, vol. 33, pp. 96–107, 2016.

[32] N. A. MacCarty and K. M. Bryden, "An integrated systems model for energy services in rural developing communities," *Energy*, vol. 113, pp. 536–557, 2016.

[33] E. García-Frapolli *et al.*, "Beyond fuelwood savings: Valuing the economic benefits of introducing improved biomass cookstoves in the Purépecha region of Mexico," *Ecol. Econ.*, vol. 69, pp. 2598–2605, 2010.

[34] I. Ruiz-Mercado, O. Masera, H. Zamora, and K. R. Smith, "Adoption and sustained use of improved cookstoves," *Energy Policy*, vol. 39, no. 12, pp. 7557–7566, 2011.

[35] T. Urmee and S. Gyamfi, "A review of improved Cookstove technologies and programs," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 625–635, 2014.

[36] E. Adkins, E. Tyler, J. Wang, D. Siriri, and V. Modi, "Field testing and survey evaluation of household biomass cookstoves in rural sub-Saharan Africa," *Energy Sustain. Dev.*, vol. 14, no. 3, pp. 172–185, 2010.

[37] M. H. Pakravan and N. A. MacCarty, "Evaluating User Intention For Uptake Of Clean Technologies Using The Theory Of Planned Behavior," in *Proceedings to ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2018. In press.

[38] J. Rosenbaum, E. Derby, and K. Dutta, "Understanding Consumer Preference and Willingness to Pay for Improved Cookstoves in Bangladesh," *J. Health Commun.*, vol. 20, no. 1, pp. 20–27, 2015.

[39] E. Puzzolo, D. Pope, D. Stanistreet, E. A. Rehfuess, and N. G. Bruce, "Clean fuels for resource-poor settings: A systematic review of barriers and enablers to adoption and sustained use," *Environmental Research*, vol. 146. 2016.

[40] R. R. Johnson, M. J. Salvo, and M. W. Zoetewey, "User-centered technology in participatory culture: Two decades 'Beyond a narrow conception of usability testing,'" *IEEE Trans. Prof. Commun.*, vol. 50, no. 4, pp. 320–332, 2007.

[41] J. Loo *et al.*, "User Perspectives of Characteristics of Improved Cookstoves from a Field Evaluation in Western Kenya," *Int. J. Environ. Res. Public Health*, vol. 13, no. 2, p. 167, Jan. 2016.

[42] J. Redish, "Expanding usability testing to evaluate complex systems," *J. Usability Stud.*, vol. 2, no. 3, pp. 102–111, 2007.

[43] S. Rosenbaum, "The Future of Usability Evaluation: Increasing Impact on Value," in *Human-Computer Interaction Series 2008*, Springer, London, 2008, pp. 344–378.

[44] D. Stanistreet *et al.*, "The role of mixed methods in improved cookstove research," *J. Health Commun.*, vol. 20, pp. 84–93, 2015.

[45] T. Clemmensen, M. Hertzum, K. Hornbaek, Q. Shi, and P. Yammiyavar, "Cultural Cognition in Usability Evaluation," *Interact. Comput.*, vol. 21, no. 3, pp. 212–220, 2009.

[46] R. Vatrapu and M. Perez-Quinones, "Culture and Usability Evaluation: The Effects of Culture in Structured Interviews," *J. Usability Stud.*, vol. 1, no. 4, pp. 156–170, 2006.

[47] C. Oyugi, L. Dunckley, and A. Smith, "Evaluation methods and cultural differences: studies across three continents," *Nord. '08 Proc. 5th Nord. Conf. Human-computer Interact. Build. Bridg.*, pp. 318–325, 2008.

[48] M. Pawson, "Extremely Rapid Usability Testing," *J. Usability Stud.*, vol. 4, no. 3, pp. 124–135, 2009.

[49] J. A. Lenker, M. Nasarwanji, V. Paquet, and D. Feathers, "A tool for rapid assessment of product usability and universal design: development and preliminary psychometric testing.," *Work*, vol. 39, no. 2, pp. 141–50, 2011.

[50] N. D. Moses, N. A. MacCarty, and M. H. Pakravan, "A Practical Evaluation for Cookstove Usability," *Energy Sustain. Dev.*, in review, 2018.

[51] UNFCCC CDM Executive Board, 2012, *Guidelines for Sampling and Surveys for CDM Project Activities and Programme of Activities*. https://cdm.unfccc.int/Reference/Guidclarif/meth/meth_guid48.pdf

[52] R. Bailis, "Controlled Cooking Test (CCT)," Shell Foundation, 2004. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/80-1.pdf

[53] J. Robinson and M. Ibraimo, "The Uncontrolled Cooking Test: Measuring Three-Stone Fire Performance in Northern Mozambique," in *The 19th international conference domestic use of energy*, 2011, pp. 2–7.

[54] R. Bailis and R. Edwards, "Kitchen Performance Test (KPT)." Shell Foundation, 2007. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/80-1.pdf

[55] *Guidelines for evaluating cookstove performance.* ISO/IWA 11: 2012. Geneva, Switzerland, 2012. https://www.iso.org/obp/ui/#iso:std:iso:iwa:11:ed-1:v1:en

[56] N. D. Moses and N. A. MacCarty, "Field and Laboratory Trials of a of Cookstove Usability Testing Protcol," *Energy Res. Soc. Sci.*, in review, 2018.